A Head-to-Head Comparison of Human-Based and Automated Text Analysis for Measuring Populism in 27 Countries[*]

Kirk A. Hawkins
Brigham Young University
kirk_hawkins@byu.edu

Bruno Castanho Silva
Central European University
paula-castanho_bruno@phd.ceu.edu

April 20, 2016

Abstract

In this paper we use holistic grading to measure the level of populism found in political parties' discourse during electoral campaigns. The technique is applied to electoral manifestos from 144 parties in 27 countries from Western Europe and the Americas. For about half the sample, we also code speeches by the candidates or party leaders, which indicate the validity of measuring populism on electoral manifestos. Our results indicate that populism is stronger in Latin America than in Europe, and that the level of populism among some European parties typically considered as examples of radical right populism might be overstated when compared with Latin American counterparts. We contrast these results with automated content analysis of the manifestos through machine learning. In this sample, the results of computerized text analysis are not satisfactory, but indicate that with more data it could be a viable option.

---

[*]An earlier version of this paper was presented at *Explaining Populism: Team Populism January Conference*, Provo-UT, January 28-30 2016. This is a draft. **Please do not cite without permission**.

1

1   Introduction

One of the main challenges in studying populism in comparative perspective is defining which political actors deserve this label. Because it is such a controversial concept, most comparative studies classify cases by *fiat*, based either on literature reviews, or relying on country specialists to decide on each case [see, for example, Doyle, 2011, Levitsky and Loxton, 2013, Bustikova, 2014, Mudde, 2007, 2014]. The problem with the first approach is that it often relies on second-hand literature instead of primary sources, and has little room for testing reliability. The problem with the second is that it depends on the different conceptions of populism experts might have, and how their perceptions are driven by the cases they know well. While it gives an idea of how populist are parties within one political system in relation to one another, the scale is not absolute across countries and renders a comparison virtually impossible. While specialists in Sweden may consider, for example, that the Sweden Democrats (*Sverigedemokraterna*) are extremely populist in that country's context, this does not mean that it is also very populist in comparison to parties in other countries.

In this paper we use a tested and validated approach to measuring populism – holistic grading [Hawkins, 2009] – and apply it to 144 parties from 27 countries in Europe and the Americas, creating the first comprehensive data set classifying entire party systems according to the level of populism in parties' discourse. By looking at campaign documents – electoral manifestos and speeches by party leaders –, from all main parties in a political system, we are able to observe how populist each actor is, and compare that to a range of international cases. With these data in hand, we

first observe how populism is distributed across the regions in this study, and which specific parties have been classified as such. Next, we introduce supervised learning methods for automated text analysis that can, theoretically, generate a model capable of classifying these texts but which, given the limitations of the data at hand, have not yielded satisfactory results.

2   Populism and its measurement

We consider populism to be in the realm of ideas, a perspective that has become prominent in recent years [Mudde and Rovira Kaltwasser, 2013, 497]. It is a discourse which sees politics as divided in moral terms, where the good is identified with "the people" and the evil is embodied by an "elite". This "people" encompasses the majority of the population and is a homogeneous, unified body that has an identifiable *will* – the General Will or *volonté générale* –, which should be guiding all decision-making in politics. The elite, on the other side, is a minority who is in power (or in risk of imminent return to), who uses its resources to exploit the people. It is morally evil, and to blame for all bad things that befall the country. Because of this division, populist discourse calls for a "systemic change", or liberation of the people from the grip of the elites. It charges the whole political system of being corrupted by a small ruling group, and pleads that overthrowing this group is the only way to enforce democratic rule by the people. Undemocratic means may be accepted to achieve this goal since, in this framing, the elites are thieves who do not deserve a fair treatment, and the enforcement of the people's will should not be blocked by

formalities and institutions.

An ideational approach along these lines lends itself to operationalization and measurement, since it identifies elements that should be present in a discourse for it to be populist. Following it, researchers have used different content analysis methods to measure populism in recent years. Jagers and Walgrave [2007] test a dictionary-based content analysis to classify populist parties in Flanders, which is extended in Rooduijn and Pauwels [2011] to three more countries. It consists in defining a dictionary of "populist" terms and classifying documents based on their frequency. Rooduijn et al. [2014] use quantitative human-based content analysis of party manifestos from five European countries. This approach has paragraphs as units of analysis, and uses trained coders to classify each one as populist or not, with the aggregated proportion of populist paragraphs being the party score. A third comparative approach has been put forward in Hawkins [2009], and consists of holistic grading. There, chief executives' speeches are coded as a whole, without breaking them down into words or paragraphs. [1]

From these alternatives, this paper uses the third. The dictionary-based technique demands a high knowledge of each specific country for the selection of relevant terms. It may be feasible in single case studies or small-n comparisons, but becomes much less so when a larger number of cases are included. Of the other two, both depart from a similar definition of populism and could potentially be used for the purposes of this study. Hawkins' approach has the upper-hand, however, for having been tested and validated across a large number of countries and time-periods. The original

---

[1]For a review of content analysis methods measuring populism, see Poblete [2015].

study [Hawkins, 2009] included 40 contemporary and historical presidents and prime-ministers from Latin America, Europe, and Asia, while a second round was done with chief executives from Eastern Europe and Central Asia [Hawkins, 2013]. The technique by Rooduijn et al. [2014] has not yet been applied outside of France, Italy, Germany, the Netherlands, and the United Kingdom.

Holistic grading was developed in educational psychology for assessing students' writing [White, 1985, Sudweeks et al., 2004]. It is a human-based coding approach that evaluates the text as a whole. Graders are trained to allocate scores based on the elements of the concept and a set of anchor texts defined as examples for the lowest, intermediate, and highest boundaries. In this case, coders are trained in English on the concept of populism, and the set of training documents are in English as well – but come from a variety of regions. Anchors texts include, for example, speeches by politicians as diverse as Robert Mugabe, Evo Morales, Barack Obama, Tony Blair, Sarah Palin, and Stephen Harper. The training emphasizes that the most important dimension of populism is the notion of a unified, homogeneous people, or the "will of the people", and that this people has to be defined in opposition to an "elite", who is powerful and oppressive. Therefore, even if there is much anti-elitism in the text, if there is no general will of the people, coders are instructed to assign a low score. As in Hawkins [2009, 1062], grades range from 0 to 2, where 0, 1 and 2 are categories defined as follows:

- 0 A speech in this category uses few if any populist elements. Note that even if a manifesto expresses a Manichaean worldview, it is not considered populist if it lacks some notion of a popular will.

5

- 1 A speech in this category includes strong, clearly populist elements but either does not use them consistently or tempers them by including non-populist elements. Thus, the discourse may have a romanticized notion of the people and the idea of a unified popular will (indeed, it must in order to be considered populist), but it avoids bellicose language or references to cosmic proportions or any particular enemy.

- 2 A speech in this category is extremely populist and comes very close to the ideal populist discourse. Specifically, the speech expresses all or nearly all of the elements of ideal populist discourse, and has few elements that would be considered non-populist.

Because graders in earlier studies reported that it was often difficult to choose between the blunt categories, this time they were instructed to give decimal scores, and told that 0.5 rounds up to a categorical 1, and 1.5 rounds up to a categorical 2, so they should consider the qualitative difference between the categories when assigning decimal points. After the training, coders are given the texts – speeches or manifestos – in their original language. One rubric is filled for each document, and each one is discussed with the other coders and the coordinator to clarify questions and check for possible misunderstandings.

