UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

A populist Zeitgeist? The impact of populism on parties, media and the public in Western Europe

Rooduijn, M.

Publication date
2013

[Link to publication](#)

Chapter 3

MEASURING POPULISM IN COMPARATIVE RESEARCH

Comparing Two Methods of Content Analysis

Much as with older accounts of populism, newer ones tend to declare certain leaders populist by fiat rather than through any kind of systematic measurement, and analyses that do offer justifications are usually single-country studies that avoid demonstrating the broad applicability or reliability of their measure.

Kirk A. Hawkins, 2009, p. 1041.

Introduction

As the previous chapter demonstrated, the term populism has been applied to a wide and diverse range of movements, politicians and parties. It has been shown that a potentially problematic aspect of populism as a scientific concept is its contextual sensitivity: populist actors in Latin America have different characteristics than populists in Western Europe, and populists in Italy are different from populists in France. Because of this contextual sensitivity, populism is in turn plagued by a lack of conceptual clarity (Barr, 2009; Canovan, 1981; Laclau, 2005; Taggart, 2000). Consequently, scholars have not yet developed systematic methods to empirically measure populism across cases and over time. There have been some empirical investigations into populism, yet most of these are single case studies (e.g., Albertazzi & McDonnell, 2008a; Mény & Surel, 2002a). A more systematic and comparative perspective is still lacking. This is problematic because the most relevant scientific questions – including the main questions of this dissertation – are empirical and comparative in nature: are we increasingly living in a populist Zeitgeist?; under which circumstances are mainstream parties becoming populist?; have the media become increasingly populist? In this chapter, I do not aim to answer these empirical questions yet. However, I do focus on a fundamental *prerequisite* for answering these questions: the methodological issue of how to measure populism empirically.

I compare two methods to measure populism: (1) a classical content analysis where coders systematically analyze texts by means of a codebook; and (2) a computerized content analysis in which an a priori designed

dictionary serves as a gauge of the degree of populism. Although this study demonstrates that both approaches can be used to measure populism across cases and over time, the classical analysis turns out to generate more valid results.

I am not the first to measure populism by means of the method of content analysis. Jagers and Walgrave (2007) were among the first to do that. However, although their analysis provided a breakthrough in measuring populism, their study included only one country, and issues of reliability and validity were not dealt with. Another content analysis of populism has been executed by Hawkins (2009). He analyzed speeches by means of 'holistic grading', in which the unit of measurement is the entire text. The main problems with Hawkins' study are the rough – and therefore possibly invalid – measurement due to the holistic grading method, and at times low reliability (Kappa = 0.44). More recently, Pauwels (2011) also measured populism by means of a computer-based content analysis. However, he only studied the Belgian case. In this chapter, I focus extensively on issues of validity (i.e., evaluating whether we measure what we think we are measuring) and reliability (i.e., the consistency of the measurement), while maintaining a comparative perspective.

Populism as a thin ideology consisting of two components

To measure a concept systematically, we first have to agree on a clear definition. I build on the definition of populism that has been provided by Mudde (2004: 543): '[populism is] an ideology that considers society to be ultimately separated into two homogeneous and antagonistic groups, "the pure people" versus "the corrupt elite", and which argues that politics should be an expression of the *volonté générale* (general will) of the people' (see also the previous chapter). Although other scholars have defined populism as a style (see Jagers & Walgrave, 2007) or an organizational form (see Taggart, 1995), I focus on this *ideational* minimal definition of populism because of its denotative clarity (Abts & Rummens, 2007; Albertazzi & McDonnell, 2008b; Stanley, 2008) and because it is an appropriate definition to compare cases

across countries and over time (Mudde & Rovira Kaltwasser, 2012). Populism is not a 'full ideology', such as liberalism or socialism, but a 'thin ideology' (Freeden, 1998), which only focuses on a confined range of concepts (e.g., nationalism, feminism, ecologism).

According to Mudde's definition the thin ideology of populism consists of two components: people-centrism and anti-elitism. The point of departure of every populist is the fundamental importance of the centrality of 'the people' (Ionescu & Gellner, 1969b; Mény & Surel, 2002b). Yet 'the people' can mean many different things in many different circumstances (Canovan, 1981; Mudde, 2004). It can refer, for instance, to peasants, the working class, the electorate, the nation, or no fixed group at all (Canovan, 1981; Taggart, 2000). Populists are anti-elitist because elites stand in the way of the centrality of the people. Elites are portrayed as corrupt and are contrasted against the general will of the people (Mudde, 2004).

