

# **Statewide and Regionalist Parties' Perspectives in the Long-Term Dynamics of Decentralization**

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2021

vorgelegt

von

Daniel Felipe Saldivia Gonzatti, M.A.

aus

Caracas, Venezuela

Referent: Prof. Dr. André Kaiser  
Korreferent: Prof. Dr. Sven-Oliver Proksch  
Tag der Promotion:

## ACKNOWLEDGEMENTS

Through the last four years, I learned that perseverance is key to accumulate knowledge and wisdom. However, perseverance is a necessary but not sufficient element. Scientific work lives based on an academic community that challenged my ideas and pushed me forward to make better arguments, write better, learn more, and engage more eagerly with my field of study. The journey of the last few years was not an easy one for me, but several people made it extremely enjoyable and pleasant in many ways. I thank all of you who helped me get through these four years.

First, I would like to thank my first supervisor, André Kaiser, for his unconditional support over the years and for believing in me. Since my first master studies' course at the University of Cologne in the winter of 2014, André Kaiser has motivated me to study political institutions and engage in the academic world. I am very grateful for his confidence over the development of my doctoral studies. Sven-Oliver Proksch, my second supervisor, improved my logical thinking about designs, text, and methods, while also taking the time to listen to my ideas in full length, acknowledging what worked well, and asking critical questions about what did not. His friendly supervision improved my work in many ways.

Two other important persons and co-authors of articles in my dissertation have accompanied me on this journey since 2015. Lea Kaftan, office-partner for more than four years, beloved friend, and extremely intelligent co-author, who had to listen to all my crazy ideas and deepest fears almost every day. Still, this did not discourage her from grabbing coffee by the canal at the beginning of the day and discussing daily politics and political science with me. Thanks for always asking the right questions, for your academic company, and your friendship, Lea.

Leonce Röth has been a great mentor, challenging boss, timely friend, and provocative discussant. He introduced me to the world of territorial politics and always offered a different perspective on my ideas, which improved my work tremendously. Working jointly on a last-longing project with Leonce has taught me that perseverance is important. He once said that our project will truly be finished once we retire. I look forward to working together for many years.

Alongside my doctoral project co-authors, I also made great friends and colleagues at the Cologne Graduate School and the Cologne Center for Comparative Politics.

Meeting you all was an enriching experience for me, and I hope to stay in touch as much as possible with you all.

One of these persons is Jan Schwalbach. Jan is probably the most patient and friendliest political scientist in Western Europe. I share with Jan the not-so-ideal characteristic of initiating amazing research projects at unfortunate times. Still, I enjoy a lot working together with you, Jan. The only thing missing will be our walks by the canal and discussions on daily German politics.

I would also like to thank Dennis Abel and Armin Mertens for their friendship and for all the deep and funny conversations we had. I am also very happy that I could share my ping pong skills with you as a teacher in our weekly training sessions, so I could help you improve your terrible techniques.

My gratitude also goes to the following colleagues for crucial personal conversations about my doctoral studies and for improving my work with excellent and critical feedback: Jan Sauermann, Dennis C. Spies, Paul Beckmann, Sarah Berens, Stephan Vogel, Michael Schwan, Arjan Schakel, and Amuitz Garmendia Madariaga. I am also very thankful for comments received at the 2019 CES conference in Madrid, the 2019 ECPR General Conference in Wrocław, the 2019 Zurich Text-As-Data workshop, the 2020 ECPR Online General Conference, the CCCP Research Seminar at the University of Cologne, and the ZZ Colloquium at the WZB in Berlin. Many thanks to Saskia Gottschalk, Nermin Abbassi, Moritz Raykowski, and Keno Röller-Siedenburg for their excellent assistance in collecting and coding data that was used in my doctoral project. I really appreciate the doctoral funding by the Cologne Graduate School. Most of this dissertation was written at the Cologne Center for Comparative Politics while I was working as a research assistant in the DFG-funded project “Decentralization and Electoral Geographies”.

Last but not least, I would like to thank my parents Ross Mary and Felipe, and my brother Luis Ignacio for their unconditional support and necessary humor. Probably the person who knows best how much effort went into this dissertation is Jenny – thank you for being there and listening to all my ideas, for your patience, and your lovely support every day.

Addendum. The engagement with one important topic of this dissertation – territorial politics in Spain – kindly forced me to further research the history of the Spanish Civil War. By chance, or perhaps not, this endeavor led me to discover my great-grandfather’s history and his important political role for the Republic during the war. I am very thankful to the *Museum of Mail and Telegraph* in Madrid for making very old documents about my great-grandfather available. The information collected from these archive documents allowed me to tell my family a new story with more details about my great-grandfather Jesús Sancristóbal.

Daniel Felipe Saldivia Gonzatti  
*University of Cologne*  
September 2021

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	A Short Definition of Decentralization . . . . .	3
1.2	Party Politics and Decentralization . . . . .	3
1.3	Framework and Contributions: A New Road Map to the Study of Decentralization . . . . .	4
1.3.1	Overview of Study in Chapter 2 . . . . .	5
1.3.2	Overview of Study in Chapter 3 . . . . .	7
1.3.3	Overview of Study in Chapter 4 . . . . .	8
1.4	Relevance and Broader Implications . . . . .	9
1.5	Publication Status of the Articles . . . . .	10
<b>2</b>	<b>The Long-run Dynamics of Decentralization: Regional Democracy and Statewide Governments' Strategies, 1950-2018</b>	<b>12</b>
2.1	Introduction . . . . .	13
2.2	The literature on decentralization and party politics . . . . .	15
2.3	Theory . . . . .	18
2.3.1	Decentralization reforms and institutional change . . . . .	18
2.3.2	Regional democracy and party-based mechanisms of decentralization	20
2.4	Empirical study . . . . .	25
2.4.1	Research design . . . . .	25
2.4.1.1	Dependent variable . . . . .	25
2.4.1.2	Independent variables . . . . .	26
2.4.1.3	Model specifications of party explanations of decentralization . . .	27
2.4.2	Results . . . . .	28
2.4.2.1	Decentralization reforms and regional democracy . . . . .	28
2.4.2.2	Party-based mechanisms of decentralization . . . . .	30
2.5	Discussion . . . . .	35
2.6	Conclusion . . . . .	38
<b>3</b>	<b>The Measurement of Issue Attention across Political Arenas: Ex- ploiting Increasing Returns with Optimized Dictionaries</b>	<b>41</b>
3.1	Introduction . . . . .	42
3.2	The logic of issue attention across political arenas in democracies .	43
3.3	Valid and comparable data on issue attention across political arenas	45

3.3.1	The measurement of attention to issues in text . . . . .	46
3.3.2	Hand-coding . . . . .	47
3.3.3	Dictionaries . . . . .	47
3.3.4	Machine learning . . . . .	48
3.3.5	Topic models . . . . .	49
3.4	The optimization of dictionaries . . . . .	50
3.5	Case study: territorial politics in Spain and the UK . . . . .	53
3.5.1	Issue attention: territorial politics in Spain and the UK . . . . .	56
3.5.2	Attention to sub-issues . . . . .	58
3.6	Conclusion . . . . .	62
<b>4</b>	<b>Mediated party positions from newspapers</b>	<b>64</b>
4.1	Introduction . . . . .	65
4.2	Party positions and newspapers . . . . .	67
4.3	Measuring positions in newspapers with sentiment analysis . . . . .	70
4.3.1	Information load . . . . .	72
4.3.1.1	Aggregation . . . . .	72
4.3.1.2	Context . . . . .	73
4.3.2	Certainty . . . . .	73
4.3.2.1	Issue certainty . . . . .	73
4.3.2.2	Direction . . . . .	74
4.3.2.3	Media source harmonization . . . . .	75
4.4	Case study . . . . .	76
4.5	Data generation . . . . .	77
4.6	Estimation and comparison analyses . . . . .	78
4.6.1	Convergence of positions with manifestos and expert surveys . . . . .	78
4.6.2	Newspaper sources: comparison, biases and (un-)certainty . . . . .	80
4.6.3	Text-analytical comparison with parliamentary speeches . . . . .	82
4.7	Qualitative validation . . . . .	83
4.8	Position estimates and application . . . . .	85
4.9	Towards an agenda of mediated party positions . . . . .	87
<b>5</b>	<b>Supplementary Material</b>	<b>90</b>
SM2	The Long-run Dynamics of Decentralization: Regional Democracy and Statewide Governments' Strategies, 1950-2018 . . . . .	90
SM2.1	Overview of theory . . . . .	90
SM2.2	Summary and descriptive statistics statistics . . . . .	91
SM2.3	Data Aggregation and Fundamental Decisions . . . . .	92
SM2.3.1	Regional and electoral aggregation . . . . .	92
SM2.3.2	Electoral results structure and missing elections . . . . .	92
SM2.4	Description of cases and reforms . . . . .	93
SM2.4.1	Cases of region/region groups by decentralization reform type . . . . .	93
SM2.4.2	Political decentralization: year of emergence of regional democracy . . . . .	94
SM2.5	Full models and interaction figures . . . . .	96
SM2.6	Fiscal Decentralization Reforms . . . . .	100
SM2.7	Test of First Reform as Hypothetical Critical Juncture . . . . .	102

SM3	The Measurement of Issue Attention across Political Arenas: Exploiting Increasing Returns with Optimized Dictionaries . . . . .	104
SM3.1	Data description . . . . .	104
SM3.2	Optimized dictionaries . . . . .	105
SM3.3	Territorial issue identification using machine learning and optimized dictionaries . . . . .	111
SM3.4	Dictionary performance compared to the performance of machine-learning algorithms . . . . .	113
SM3.5	Transferability of optimized dictionary to new sources and political arenas . . . . .	114
SM3.6	STM selection . . . . .	116
SM3.7	Territorial topics without aggregation across time and arenas . . . .	118
SM3.8	Representative texts of the topics without aggregation . . . . .	119
SM4	Mediated party positions from newspapers . . . . .	123
SM4.1	Established word-scaling methods in political science . . . . .	123
SM4.2	Regional newspaper readership in Spain . . . . .	124
SM4.3	Data description: party sentences over time . . . . .	124
SM4.4	Correlation of aggregated position with manifestos and expert surveys	126
SM4.5	Manifestos, expert surveys, media positions based on word-scaling .	127
SM4.6	Model performance analysis for STM . . . . .	128
SM4.7	Sentences reduction for the increase of certainty . . . . .	129
SM4.8	Convergence with parliament - extension . . . . .	129
SM4.9	Qualitative validation - extension . . . . .	130
SM4.10	Regional voter transition and party policy shifts . . . . .	132
SM5	Statistical Software . . . . .	134

<b>Bibliography</b>	<b>136</b>
---------------------	------------

## LIST OF FIGURES

1.1	Regionalist Parties at Regional and Statewide Elections in 19 Democracies . . . . .	6
2.1	Trends of regional authority across time and by status of political decentralization . . . . .	29
2.2	Effect of Ideological Incentives and Degree of Decentralization on Further Reforms Moderated by Political Decentralization (0 — 1) .	32
3.1	Attention to territorial politics across political arenas in Spain . . .	57
3.2	Attention to territorial politics across political arenas in the UK . .	58
3.3	Attention to territorial sub-issues in Spain . . . . .	60
3.4	Attention to territorial sub-issues in the UK . . . . .	61
4.1	Simple top-down model of party position transmission: parties, media, and voters . . . . .	69
4.2	Party positions from newspaper sentiment estimates, expert survey and electoral manifestos . . . . .	80
4.3	Convergence of party positions by newspaper outlet with different measurements and minimum sentence thresholds Left panel: w/o newspaper adjustment. Right panel: w/ media source harmonization	82
4.4	Correlation between sentiment-based positions in mediated estimates and parliament estimates for territorial sub-issues . . . . .	84
4.5	Mediated party positions on territorial sub-issues in Spain . . . . .	85
4.6	Voter transition and party-position shifts in Spanish territorial politics – VAR models . . . . .	87
SM2.1	Statewide Ideological Proximity, Political Decentralization and Reforms . . . . .	91
SM2.2	Decentralization Reforms across Time and Cases . . . . .	95
SM2.3	Effect of Autonomist and Secessionist Parties' Electoral Support Before and After Political Decentralization . . . . .	98
SM2.4	Effect of Ideological Incentives, Decentralization Degree and Political Decentralization on Fiscal Authority Transfers . . . . .	101
SM2.5	Effect of Autonomist and Secessionist Parties' Electoral Support Before and After Political Decentralization on Fiscal Authority Transfers	101