2.1 Sampling

Two innovations were introduced in this study in relation to the previous uses of holistic grading: first, it has been expanded from chief executives to candidates

of most parties to the highest executive office. The second change is that party manifestos are coded, instead of only speeches. The option for including manifestos is that these documents help to explore a party's discourse as an institution, which may be distinct from that of its candidate. Also, speeches and manifestos are the documents most comparable across countries: almost everywhere parties produce some kind of election program and party candidates deliver speeches. This means we are looking for populist discourse in documents that are produced and made public with similar purposes across cases. Speeches used are all from national election campaigns – this means that it is possible to find texts for all parties of interest. [2] While for the manifesto we effectively use a census sample (there is usually only one manifesto), for speeches we use a quota sample that selects one speech from the beginning of the campaign and one from the end. The one from the beginning is ideally the one where the candidate is announced by the party, or confirms her candidacy, frequently done in a large party event with significant media coverage. The second speech comes from the end of the race, a few days before the election, often given in the context of a large rally closing the campaign that also has significant media coverage. The reasoning behind these choices is to, first, capture the discourse in distinct moments in the race. Also, if it is possible to have speeches from events that received large coverage, we are looking at those which have the most potential to be heard by the largest number of voters. If a politician is to use a populist discourse, these are the moments when she would most likely have been heard. Furthermore, while not all countries have a tradition of parties holding large rallies to end the

---

[2]If we used speeches in parliament, for example, new parties would be excluded.

7

campaign, most hold party conventions where the leading candidate is acclaimed. Limiting the number of speeches to two is dictated by a practical reason: it is very difficult to find more than two campaign speeches for several candidates. [3]

In terms of coverage, the sample includes 144 parties from 27 countries in the Americas and Western Europe. The selection of countries was partly dictated by convenience: we had to include those for which there were coders who spoke the language. This was less of a problem in the Americas: most of South America plus the whole North America are included. In Western Europe the sample is more limited, but we could not identify any evident biases: there are countries where populism is often said to be high, and others where it is usually off the radar. Also, there are both Southern and Northern countries. What are completely absent, though, are post-socialist Central and Eastern European cases. For half the countries in our sample, parties were included if they got more than 1% of the vote in the national election of interest. In the other half, the cutoff was 5%. The reasons were practical: first the availability of documents, which in some cases could not be found for the smallest parties; and second, the amount of resources and the number of coders. For instance, a 5% line in Sweden already included seven parties[4]. The manifestos and speeches all come from the most recent national elections up to March 2015 in which

---

[3]In Hawkins [2009] it was suggested that three to four speeches were enough for a reliable grade. However, there a politician's discourse was studied for all her time in office. Because we are limiting it to how populist are political campaigns – shorter in time –, it may be expected that there is less variance, and fewer speeches may be needed.

[4]Countries where the 1% rule applied: Belgium (vote shares considered for each electoral college), Brazil, Canada, Colombia, Germany (Pirates, with two percent of list votes, were not included), Mexico, Peru, Spain, United Kingdom, Uruguay, United States, and Venezuela. Countries with a 5% cutoff: Argentina, Austria, Bolivia, Chile, Ecuador, France, Greece, Italy, Netherlands, Paraguay, Portugal, Sweden, and Switzerland. In some of these, one or more smaller parties were still included, when sources were available. From Norway, only the three largest parties are present in the sample.

the chief executive was defined[5]. Regarding documents, we had party manifestos for 142 of the 144 parties in our sample[6]. The coverage of speeches is smaller. We have not coded speeches for nine countries[7], what brings the total of texts coded to 307.

3   Description of results

3.1   Manifestos versus speeches

This is the first time that holistic grading has been applied on such a large scale to party manifestos, and some issues of adapting it to this kind of document emerged. First, as coders started to report results, many indicated that there were two very different tones in some manifestos, where the preamble, or introduction, contained high levels of populism, while the rest (always a list of policy proposals) had a more pragmatic or technical feel. We decided to ask coders to give separate scores for the preamble/introduction, where it existed, and the list of policy positions. The mean level of populism in preambles is 0.32, while that of the list of positions is 0.26. The scores for all parties coded are found in Appendix A, in Table 4. In it, the Manifesto column is a simple mean of the preamble and the list of issues scores. [8] Because the preamble is always shorter than the list of positions [9], the net result is to weight the

---

[5]With the exception of Canada, where documents from 2006 were used, Austria, where we coded the 2008 legislative election, and Spain, where we included the most recent December 2015 elections.

[6]Greek parties did not produce proper electoral manifestos for the January 2015 elections. However, they all had editorials which outlined party policies, and these documents were used as the most comparable we had to manifestos.

[7]Austria, Belgium, Bolivia, Canada, Germany, Ireland, Netherlands, Norway, and Switzerland.

[8]This issue also emerged in Rooduijn et al. [2014], and the authors decided to count each paragraph of the preamble twice.

[9]The length of manifestos ranged from 4 pages, from the German *Alternative für Deutschland* (AfD), to 810, from the Walloon *Ecolo*. The length of preambles span from two paragraphs to five

preamble more heavily.

The results confirm an intuitive expectation: manifestos are less populist than speeches. Given their nature as formal party documents for elite consumption, it is not surprising that the tone is more sober. The mean grade for manifestos is 0.29, while that of speeches is 0.47, with no difference between speeches from the beginning and end of the campaign (means of 0.47 and 0.46, respectively). In categorical terms, this means the average campaign speech is almost at the 0.5 threshold that indicates the presence of necessary elements of populism, but weak or used inconsistently throughout the speech. This shows that populism in political campaigns might not be dominant, but still has a non-negligible presence.

The correlation between speeches' and manifestos' final scores is presented in Figure 1, on the left-hand side. It is 0.59, and there are only two cases of parties where one kind of document received a categorical 2 (a decimal score equal or above 1.5) and the other a categorical 0 (a decimal below 0.5). These are the *Rivoluzione Civile* (RCI), in Italy, with an average for speeches of its leader, Antonio Ingroia, of 1.5, while the manifesto scored 0.4, and the *Solidaridad Nacional* (SN), from Peru, whose manifesto received a 0.0 and the candidate's speech 1.6 [10]. These results indicate that, when possible, it is ideal to have both manifestos and speeches coded to give a more complete picture of how populist a party is but, in the absence of

_____

pages.

[10]Both might be explained by two shortcomings in the data available from these parties. The RCI manifesto was one of the few that had no preamble, only a list of policy positions, what may have contributed to its lower score. For the SN, only one speech from its presidential candidate, Luis "Lucho" Castañeda Lossio, was found and coded, and the version of the speech had been edited before it was made available. This score, therefore, is more sensitive to the possibility of a single non-representative speech or of non-representative fragments that have been kept.

speeches, manifestos still give a reasonable approximation.

Figure 1 shows a few parties close to the top, some more or less scattered in the middle, and many in the lower-left corner. Focusing on the specific parties, we see that those on the top are PSUV[11] – Chávez' party in Venezuela –, two other Latin American ones that closely follow his discourse – Alianza PAIS, in Ecuador, and Partido Igualdad, in Chile, and the Greek SYRIZA. These are the only parties where it is possible to find the use of very strong populism consistently across different kinds of texts, and also give an informal validation check to the results: there is little discussion that these four parties should be among the top in any populism classification.

3.2   Intercoder reliability

After the coding efforts in Hawkins [2009, 2013] showed high intercoder reliability, it seemed possible to have only one coder doing some cases, in order to increase the number of countries and parties covered. Part of this sample, therefore, is based on the grades assigned by only one coder. For the other part (99 documents out of a total of 308 in total), two were kept, and the results of intercoder reliability checks confirm that the method is reliable. Krippendorff's alpha is high, 0.87, showing that using only one coder for part of the sample should not bring major measurement errors. Figure 2 shows the correlations between scores assigned by two coders to each document, with dot sizes representing the number of documents in each point coordinate.