Research strategy

The unit of analysis in my content analyses is the election manifesto.[20] This unit was chosen for two reasons. The first substantive reason is that an election manifesto can be seen as the document that gives the clearest overview of what a party stands for at a certain point in time. In most cases, politicians are bound to the policy promises laid down in an election manifesto. 'As an official document, it will be difficult for party members to resile from policies in the party manifesto, while party leaders can be charged with failure to implement published manifesto pledges when given the chance to do so' (Laver & Garry, 2000: 620). The second, more practical, reason is that election manifestos are appropriate documents for a cross-national study

---

[20] The unit of analysis should not be confused with the unit of measurement. The classical and the computerized content analysis approach have the same unit of analysis (election manifestos) but different units of measurements (paragraphs and words respectively). This is due to the different points of departure of the two methods. This is discussed further in the next sections of this chapter.

because they are reasonably comparable across countries and over time (Klemmensen et al., 2007).

I focus on the election manifestos of parties in Western Europe because 'the main area of sustained populist growth and success over the last fifteen years in established democracies has been Western Europe' (Albertazzi & McDonnell, 2008: 1). More specifically, I selected France, Germany, Italy, the Netherlands and the United Kingdom because these countries accommodate a wide range of different kinds of allegedly populist parties: the *Front National* (FN) (Rydgren, 2008) in France; the *Partei des Demokratischen Sozialismus* (PDS, later: *Die Linke*) (Hough & Koß, 2009) in Germany; *Forza Italia* (FI), the *Lega Nord* (LN) and the *Alleanza Nazionale* (AN) in Italy (Tarchi, 2008); the *Partij voor de Vrijheid* (PVV) (Vossen, 2010), the *Lijst Pim Fortuyn* (LPF) (Lucardie, 2008), the *Socialistische Partij* (SP) (March, 2007), and the *Centrumdemocraten* (CD) (Mudde, 2007) in the Netherlands; and the *British National Party* (BNP) and the *United Kingdom Independence Party* (UKIP) (Fella, 2008) in the UK. They range from the left (e.g., the SP) to the right (e.g., the PVV), and from electorally very successful (e.g., FI) to relatively unsuccessful (e.g., the BNP).[21] My focus is not only on the 'usual suspects' but also on the mainstream parties in each country – i.e., liberal, conservative/Christian-democratic and social-democratic parties. The reason for doing so is to explore to what extent my measurement distinguishes allegedly populist parties from mainstream parties. Moreover, some mainstream parties, such as the UK *Labour* party under Tony Blair, have also been labeled as populist (Mair, 2002).

I have not only compared parties across space but also over time. Four election years (between 1988 and 2008) have been selected in each of the countries under investigation. This resulted in 83 election manifestos, which

---

[21] All these allegedly populist parties are commonly classified as either radical right-wing or radical left-wing. Although many other radical left-wing parties, such as the DS (IT) and the PCF (FR), are usually not categorized as populist, I have included them in my case selection as well because they share many characteristics with their ideological brethren Die Linke (GE) and SP (NL), which are commonly categorized as populist.

were included in the classical content analysis. For an overview, see Appendix A. However, because 13 of these manifestos could not be translated into a legible digital format, the computerized content analysis is based on 70 election manifestos.[22]

The classical content analysis

In my classical content analysis of populism, the manifestos were analyzed by extensively trained coders by means of a codebook. In this codebook, people-centrism was operationalized by the following question: 'Do the authors of the manifesto refer to the people?' The coders were instructed to look at every possible reference to the people. It did not matter whether this reference concerned, for instance, 'citizens', 'our country', 'society' or 'we' (as in 'we the people'). The coders were also instructed to interpret the broader context in deciding whether to code people-centrism or not. To assist the coders, I provided an extensive list of words and combinations of words that *could* refer to the people.[23]

Anti-elitism was measured by means of the question: 'Do the authors of the manifesto criticize elites?' The critique had to concern political elites *in general*. A critique of a specific party or a particular politician was not general enough and was therefore not coded. Because anti-elitism can be expressed in many different ways, the coders were again instructed to interpret the context while coding.[24]

---

[22] The excluded manifestos are: PS 1993, RPR 1993, UDF 1993, PCF 1993, PS 1997, RPR 1997, and FN 1997 in France; CDU/CSU 2002 in Germany; AN 1992, LN 1992, AN 1994, and DS 1994 in Italy; CD 1994 and VVD 2006 in the Netherlands; and Libdem 1992 and BNP 1992 in the United Kingdom.