SM2.6	Effect of Ideological Incentives and Decentralization Degree Before and After the First Reform . . . . .	103
SM3.7	Correlation of territorial issue identification within articles in Spain, El País 1976-2012 . . . . .	111
SM3.8	Article correlation between methods' predictions in the UK case, The Times 1900-2013 . . . . .	112
SM3.9	STM optimization parameters for territorial sentences in newspapers and parliament, Spain . . . . .	116
SM3.10	Comparison of exclusivity and semantic coherence of STMs with K = 9 and 11, Spain . . . . .	116
SM3.11	STM optimization parameters for territorial sentences in newspapers and parliament, UK . . . . .	117
SM3.12	Comparison of exclusivity and semantic coherence of STMs with K = 5, 11, and 19, UK . . . . .	117
SM3.13	Territorial topics without aggregation, Spain 1976-2019 . . . . .	118
SM3.14	Territorial topics without aggregation, UK 1900-2019 . . . . .	118
SM4.15	Share of party sentences within territorial sentences in newspapers, 1976-2019 . . . . .	124
SM4.16	Sub-issues by newspaper over time (predicted topic prevalences), 1976-2019 . . . . .	125
SM4.17	Share of party sentences within territorial sentences in newspapers by topic sub-issue . . . . .	125
SM4.18	Mediated party positions, expert survey and electoral manifestos . .	126
SM4.19	Party positions of newspaper sentiment estimates, expert surveys, and electoral manifestos with different adjustments . . . . .	126
SM4.20	Single territorial issue position estimates by wordshoal and CCA compared with CHES and MARPOR/CMP . . . . .	127
SM4.21	Wording and positions of parties according to CCA . . . . .	128
SM4.22	STM optimization parameters for territorial sentences in newspapers	128
SM4.23	Yearly sentences' distribution by condition and newspaper outlet .	129
SM4.24	Convergence with parliament by party and measurement . . . . .	129
SM4.25	Convergence with parliament by territorial sub-issue and party . . .	130
SM4.26	Qualitative validation - two samples . . . . .	130
SM4.27	Qualitative validation - joint samples . . . . .	131

## LIST OF TABLES

2.1	Decentralization and re-centralization reforms in different institutional settings . . . . .	30
2.2	Ideological Incentives and Degree of Decentralization <i>before</i> and <i>after</i> Political Decentralization explaining Reforms . . . . .	31
2.3	Influence of Regionalist Parties across Time . . . . .	33
2.4	Multi-level Incentives <i>after</i> Political Decentralization . . . . .	34
3.1	Properties of measurement approaches for issue identification in text corpora . . . . .	50
3.2	F1 scores: performance of dictionaries throughout the steps . . . . .	55
3.3	Transferability . . . . .	55
4.1	Sentiment measurement adaptations for party positions in newspapers	75
SM2.1	Summary of theoretical hypotheses and arguments' references . . . . .	90
SM2.2	Summary Statistics . . . . .	91
SM2.3	Ideological Incentives and Decentralization Degree <i>Before</i> and <i>After</i> Political Decentralization . . . . .	96
SM2.4	Influence of Regionalist Parties Across Time . . . . .	97
SM2.5	Multi-Level Incentives <i>After</i> Political Decentralization . . . . .	99
SM2.6	Main models adapted to only predict fiscal decentralization . . . . .	100
SM2.7	Ideological Proximity, the First Reform and Posterior Decentralization	102
SM3.8	Territorial articles distribution . . . . .	104
SM3.9	Territorial politics prediction confusion matrix and performance parameters of different dictionary versions and different ML algorithms in the Spanish case, El País (N = 2,535) . . . . .	113
SM3.10	Territorial politics prediction confusion matrix and performance parameters of different dictionary versions and different ML algorithms in the UK case, The Times (N = 1,368) . . . . .	113
SM3.11	Transferability of an optimized dictionary of territorial politics to a new media source, Spain (El Mundo) . . . . .	114
SM3.12	Transferability of an optimized dictionary of territorial politics to a different arena, Spain (Congreso de los Diputados) . . . . .	114
SM3.13	Transferability of an optimized dictionary of territorial politics to a different media source, UK (the Guardian) . . . . .	115

SM3.14 Transferability of an optimized dictionary of territorial politics to a  
different arena, UK (House of Commons) . . . . . 115

SM4.15 Party Position Shifts - VAR Models . . . . . 132

SM4.16 Regional Voter Transitions - VAR Models . . . . . 133



## INTRODUCTION

*“Almost every territory in Europe has been combined at some time or other with almost every one of its neighbors, but only certain of these combinations have endured and kindled the loyalties and imaginations of their peoples.”*

– Karl W. Deutsch (1953)

*“This is the only way to achieve true decentralization - since power will be handed over to elected officials - and also true deconcentration - since ministerial departments will no longer depend on the ministries in Paris.”*

– Gaston Defferre (1981)<sup>1</sup>

Territorial politics have been fundamental for the design of the modern state, and political disputes over territory span from Catalan separatism efforts during the 1930s Spanish Republic to the Irish troubles in the second half of the 20th century. Yet territorial politics are not always linked to hostile relationships or political violence in specific regions of a country. During the 19th century, democratic party forces in Western Europe were fiercely contending for different architectures of the nation-state. Together with the question of the role of the Church, political parties in national governments from different ideological fields had very divergent preferences on whether to centralize or not different policy areas, and expand statewide responsibilities towards a stronger state (Ansell and Lindvall 2020). These political conflicts on the multi-level

---

<sup>1</sup>Original text: “C’est même le seul moyen d’arriver à une véritable décentralisation - puisqu’on va remettre le pouvoir aux élus - et aussi à une véritable déconcentration - puisque les services ministériels ne dépendront plus des ministères de Paris.” Gaston Defferre was French Minister for the Interior and Decentralization, Parti Socialiste.

allocation of authority raised dilemmas of centralization vis-à-vis decentralization of governance that remain relevant today.

How to organize the state is not a technical issue whatsoever. In Rokkan's perspective, how nations were built was crucial to define political communities, and trace boundaries by accommodating cultures, languages, and socio-economic differences (Rokkan and Flora 2000; Deutsch 1953, p. 187). How these nations and communities were designed would strongly impact how cultural and economic conflicts were to be solved, or not. Inversely, cultural and economic tensions – so-called political cleavages – were prominent factors affecting nation- and community-building (Lipset and Rokkan 1967) – still in today's Europeanization process (Kriesi et al. 2012).

Since World War II, different nation-states around the world have experienced a decentralization trend, especially in Europe and Latin America (Hooghe et al. 2016; Niedzwiecki et al. 2018). For a long time, it was implicitly thought that decentralization of the state was aimed to fulfill a function: to better manage the allocation of economic resources (Oates 1999); to bring the government closer to the local and regional communities (Faguet 2014); or to pacify ethno-territorial and regional conflicts (Keil and Anderson 2018). Economic and democratizing motivations were especially prominent among neo-liberal reforms in Latin American countries during the 1990s (Eaton 2012).

However, decentralization, and with it a diversification of political systems, has also fueled state corruption in certain occasions by growing the size of government across governance levels (Bardhan and Mookherjee 2006), and has generated complex multi-level structures that make effective policy-making more difficult. A post-functionalist view postulates that political causes and consequences of decentralization are fundamental for the understanding of how institutional reforms occur, and how multi-level institutions affect politics (Marks and Hooghe 2000). In the end, how the community is built and at which state-level decisions are taken are politically contested issues. Decentralization has led to party system fragmentation, the rise of regionalist parties as well as the rise of new regional actors. Politically, statewide governments – and within them political parties – are the key actors driving decentralization by initiating legislation. In this sense, several scholars (O'Neill 2003, 2005; Falleti 2005, 2013; Alonso 2012; Meguid 2015; Röth et al. 2016; Röth and Kaiser 2019; Toubreau 2017; Toubreau and Massetti 2013; Toubreau and Wagner 2016; Convery and Lundberg 2017) have asked: Why do political parties in government decentralize? Which positions and preferences have parties on decentralization? These questions have no obvious answers. If we believe that parties are particularly rational and want to accumulate power by

maximizing votes, offices, or policy interests, why should they distribute power? The present dissertation aims to systematically offer a comprehensive explanation to these dilemmas.

## 1.1 A Short Definition of Decentralization

Decentralization is conceived as the transfer of authority from upper levels of government, normally the statewide national level, downwards to sub-national governance tiers (Dardanelli et al. 2019a; Hooghe et al. 2016). This dissertation focuses exclusively on decentralization from the statewide arena towards the regional arena, usually the NUTS-1 or NUTS-2 administrative level, depending on the country. The literature normally differentiates between self-rule and shared-rule. Self-rule consists of decentralization of political autonomy, policy area competences, fiscal authority, and administration, whereas shared-rule embraces the transfer of joint rights for the regions to act together in influencing statewide policies and also grants regions with veto rights for constitutional reforming (Elazar 1987).

A distinctive perspective of importance in this dissertation is the asymmetry-symmetry divide in decentralization. Symmetric decentralization is characterized by the transfer of authority to many regions simultaneously as a homogeneous policy package or institutional reform, as is the case of the German *Länder* or the Mexican *Estados*, often implying a joint negotiation with several regions simultaneously (Hernández Rodríguez 2008). Asymmetric decentralization encompasses however a transfer of authority to a single region that enjoys – for diverse reasons, e.g. linguistic or cultural – a special status within the nation-state (see Zuber 2011; Zuber and Szöcsik 2021).

## 1.2 Party Politics and Decentralization

The link between political parties and decentralization is complex. On the one hand, parties in statewide governments decide whether to decentralize or not, i.e. they take policy-making and institutional decisions. On the other hand, decentralization's institutional changes, in turn, affect party systems and party competition in many different manners. For example, decentralization and accommodation of regionalist demands might lead to sub-national party system fragmentation (Amat and Falcó-Gimeno 2014; Harbers 2010) and radicalization of regionalist parties' positions (Masseti and Schakel 2017; Sorens 2018), especially through reinforcement of regionalist identities (Brancati 2006). Interestingly, since the 1950s, regionalist parties at statewide and regional

elections have gained popular support.

Within the framework of political motivations of decentralization, statewide parties have been the center of attention to explain incentive-oriented strategies of decentralization more strongly based on rational choice theory (Toubeau and Massetti 2013; Riker 1980). The literature has focused on electoral and ideological party strategies. When decentralizing, parties might aim to improve electoral fortunes by securing the sub-national arena by gaining some regional government offices, especially if they expect to lose power at the national level (O'Neill 2003, 2005). A complementary explanation is that statewide parties could be intending to empower ideological allies and strongholds to still be able to pass and implement policy projects even when losing some electoral support (Röth and Kaiser 2019). An alternative theory argues that statewide governments decentralize to satisfy regionalist demands and in turn gain support at statewide elections (Meguid 2015). This literature has offered an excellent start in the study of political explanations of decentralization. But most studies have focused on asymmetric regions or Latin American countries in their case selections. A systematic investigation across political systems, decentralization forms (asymmetric or symmetric), and considering both multi-level electoral arenas has been missing.

### **1.3 Framework and Contributions: A New Road Map to the Study of Decentralization**

Previous studies have exclusively focused on party-based explanations of asymmetric decentralization (Röth and Kaiser 2019), symmetric decentralization in Latin America (O'Neill 2003, 2005; Falleti 2005), or symmetric decentralization and the regional arena (Meguid 2009). A comprehensive perspective including a large range of different political systems, covering both asymmetric and symmetric decentralization over time can help uncover new insights into the long-run dynamics of decentralization. For example, it is not clear how explanations of asymmetric decentralization travel to cases of symmetric decentralization, a rather more complex type of state reform (Dardanelli et al. 2019b).