---

[11]The list of abbreviations may be found in Appendix B

Figure 1: Speeches and manifestos scores

Party manifestos

Candidates speeches

a Americas

a Europe

2.0

1.5

1.0

0.5

0.0

0.0   0.5   1.0   1.5   2.0

IGUAL
PSUV
Pais
SYRIZA
UCR
UP
IU−UP PSOL
DI
PRD
FG_FR
MUD
PS_CHI
PCP Podemos
CREO PD
FpV
PNP BNP ANEL
PRD TAMIND FN
BE
KKE RCI
CD
PRO
LabR_UK
FABIAN
FP AP_PY
PSP M5S
CF
PDA
Parisi PTP
SN

r = 0.59, n = 84

12
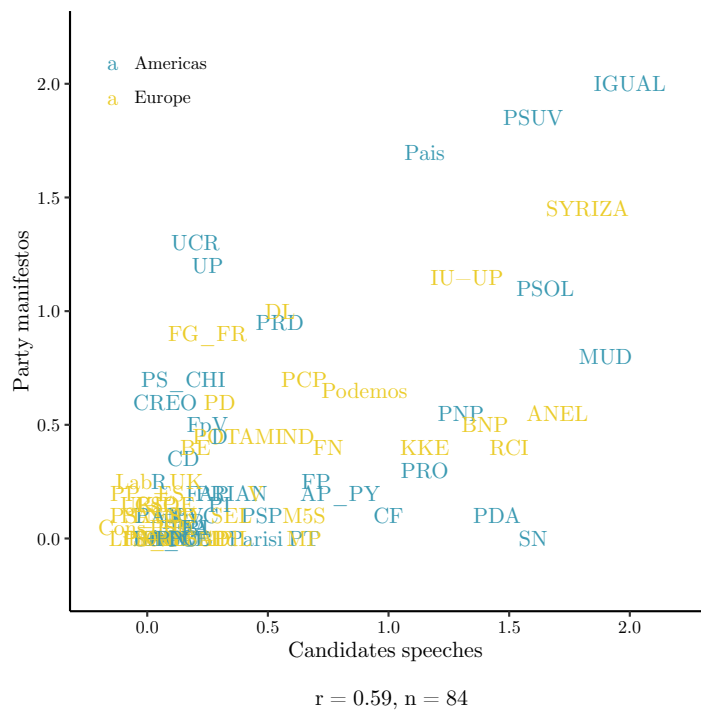
Figure 2: Correlation between coders' scores

Krippendorff's alpha = 0.87, n = 99
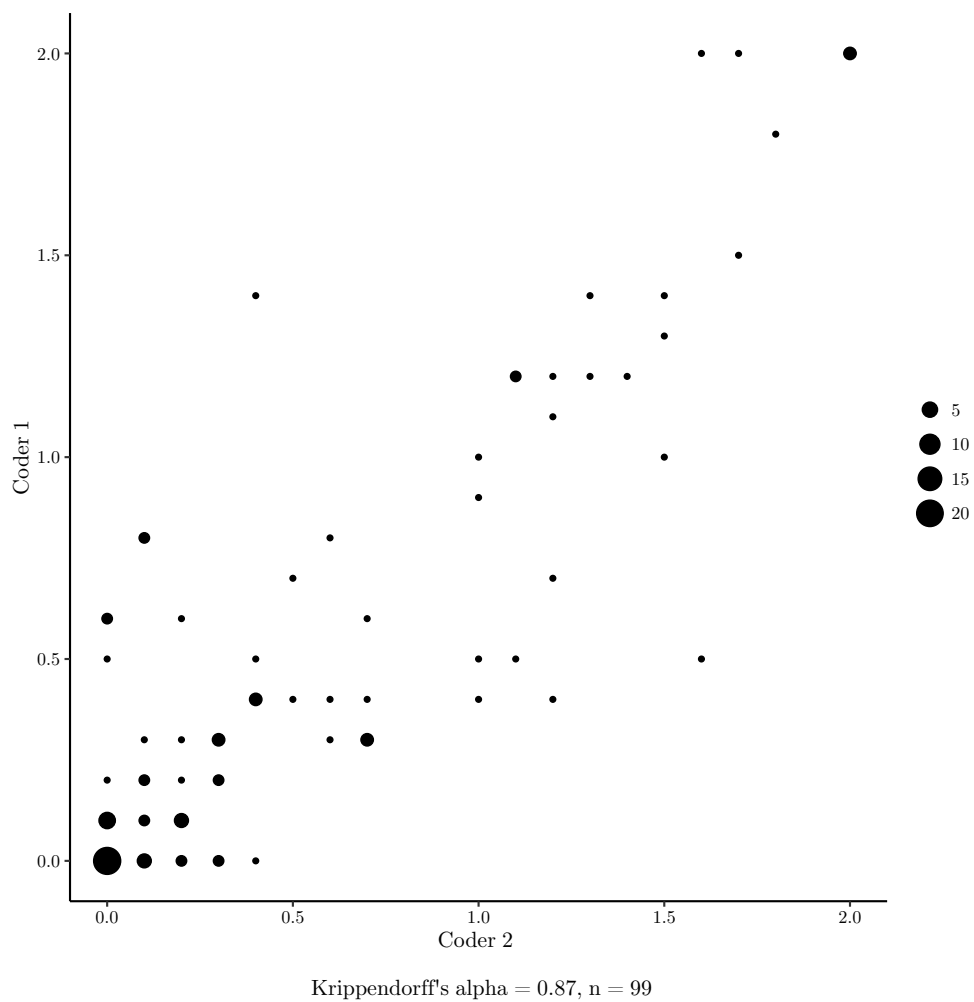
4 Comparing results

4.1 Regional differences

Table 1 indicates all parties whose manifestos surpassed the categorical threshold of 0.5, indicating there is a non-negligible component of populism in their discourse, even if not consistently used. The four most populist parties identified, all of which received a grade above 1.5, are from Latin America. They are the *Partido Socialista Unido de Venezuela* (PSUV), founded by the late Hugo Chávez; the *Movimiento al Socialismo*, Evo Morales' party in Bolivia; the *Alianza Pais*, from president Rafael Correa of Ecuador; and a small party from Chile, *Partido Igualdad*, whose leader in the 2013 presidential campaign, Roxana Miranda, followed closely the radical-left ideology of the previous three. Bolivarianism is found here to be strongly associated with populism: of the three Latin American parties that have scores between 1.0 and 1.5, two hold a similar ideology as well. The one exception is the *Union Cívica Radical*, a center-left Argentinean party, not usually associated with populism.

As the absence of cases in the top-right box hints, populism is stronger in Latin America than in Europe. While we identified 13 parties with a grade higher than 0.5 in Latin America, meaning 28.9% of the regional sample, there were only 16 European ones in this range, out of 92, or 17.4%. Therefore, not only the highest levels of populism are found in Latin America, the proportion of populist parties to the total in the region is substantively larger than in Europe. This point is made in Figure 3, which compares the means of populism in parties in the two regions. The dark blue bars are unweighted mean scores of populism, while the lighter blue

Table 1: Populist manifestos in Europe and Americas

	Americas	**Europe**
Party score ≥ 1.5	Partido Igualdad, Chile (2.0) Partido Socialista Unido de Venezuela, Venezuela (1.85) Alianza Patria Altiva i Soberana, Ecuador (1.7) Movimiento al Socialismo, Bolivia (1.55)	
Party score ≥ 1.0	Unión Cívica Radical, Argentina (1.3) Unidad Popular, Uruguay (1.2) Partido Socialismo e Liberdade, Brazil (1.1)	Coalition of the Radical Left (SYRIZA), Greece (1.45) Nationaldemokratische Partei Deutschlands, Germany (1.4) Die Linke, Germany (1.3) Partij voor de Vrijheid, Netherlands (1.25) Izquierda Unida-Unidad Popular, Spain (1.15) Schweizerische Volkspartei, Switzerland (1.0) Freiheitliche Partei Österreichs, Austria (1.0) Democràcia i Llibertat, Spain (1.0)
Party score ≥ 0.5	Partido de la Revolución Democrática, Mexico (0.95) Mesa de la Unidad Democratica, Venezuela (0.8) Conservative Party, Canada (0.8) Bloc Québécois, Canada (0.75) Partido Socialista, Chile (0.7) Creando Oportunidades, Ecuador (0.6) Partido Nacionalista Peruano, Peru (0.55) Frente para la Victória, Argentina (0.5)	Front de Gauche, France (0.9) Esquerra Republicana de Catalunya, Spain (0.8) Partido Comunista Português, Portugal (0.7) Podemos, Spain (0.65) Partito Democratico, Italy (0.6) Independent Greeks (ANEL), Greece (0.55) Parti Populaire, Belgium-WAL (0.55) British National Party, UK (0.5)

Party score refers to the electoral manifesto score.

present scores weighed by parties' vote-share. The two panels show Latin American averages higher than European ones, as well as the difference between speeches and manifestos. What the figures also show is that, not only are Latin American parties on average more populist than Europeans, but populism in Latin America is used by electorally stronger parties – when we weigh in the vote-shares, the level of populism in manifestos goes up. In other words, populism is a more mainstream discourse in Latin America.