[23] These words are: people, citizen(s), community, society, public, population, nation(al), all of us, each of us, everyone, our, we, voter(s), electorate, referenda, direct democracy, public opinion, country. And words such as: United Kingdom, Britons, Netherlands, Dutch, Italians, Gemany, etc. (depending of course on the country under analysis).

[24] For the question wording see Appendix B.

The unit of measurement is the paragraph because paragraphs are objectively traceable distinctions between arguments.[25] I have selected those paragraphs as populist in which both people-centrism and anti-elitism were present. Eventually, the percentage of populist paragraphs in every election manifesto was computed.

The computer-based content analysis

Because a classical content analysis is a very time-consuming and therefore expensive enterprise, I have also employed a much easier applicable measurement of populism, drawing on a computer-based content analysis. This measurement relies on the dictionary approach in which a computer counts the proportion of words that can be considered to be indicators of populism.[26] This means that words instead of paragraphs are the unit of measurement. Although skeptics might argue that the same word can have different meanings depending on the context, it is often possible to code words unambiguously (Laver & Garry, 2000).[27]

---

[25] Because populist claims are usually presented in multiple sentences, the sentence is not an appropriate unit of measurement in this study (cf. Guthrieet al., 2004). 'Themes', also referred to as 'appeals' or 'statements', do represent clearly delineated arguments. It is, however, difficult to extract them from texts (Weber, 1990: 22), which makes it difficult to obtain reliable results when using the theme as unit of measurement.

[26] While there are different approaches available in computerized textual analyses – such as *Wordscores* or *Wordfish* – I draw on a dictionary approach (Laver & Garry, 2000). A drawback of *Wordscores* is that it requires scores to be computed by other methods such as expert surveys (Laver et al., 2003). *Wordfish* works well for extracting single left-right dimensions (Slapin & Proksch, 2008), while it is less suited to explore a specific ideological aspect such as populism.

[27] The word 'taxes', for instance, might be associated with cutting taxes, but it can equally be used to indicate that a party wants to increase taxes. In practice, however, this latter meaning will hardly be found in party manifestos, and the word taxes is hence a good indicator for the category 'reduce state involvement in the economy', identifying socio-economic rightist parties.

After a first explorative analysis, it turned out that a measurement of people-centrism by means of individual words only is nearly impossible. In many instances, 'the people' is referred to by the words 'we' and 'our' (e.g., *we* [the people] need to raise *our* voice'). Yet not every mentioning of the words 'our' or 'we' is a reference to the people. Often, these words refer to the political party instead of the people (e.g., '*we* [the party] propose *our* plans in the next chapter'). I therefore decided to only focus on words that refer to anti-elitism. Although the computer-based measurement is therefore likely to be less valid, I believe that anti-elitism is a reasonable indicator of populism because the classical content analysis taught that criticism towards elites is mostly motivated by the argument that elites betray the ordinary people. The argument cannot be reversed, however, because many political parties will centralize the people without being negative towards elites.[28] Whether anti-elitism alone is indeed a good indicator of populism can only be concluded from the comparison of the measurements.

My selection of words for the dictionary was based on both empirical and theoretical reasoning. For inspiration, I used empirical examples (election manifestos of allegedly populist parties that I did not actually analyze) to make a list of words that such parties have used to express their negativity towards elites. However, my final decision of whether to include these words in the dictionary was based on theoretical reasons: only those words have been selected that were explicitly used to position the bad elites against the good people. The development of a dictionary is not an easy task, however. Not every word that could refer to anti-elitism always refers to it, while at the same time, many instances of anti-elitism can easily be missed because it is impossible to formulate every word that could refer to anti-elitism beforehand. Theoretically deduced words that never seemed to appear in the manifestos of any party were excluded. I attempted to translate the dictionary

---

[28] The classical content analysis empirically confirmed this: there is only a weak correlation between people-centrism and anti-elitism (r = 0.04, not significant at p < 0.05), whereas – the other way around – almost every anti-elitist paragraph also contains a reference to the people.

for the five languages as accurately as possible. However, in addition to the translated 'core words', I also added some 'context-specific words'. The context-specific words are words that are too context-specific to be translated from one language to another.[29]

Although I am aware of the possible pitfalls when translating the dictionary, I argue that the theoretical argument of anti-elitism is generally similar across cases and over time. Whatever the specific context, populists in every country and every time-period do essentially the same thing: they position the good people against the bad elites. Because they make this same argument, I assume that they also use similar words. I used the open software program Yoshikoder to measure the populism score, which is the percentage of dictionary words. For a complete overview of the dictionary, see Appendix C.