Across time, additionally, the institutional setting of multi-level democracies changes, and can affect how subsequent decentralization reforms occur (see Falleti 2005). Especially, whether regions have autonomous governors and assemblies or not, and how these can, in turn, push statewide governments towards implementing subsequent reforms (Hernández Rodríguez 2008), plays a fundamental role in the dynamics of decentralization. In this regard, the sequence of decentralization types is also key (Falleti



2005). For example, whether fiscal authority is transferred previous to administrative decentralization affects regional actors' bargaining power against or dependency on the *center*. So far, how the rise of regional democracy – a major institutional decentralization reform – affects long-run dynamics and interacts with established party-based explanations of decentralization has not been investigated comparatively.

A long-run perspective on decentralization dynamics is also necessary, since regionalist demands are not static in time. The role of regionalist parties is key for decentralization since we know that regionalist parties can exercise pressure on and at the same time create risks for statewide governments to decentralize (Massetti and Schakel 2016; Zuber and Szöcsik 2021; Brancati 2006). The strength of regionalist parties is however very dynamic across decades, but a slight incremental trend has prevailed for the period since the 1950s (see Figure 1.1).

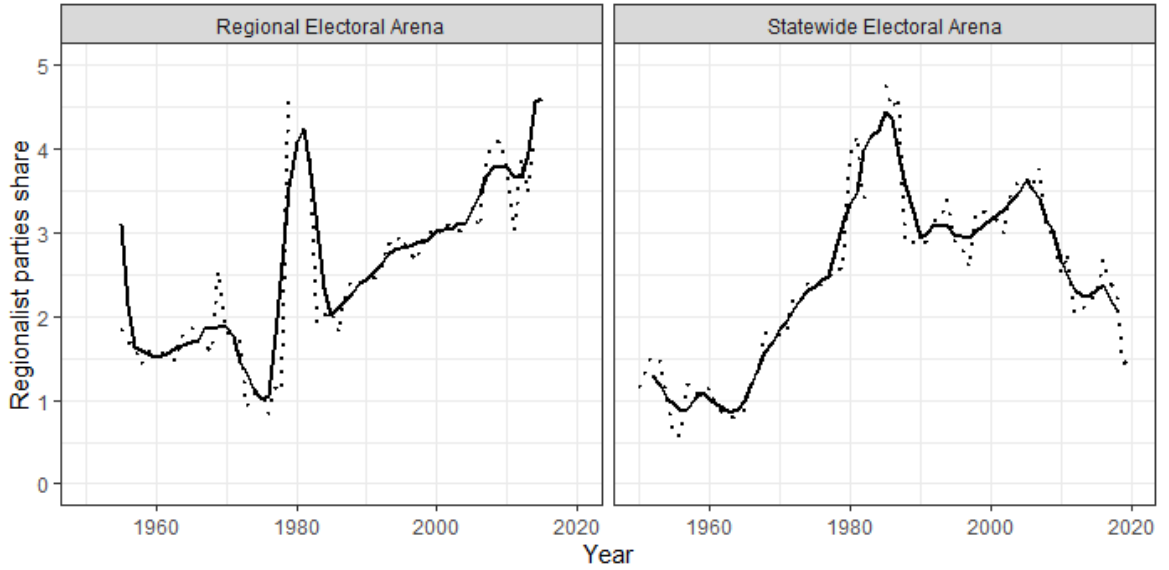
One major issue in the study of territorial politics and decentralization is the artificially low-societal salience of the issue, and therefore the lack of data to conduct systematic analyses linking these phenomena with party politics. The low salience of territorial politics is often artificially generated in political science measurements since researchers often prioritize other dimensions over the territorial one. However, if we take Meguid's (2005; 2008) interacting framework of party competition seriously, we must pay here more attention to the notions of salience, position, and issue ownership. The backbone of this dissertation is inspired by the necessity to complement this picture by offering a more solid link between party politics and decentralization.

Furthermore, territorial politics is a political issue with high sub-issue diversity. Statewide parties may embrace policy and administrative decentralization but oppose political and fiscal authority transfers (Röth et al. 2016; Meguid 2009). This notion is diametrically contrary to conventional political dimensions like welfare state or cultural issues, where sub-issue preferences are normally more aligned. For example, supporters of public health care expansion are usually also in favor of higher unemployment support. Therefore, aggregated party positions on territorial politics can be deceptive and biased. So far, political science resources have only measured party preferences on decentralization without differentiating among different sub-issues. The following section describes how the three particular investigations in this dissertation make innovative contributions to the literature and the above-mentioned gaps.

### **1.3.1 Overview of Study in Chapter 2**

In chapter 2, the study “The Long-run Dynamics of Decentralization: Regional Democracy and Statewide Governments' Strategies, 1950-2018” focuses on the above-mentioned

**Figure 1.1:** Regionalist Parties at Regional and Statewide Elections in 19 Democracies



Note: Figure shows the average party vote share of regionalist parties, averaging across regions within countries and across countries per year. The darker line represents 5-year moving averages, the dotted line represents averages per year. Extreme changes from one year to another are driven by the inclusion of available countries across time and should not be over-interpreted. Data source: Kaiser et al. 2021 *a,b*

dilemma of why parties decentralize. To answer this question, I combine an institutionalist perspective with party-based mechanisms of decentralization by looking at a period of almost 70 years.

First, I develop a neo-institutionalist theory that conditions the party politics of decentralization. Since institutions affect how political actors behave, parties' decentralization preferences will depend on the multi-level institutional setting in place. I argue that the rise of regional democracies – namely when regions acquire full political autonomy through political decentralization – generates a critical juncture in the long-run dynamics of decentralization. Autonomous regional governments and parliaments can then make use of institutional channels and exercise pressure on statewide governments in their own interests. This counterbalance changes the action space of statewide parties to subsequently decentralize or not. Here, I discuss the different rational choice considerations of political parties, especially concerning electoral and ideological strategies.

Second, I expand the study of decentralization dynamics and systematically compare both symmetric and asymmetric decentralization – a research approach not undertaken so far. Additionally, I look at both the statewide and regional electoral arena to

overcome methodological nationalism bias (Schakel 2018) – the tendency to associate regional arena dynamics to disaggregated statewide elections results. To the best of my knowledge, this is the first study looking at different patterns of decentralization (asymmetric and symmetric) across so many countries, covering such a long period of time, and looking systematically at party competition both at the statewide and regional level. The results highlight the impact of regional democracy in asymmetric decentralization and the importance of fiscal authority transfers in the sequential logic of decentralization reforms. At the same time, the study shows the idiosyncrasy of symmetric decentralization and why research must engage more intensively with symmetric decentralization patterns.

Additionally, this research benefits from two innovative and unique data-set resources partly created by myself: the Regionally Disaggregated Statewide Elections Data-Set (RD|SED) and the Regional Elections Data-Set (RED) (Kaiser et al. 2021*a,b*). These data-sets cover statewide and electoral results in 19 democracies from 1941 through 2019 and link electoral performance statistics with parties’ ideologies to track region’s ideology shifts over time.

### **1.3.2 Overview of Study in Chapter 3**

The second study in chapter 3 is titled “The Measurement of Issue Attention across Political Arenas: Exploiting Increasing Returns with Optimized Dictionaries” and is based on an article co-authored with Leonce Röth and Lea Kaftan, both from the University of Cologne. The study engages with the measurement of attention in politics across political arenas. We focus on the issue of territorial politics in the United Kingdom (1900-2018) and Spain (1976-2019), two prominent cases of complex decentralization over time. There is excellent qualitative work on territorial politics in these two countries (see Convery and Lundberg 2017), but it lacks systematic evidence on political attention dynamics on territorial politics across time. Mass media and parliament serve as suitable sources to capture dynamics of attention. We use newspaper data and transcripts from parliamentary debates to measure the level of attention that territorial politics and its sub-issues – such as region-specific decentralization, fiscal authority, or political autonomy – obtain in daily politics.

Attention in politics is a scarce resource and has an important influence on political parties’ decisions on how policy-making and institution-building are decided upon. The procedure allows researchers to efficiently develop optimized dictionaries to implement on texts to identify societal phenomena of interest. Optimized dictionaries are important to avoid biases driven by ideology or researchers’ historical perspectives.

Moreover, we carefully consider the different logics of communication in mass media and parliament, and how optimized dictionaries can help compare these different arenas based on text-as-data methods. We show that optimized dictionaries travel better across different arenas than machine learning methods. The substantial results of the investigation unveil the main differences between the logics of communication in media and parliament. Newspapers in both countries tended to highlight conflicting and negative territorial sub-issues, such as the Basque Country-based terrorism by ETA in Spain or the Northern Ireland conflict in the UK. Representatives in parliament rather emphasized technical territorial sub-issues related to legislation, such as fiscal and administrative authority, as well as political autonomy reforms. These differences have important implications for the politics of attention and representation dynamics between representatives and the citizenry – the latter acquiring daily political information from mass media. We are strongly convinced that such insights and procedural advancements in the measurement of attention across political arenas are highly relevant for political communication research and public opinion studies.

### 1.3.3 Overview of Study in Chapter 4

The third study in chapter 4, “Mediated party positions from newspapers” is also based on co-authored work with Leonce Röth and Lea Kaftan. Building on chapter 3, we elaborate on how measures of attention can be complemented by party positions on territorial politics and more importantly sub-issues of decentralization. Since established political science resources such as electoral manifestos or expert judgments do not track sub-issues of territorial politics, we develop a methodological procedure to gain *mediated* party positions from newspapers.

Mediated party positions complement self-placements by representatives or position assignments by political scientists in that ordinary citizens inform themselves through mass media and are rather receptive to mediated positions. Mediated positions can insofar fill the connection between political elites and the citizenry, and help answer questions of representation and democratic delegation. Citizens can only feel represented depending on how they are informed. Before turning to methodological considerations, we consider how newspapers impact mediated party positions through their own logic of communication, e.g. emphasizing conflicting and negative issues more than others, and eventually altering party dynamics of issue-ownership and issue-salience.

The methodological procedure lays out that sentiment signals are comparable for indirect measures of party positions since they can travel across mass media sources in contrast to wording-estimation. The latter is more likely to be biased by language

driven by journalists’ idiosyncrasies and outlets’ ideological leaning. Subsequently, aggregation of sentiment signals improves randomness within few text passages and allows researchers to acquire fine-grained party positions for low-level time units. Additionally, we introduce a measurement of sub-issue certainty – technically speaking, topic prevalence loads – which informs on clearer party sentiment signals based on the sub-issue content of news’ text passages. We show that our approach converges with established measures to a certain degree (which supports our theory of mediated party positions) and arrives at satisfying levels of internal validity. The study concludes with a practical application of mediated party positions on the dilemma of voter transitions and strategic elites’ policy shifts, as well as with a description of a future research agenda for mediated party positions.

## 1.4 Relevance and Broader Implications

This dissertation has an important impact on territorial politics studies and beyond. Chapter 2 elaborates how decentralization dynamics change across time depending on institutional contexts. The most important insight gained from this investigation is the idiosyncrasy of symmetric decentralization. Although there has existed a focus – quantitatively as well as qualitatively – on asymmetric decentralization for different reasons, symmetric decentralization seems to function under a very different logic. Additionally, Chapter 2 sets the ground to take the rise of regional democracies and by default of autonomous regional executives and assemblies seriously. In this sense, political decentralization appears to be a critical juncture in the long-run dynamics of asymmetric decentralization, but the mechanisms behind it remain blurry. A regional democracy perspective should be included in future work also to uncover the mechanisms behind such dynamics. Chapter 2 also makes use of two innovative and unique data-sets – RD|SED and RED (Kaiser et al. 2021a,b) – that can serve to study electoral geographies, regions, decentralization, and territorial politics more systematically. Additionally, these data resources can be of much interest for other sub-disciplines of political science, such as electoral and representation studies, party politics, and political economy.

Chapters 3 and 4 present groundbreaking work on the analysis of public attention on territorial politics and party positions on territorial sub-issues. The insights and the data generated for Spain and the UK in these issues can establish a political science resource for further work on two of the most prominent cases of decentralization in historical perspective. A systematic benchmark for both saliences and party positions in both cases enables to opening a systematic research agenda to develop accumulative

knowledge across cases.