This pattern is an important finding in itself. While this has not been openly expressed or argued for, the common assumption in studies comparing populist parties in both regions is that they are populist to the same extent. Mudde and Kaltwasser [2012], for example, study the difference between left and right populism, concluding that one is inclusionary (left) while the other is exclusionary (right). They select typical cases of each, which are the French National Front and the Freedom Party of Austria for right-wing populism, and Evo Morales and Hugo Chávez for left-wing. The underlying assumption is that the only thing that differentiates these movements, at least in terms of their ideas, is their "thick" ideology – populism is treated as a constant. What we see here, however, is that the National Front, with a score for its manifesto of 0.4, is not nearly as populist as Morales' MAS (1.55) or Chávez's PSUV (1.85), while the Freedom Party of Austria might come closer (1.0) but is still one category below the two Latin American ones.

Figure 3: Levels of populism by region
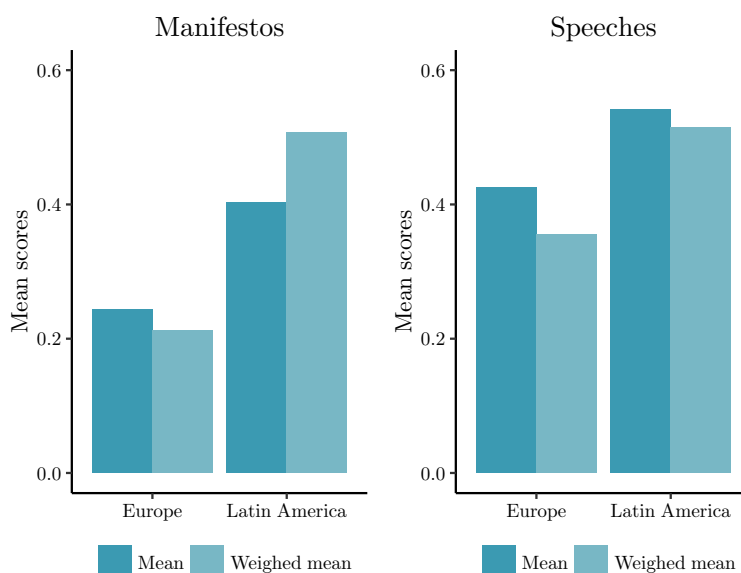
5 Automating

Automated content analysis has been gaining traction in the social sciences, to measure the most varied concepts with the assistance of modeling methods and large bodies of data. Considering that what we present in this paper is, to date, the largest classification of political documents on how populist they are, it seemed appropriate to test whether it is possible to train a model that is able to reproduce the results we obtained with human coding.

In the attempt of capturing populism through computerized content analysis, we use machine learning techniques to perform supervised classification of documents.[12] This kind of approach involves three steps: first, splitting the sample into a training

---

[12]See Grimmer and Stewart [2013] for a comprehensive assessment of automated text analysis methods.

and a test set; second, use the training set to develop a model that predicts classification scores in this subsample; and third, apply the model to the test set and compute the error rate, to see if results are acceptable. There are different ways of calculating error, but all of them assume that the closest we can get to the "true scores" of a text is the human classification. Therefore, each time the computer classifies a document with a different grade than human coders, it counts as an error. There are multiple techniques that can be used in the second step, to train the model, about which we talk in more detail below.

5.1   Preparing the data

We report here the application of models to manifestos alone. A large part of the speeches used for coding in this paper is not available in text, but only in video or audio. We had only 72 transcripts, a majority of which (57) categorical 0's with no populism in them, and this number turned out to be too small for the methods applied here. [13] For the manifestos analysis, we include also the anchor documents used in the training of coders. Our total sample size is 154.[14]

The corpus was prepared in the following way: first, all documents that were in pdf format – most manifestos – were automatically converted to plain text files. A review of the results shows that formatting was obviously lost, and in some cases the order of pages was shuffled. However, this is not a problem because the methods

---

[13]We did perform all the same analyses as those reported for the manifestos, but results showed no method that classified speeches much better than a coin toss would. At this point, all that can be said is that more coded speeches would be necessary before judging any results of automation to classifying them.

[14]The editorials used as manifestos in Greece are not included.

applied here use the "bag of words" assumption, meaning that the order of words in a text is not taken into consideration, but only their individual frequency. [15] The texts were then translated into English with Google Translator, and an inspection of the output shows satisfactory results. [16]

After that, the documents are pre-processed to make them more practical for computerized text analysis. The procedures applied are standard: turning all words to lower case, removing punctuation, numbers, and stopwords (such as prepositions and articles), stemming words, and removing unnecessary space (meaning, all else than a single space between each word). The documents are then transformed into a document-term matrix, where each row is one document, and each column is one *unigram* (a word or stem) that appears at least once in the whole corpus. Cells are filled with the number of times that the unigram i appeared in text t. Sparse terms are then removed with a .7 cutoff, meaning that words present in less than 30% of the texts were removed, and common terms, present in more than 75% of the texts, were also cut. This reduces the demand for computational power and reduces noise, by removing information that is unlikely to help in classifying. In fact, we tested different levels of ratios for sparsity – including no removal of common or sparse words –, and the numbers used presented better performance than alternative proportions of removal for sparse or common terms. All transformations were done using the 'tm' package for R [Feinerer et al., 2008, Feinerer and Hornik, 2015].

---

[15]Issues identified with the conversion were: words divided at the end of lines were not rejoined; in some cases titles and graphical words/sentences were not properly recognized. These problems, however, were few and there is no reason to expect them to be systematic and bias our results.

[16]This choice is due to the fact that automatic translators are usually considered to work better with English as a target, one of the reasons being that for many pairs of languages the translation is not done directly, but through English.

Our sample presents a couple of major challenges for automated text analysis: first, it is relatively small, compared to common applications of classifiers to Big Data. Second, the distribution is very skewed towards low populism. Only four of the manifestos coded for this paper received a categorical 2 – meaning high and consistent populism –, and there were four more 2's in the training set. 27 of the 154 are categorical 1's, and all the rest, more than 75% of the sample, is composed of 0's. For this reason, we coupled 1's and 2's into the same category, making populism a binary variable, what may give the models more information to identify populist texts.

5.2 Methods

We employ a penalized least squares method with an elastic net penalty [Zou and Hastie, 2005]. It combines two kinds of regularized regression, the ridge [Hoe, 1970], and the lasso [Tibshirani, 1996]. These methods have in common that they shrink regression coefficients, following the value of a certain parameter λ, so that those with lower predictive power are penalized. The difference between the ridge and the lasso is that the first only reduces the value of smaller coefficients, while the second effectively forces some (the number of which depending on λ) to zero, in effect proceeding with variable selection, removing from the model those with lower predictive power.[17] While the lasso is attractive for text analysis where there are thousands of variables (each word being one), given the small number of texts we have, we run into a limitation where the number of observations is smaller than the

---

[17]No one outperforms the other absolutely, and each has its advantages over specific kinds of data [James et al., 2013]

number of variables. To solve this issue of $n < p$, Zou and Hastie [2005] propose the elastic net, which combines both ridge and lasso regressions through an α coefficient that balances the two. The formula that is minimized in elastic net regression is

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} (\alpha \beta_j^2 + (1-\alpha)|\beta_j|) \tag{1}$$

where the left-hand side is the residual sum of squares, which is minimized in common linear regression. The right hand side adds the elastic net penalizer. β_j^2 refers to the ridge penalizer, whereby smaller β are shrunk towards zero by squaring, while $|\beta_j|$ is the lasso penalizer. α indicates the proportion of each one in the final estimation. An α of 1 excludes the lasso, and makes it a ridge regression. Conversely, an alpha of 0 removes the ridge term and makes it a lasso. For $0 < \alpha < 1$, a combination of both is used. The value of λ will determine the weight of this whole term on the estimation, so that larger values increase the penalties, and smaller values approach the final estimation to that obtained with OLS regression. Our models are run as a binomial regression, classifying documents as 1's or 0's.