Results

Validity

I focus on three types of validation: content validation, face validation, and concurrent validation. A measurement is *content valid* if the systematized concept is adequately captured by its indicator or indicators (Adcock & Collier, 2001). I have argued that populism consists of the combination of people-centrism and anti-elitism. In the classical analysis I have measured it accordingly because I verified for every measurement unit whether there was a reference to 'the people' combined with a critique on elites in general. Yet the computer-based analysis is less content valid. After all, here the systematized concept of populism is not adequately captured by its indicators.

---

[29] For instance, allegedly populist parties in the Netherlands sometimes talk about 'regenten' to express anti-elitism. This word refers to the Dutch political rulers in the sixteenth, seventeenth and eighteenth century. Although the 'regenten' did not form a hereditary class, they did form a closed group that reserved government offices for themselves. This specific word is not used by allegedly populist parties in countries other than the Netherlands.

Instead of measuring both people-centrism and anti-elitism, the computerized content analysis only focused on anti-elitism.[30]

A measurement has *face validity* when it appears to be measuring the concept that it intends to measure (Weber, 1990). In my case, the measurement of populism has face validity if allegedly populist parties have high values on this measure, while other parties score low(er). The results indicate that the allegedly populist parties have, on average, much higher values than the other parties on both measures of populism (see Table 3.1).

Table 3.1

Mean populism scores (standard deviations between brackets)

|	Allegedly populist parties	Other parties	Overall
Classical content analysis (N = 87)	7.739 (6.785)	1.150 (2.106)	2.912 (4.874)
Classical content analysis (N = 70)	6.080 (6.194)	0.936 (1.858)	2.203 (4.075)
Computerized content analysis (N = 70)	0.099 (0.091)	0.036 (0.038)	0.051 (0.062)

According to the classical content analysis, there are some allegedly populist parties that are less populist than expected (i.e., less populist than the mean populism score): the FN (2002) in France, PDS/Die Linke (1990, 1994, 2002 and 2005) in Germany, LN (1992 and 1994) and the CdL/PdL (2001 and 2008) in Italy, and the SP (2006) in the Netherlands (see Figure 3.1). The most remarkable of these outcomes is that PDS/Die Linke turned out to be only marginally more populist than mainstream parties in all German election

---

[30] It needs to be admitted that a mere reference to the people does not necessarily imply that the people is conceived of as a homogeneous actor. Yet, a close look at the results of the analysis shows that when a reference to the people coincides with an anti-elitist claim, the writers of the manifesto almost always argue that the people are betrayed by the elites. This implies that the people are seen as a unified actor, at least with regard to their relationship with the elites.