The studies in chapters 3 and 4 do not just make an impact on the study of territorial politics, but also offer methodological contributions. First of all, the investigation “The Measurement of Issue Attention across Political Arenas: Exploiting Increasing Returns with Optimized Dictionaries” (chapter 3) elaborates how different logics of communication – in mass media and parliament – can be systematically compared, which deals with an extremely relevant interaction between political science and communication studies. Furthermore, we present a procedure to develop optimized dictionaries in chapter 3, e.g. to capture social or political phenomena. Dictionary development is a true challenge in text-as-data scholarship since different types of biases can influence keyword selection and that way bias analysis. The procedure proposes an easy guide for social scientists without a deeper understanding of text methods. This contribution improves the set of resources comparative political scientists can benefit from, especially with a historical focus.

The study “Mediated party positions from newspapers” (chapter 4) develops a procedure to grasp mediated party positions on fine-grained sub-issues from mass media. It introduces a methodological innovation by exploiting the advantages from both sentiment analysis and topic models to pinpoint certain positional signals. Similarly, this promising contribution is accessible for generalist political scientists. Moreover, this methodological investigation to measure indirect party placements in newspapers advances the overall study of party politics and representation studies. It enables these literature strands to fill the gap of *media* in the connection between political elites and the citizenry. I argue that this is probably the beginning of a future research avenue that takes mediated party positions into consideration and compares these with direct statements by parties. From a communication science perspective, it is highly relevant how the portrayal of political parties in mass media alters actors’ own presentations.

## 1.5 Publication Status of the Articles

The solo-authored investigation “The Long-run Dynamics of Decentralization” (chapter 2) has been presented at the Research Seminar of the Cologne Center for Comparative Politics (CCCP, University of Cologne) in November 2018, at the 2019 Conference of Europeanists (CES) in Madrid in June 2019, at the ECPR General Conference in Wrocław in September 2019, and at the Center for Civil Society Research’s Colloquium at the WZB Berlin Social Science Center in June 2021. The chapter version in this dissertation was submitted to the journal Comparative Political Studies (CPS) in

September 2021. However, it has not been accepted for the review process, although the editors offered some short helpful comments. A shorter version will be adapted for a future journal submission.

The study “The Measurement of Issue Attention across Political Arenas: Exploiting Increasing Returns with Optimized Dictionaries” (chapter 3), based on co-authored work with Leonce Röth and Lea Kaftan, has been presented at the 2019 Conference of Europeanists (CES) in Madrid in June 2019, at the 2019 Zurich Text-As-Data workshop in October 2019, and at the CCCP Research Seminar (University of Cologne) in November 2020. The study version in this dissertation has been submitted to the American Journal of Political Science (AJPS) in August 2021 and is currently under review. All authors contributed equally to this investigation.

The manuscript in chapter 4 of this dissertation, “Mediated party positions from newspapers”, which is also based on joint work with Leonce Röth and Lea Kaftan, was presented at the 2019 CES in Madrid in June 2019, at the 2019 Zurich Text-As-Data workshop in October 2019 (in both cases together with the study of chapter 3), and separately at the 2020 ECPR General Conference (online). This investigation is being prepared for submission at the end of the year 2021. All authors contributed equally to this investigation.

# THE LONG-RUN DYNAMICS OF DECENTRALIZATION: REGIONAL DEMOCRACY AND STATEWIDE GOVERNMENTS' STRATEGIES, 1950-2018

*Enric Company (Journalist, El País): “The president of the PP in Catalunya, Josep Piqué, affirmed last Saturday that if the Catalan statute [of autonomy] were approved under the known terms, once the Government changes, the statute would be changed.”*

*Pasqual Maragall (President of the Catalan Generalitat, PSC): “What do you want me to say? That is impossible.”*

– El País (2006)<sup>1</sup>

## Abstract

Understanding the causes and consequences of decentralization is crucial for the study of political systems. Typically, decentralization reforms have been studied independently of their institutional setting, as if they were all the same. I argue that the dynamics of decentralization change following the implementation of political decentralization, and subsequent reforms are driven by a very different causal logic. After political decentralization, the regional arena gains relevance and regional actors can access formalized politics more easily. Incentives therefore change for both statewide and

---

<sup>1</sup>Original text: Enric Company (Journalist, El País): “Pues el presidente del PP catalán, Josep Piqué, afirmó el sábado que, si se aprueba en los términos conocidos, cuando cambie el Gobierno habrá que cambiar el Estatuto [de autonomía] catalán.” Pasqual Maragall (President of the Catalan Generalitat, PSC): “¿Qué quiere que le diga? Eso es imposible.”



regional(-ist) actors. My theoretical argument is located in a historical institutionalist framework. First, I discuss the changing dynamics and pay-offs from initial political decentralization reform to those following. The updated incentive structures make it more reasonable for statewide parties to decentralize under regional political pressures, irrespective of their electoral and ideological relations to the different regions. I test the theoretical argument with data from 19 democracies between 1950 and 2018, using logistic panel-regression models. I find that the emergence of regional democracies is a critical juncture only for governments' decisions on asymmetric decentralization. Symmetric decentralization is insufficiently explained by electoral-ideological incentives, although reforms are more likely after political decentralization. This study has important implications for the study of decentralization and multi-level politics across institutional contexts and time.

## 2.1 Introduction

Decentralization is a political phenomenon fundamentally affecting multi-level politics. When and to what degree authority is transferred to sub-national units of governance defines how states react to crises, how optimal policies can be designed, how citizens in different territories are democratically represented, and how fiscal equalization between regions works (or not). Understanding the causes of decentralization is therefore key to foreseeing the future political developments of the nation-state. However, institutional explanations of decentralization have often lost sight of the long-run dynamics of multi-level institutional changes and the order and kind of reforms across time (Collignon 2019; see Falleti 2005 for an exception).

The neo-institutionalist argument of this study is that political decentralization, and with it the emergence of regional democracies, functions as a critical juncture in the long-run dynamics of decentralization. The institutional change at the regional level generates autonomous elections and autonomously elected governments, and their new formalized access to politics alters the multi-level politics of decentralization (cf. Deschouwer 2009). Further, political parties play a crucial role in these dynamics. Parties in government and parties with parliamentary majorities are the key actors promoting decentralization reforms – and according to more recent studies, parties follow a post-functional logic by attending to their own partisan and ideological interests (Meguid 2015; O'Neill 2003, 2005; Röth and Kaiser 2019).

At the macro level, historical institutionalism helps to explain how political decentralization generates increasing returns through empowered regions after the emergence

of regional democracy, so that subsequent decentralization becomes more likely (Pierson 2000). At the micro level, and according to rational choice theory, parties only decentralize in order to maximize electoral gains. Within this framework, political decentralization generates policy feedback mechanisms endogenously affecting statewide parties' interests vis-à-vis subsequent reforms (Falleti 2005; Shepsle 2006). Within this framework, parties follow electoral and ideological incentives at the time of decentralization, and the multi-level institutional design will in turn affect these incentive structures. Accordingly, regional electoral arenas and the emergence of established regional actors potentially create cross-pressure for statewide electoral parties to relax their electoral maximization logic.

This study investigates how long-run decentralization dynamics change with the emergence of regional democracy and how party incentives to decentralize adapt to the initial major reform of political decentralization, when autonomous regional arenas emerge. The theoretical contribution is the interplay of the historical institutionalist perspective on the general dynamics of decentralization (Falleti 2005) with an analysis of rational choice party mechanisms at the micro level (see Toubreau 2017 for a qualitative approach). After political decentralization, the salience of territorial politics decreases (see chapter 3 in this dissertation), party competition unfolds in a multi-level context where dominant statewide parties begin to lose electoral support due to a division of labour across state levels (Thorlakson 2009), and regional political actors can exercise pressures on the national arena after winning regional elections. Party incentive structures adapt to different levels of decentralization that carry diverse institutional effects on further reforms (Toubreau and Massetti 2013; Verge 2013). The following research questions arise:

*How does the rise of regional democracy affect further decentralization?*  
*How do statewide parties' incentives to further decentralize change after the major political decentralization?*

The innovative contribution of this study to territorial politics scholarship entails a comparison of symmetric and asymmetric reforms (Zuber 2011) in a long-run perspective of almost seven decades, simultaneously accounting for statewide and regional electoral dynamics in order to avoid "methodological nationalism" (Schakel 2018; Schakel and Jeffery 2013). Additionally, I test O'Neill's (2003) and Falleti's (2005) theories of decentralization in a more general context beyond Latin America.

The investigation is structured as follows. First, I link previous research on decentralization and party politics to the neo-institutionalist perspective, and to the major reform of political decentralization and its effects on subsequent decentralizing reforms.

I advance hypotheses on how decentralization and statewide parties' incentives to decentralize change across time and institutional settings. I then describe the research design and quantitative analyses. The first empirical analysis shows overall trends of decentralization in more than 70 countries, using data from the Regional Authority Index (Hooghe et al. 2016). The second empirical analysis focuses on party-based mechanisms of decentralization, based on 19 democracies with large institutional variation from 1950 through 2018, and makes use of a comprehensive dataset of regional electoral data. The results highlight the importance of regional democratic institutions for the politics of decentralization and demonstrate that a different perspective from asymmetric decentralization is necessary to study symmetric reforms. The study concludes with a discussion of the role of regional democracy in the long-run dynamics of decentralization.

## **2.2 The literature on decentralization and party politics**

Decentralization is conceived as the transfer of authority from upper levels of government, normally the nation, downwards to sub-national governance tiers (Dardanelli et al. 2019a; Hooghe et al. 2016). We can distinguish between self-rule reforms – decentralizing political power, policy competences, fiscal rights, and administrative responsibilities – and share-rule reforms – migrating joint veto rights for constitutional decision-making at the national level to sub-national governments, as with federal second chambers of parliaments (Elazar 1987). Since the 1950s, decentralization has become a political trend in established and new democracies (Hooghe et al. 2016; Niedzwiecki et al. 2018).

A vast body of research has focused on how decentralization reforms affect economic efficiency (Alesina, Baqir and Easterly 1999; Oates 1999), state corruption (Bardhan and Mookherjee 2006), democratic quality (Faguet 2014; Stepan 1999), and resolution of territorial conflicts (Brancati 2006, 2008; Keil and Anderson 2018), among many other topics.

Comparative electoral studies investigated how territorial state organization affects party competition and creates multi-level cross-pressures for political parties in both statewide and regional arenas (Alonso 2012, p. 48; van Biezen and Hopkin 2006, pp. 35-36; Swenden and Maddens 2009; Thorlakson 2006). Case-study insights have shown how political decentralization fosters the channelling of regional interests through regionalization of the statewide party system (Pallarés and Keating 2003; Deschouwer

2009).

A different line of research focuses on explanations of decentralization. Jurado and León (2021) demonstrate how decentralization to regions is accompanied by globalization as upward economic integration into global markets. More broadly, Dardanelli et al. (2019b) find that a conjunction of socio-economic and cultural factors better explain decentralization dynamics in federations, and that fine-grained centralization is more common than previously thought. Institutionalist approaches have focused on structural dynamics and short-term actor-based perspectives by qualitatively studying emergent architectures of federations after critical junctures (Toubeau 2017). Falletti's qualitative work (2005; 2013) profoundly considers institutional context and differentiates between types of decentralization reforms and the sequences of these reforms. A central concept is political decentralization, which fuels the emergence of regional democracy with the transfer of political power to autonomous sub-national governments. Sub-national governments are no longer designated by or dependent on central government, but instead are democratically elected in fair elections by a regional demos (see Treisman's (2007) discussion of appointment and electoral decentralization). This allows them, for the first time, to exercise pressure as established and formal political actors and bargain with statewide governments in their own interests. Other decentralization reforms (policy, administrative, fiscal) play a lesser role in explaining procedural sequences.<sup>2</sup> Yet, historical perspectives focusing on *change* and *persistence* are difficult to transpose to dichotomous phenomena such as decentralization reforms across time (Broschek, Petersohn and Toubeau 2018).

More recent contributions have responded to the post-functionalist call to bring politics back in (Hooghe and Marks 2018) and focus on party-based mechanisms explaining decentralization. Statewide political parties enact decentralization reforms according to a rational choice strategy of expansion and preservation of electoral and office benefits. The literature identifies two complementary mechanisms that bridge party politics with decentralization. On the one hand, parties might aim to maximize their electoral support in the emerging regional arena by decentralizing (O'Neill 2003, 2005).<sup>3</sup> Governing statewide parties that expect less national-level electoral support in the future will politically decentralize to gain and hold office in, at least, certain regions. With this trade-off between national and regional elections, statewide parties

---

<sup>2</sup>In the following, unless indicated otherwise, "decentralization reforms" will refer to fiscal, policy competence and administrative (non-political) transfer of authority in contrast to "political decentralization".