To tune the model and reduce overfitting, we use k-fold cross-validation with $n/2$ folds of size $n-2$.[18] This technique involves sequentially splitting the sample into a subsample with 2 documents as the test set, and the remaining $n-2$ as the training set. This is applied $n/2$ times, so that every document has been once and only once on the test set. In this application, several values of λ and α are tested, in order to identify the optimal that leads to the lowest mean error. The optimal model selected

---

[18]The minimal mean misclassification rate is stable across 10-fold, 5-fold, and leave-one-out cross validation, at around 0.12–0.14. The models are fit using the R package glmnet [Friedman et al., 2010].

is that which gives the minimal mean cross-validation misclassification rate. This is the amount of times that the predicted value for each document in the test sets ($n = 2$ in each) is different from that assigned by coders.

Elastic net regression is expected to be the best fitting model for our data, because it can deal both with sparse matrices and with the problem of having more variables than cases. Further, it returns an interpretable model and we can see what words are attributed a higher coefficient for being more predictive of populism. This helps validating the results. However, following the saying that in automated text classification no model is good, but some work, we also apply other classifiers to our data: support vector machine (SVM) [Karatzoglou et al., 2006], classification tree with random forests [Liaw and Wiener, 2002, Ripley, 2015], logistic boosting [Friedman et al., 2000], and scaled linear discriminant analysis [Läuter et al., 1998, Peters and Hothorn, 2015].[19] We apply leave-one-out-cross validation for each one to obtain the best predictive model for all classifiers, and use Krippendorff's alpha [Krippendorff, 2004] between each individual classifier and the manual code to evaluate the quality of results.[20]

Table 2: Accuracy of elastic net regression classification

		Classifier	
		0	1
Manual	0	115	4
	1	19	16
Error rate:		0.149	
K. alpha		0.493	

5.3 Results

Table 2 shows the results of elastic net regression classification with $\lambda = 0.008444$ and $\alpha = 0.5$, identified in the cross-validation as the optimal values for these two hyperparameters. The misclassification rate, which is the proportion of documents miscoded by the classifier in relation to the original hand-coding, is 0.149. This is a low number, indicating that the classifier got 85% of the scores correctly. Krippendorff's alpha, however, shows that the classifier did not perform so well: it is only 0.493. The low rate is caused by the classifier attributing too many 0's – while these were a majority in the data, and thus it was often right, the elastic net regression had difficulties in identifying populist manifestos. Of the 35 populist documents, it coded correctly only 16. However, a further look into the texts shows an interesting feature on the breakdown of these numbers: if we compare the performance based on the original scores, the classifier identified correctly 6 out of the 8 manifestos originally coded as 2 (very populist). Of those "in-between", which had an original

---

[19]For these methods with use off-the-shelf software with default settings from the R package 'RTextTools' [Jurka et al., 2014].

[20]Similar to evaluating intercoder reliability, there is not a consensual critical value, but alphas around or above 0.8 are safely considered to indicate a reliable coding, and higher than 0.7 could be accepted. In this case we should look for something similar: if there is a high alpha, it means that the agreement between the automated method and the manual scores is comparable to the agreement level we would consider acceptable if two human coders were doing it.

score of 1, it only coded as populist 10 out of 27. In sum, this classifier was efficient in identifying very non-populist and very populist manifestos, but did a poor job with those that, following the description, "include strong, clearly populist elements but [...] not consistently".

The elastic net variable reduction resulted in 128 non-zero coefficients. Figure 4 plots all these 128 stems, aligned according to the size of their coefficient. Those below zero (blue) are predictive of classification as not-populist (0), while those above zero (red) are predictive of populism (1). The models seem to be valid in capturing this concept: we can identify among the largest coefficients, for instance, *elit, domin, destruct, reconstruct, struggl, lie, privileg* and others that might indicate a belligerent kind of speech against an elite. Also, there are some which may be seen as references to "the people", such as *mass, referendum, street, sovereignti, popular*. Another interesting point is the presence of *IMF*, referring to one of the most common targets of Latin American populists. On the non-populist side, there seem to be more stems referring to material issues, such as *countrysid, bureaucrat, recess, viabl, poorest, outlin, percentag, fossil, brand, tuition*.

Figure 5 shows the value for Krippendorff's alpha calculated between the results obtained with the other classifiers and manual codes. One of the alternative classifiers in fact did slightly better than the elastic net: logit boosting, with a kalpha of 0.506. This method starts by fitting a model, checking what observations were wrongly coded, and then reassigns weights to all observations so that those miscoded have a higher weight, and another model is fit. This process is repeated until an optimal model which minimizes classification error is accepted. The other methods

Figure 4: Words predictive of populist and not-populist classification
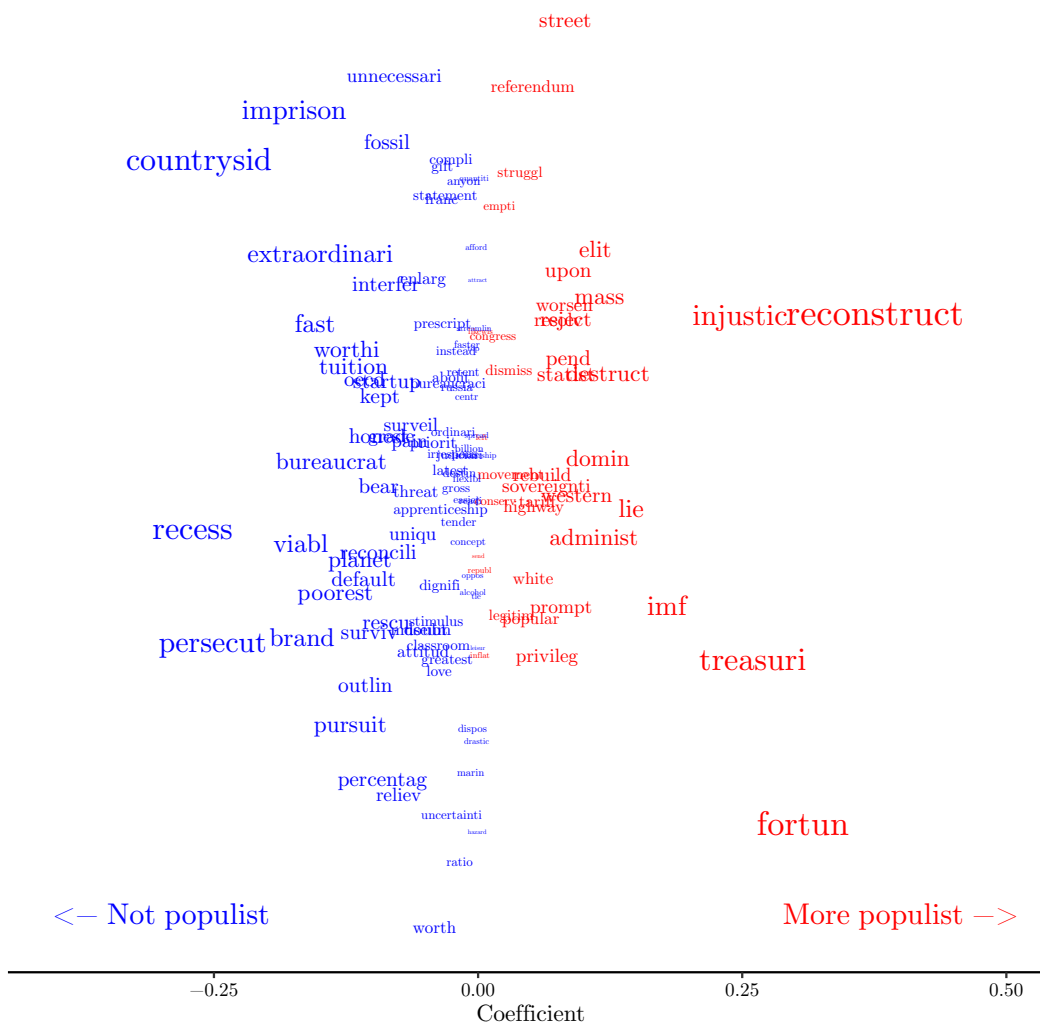
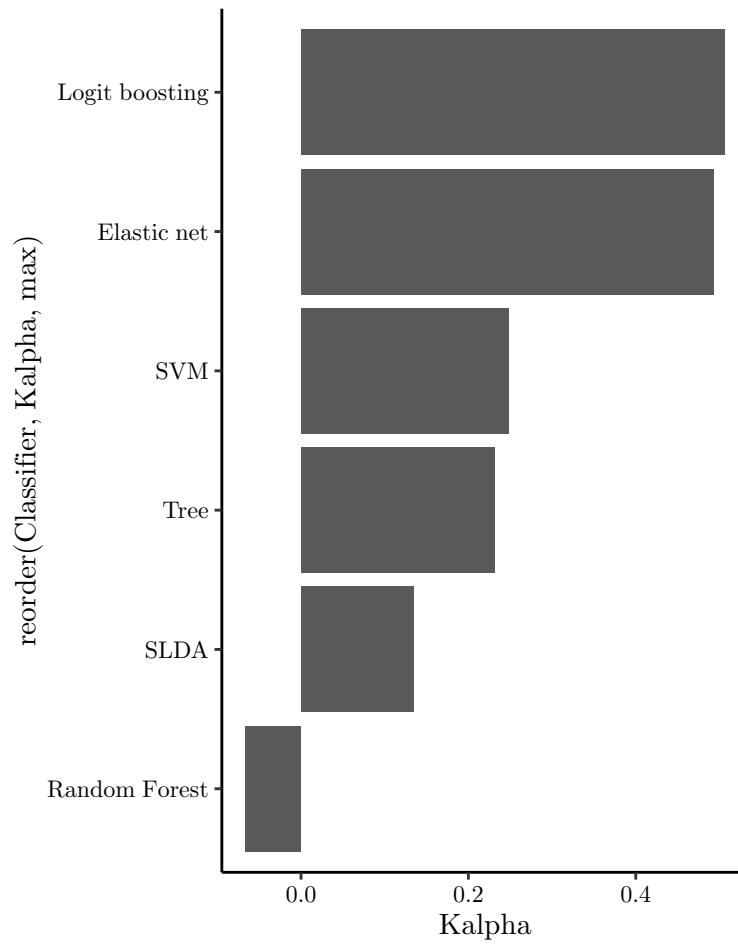Figure 5: Krippendorff's alpha for each individual classifier