years under investigation. This might be explained by the moderate political climate in Germany (Art, 2006). The low populism score for the SP in 2006 is in line with the claim that this party has become less populist over time (Lucardie & Voerman, 2012). That the CdL/PdL has relatively low populism scores might be due to the party being the main right-wing party (block) during the elections of 2001 and 2008 and wanted to present itself as a mainstream political movement (Tarchi, 2008). The low scores of FN and LN are more difficult to explain. My suspicion is that these parties scored low on the populism scale because they focused more strongly on migration and integration than on the core elements of populism (i.e., the people versus the elites).
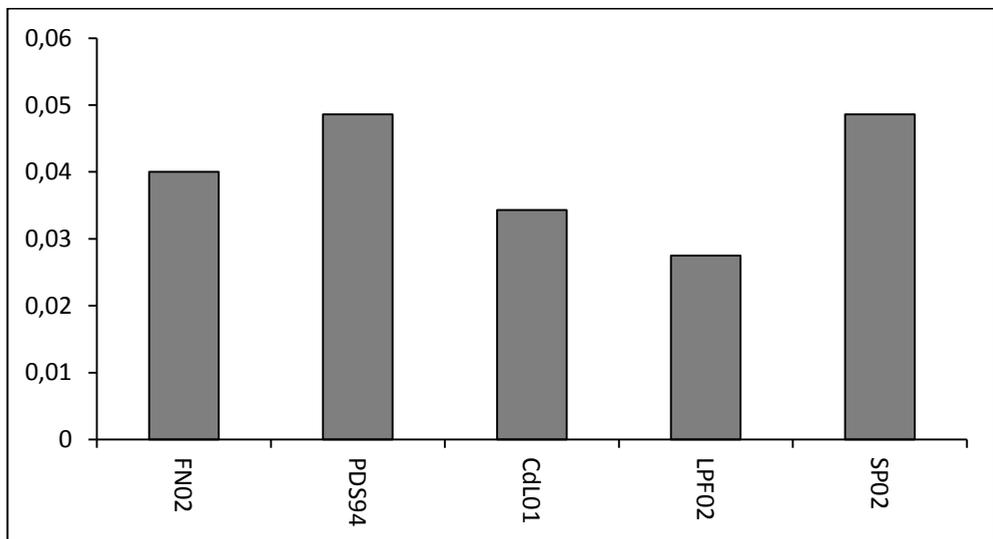
Figure 3.1

Classical content analysis: allegedly populist parties with populism scores below the mean (2.91)
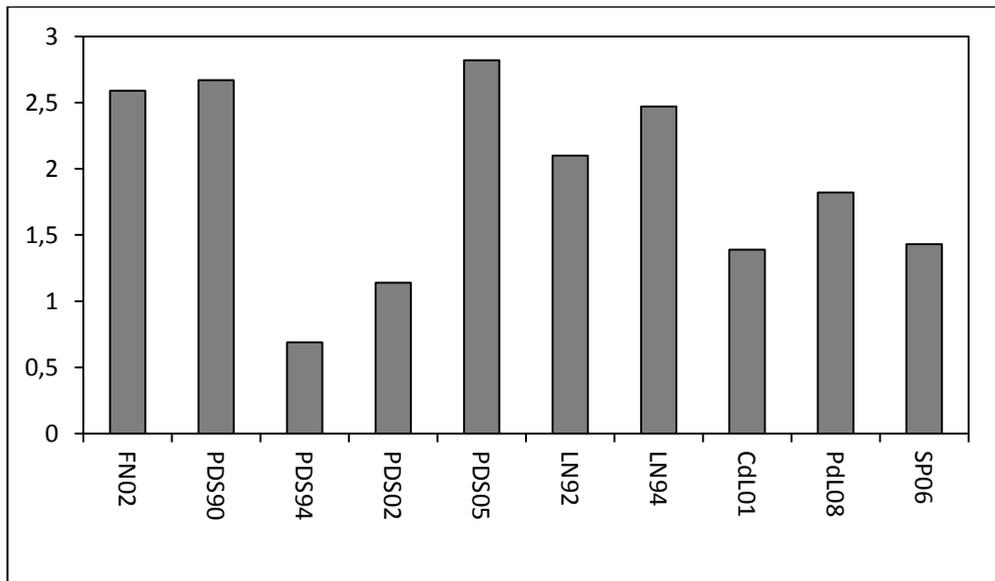
The computer-based content analysis yields similar results (see Figure 3.2): the FN, the DS, the CdL and the SP are, just as in the classical analysis, identified as less populist than expected.[31] A striking result of the computerized method is that the LPF is only marginally populist. This contrasts with the existing literature (see Koopmans & Muis, 2009; Mudde, 2007; Van der Brug & Mughan, 2007) and with the results of the classical content analysis.