<sup>3</sup>A fundamental feature of O'Neill's work is her case selection of presidential democracies in Latin America during the 1990s, which have specific party system structures systematically different from West European parliamentarism.

expect to dominate the emergent regional arena and reinforce regional support there in exchange for imminently loosing national office.

On the other hand, statewide governments can intend to neutralize regionalist parties' threats at the national level by gaining support from satisfied regionalist voters after decentralizing (Meguid 2009, 2015; Toubreau and Wagner 2016; Sorens 2009).<sup>4</sup> This explanation perceives regionalist parties as demanding decentralization, not just statewide parties promoting reform in isolation. Furthermore, territorial electoral vulnerability matters; since statewide parties will weigh regions by their electoral vulnerability when considering decentralization, in order to maximize support at national elections (Alonso 2012). In addition, ideological incentives can partially explain when decentralization occurs (Massetti and Schakel 2016). When the national government and the regional level share similar ideologies, statewide parties are more prone to decentralize authority strategically to gain ideological allies as cooperative veto players or as coalition partners for multi-level policy-making or shared-rule constellations (Röth and Kaiser 2019; Toubreau and Massetti 2013).

Moreover, statewide parties choose the positioning on, the tempo and the type of decentralization reform rationally but within specific political environments: regionalist threats, opposition dynamics and intra-party pressures are fundamental (cf. Meguid 2005; 2008, p. 102). Monocausal theories of conservative ideology have not successfully explained the popular cases of the British Conservatives' pro-periphery attitude towards Scotland in the 1960s or Partido Popular's (PP) changing strategy on decentralizing the Spanish autonomous communities (Convery and Lundberg 2017). For example, PP changed its preferences for decentralization between its period in opposition in the 1980s and when in minority government supported by regionalist parties in 1996 (Pallarés and Keating 2003, 241; van Biezen and Hopkin 2006).

Studies on party-based explanations of decentralization have mostly focused on comparatively investigating asymmetric decentralization reforms (Alonso 2012; Amat and Falcó-Gimeno 2014; Meguid 2009; Röth and Kaiser 2019) and to a lesser extent symmetric decentralization (Meguid 2015; in Latin America: Falleti 2005; O'Neill 2003, 2005; see Dardanelli et al. 2019b). The former reform type consists of decentralization towards a specific region or a small number of regions, as with Catalonia, Scotland or the Italian regions, with special statutes and bilateral relationships with the center. The latter reform types include reforms that decentralize authority to all regions of a

---

<sup>4</sup>Statewide parties refer to political parties running for election in the whole national arena without specific regional foci. If not indicated otherwise, by "regionalist parties" I mean ethno-territorial and non-ethno-territorial parties with regionalist demands running for office in either one or a few regions.

country like with the Länder in Germany or the French regions (excluding Corsica). So far, no systematic work has investigated O’Neill’s regional electoral expectations beyond Latin American democracies, and Meguid (2015) does not comprehensively elaborate further on national arena incentives beyond attending to regionalist parties’ threats. Explanations of symmetric decentralization remain under-theorized. The symmetric-asymmetric divide can be driven by different dynamics, especially since electoral incentives based on one or many regions are politically different in substance and conceptually challenging for comparisons. The following section theoretically integrates macro-dynamic expectations based on Falleti’s neo-institutionalist arguments with party-based mechanisms explaining decentralization reforms.

## 2.3 Theory

### 2.3.1 Decentralization reforms and institutional change

The territorial organization of the state largely determines how policies and politics are shaped. At the same time, politics and policy influence the design of multi-level governance. In the 19th century, the rise of Western and modern nation-states foreshadowed a centralization of policies across state levels. The process was strongly driven by rising liberal and socialist political forces that aimed to universalize public services offered by a functional state (Ansell and Lindvall 2020). In the 20th century the eclipse of authoritarian regimes in Western Europe allowed established democracies to be created and become established. Democratization, as well as new economic risks, diversified political conflict lines. The center-periphery political cleavage re-emerged and its salience increased in specific contexts by structuring the tension between regional demands and central power (Rokkan and Flora 2000).

From the 1950s the pendulum swung away from centralization (Hooghe et al. 2016; Niedzwiecki et al. 2018). Yet, most of the conceptual literature on these trends and their causes and effects has not explicitly elaborated on *how* the dynamics of decentralization fundamentally change across time (O’Neill 2003, 2005; Röth and Kaiser 2019, among others; see exceptions by Falleti 2005, 2013). For this purpose, historical institutionalism enriches the analysis of how institutional design affects decentralization patterns in the long-run.

The argument is simple: not all decentralization reforms are equal. Falleti (2005) outlines how different decentralization reform sequences differ. Substantially, political decentralization is a game-changer, unlike transfers merely of policy competences, fiscal or administrative rights. Political decentralization is itself a fundamental institutional

change. Other decentralization reforms carry little substantial political weight when the system has not been politically decentralized, since political power remains *in the center*. However, after political decentralization, established regional governments can, as autonomous actors, exercise pressure to obtain benefits from further decentralization reforms. What kind of transfers follow political decentralization is crucial. For example, fiscal decentralization empowers regional governance by reinforcing its financial autonomy. Conversely, transfers of policy competences and administrative authority can burden regions by generating responsibilities and costs.

In Pierson's (2000) view, political decentralization promotes the emergence of regional democracies introducing increasing returns and thus generates a path dependence:

To put it a different way, the costs of exit – of switching to some previously plausible alternative - rise. Increasing returns processes can also be described as self-reinforcing or positive feedback processes (Pierson 2000, p. 252).

Insofar as this is true, political decentralization is *a critical juncture* in multi-level politics.

Political decentralization can also affect national party competition and generate subsequent endogenous institutional changes (Thorlakson 2009; van Biezen and Hopkin 2006, p. 35; cf. Capoccia and Kelemen 2007). The reinforcement of democratic regional institutions and the establishment of power asymmetries between governance levels can become irreversible, producing a ratchet effect along a new path (Kaiser 1997, pp. 434-435). In this framework, recentralization becomes an unpopular measure that can trigger regional resistance or even fuel secessionist demands (Collignon 2019, p. 179). Major institutional reforms and crises in the Europeanization processes have similarly acted as critical junctures, causing comparable transformations of political dimensions and reorganizing preferences in the conflict space (Hooghe and Marks 2018; Kriesi et al. 2012).

Historically, political decentralization produced a major institutional shift towards the regionalization of territorial politics, for example in Belgium, Italy, Spain, Switzerland and the United Kingdom even at the end of the 20th century, the century of nationalized politics (Caramani 2004, pp. 37, 93). After political decentralization, the regionalization trend implies only that the regional arena becomes relatively more important for the organization of political conflicts and statewide governments' interests than the national arena. Politics remains organized around the nation-state, but

responses are driven partly by the regional or European arenas (Caramani 2004, p. 37; 2015; Hooghe and Marks 2018; see also Kriesi et al. 2012). Within this process, once autonomous regional elections emerge, political actors shift part of their attention to the regional electoral arena, the statewide arena having previously been the only focus of attention.

The French decentralization reforms of the 1980s illustrate how political decentralization is a critical juncture with irreversible consequences that generate increasing returns. After a long democratic period with a strong central state, the Parti Socialiste entered government and enacted political decentralization with the Defferre laws:

Once the institutional reform had been passed, Defferre felt secure that the process of decentralization would become irreversible; indeed, he anticipated that the demands of local authorities would create pressures for further reforms (Levy 1999, p. 137).

Attaining government was rather surprising even for the party itself and it expected to lose office to the conservatives relatively soon. Political decentralization was an institutional reform that its promoter Gaston Defferre believed would be extremely hard to reverse, and would benefit the Socialists by securing at least some of the emerging regional governments. Far from being a technical or consensual decision, the reform was explicitly intended to transform the power equilibrium (Keating 1983). According to the theoretical arguments of endogenous institutional change and increasing returns, political decentralization should then promote further transfers of authority to the regional level. From this, the further expectation can be derived:

**H1:** *After political decentralization, subsequent decentralization reforms become more likely.*

The effect of political decentralization on further decentralizing reforms has till now been elaborated as a macro phenomenon. However, many previous explanations involve, in different manners, the roles of political parties and statewide governments – the actors passing such reforms. The following section focuses on the party-based mechanisms behind decentralization reforms and how the mechanisms change once political decentralization stimulates the rise of regional democracies.

### **2.3.2 Regional democracy and party-based mechanisms of decentralization**

While party-based explanations of the long-run dynamics of decentralization have treated all reforms equally (Meguid 2015; Röth and Kaiser 2019), in fact the effect



of political decentralization on historical dynamics will influence statewide parties' decisions on *when* to decentralize.

The critical juncture of political decentralization can generate endogenous institutional changes. These endogenous changes condition how rational actors structure their preferences for institutional or policy choices across decision periods (cf. Riker 1980; Shepsle 2006). When regions acquire a minimum level of political autonomy, and political attention shifts to the regional arena, regional actors can detach from the national arena and generate multi-level cross-pressures for statewide parties (Alonso 2012, p. 48; Swenden and Maddens 2009; Thorlakson 2006; van Biezen and Hopkin 2006, pp. 35-36). Regional actors with effective autonomy then improve their bargaining position vis-à-vis the statewide government in order to negotiate further decentralization in their own interest (Falleti 2005; Field 2016).

How does this mechanism work in detail? In initial period, statewide parties will assess their preferences on decentralization in terms of maximizing electoral support or benefitting ideological allies. Their calculations are based on benefits they anticipate *after* the institutional or policy reform they aim to enact (Amat and Falcó-Gimeno 2014; see Tsebelis 1990).

But in the second period, that following political decentralization, unpredicted changes can occur (Shepsle 2006, p. 1046; see also Stewart and Weingast 1992). The different components of strategic party competition – such as salience, position, and issue-ownership – might be reshaped (Meguid 2005; Toubreau and Massetti 2013; Verge 2013) and territorial politics, as well as core-periphery conflicts, may become more important (Mazzoleni 2009; see chapter 3 in this dissertation). Statewide parties will confront relatively large and uncertain fixed costs. Loss of electoral support, caused by a division of labor across state levels, is a threat, and learning, adaption, and coordination will affect statewide parties' cost-benefit calculations for subsequent reforms (cf. Falleti 2005; Levi 2004; Pierson 2000; Thorlakson 2009).

The addition of established regional actors will also alter multi-level politics. Reactions to decentralization by third actors might provoke so-called feedback mechanisms (Shepsle 2006; cf. Bawn 1993; Zuber 2011).<sup>5</sup> Decentralization can lead to regionalist parties ramping up their demands for further autonomy, which in turn put pressure on statewide governments (Massetti and Schakel 2013; 2016; 2017). Consequently, the governing statewide parties' pay-offs from decentralizing competences will be changed

---

<sup>5</sup>Zuber (2011) discusses unstable bargaining situations in symmetric and asymmetric decentralization strategies and between decentralized and non-decentralized regions, and how such configurations foster further reforms.

by to the new context of the relocation of the power equilibrium towards the regions (cf. Greif and Laitin 2004; Pierson 2000, p. 252). A similar dynamic occurred in Mexico after a long period of one-party hegemony across governance levels. When the PRI government decided to effectively allow regional electoral competition in 1989, PRI and opposition parties started sharing regional executives. Regional governments – even PRI ones – began to counterbalance the statewide government and press for more autonomy (Hernández Rodríguez 2008). Eventually, the statewide government faced a fragmented, heterogeneous multi-level party system. PRI lost the presidential elections in 2000 after more than seven decades in power – against a regional governor as candidate.