Table 3: Consensus results of elastic net regression and boosting classifiers

		Consensus	
		0	1
Manual	0	111	1
	1	12	6
	2	1	5

Krippendorff's alpha: 0.553
Misclassification rate: 0.103
Coverage: 0.88

performed poorly, not much better than random guessing. This should be expected: SLDA, for instance, assumes multivariate normal data, what is known not to be the case in our sample. Classification trees and random forests work better when there are a few strong and accurate predictors of the outcome. The stems that predict classification as 1 are likely to be present in manifestos coded as 0 as well, even if not so common. The differences between categories are nuanced, and unlikely to be captured accurately by individual stems or specific combinations of them.

Combining the results of the two best performing methods gives a small improvement in accuracy. Of the original 154 documents, elastic net and boosting agreed on 136 classifications, or 88% of the sample. The results of their consensus scores are in Table 3. Krippendorff's alpha is .553, above any of the two alone. We see that the problem with "in between" cases persist. The two are very accurate with identifying non-populist manifestos, and did well with those that were original 2's. Of the original 1's, however, 12 were seen as not-populist, and only 6 as populist.

6 Discussion

This paper started by introducing what is, up to date, the most comprehensive data set of political parties classified by how populist their discourses are. We applied a technique of holistic grading to party manifestos and to candidate speeches from Western Europe and the Americas, to see how much populism they displayed in these texts. From the methodological perspective, it was shown that the method can transition to manifestos – a novelty in its application – and that it is possible to use only one coder if resources are scarce, since intercoder reliability proved to be high. We also have presented results on automated classification of these documents, and shown that computerized text analysis could potentially be successful in identifying populism if there is enough data for training models.

The classification data on their own are a relevant contribution that may be used for a wide array of future research. Models trying to explain support for populist parties, for instance, have up to now mostly relied on dichotomous divisions based on literature reviews or experts' classifications of parties [see, for instance, Bustikova, 2014, Doyle, 2011, Remmer, 2011]. It is now possible to model these preferences based on data on parties' levels of populism in comparative perspective, across different regions, derived directly from party communication and coded using a single definition of the term. This reduces measurement error, increases precision (by using a continuous measurement instead of a dichotomous one), and increases the comparability of results. If country experts overestimate the level of populism in parties they know well, as do those who include the Sweden Democrats in lists of populists, for example, this may now be corrected by having a scale that puts parties into fully

28

comparative perspective.

A clear regional cut stands out from the results. While there are populists in both regions, the level of populist discourse found among Latin American parties and politicians is much higher than that found among their European counterparts. A prototypical European populist, the National Front, had a score of 0.4 for their manifesto, and Marine Le Pen's 2012 campaign speeches received an average of 0.75 out of 2. Other typical European populists did not fare much higher – Beppe Grillo's speeches averaged 0.65, Berlusconi's 0.35, and Nigel Farage's a round zero. The Sweden Democrats' score, whose recent electoral success has spurred much debate in Europe about the rise of yet another radical right populist party, shows that it might be radical right, but it is definitely not populist – it is not even the most populist party in Sweden. This does not mean populism is not found in all cases expected. Nick Griffin and the British National Party do get moderately high scores, as does the German extreme-right NPD, and Geert Wilders' PVV in the Netherlands. However, if we isolate only the populist portion of their discourse, they are still not as radically populist as a couple of Latin American examples.

A skeptical reader might say that low scores for some parties may be a result of poor selection of texts, rather than parties' lack of populism. It is possible to find quotes by Nigel Farage, for example, which sound very populist. Two counterarguments may be offered to this. First, while three documents might not be ideal, and there may be large variation in a politician's discourse, the odds of all three being the few non-populist exceptions from a very populist candidate are quite small. As we have seen, strongly populist actors have this discourse even in the list of policy

proposals of their party manifestos. Second, our selection of speeches has a partial bias for high profile speeches: the opening and closing of campaigns. The simple fact that speeches were available online often indicates that they received at least some attention, and were not obscure talks to small audiences. Therefore, we are capturing political discourse in moments when it has a better chance of being heard by the public. If populism is to matter in a politician's discourse, those are the right times. It might be that a politician's tone is more populist to her partisans than to the general public, but then the question goes to another level: is intra-party populism more relevant than the part of the party's discourse made for mass consumption? If our larger concern is with the causes of support for, and consequences of populism at the society in which it is embedded, the answer is probably negative.

These findings have important consequences for future research comparing populist experiences in the two regions, a topic which is currently flourishing. Knowing that Latin American populists are more radical in their populism than Europeans may have implications for explaining their support, as well as for the parties' and politicians' actions in office. For example, this might be one of the keys for explaining the openly anti-democratic actions taken by some Latin American populists in recent years [see Levitsky and Loxton, 2013, Huber and Schimpf, 2015]. Such a high degree of populism comes together, by definition, with a high level of demonization of the opposition, which is used to justify their persecution as illegitimate actors. At the same time, it reinforces the argument made by Mudde and others [Mudde, 2013, 2014] that public fears about the rise of radical right populism in Europe may be exaggerated. Not only they have not been as successful as one might think from

30

reading the news but, in general, they are not as populist as generally thought. While concerns may be raised over their radical ideology, their populism may not be as important an issue.

If these findings may call for a change in how we see populism in Europe, they are also of practical concern for the state of Latin American politics. While for some it is sobering to see that Marine Le Pen is far from being as populist as Evo Morales or Nicolas Maduro, the fact that her score is similar to that of moderate Latin American leaders, who are not usually associated with populism, shows just how deeply ingrained into the region's political culture this kind of discourse is. In this paper, the European "surprises" were mostly cases that were expected to be very populist and turned out not to be. In Latin America, the "surprises" were rather parties that unexpectedly had moderate or high levels of populism (even rising above prime European examples) and are commonly not thought to be so. For example, the Peruvian President, Ollanta Humala, who is usually seen as having dramatically moderated his tone during his successful electoral bid in 2011. Or Henrique Capriles, leader of the opposition to Chavismo in Venezuela and who, as our findings indicate, has adopted much of his opponent's populist discourse for his own campaign in 2013. This observation, coupled with the finding that populism today, in Latin America, is the language of stronger parties, gives cause for concerns over how its developing democracies will keep on dealing with the divisive and anti-conciliatory aspects of such a discourse.