Figure 3.2

Computerized content analysis: allegedly populist parties with populism scores below the mean (0.05)
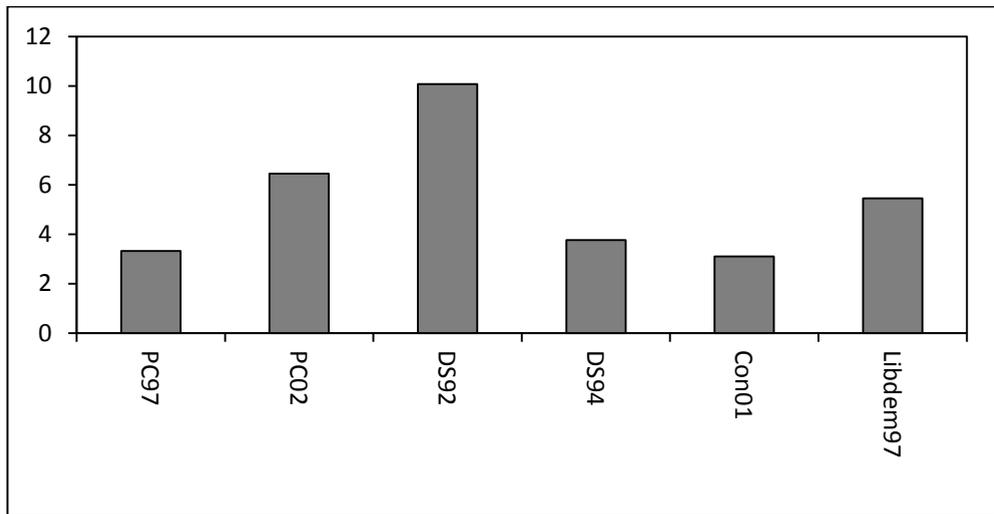

There are some non-populist parties that turned out to be rather populist (i.e., more populist than the mean populism score) according to the classical content analysis (see Figure 3.3). These parties are the PCF (1997 and 2002) in France, the DS (1992 and 1994) in Italy, and the Libdems (1997) and Conservatives (2001) in the United Kingdom. It is not so surprising that the French PCF and the Italian DS are rather populist. After all, both belong to the

---

[31] Note that LN 1992 is not included in the computerized analysis.

far left party family (e.g., Budge et al., 2001; Klingemann et al., 2006), and it has been argued that populism is highly compatible with radical left ideologies (March, 2007; March & Mudde, 2005). Moreover, Italy also witnessed a large-scale corruption scandal in the early nineties, which might well have fuelled DS's populism (Tarchi, 2002). The high score of mainstream parties in the United Kingdom can perhaps be explained by the UK's two party system in which fierce opposition campaigns can come close to populism. Indeed, various scholars have argued that mainstream parties in Britain are rather populist (e.g., Mair, 2002; Mudde, 2004).

Figure 3.3

Classical content analysis: other parties with populism scores above the mean (2.91)


According to the computer-based content analysis, there are more non-populist parties that have rather high populism scores (see Figure 3.4): Libdem (1997 and 2005), the Conservatives (2001 and 2005) and Labour (2005) in the United Kingdom; the FDP (2005) in Germany; the DC (1992 and 1994) and the DS (1992) in Italy; and the PvdA (1994 and 2006) in the Netherlands.[32]

---

[32] Note that DS 1994 (Italy) is not included in the computer-based analysis.

Figure 3.4

Computerized content analysis: other parties with populism scores above the mean (0.05)
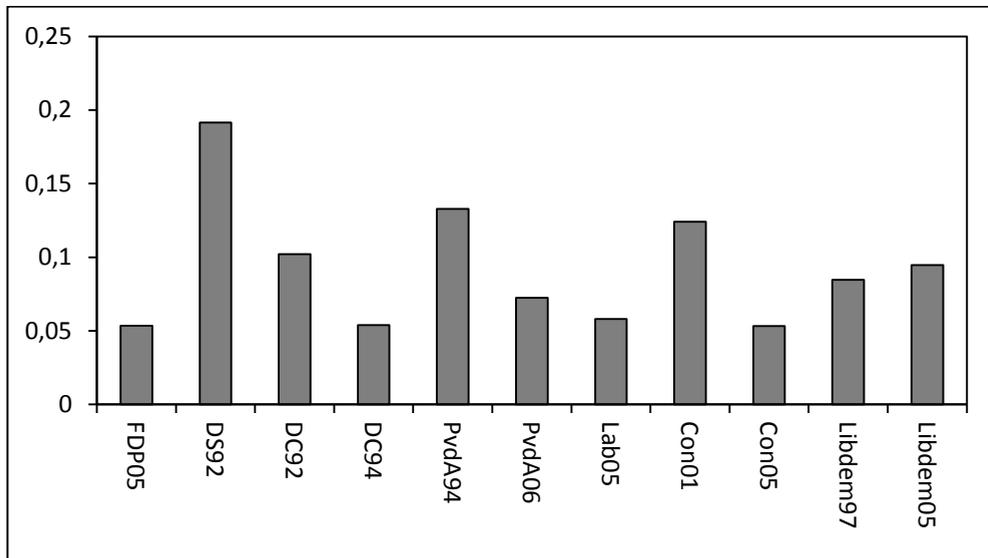
Although the high scores for the Libdems, the Conservatives and Labour might be explained by the British party system, the results for the German FDP, the Italian DC and the Dutch PvdA are more difficult to explain. Because the classical content analysis generated fewer results which are difficult to explain (3) than the computerized analysis (8), [33] it can be concluded that the classical content analysis is more face valid than the computerized method.