Political decentralization, then, will affect *how* statewide parties decide on whether to embark on further decentralization reforms or not. They will change their logic of attention towards regional resources and electorates in order to secure them (Collignon 2019; Sorens 2009; Swenden and Maddens 2009, p. 5). Previous studies suggest that statewide parties opt for regional office in exchange for national government; seek to increase support from regional voters at national elections; or strategically search for ideologically adjacent partners (Meguid 2009; Röth and Kaiser 2019; O’Neill 2005). Since O’Neill (2005) focuses on Latin American presidencies and Meguid (2009) only on the statewide arena, I take Röth and Kaiser’s (2019) notion as the theoretical benchmark. Accordingly, statewide parties will decentralize when they benefit from large electoral support in the regions or share similar ideologies with the regional actors. This incentive structure will adapt to the processes triggered by political decentralization, while their main goal – maximization of electoral net gains and ideological allies – remains constant. Nevertheless, the balance of seeking power at either the national or the regional level is shifted towards the latter – in other words shared between both levels (Aalen and Muriaas 2017). Hence, regional electoral support and ideological proximity to regional actors in statewide elections should be less relevant after political decentralization, since statewide parties are under pressure and have to *lower their expectations*.<sup>6</sup> Meanwhile, regional elections’ incentives grow more important. By rational choice logic, statewide parties will then shift their attention and decentralize, based on large electoral support and ideological proximity to the regional arena. Put simply, after political decentralization and the emergence of regional democracies, regional actors and elections become essential for statewide parties (see also Field 2016).

## **H2.1:** *The greater governing parties’ electoral support in and ideological*

---

<sup>6</sup>An additional argument emphasizes that after political decentralization, subsequent reforms may be rather fine-grained concessions. Fiscal, administrative, and policy reforms are easily reversed and involve less political risk.

*proximity to the regions in statewide elections, the more likely are decentralization reforms.*

**H2.2:** *After political decentralization, governing parties' electoral support in and ideological proximity to the regions in statewide elections are not related to subsequent decentralization reforms.*

**H2.3:** *After political decentralization, the greater governing parties' electoral support in and ideological proximity to the regions in regional elections, the more likely are subsequent decentralization reforms.*

However, an alternative dynamic is plausible if regional governments are powerful enough to influence the process. Political decentralization enables regional and regionalist parties to raise attention and shape the statewide public agenda. These actors benefit from more bargaining power, permeate formalized politics more easily, confront statewide parties in the electoral arena, exercise coalition pressures, veto legislation through shared-rule and exercise agenda-setting powers (Field 2009; Freidenberg and Cao 2014; Harbers 2010, p. 607; Klesner 2005; León 2006, pp. 155-156; Meguid 2005, pp. 158, 165; Pallarés and Keating 2003, p. 241; Thorlakson 2006, p. 51).<sup>7</sup> These developments trigger two new phenomena that are relevant for the process of decentralization and are worthy of discussion.

First, the sequence and type of reform (fiscal, administrative, policy) in the dynamics of decentralization substantially influence how regional actors improve their strategic situation when negotiating with the national government (Falleti 2005, 2013). This can reinforce and catalyze further decentralization reforms through policy-feedback effects, assuming regional governments want to obtain more authority with benefits. After political decentralization, regional actors have a higher power benchmark vis-à-vis statewide parties. It is therefore important to account for the level of authority *already* transferred in order to explain decentralization reforms. According to Falleti (2005), the more authority and competences powerful regional actors gained from political decentralization and the more pressure they can exercise, the more likely statewide parties are to subsequently decentralize fiscal authority (beneficial for regions) and the less likely they are to subsequently decentralize only policy and administrative authority (costly for regions). This assumption stands as an alternative and can cancel out the expected effect of H2.3 at high levels of decentralization. Similarly, in 1990s Mex-

---

<sup>7</sup>This dynamic can be found in Spain from the 1990s, when regionalist parties began to support statewide minority governments (Field 2016; Pallarés and Keating 2003, pp. 250, 252). Similarly, intra-party pressures from regional branches become more influential in less integrated multi-level systems, where powerful regional leaders who outperform the national party can demand more competences (Hopkin 2009; Meguid 2005, p. 277; Swenden and Maddens 2009; van Biezen and Hopkin 2006).

ico after political decentralization, regional governors were eventually able to organize institutionally to jointly influence the statewide government's decisions and partisans in congress (Hernández Rodríguez 2008, pp. 269ff.). It remains an empirical question how these contrary expectations play out across cases.

Second, formal access to regional politics and regional policy influence also affect regionalist parties (Pallarés and Keating 2003, p. 250). Political decentralization should moderate regionalist parties' demands for subsequent decentralization, since their major demands are largely satisfied. In this case, regionalist parties diversify ideologically towards mainstream positions while keeping territorial demands salient (Alonso 2012). Regionalist parties that follow the moderation strategy can successfully negotiate further authority concessions in order to benefit electorally.

Yet, regionalist parties are very heterogenous (Chandra 2001; cf. Meguid 2005; Sorens 2009). New regionalist parties may even emerge long after political decentralization – like Lega Nord in Italy and Coalición Canaria in Spain. In contexts of territorial polarization, secessionist movements and parties can arise with radicalized autonomy demands (Masseti and Schakel 2016, 2017), especially in asymmetric regions (Zuber and Szöcsik 2021). The radicalization of regionalist parties can carry ceiling effects in the decentralization dynamics. Statewide governments might fear to fuel the fire of secessionism by accommodating radical regionalist demands and that way risk democratic stability and the territorial integrity of the state (see Brancati 2006). It is therefore important to account for different regionalist parties electoral strength. Stronger autonomist parties with moderate demands might foster decentralization reforms, especially after autonomous regional governments emerge. Conversely, stronger secessionist parties with radicalized demands might brake decentralization, especially after political decentralization, when statewide parties might most fear state fragmentation.

To sum up, political decentralization will affect how and when statewide parties decentralize since the relative equilibrium of power shifts to the regional level. In turn, regional actors acquire more power and increase their capacity to demand further reforms, assuming this is in their interest. Electoral and ideological incentives articulated at regional elections therefore gain relevance compared to statewide election incentives. The following section empirically tests these expectations in comparative perspective.

## 2.4 Empirical study

### 2.4.1 Research design

This section outlines the two quantitative analyses conducted to test the theoretical expectations. A first descriptive part focuses on macro-level decentralization dynamics (the dependent variable) in a large sample of 75 countries with 194 region groups, and a second part focuses on party-based explanations of decentralization comprising 19 countries: 14 Western, three South-East European, and two Latin American democracies.<sup>8</sup> The period under investigation ranges from 1950 to 2018. The country variance of the party mechanism analysis aims to make inferences about decentralization and institutional characteristics more generalizable. The cases include 39 where some regions – like Catalonia, French Corsica, Mexico City, and Scotland – experience asymmetric decentralization, and 20 characterized by symmetric decentralization, for example in the Turkish Iller, Australian States, and German Länder.<sup>9</sup> In order to empirically test the theory on party explanations of decentralization, the unit of analysis is yearly dyads of statewide governments and their regional performance, reflecting cabinet parties’ ideological incentives at the sub-national level (independent variable) and whether they enacted a decentralization reform or not (dependent variable).

#### 2.4.1.1 Dependent variable

The dependent variable is decentralization reforms measured by the Regional Authority Index (RAI) (Hooghe et al. 2016; Shair-Rosenfield et al. 2021). RAI is an additive 27-point indicator composed of self-rule and share-rule in the regional arena. Self-rule consists of 15 points distributed among fiscal authority, institutional depth, policy competences, borrowing autonomy, and representation; and share-rule of 12 points covering shared national legislation, fiscal, executive and borrowing national coordination, and constitutional reform rights (see Elazar 1987 for conceptualization). A reform event is binary registered if any component of the RAI value increases, indicating an authority transfer to the regions by the statewide government in a specific year.<sup>10</sup>

---

<sup>8</sup>Australia, Austria, Belgium, Bosnia and Herzegovina, Canada, Denmark, France, Germany, Italy, Mexico, Nicaragua, Norway, Portugal, Serbia, Spain, Sweden, Switzerland, Turkey, United Kingdom.

<sup>9</sup>The symmetric-asymmetric divide follows the categorization by Hooghe et al. (2016). See sections SM2.3 and SM2.4 for methodological details and distribution of the cases.

<sup>10</sup>Overall, the RAI has been criticized for missing details in the policy dimension and for not emphasizing recentralization strongly enough (Dardanelli et al. 2019a). It nevertheless provides the most comprehensive data available for the comparative study of decentralization.

#### 2.4.1.2 Independent variables

The independent variables are lagged by one year to account for the Granger causal direction and for the timeframe of legislative procedure from initial proposal to actual reform. The measurement of incentive structures relies on the Regionally Disaggregated State-wide Elections Data-Set (RD|SED) and the Regional Elections Data-Set (RED) (Kaiser et al. 2021*a,b*), which cover disaggregated statewide and regional electoral results and provide information on cabinet participation and party information. I make use of the *Centers of Gravity* to determine the ideological balance of a region based on election results (Kaiser et al. 2021*a,b*; Röth and Kaiser 2019), which is the mean party ideology weighted by the sub-national electoral vote share. The advantage of this measurement is that it captures inherent electoral support, since the higher the vote share of the statewide governing party, the more proximate is the region ideologically. It then works as a continuum accounting for the electoral performance of the rest of the parties.

In order to account for ideology, I focus on the economic dimension of state interventionism-market liberalism (Röth 2017), which is more stable and therefore more readily comparable across time and different contexts than a left-right dimension or a cultural dimension. The state-market ideological estimates are based on an item-response model using economic components of the Comparative Manifesto Project (Volkens et al. 2020). Ideological proximity to a region is then built in reference to the statewide cabinet’s ideology, a continuous variable from the most distant (0) to the most proximate (1). For coalition governments, ideological proximity to a certain region builds on a weighted average formed by the proportion of seats in national parliament held by parties in governing cabinet (see Döring and Schwander 2015). For symmetric decentralization involving several regions, incentives – ideological proximity weighted by electoral results – are weighted by national parliament seat distribution of the regions to mirror territorial electoral vulnerabilities (Alonso 2012). For asymmetric decentralization reforms, the relationship to the region is straightforward, and incentives are calculated without weights.

To assess whether the ideological incentive structures of statewide parties concerning decentralization change with political decentralization – that is, the emergence of regional democracies with autonomous and competitive elections – the binary variable “politically decentralized” indicates regions fulfilling this condition. Defining the time juncture of political decentralization is guided by information in the *representation* dimension of self-rule in RAI, although there are some disagreements. For example, the index does not recognize French regions as politically autonomous because of their cen-

trally appointed *préfets* with legal oversight functions over regional executives; it deems that, the *préfets*' existence notwithstanding, political decentralization in France dates from 1986. Generally, though, political decentralization in France is marked starting in 1986. Withal, if the statewide government appoints or co-selects powerful regional executives, the fundamental condition of political decentralization remains unfulfilled. The timelag between the reform conceding political autonomy and holding the first regional elections also adds complications. The case selection encompasses different period coverage of political decentralization: from the beginning in some federations (e.g. Canada, Germany) or decentralized, non-federal states (e.g. Italy's special statute regions); within the period of observation, as a critical juncture (e.g. France, Denmark, Spain); and negative cases still lacking political decentralization (e.g. the Turkish *İller* and Nicaraguan *departamentos*).

To track the strategic evolution of regionalist parties, Massetti and Schakel (2016) identify parties along the centre-periphery dimension as moderate autonomist or radical secessionists. The aggregated vote shares of each group of regionalist parties capture that group's political influence in its region, based on electoral strength and leverage to push further decentralization.<sup>11</sup> Finally, all models include controls for whether decentralization had already occurred during the current legislature, since such reform processes tend to be unique for each legislative period (reform passed, dummy); duration of main cabinet party in government, reflecting expectations regarding electoral stability and experience with regional dynamics (government participation, logarithm of years count); the number of regions in case of symmetric decentralization (number of regions, count number); whether the case is an EU member (dummy); and the level of electoral democracy assigned by the polyarchy indicator of the V-DEM project to account for democratic and competitive party dynamics (Coppedge et al. 2020).