6.1 An automated future in sight?

The results observed with machine learning classifiers should give a moderate optimism on the possibility of automating the coding of manifestos for populism. The sample at hand is not one where classifiers would be expected to excel: there are relatively few documents, these were long – sometimes a few hundred pages –, from a variety of countries and languages, all converted automatically into English. Moreover, the concept we try to capture is known to be elusive and subtle. To make matters more complicated, when applying holistic grading to manifestos, it is often the case that populism can be found only in a few important paragraphs. That, for a human coder interpreting the context, would be enough to say that the text has indeed elements of populism. However, for an algorithm looking for words with which to do the classification, that is just a small fraction of the total amount of information and may easily go unnoticed.

Given these difficulties, elastic net regression and logit boosting might be said to have gone beyond expectations, showing an ability to reliably recognize manifestos that are clearly non-populist or clearly populist. Where they do not perform well are those in which human coders saw the necessary elements of populism, but used inconsistently and not so strongly. At this point, there are two ways of explaining the inaccuracy of classifiers with this category. First, it may be that this is the limit of automated classification: populism in these texts is such a subtle topic that it requires human interpretation and contextual information to be identified, and there is no counting of words that can replace that. Or second, it might just be a matter of adding more data. With a large enough sample of manifestos, the training models

would be able to capture the small differences that make for a 1 manifesto. However, knowing which one of the two is correct is beyond the possibilities in this paper.

References

Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12 (1):55–67, 1970.

L. Bustikova. Revenge of the Radical Right. *Comparative Political Studies*, 47(12): 1738–1765, February 2014. ISSN 0010-4140. doi: 10.1177/0010414013516069.

D. Doyle. The Legitimacy of Political Institutions: Explaining Contemporary Populism in Latin America. *Comparative Political Studies*, 44(11):1447–1473, May 2011. ISSN 0010-4140. doi: 10.1177/0010414011407469.

Ingo Feinerer and Kurt Hornik. *tm: Text Mining Package*, 2015. URL http://CRAN.R-project.org/package=tm. R package version 0.6-2.

Ingo Feinerer, Kurt Hornik, and David Meyer. Text mining infrastructure in r. *Journal of Statistical Software*, 25(5):1–54, 2008. URL http://www.jstatsoft.org/v25/i05/.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28(2):337–407, 2000.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for

generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

Justin Grimmer and Brandon M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3): 267–297, 2013.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Verlag, New York, 2009.

Kirk Hawkins. Is Chavez Populist?: Measuring Populist Discourse in Comparative Perspective. *Comparative Political Studies*, 42(8):1040–1067, February 2009. ISSN 0010-4140. doi: 10.1177/0010414009331721.

Kirk Hawkins. Measuring Populism in Comparative Perspective. In *XXXI International Congress of the Latin American Studies Association, May 29 – June 1, Washington D.C.*, 2013.

Robert a. Huber and Christian H. Schimpf. Friend or Foe? Testing the Influence of Populism on Democratic Quality in Latin America. *Political Studies*, pages 1–18, 2015. ISSN 00323217. doi: 10.1111/1467-9248.12219.

Jan Jagers and Stefaan Walgrave. Populism as Political Communication Style: An Empirical Study of Political Parties' Discourse in Belgium. *European Journal of Political Research*, 46(3):319–345, May 2007. ISSN 0304-4130. doi: 10.1111/j. 1475-6765.2006.00690.x.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer, New York, 2013.

Timothy P. Jurka, Loren Collingwood, Amber E. Boydstun, Emiliano Grossman, and Wouter van Atteveldt. *RTextTools: Automatic Text Classification via Supervised Learning*, 2014. URL https://CRAN.R-project.org/package=RTextTools. R package version 1.4.2.

Alexandros Karatzoglou, David Meyer, and Kurt Hornik. Support Vector Algorithm in R. *Journal of Statistical Software*, 15(9):1–28, 2006.

Klaus Krippendorff. *Content Analysis: an Introduction to Its Methodology*. Sage, Thousand Oaks, CA, 2 edition, 2004.

Jürgen Läuter, Ekkehard Glimm, and Siegfried Kropf. Multivariate Tests Based on Left-Spherically Distributed Linear Scores. *The Annals of Statistics*, 26(5): 1972–1988, 1998.

Steven Levitsky and James Loxton. Populism and competitive authoritarianism in the Andes. *Democratization*, 20(1):107–136, 2013.

Andy Liaw and Matthew Wiener. Classification and Regression by randomForest. *R news*, 2(3):18–22, 2002.

David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2015. URL https://CRAN.R-project.org/package=e1071. R package version 1.6-7.

Cas Mudde. *Populist Radical Right Parties in Europe*. Cambridge University Press, Cambridge, 2007.

Cas Mudde. Three decades of populist radical right parties in Western Europe: So what? *European Journal of Political Research*, 52(1):1–19, January 2013. ISSN 03044130. doi: 10.1111/j.1475-6765.2012.02065.x. URL http://doi.wiley.com/10.1111/j.1475-6765.2012.02065.x.

Cas Mudde. The Far Right and the European Elections. *Current History*, 113(761): 98–103, 2014.

Cas Mudde and Cristóbal Rovira Kaltwasser. Exclusionary vs. Inclusionary Populism: Comparing Contemporary Europe and Latin America. *Government and Opposition*, 48(2):147–74, December 2012. ISSN 0017-257X. doi: 10.1017/gov. 2012.11.

Cas Mudde and Cristóbal Rovira Kaltwasser. Populism. In Michael Freeden, Lyman Tower Sargent, and Marc Stears, editors, *Oxford Handbook of Political Ideologies*, number January, pages 493–512. Oxford University Press, Oxford, 2013.

Andrea Peters and Torsten Hothorn. *ipred: Improved Predictors*, 2015. URL https://CRAN.R-project.org/package=ipred. R package version 0.9-5.

Mario E. Poblete. Review article: How to assess populist discourse through three current approaches. *Journal of Political Ideologies*, (May):1–18, 2015. doi: 10. 1080/13569317.2015.1034465.

K. L. Remmer. The Rise of Leftist- Populist Governance in Latin America: The Roots of Electoral Change. *Comparative Political Studies*, 45(8):947–972, December 2011. ISSN 0010-4140. doi: 10.1177/0010414011428595.

Brian Ripley. *tree: Classification and Regression*. R Package version 1.0-36, 2015. URL https://cran.r-project.org/web/packages/tree/tree.pdf.

M. Rooduijn, S. L. de Lange, and W. van der Brug. A populist Zeitgeist? Programmatic contagion by populist parties in Western Europe. *Party Politics*, 20 (4):563–575, April 2014. ISSN 1354-0688. doi: 10.1177/1354068811436065.

Matthijs Rooduijn and Teun Pauwels. Measuring Populism: Comparing Two Methods of Content Analysis. *West European Politics*, 34(6):1272–1283, 2011.

R. R. Sudweeks, S. Reeve, and W. S. Bradshaw. A Comparison of Generalizability Theory and Many-Facet Rasch Measurement in an Analysis of College Sophomore Writing. *Assessing Writing*, 9(3):239–61, 2004.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–88, 1996.

Edward M. White. *Teaching and Assessing Writing: Recent Advances in Understanding, Evaluating, and Improving Student Performance.* Jossey-Bass Publishers, San Francisco, 1985.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.