    Concurrent validity entails that a measure is valid if the results of a measurement of the systematized concept in one study are empirically related to the results of a different measurement of the same systematized concept in another study (Adcock & Collier, 2001). I can test whether my measurements of populism are valid by comparing their results. The results are generally

---

[33] These cases that are difficult to explain are FN 2002, FN 2004 and LN 1992 with regard to the classical analysis, and FN 2002, DS 1994, LPF 2002, FDP 2005, DC 1992, DC 1994, PvdA 1994 and PvdA 2006 with regard to the computerized analysis.

concurrent. There is a strong correlation between the results of the two measurements: Pearson's r = 0.81 (significant at p < 0.01).

There are also some differences between the two measurements, which are visualized by the scatterplot in Figure 3.5. For the Dutch LPF (in 2002), the results of the two methods differ strongly from each other: according to the classical content analysis, this party is much more populist than according to the computerized method.[34] Because many experts describe the LPF as a populist party – see Mudde (2007), Lucardie (2008) and Van der Brug (2003) – it is likely that the classical content analysis generated a better estimation of its true degree of populism than the computerized analysis.

Figure 3.5

The classical and computerized content analyses compared

---
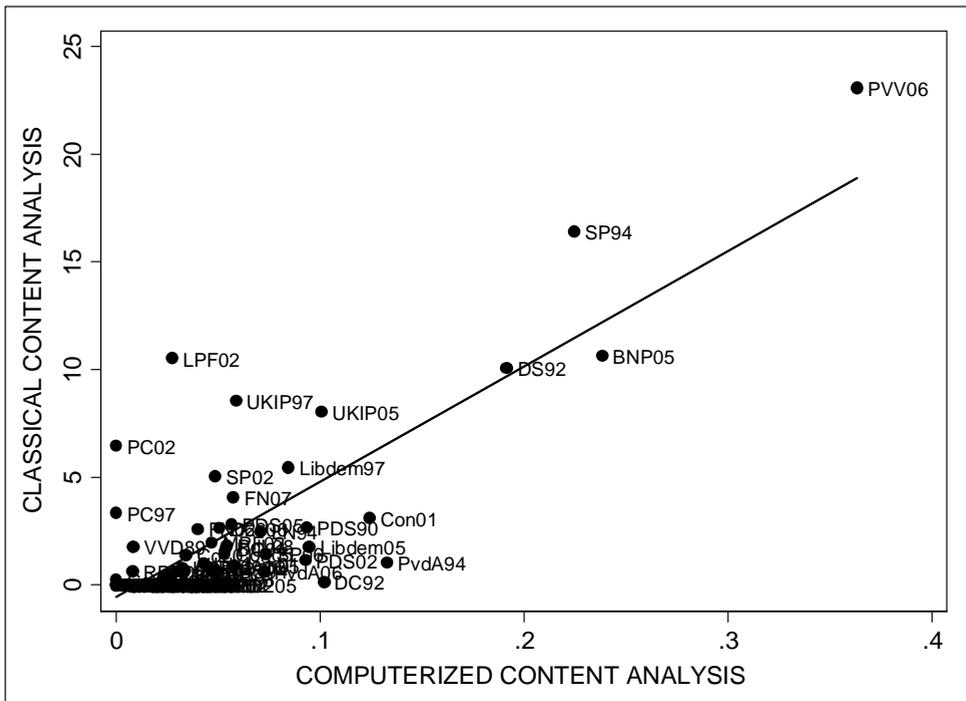
[34] This becomes even more apparent when we regress the results of the two methods on each other and look at the standardized residuals: the standardized residual of the LPF in 1994 is larger than 2. (All the other standardized residuals are below 2.)