#### **2.4.1.3 Model specifications of party explanations of decentralization**

The statistical models are logit regressions considering the panel structure of the data by enabling random intercepts at the level of the country-region dyads. A logit estimator seems suitable, since decentralization reforms (Y) are "common" enough with 5.5 per cent occurrence and specific models for rare events are not necessary (e.g. bias reduced or penalized maximum likelihood estimation) (see King and Zeng 2001). Random effects specification serves to model institutional effects across units rather than

---

<sup>11</sup>It is important to highlight that, depending on the analysis and variable selection, the data sample changes, because drawing from different data sources on comparative political institutions omits certain periods or cases due to periods without overlap. In particular, data based on expert judgments on regionalist parties is hard to code on one's own.

within units, especially since institutionalized incentive structures and their measurements are comparable across contexts and time, and since the samples in the regression analyses are relatively large for TSCS data. Besides, no inferences at the level of the units of analysis are intended (Bell and Jones 2015; Clark and Linzer 2015, p. 403). Due to the model specification, many variables seem to be endogenous. However, the measurement avoids auto-correlation issues, since the dependent variable of a *reform* is conceptualized as a recurrent event, the lagged level of RAI depicts the *degree* of decentralization, and political decentralization (*dummy*) can be interpreted as the independent variable with constant influence (see Amat and Falcó-Gimeno (2014) for an auto-regressive approach in decentralization research). The following section presents and describes the results.

## 2.4.2 Results

### 2.4.2.1 Decentralization reforms and regional democracy

A first description of decentralization dynamics is mandatory to test whether decentralization reforms are more likely, in general, after regional democracies emerge with political decentralization (H1). I look first at global trends of the RAI across 75 countries with 194 region groups, which represent all cases available that achieve a minimum level of electoral democracy as measured in the V-DEM project (polyarchy indicator  $>0.3$ ).<sup>12</sup> Figure 2.1 depicts simple linear trends of decentralization for different political decentralization statuses across time and since the beginning of the democratic period of each case. Overall, cases having regional democracy from the beginning, normally federations, experience more decentralization than those that have never decentralized politically. However, the within-status comparison shows a different picture (right panel). Cases that experience political decentralization after a certain period of democratic rule do not dramatically change their patterns of overall decentralization. Decentralization even diminishes compared to the previous period, although strong self-selection into a decentralization track is plausible.

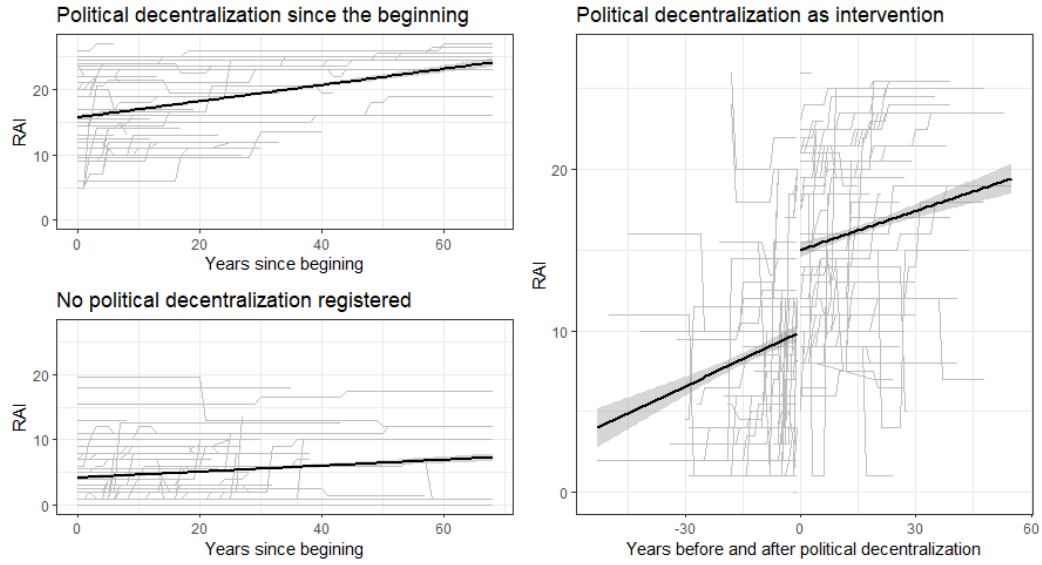
Table 2.1 presents a more differentiated picture by looking at the federation-union and asymmetric-symmetric divides in separate. The table takes the average number of reforms every 10 years, and not on the level or intensity of the decentralization reforms. This binary logic offers a more comparable assessment, since often the RAI assigns the same degree of decentralization for non-comparable reforms across countries. It is clear that decentralization reforms are passed more often *after* regional democracy arises only for symmetric non-federations, like France or Norway. Yet, the more strictly the demo-

---

<sup>12</sup>India is excluded due to complex aggregation of region groups and status changes across time.



**Figure 2.1:** Trends of regional authority across time and by status of political decentralization



Notes: Black lines depict linear trends of RAI across time; light grey lines represent one case each. 75 democratic countries with 194 region groups, 1950-2018.

cratic criteria are set (increasing polyarchy indicators), the clearer is a differentiated pattern showing more decentralization for both symmetric and asymmetric reforms in unitary systems (values not shown). The federations' pattern remains stable even at higher levels of democracy. In federations, more decentralization reforms occur *before* political decentralization.

This simple fact apparently confounds with H1, which expects that after political decentralization further reforms will be more likely. It is very plausible that federations have already reached a ceiling if regional democracies are in place, and few "potential" possibilities for decentralization remain. Political decentralization might thus affect party-based mechanisms of decentralization differently in federations and unitary systems. This further institutional factor – unions vs. federations – seems important in order to understand the long-run dynamics of decentralization. Furthermore, re-centralization reforms are less frequent across all institutional configurations than decentralization reforms. Also, within institutional configurations, re-centralization reforms are always less frequent *after* political decentralization has occurred, in line with the irreversibility argument of generating increasing returns. In short, the expectation that political decentralization, in the form of emerging regional democracies, catalyzes further decentralizing reforms seems to hold only for unitary countries. Federations may experience high levels of decentralization with their instantiation or alongside

their major political decentralization reform.

**Table 2.1:** Decentralization and re-centralization reforms in different institutional settings

			Number of reforms every 10 years		
			<i>N</i>	DEC	REC
No federation	Asymmetric	No political decentralization	1,652	0.430	0.085
		Political decentralization	1,390	0.403	0.058
	Symmetric	No political decentralization	930	0.366	0.129
		Political decentralization	369	0.407	0.081
Federation	Asymmetric	No political decentralization	828	1.200	0.556
		Political decentralization	1,054	0.683	0.209
	Symmetric	No political decentralization	78	1.280	0.385
		Political decentralization	612	0.343	0.163

75 democratic countries with 194 region groups, 1950-2018.

Note: The category "No political decentralization" includes the original political decentralization reform establishing regional democracy, since that took place in a non-politically decentralization context.

#### 2.4.2.2 Party-based mechanisms of decentralization

This sub-section investigates party-based mechanisms exclusively explaining decentralization and accounting for the general dynamic found here. It focuses on electoral-ideological predictors of decentralization reforms in 19 countries and 59 region groups. Table 2.2 shows the main results, finding differences between asymmetric and symmetric reforms. Ideological proximity is positively related to asymmetric decentralization reforms, but this is not statistically significant. The interaction with political decentralization finds that the explanatory power of ideological incentives even diminishes (see Figure 2.2). Symmetric reforms reveal the opposite pattern. Here, ideological proximity is positively associated with decentralization reforms *only after* political decentralization, whereas a negative relationship previously existed. Yet, neither interaction is statistically significant. To connect the hypothesis test with the type of pattern that fits H1 (symmetric reforms being more likely after political decentralization; see previous sub-section), separate models specifying effect on unitary systems in isolation were estimated (not shown here). These models show a stronger positive effect, but are still not statistically significant. The evidence supports neither H2 nor H2.1. Statewide governments' electoral-ideological calculations do not seem to systematically predict decentralization reforms; nor does political decentralization influence this patterns.

As indicated above, different types of reform – fiscal vs. policy/administrative – involve different levels of benefits and costs. Falleti (2005) argues that powerful re-

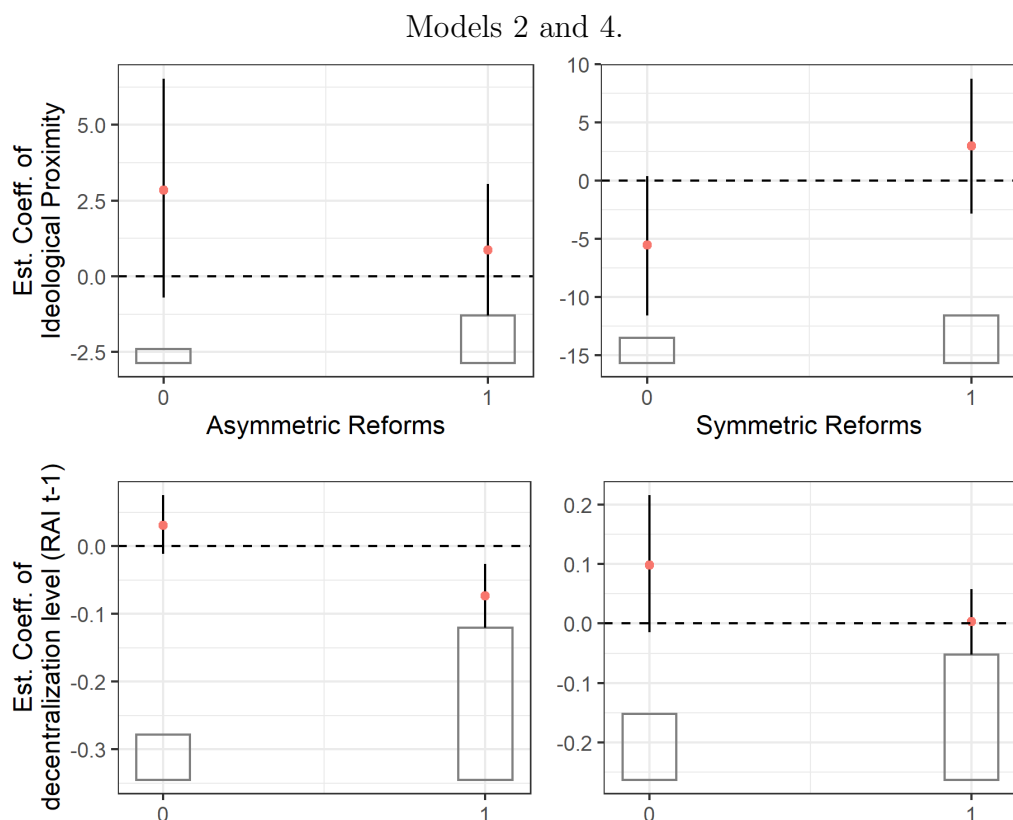
gional actors would only demand more fiscal authority to spare responsibility burdens. Why would an autonomous region demand more policy responsibilities under the same financial conditions? In order to avoid the statewide government offloading administrative tasks (burden-shifting), regional actors would resist policy and administrative decentralization but welcome fiscal authority decentralization. Refocusing Table SM2.6 to fiscal decentralization reforms alone confirms the notion that different reforms weigh in differently for statewide parties (see section SM2.6 in the Supplementary Material). Before political decentralization, ideological proximity is positively and significantly associated with fiscal decentralization reforms. This pattern only applies for asymmetric decentralization, which again demonstrates that a different logic of decentralization applies for symmetric dynamics.

**Table 2.2:** Ideological Incentives and Degree of Decentralization *before* and *after* Political Decentralization explaining Reforms

	DV: Decentralization Reform (0 — 1)			
	Asymmetric Reforms		Symmetric Reforms	
	(1)	(2)	(3)	(4)
Ideological proximity	1.489 (1.127) [0.187]	2.847 (2.185) [0.193]	−0.360 (2.685) [0.894]	−5.478 (3.676) [0.137]
RAI <sub>t-1</sub>	−0.030 (0.016) [0.066]	0.031 (0.027) [0.250]	0.037 (0.023) [0.109]	0.097 (0.070) [0.166]
Regional democracy		2.773 (2.366) [0.242]		−6.197 (4.171) [0.138]
Regional democracy x Ideological proximity		−1.963 (2.624) [0.455]		8.447 (4.837) [0.081]
Regional democracy x RAI <sub>t-1</sub>		−0.104 (0.038) [0.007]		−0.095 (0.076) [0.211]
Controls	✓	✓	✓	✓
Random effects	✓	✓	✓	✓
N	1,673	1,673	1,003	1,003
Regional level units	39	39	20	20
Countries	13	13	18	18
Log Likelihood	−402.266	−398.117	−167.089	−164.855
AIC	820.533	818.233	352.178	353.710
BIC	863.912	877.879	396.375	412.639

Notes: Standard deviation in brackets; p-values in square brackets.  
Full model in section SM2.5 in the Supplementary Material.