Appendix A Complete coding results

Table 4: Populism in party manifestos and candidates' speeches

Country	Year	Party	Manifesto	Speeches	Party score	Party score 2
Argentina	2011	FAP	0.2	0.25	0.225	0.23
Argentina	2011	FpV	0.5	0.25	0.375	0.33
Argentina	2011	FP	0.25	0.7	0.475	0.55
Argentina	2011	UCR	1.3	0.2	0.75	0.57
Argentina	2011	CF	0.1	1	0.55	0.7
Austria	2008	BZÖ	0.2		0.2	0.2
Austria	2008	FPÖ	1		1	1
Austria	2008	Grünen	0.1		0.1	0.1
Austria	2008	ÖVP	0		0	0
Austria	2008	SPÖ	0.1		0.1	0.1
Belgium-WAL	2014	cdH	0.05		0.05	0.05
Belgium-WAL	2014	Ecolo	0		0	0
Belgium-WAL	2014	FDF	0		0	0
Belgium-WAL	2014	MR	0		0	0
Belgium-WAL	2014	PP	0.55		0.55	0.55
Belgium-WAL	2014	PS	0.15		0.15	0.15
Bolivia	2014	MAS	1.55		1.55	1.55
Bolivia	2014	PDC	0.3		0.3	0.3
Bolivia	2014	UD	0.25		0.25	0.25
Brazil	2014	PSDB	0	0.15	0.075	0.1
Brazil	2014	PSB	0.075	0.15	0.1125	0.125
Brazil	2014	PSOL	1.1	1.65	1.375	1.47
Brazil	2014	PT	0	0.65	0.325	0.43
Canada	2006	BQ	0.75		0.75	0.75
Canada	2006	Cons	0.8		0.8	0.8
Canada	2006	Green	0.2		0.2	0.2
Canada	2006	Lib	0		0	0
Canada	2006	NDP	0.3		0.3	0.3
Chile	2013	PS	0.7	0.15	0.425	0.33
Chile	2013	UDI	0	0.025	0.0125	0.017

Table 4: Populism in party manifestos and candidates' speeches

Country	Year	Party	Manifesto	Speeches	Party score	Party score 2
Chile	2013	PRO	0.3	1.15	0.725	0.87
Chile	2013	Parisi	0	0.45	0.225	0.3
Chile	2013	IGUAL	2	2	2	2
Colombia	2014	CD	0.35	0.15	0.25	0.22
Colombia	2014	C	0	0.1	0.05	0.07
Colombia	2014	PDA	0.1	1.45	0.775	1.0
Colombia	2014	PVC	0.1	0.2	0.15	0.17
Colombia	2014	U	0	0.1	0.05	0.07
Ecuador	2013	CREO	0.6	0.08	0.34	0.25
Ecuador	2013	Pais	1.7	1.15	1.425	1.33
Ecuador	2013	PRIAN	0.2	0.35	0.28	0.3
Ecuador	2013	PSP	0.1	0.475	0.29	0.35
France	2012	FG	0.9	0.25	0.575	0.47
France	2012	FN	0.4	0.75	0.575	0.63
France	2012	MoDem	0		0	0
France	2012	PS	0.1	0	0.05	0.03
France	2012	UMP	0	0.25	0.125	0.17
France	2012	Verts	0.15		0.15	0.15
Germany	2013	CDU	0		0	0
Germany	2013	Grüne	0.2		0.2	0.2
Germany	2013	FDP	0		0	0
Germany	2013	SPD	0		0	0
Germany	2013	AfD	0		0	0
Germany	2013	CSU	0		0	0
Germany	2013	Linke	1.3		1.3	1.3
Germany	2013	NPD	1.4		1.4	1.4
Ireland	2011	FF	0.025		0.025	0.025
Ireland	2011	FG	0.25		0.25	0.25
Ireland	2011	Lab	0.3		0.3	0.3
Ireland	2011	SF	0.325		0.325	0.325
Italy	2013	M5S	0.1	0.65	0.375	0.47
Italy	2013	LN	0.1	0	0.05	0.05
Italy	2013	PD	0.6	0.3	0.45	0.45

Table 4: Populism in party manifestos and candidates' speeches

Country	Year	Party	Manifesto	Speeches	Party score	Party score 2
Italy	2013	PdL	0	0.35	0.175	0.23
Italy	2013	SC	0	0	0	0
Italy	2013	RC	0.4	1.5	0.95	1.13
Italy	2013	SEL	0.1	0.35	0.23	0.27
Mexico	2012	PAN	0.1	0.05	0.08	0.067
Mexico	2012	PRI	0.05	0.18	0.11	0.13
Mexico	2012	PRD	0.95	0.55	0.75	0.68
Mexico	2012	PNA	0.05		0.05	0.05
Netherlands	2012	CDA	0.05		0.05	0.05
Netherlands	2012	D66	0		0	0
Netherlands	2012	PvdA	0		0	0
Netherlands	2012	PVV	1.25		1.25	1.25
Netherlands	2012	SP	0.2		0.2	0.2
Netherlands	2012	VVD	0		0	0
Norway	2013	A	0		0	0
Norway	2013	FrP	0		0	0
Norway	2013	H	0		0	0
Paraguay	2013	APA		0.1	0.1	0.1
Paraguay	2013	AP		0.8	0.8	0.8
Paraguay	2013	ANR-PC		0.05	0.05	0.05
Peru	2011	AGC	0	0.1	0.05	0.067
Peru	2011	F11	0	0.05	0.025	0.034
Peru	2011	PNP	0.55	1.3	0.925	1.05
Peru	2011	PP	0	0.1	0.05	0.07
Peru	2011	SN	0	1.6	0.8	0.8
Portugal	2011	BE	0.4	0.2	0.3	0.267
Portugal	2011	CDS-PP	0.05		0.05	0.05
Portugal	2011	PCP	0.7	0.65	0.675	0.667
Portugal	2011	PS	0	0.05	0.025	0.033
Portugal	2011	PSD	0.05	0.1	0.075	0.083
Spain	2011	CiU	0.25		0.25	0.25
Spain	2011	IU	1		1	1
Spain	2011	PNV	0.25		0.25	0.25

Table 4: Populism in party manifestos and candidates' speeches

Country	Year	Party	Manifesto	Speeches	Party score	Party score 2
Spain	2011	PP	0.4		0.4	0.4
Spain	2011	PSOE	0		0	0
Spain	2011	UPyD	0.1		0.1	0.1
Sweden	2014	M	0	0.1	0.05	0.05
Sweden	2014	C	0	0.1	0.05	0.07
Sweden	2014	FP	0.05	0.1	0.075	0.075
Sweden	2014	KD	0	0.1	0.05	0.07
Sweden	2014	MP	0	0.65	0.325	0.4
Sweden	2014	SAP	0	0.25	0.125	0.13
Sweden	2014	SD	0.1	0.15	0.125	0.17
Sweden	2014	V	0.2	0.45	0.325	0.37
Switzerland	2011	BDP	0.05		0.05	0.05
Switzerland	2011	CVP	0.1		0.1	0.1
Switzerland	2011	FDP	0.1		0.1	0.1
Switzerland	2011	GPS	0		0	0
Switzerland	2011	SP	0.3		0.3	0.3
Switzerland	2011	SVP	1		1	1
UK	2010	BNP	0.5	1.4	0.95	0.95
UK	2010	Lab	0.25	0.05	0.15	0.117
UK	2010	LibDem	0	0	0	0
UK	2010	C	0.05	0	0.025	0.017
UK	2010	UKIP	0.15	0	0.075	0.075
Uruguay	2014	FA	0.05	0.2	0.125	0.15
Uruguay	2014	PC	0	0.15	0.075	0.1
Uruguay	2014	PI	0.15	0.3	0.225	0.25
Uruguay	2014	PN	0	0.1	0.05	0.07
Uruguay	2014	UP	1.2	0.25	0.725	0.57
United States	2012	D	0.45	0.3	0.375	0.32
United States	2012	R	0.25	0.05	0.15	0.12
Venezuela	2013	PSUV	1.85	1.6	1.725	1.68
Venezuela	2013	MUD	0.8	1.9	1.35	1.53

Notes: **Manifesto** refers to the average between the preamble's and list of issues' grades in all countries except for Chile, Germany, Spain, and UK, where coders still gave one score for the whole document. **Party score** is the average between the manifesto and the mean of speeches; **Party score 2** is the average of all documents.

Appendix B