Reliability

In a classical content analysis, the inter-coder reliability is the extent to which different coders code the same text in the same way (Krippendorff, 2004). To prevent low inter-coder reliability, I have extensively trained my 13 coders (4 from the Netherlands, 3 from the United Kingdom, and 2 from France, Germany and Italy). Every coder attended three training sessions in which the codebook was explained and in which coding examples were discussed. Between the training sessions, the coders had to complete take-home exercises. After the training sessions, I assessed the inter-coder reliability. The coders had to complete two reliability tests. First, all coders had to analyze a sample of paragraphs from British election manifestos (all coders speak English), so I could calculate whether the *cross-national* inter-coder reliability was sufficient. I have calculated the inter-coder reliability using Krippendorff's alpha. The results for cross-national reliability are $\alpha = 0.72$ for people-centrism, and $\alpha = 0.69$ for anti-elitism. Second, all coders had to analyze another sample of paragraphs from the election manifestos of parties from their own countries, so I was able to assess the *national* inter-coder reliability coefficients. The Krippendorff's alpha's range from 0.66 to 0.89. The statistics in general are satisfactory.[35]

One of the advantages of the computer-based content analysis is its accuracy. Because a computer produces the exact same results no matter how many times one runs the analysis, Laver and Garry (2000: 625) claim that '[c]omputer coding is 100 percent reliable […].' This is, however, a rather one-sided way to look at reliability. It must be kept in mind that different researchers of populism would most likely end up with different dictionaries to measure the concept, which in turn would impact the results. I have performed a split-half test to shed some light on this issue. First, the words of the dictionaries for each country were randomly divided into two groups. For

---

[35] The sample of paragraphs in the reliability tests contained approximately five per cent of the total amount of paragraphs. The results for people-centrism are: $\alpha = 0.75$ (FR), $\alpha = 0.74$ (GE), $\alpha = 0.89$ (IT), $\alpha = 0.78$ (NL), and $\alpha = 0.73$ (UK). The results for anti-elitism are: $\alpha = 0.69$ (FR), $\alpha = 0.79$ (GE), $\alpha = 0.84$ (IT), $\alpha = 0.84$ (NL), and $\alpha = 0.66$ (UK).

each country this resulted in two 'half' dictionaries. In a second step, I explored the Pearson correlation coefficients between the results computed by the two 'half' dictionaries in each country. The results are as follows: 0.54 for France, 0.24 for Germany, 0.51 for Italy, 0.41 for the Netherlands, and 0.48 for the United Kingdom. To control for analyzing 'half' dictionaries only, I adjusted the split-half correlations by means of the Spearman-Brown prophecy formula, which resulted in reliability scores of 0.70 for France, 0.39 for Germany, 0.68 for Italy, 0.58 for the Netherlands, and 0.65 for the United Kingdom.[36] Except for Germany, which might be explained by the lack of allegedly populist parties and hence variation in the scores, these statistics seem sufficient. Therefore, even when the split-half test provides a different picture than that suggested by Laver and Garry (2000), I would nonetheless argue that the computerized approach is reliable enough to be employed in empirical research. Whether it is more or less reliable than the classical content analysis is difficult to assess because of the different approaches in reliability testing. Yet the low reliability for Germany suggests that the classical content analysis might be more reliable.

Conclusion

The measurement of populism, particularly over time and space, has not received much attention. In this chapter, I have paid extensive attention to the measurement of populism over time and across countries. By means of both a classical content analysis as well as a computerized method, I investigated the degree of populism of political parties in the France, Germany, Italy, the Netherlands and the United Kingdom. The validity of the computerized method thus turns out to be lower than that of the classical approach. The main problem with this method is that the indicator of anti-elitism alone does not cover the whole concept, which consists of *both* people-centrism and anti-elitism. Yet, empirically, the effect of this theoretical shortcoming is limited.

---

[36] Reliability = $\dfrac{n * r}{1 + (n-1)r}$.

Face validation and concurrent validation show that anti-elitism alone is a relatively good indicator of populism and that the computerized content analysis method can be employed to measure populism across cases and over time.

Nevertheless, because the classical content analysis method is a more valid approach to measuring populism and because it might be argued that it is also a more reliable approach, I employ the classical analysis of populism in the remainder of this dissertation.