**Figure 2.2:** Effect of Ideological Incentives and Degree of Decentralization on Further Reforms Moderated by Political Decentralization (0 — 1)



Note: Error bars reflect .90 level of confidence intervals.

The regression models in Table 2.2 also account for the lagged level of decentralization ( $RAI_{t-1}$ ). Asymmetric and symmetric reforms show a similar pattern, but one more pronounced for the former (statistically significant). *Only after* political decentralization, the more decentralized a region or a group of regions is, the less likely does subsequent decentralization become, contrary to expectations from the literature.<sup>13</sup> More powerful regions put the brakes on decentralization after the emergence of regional democracy – driven in federal systems, as mentioned above, by a ceiling effect.

The regression analyses in Table 2.3 focus on the role of autonomist and secessionist parties' electoral support. Contrary to expectations, autonomist and secessionist parties are inadequate to explain decentralization reforms, both alone or in interaction

<sup>13</sup>Section SM2.7 in the Supplementary Material explores whether the pattern found is an artifact of time or of long-run decentralization strategies by assuming the first reform in the period of coverage within the regional unit as the *hypothetical* critical juncture, instead of political decentralization. In this hypothetical scenario, the pattern found for political decentralization totally disappears.

**Table 2.3:** Influence of Regionalist Parties across Time

	DV: Decentralization Reform (0 — 1)			
	Asymmetric Reforms		Symmetric Reforms	
	(5)	(6)	(7)	(8)
Ideological proximity	4.948 (1.873) [0.009]	15.921 (4.632) [0.001]	−2.094 (4.088) [0.609]	−14.123 (9.090) [0.121]
Autonomist parties	−0.008 (0.008) [0.294]	−0.016 (0.023) [0.487]	−0.048 (0.075) [0.519]	−0.791 (2.350) [0.737]
Secessionist parties	0.003 (0.010) [0.770]	0.006 (0.026) [0.806]	0.256 (0.108) [0.018]	−17.265 (264.397) [0.948]
Regional democracy	−0.642 (0.485) [0.186]	14.281 (4.738) [0.003]	−0.240 (0.815) [0.769]	−13.586 (8.259) [0.100]
Regional democracy x Ideological proximity		−16.305 (5.108) [0.002]		15.668 (9.938) [0.115]
Regional democracy x Autonomist parties		0.006 (0.025) [0.796]		0.740 (2.346) [0.753]
Regional democracy Secessionist parties		0.003 (0.030) [0.931]		17.501 (264.397) [0.948]
Controls	✓	✓	✓	✓
Random effects	✓	✓	✓	✓
N	1,090	1,090	489	489
Regional level units	25	25	11	11
Countries	7	7	9	9
Log Likelihood	−247.498	−240.596	−90.037	−88.163
AIC	516.997	509.191	204.075	206.325
BIC	571.930	579.106	254.383	269.211

Notes: Standard deviation in brackets; p-values in square brackets.

Full model in Supplementary Material section SM2.5.

Country sample (excludes Sweden): Belgium, Canada, Denmark, France, Germany, Italy, Switzerland, United Kingdom, for 1950-2009.

with political decentralization (see also interactions in Figure SM2.3 in SM2.5). Yet, controlling for both types of regionalist parties unveils the predictive and statistically significant power of statewide governments' electoral-ideological incentives in asymmetric decentralization.<sup>14</sup> This mechanism cannot be found in symmetric countries. Put simply, governments take electoral and ideological considerations seriously in order to asymmetrically decentralize when regionalist demands are salient. However, governments seem to give up on this incentive structure in order to decentralize *after* political

<sup>14</sup>The model specification is also the most like previous studies concerning country selection and accounting for regionalist parties (Röth and Kaiser 2019).

decentralization has taken place and autonomous regional actors are established. Different explanations are possible here. Statewide governments might consider regionalist and ideological constellations in order to decentralize towards non-autonomous regions or to empower such regions with autonomy (political decentralization). Afterwards, autonomous regions with regionalist parties might activate the ceiling effect of slowing down the decentralization path, in order not to fuel possible secessionism. In the case of symmetric reforms, ideological considerations seem irrelevant regardless of regionalist actors.

**Table 2.4:** Multi-level Incentives *after* Political Decentralization

	DV: Decentralization Reform (0 — 1)	
	Asymmetric Reforms	Symmetric Reforms
	(9)	(10)
Ideological proximity statewide elections	1.945 (2.460) [0.430]	4.066 (4.885) [0.406]
Ideological proximity regional election	−0.084 (1.685) [0.961]	−0.393 (3.344) [0.907]
RAI <sub>t-1</sub>	−0.131 (0.038) [0.001]	0.0004 (0.041) [0.992]
Controls	✓	✓
Random effects	✓	✓
N	950	578
Regional level units	25	12
Countries	7	12
Log Likelihood	−202.161	−107.416
AIC	422.322	234.832
BIC	466.030	278.428

Notes: Standard deviation in brackets; p-values in square brackets.  
Country sample with political decentralization. Sample excludes regional elections in Mexico and the asymmetric regions of Bosnia and Herzegovina, Nicaragua, Portugal, and Serbia.

Finally, Table 2.4 presents the analyses on the decentralization incentives of the regional electoral arena *only after* political decentralization. In politically decentralized scenarios, party-electoral and ideological incentives at neither the statewide nor the regional level can predict subsequent decentralizing reforms of fiscal, policy or administrative authority. Interestingly, statewide ideological proximity's effect is relatively large for symmetric reforms in the non-expected direction, but not statistically significant. This whole set of patterns confirms the previous finding that statewide incentives are not systematically relevant for symmetric or asymmetric reforms *after* political de-

centralization. The evidence indicates that the rise of regional democracy does not *always* shift statewide governments' incentives to the regional arena in order to decentralize (H2.3). Models assessing the influence of statewide and regional incentives separately confirm that the coefficients shown here have not been cancelled out by each other due to collinearity. Overall, we do not find the expected effect of the political power equilibrium shifting towards the regional arena after the critical juncture of political decentralization across institutional settings. However, the results cannot tell us whether this dynamic is shaped by autonomous regional executives' new empowerment, how they put statewide governments under pressure and make them resign to electoral-ideological calculations. The next section discusses these results within the literature and the theoretical framework, and introduces alternative viewpoints worth considering.

## 2.5 Discussion

How do the general dynamics of decentralization change in the long-run? What role does party politics play in decentralization and how does it change across time? The institutional rise of regional democracy generated by political decentralization does not catalyze further reforms across all institutional configurations. Political decentralization fuels subsequent reforms only in unitary states. In unitary states, the idea that the rise of regional democracies works as a critical juncture generating increasing returns from subsequently decentralizations seems to hold. However, within federations, statewide governments pass fewer reforms after political decentralization. This is probably linked to the fact, perhaps idiosyncratic of federal systems, that federations are already very decentralized at the moment of political decentralization. The different dynamics across institutional settings – symmetric and asymmetric – affect which mechanism are in place when decentralization reforms are passed.

The empirical analysis does not confirm all theoretical expectations. Overall, ideological considerations are not systematically associated with decentralization reforms across settings. Ideological proximity explains asymmetric reforms, but only while accounting for regionalist party support, in line with previous research (cf. Massetti and Schakel 2016; Röth and Kaiser 2019). In other words, statewide government parties consider ideological similarities with the regions in order to asymmetrically decentralize, *but only when* regionalist mobilization demands more authority. Further, the emergence of regional democracies as the critical juncture reduces the overall influence of ideological incentives on subsequent asymmetric reforms. There is not clear picture as to whether statewide governments' electoral-ideological incentives can explain

decentralization reforms in symmetric systems. This finding highlights an important pattern. Asymmetric and symmetric decentralization are motivated by different incentive structures for statewide parties in government.

Following political decentralization, electoral and ideological incentives to decentralize were expected to form in the regional arena. The shift of power towards the regions and the potential for pressure to be applied to statewide governments by regional actors aiming to obtain more authority should make further decentralization more likely. This study finds that regional ideological incentives for statewide governments do not drive further decentralization, irrespective of whether the reform is symmetric or asymmetric. The idea that statewide governments follow the logic of decentralizing towards electoral-ideological strongholds systematically holds only under three conditions: before political decentralization; for asymmetric regions; and in conjunction with regionalist demands. These conditions show that the broad institutionalist theory of decentralization – that *after* political decentralization statewide parties will try to decentralize *more* authority towards the regions according to regional electoral considerations – does not hold in reality. Interestingly, additional models support Falleti’s (2005) findings.

Statewide government parties decentralize fiscal authority that mostly involves more benefits than costs when asymmetric regions without autonomy are ideologically more proximate. In the case of costly decentralization reforms of administration or policies, statewide parties do not strategically target ideological allies. After political decentralization, this is no longer the case, perhaps due to regional executives’ pressures on statewide governments to renounce such strategic maneuvers.

Furthermore, party-based mechanisms do not seem to explain symmetric decentralization, either before or after political decentralization. Even a sub-set analysis of unitary states, where reforms are more likely to occur after political decentralization, shows no significant effect. Whether party politics plays a role in symmetric decentralization thus remains contested. Overall, the finding once again highlights the idiosyncrasy of symmetric decentralization. According to O’Neill (2003; 2005), statewide parties in Latin America expect to gain support in the regional arena following political decentralization. Yet, regional arena incentives do not inform our models for symmetric decentralization either. It is unclear what mechanisms are in place here. Dardanelli et al. (2019b) highlight the multi-dimensionality of symmetric decentralization and how mono-causality fails to explain it.

Earlier analysis points to the electoral and ideological incentives of statewide parties being derived from vulnerabilities driven by electoral geography (Alonso 2012) –



i.e. not all regions are equal. The present study cannot further uncover the mechanisms which prevail in the symmetric reform dynamics. Central government could – following the the burden-shift and responsibility-fiscal argument – be negotiating with reliable partners in a large part of the regions to transfer policy competences and administrative responsibilities that will be backed up with financial transfers or fiscal authority concessions (León 2006). Furthermore, the analysis assumes that in symmetric decentralization all regions are crucial. Alternatively, since political decentralization generates fragmentation and regionalization of the multi-level party system, statewide governments could be under cross-pressure from heterogeneous regional voting patterns (León, Jurado and Garmendia Madariaga 2018). These alternative mechanisms should be investigated in future research in order to better understand how the critical juncture of political decentralization works and whether and how subsequent increasing returns influence statewide governments' decentralization decisions.

A further theoretical expectation stipulated that, after political decentralization, the higher the level of decentralization, the more likely further reforms, due to the pressure of regional actors on statewide governments. However, empirically, once regional democratic institutions are established, the more authority already decentralized, the less likely are further reforms, especially towards asymmetric regions. This dynamic can be linked to Falleti's argument that competences and responsibilities carry costs. In that case, autonomous regions would avoid receiving more authority if they are powerful enough to resist the statewide government. A further possibility is that, if regional actors are either demanding more authority or else aiming to avoid more policy and administrative responsibilities, the central government can "buy them off" with financial transfers (cf. Collignon 2019). Either way, the substantial meaning of the decentralization level for its endogenous further development is different after political decentralization in asymmetric regions. No systematic pattern is found for symmetric cases.

Finally, and contrary to the literature, regionalist parties are not key drivers of decentralization reforms. The support that regionalist parties retain shows no direct connection with the enactment of decentralization reforms, regardless of reform type and of whether political decentralization has been in place. It is worth mentioning, though, that only when accounting for regionalist parties support, is the theorized association between electoral-ideological incentives and decentralization reforms found. Regionalist parties seem to play a role in asymmetric decentralization, as they are necessary for the ideological considerations of statewide parties to matter. When further autonomy is demanded by the regions, statewide governing parties strategically decide when and which regions to accommodate. Accordingly, symmetric cases mostly

lack regionalist parties, usually by default. If there are no regionalist demands, the statewide government has no need to establish a special *asymmetric* relationship with these regions (Rokkan and Flora 2000, pp. 234ff.). The empirical evidence shows that party-based mechanisms of asymmetric decentralization do not travel well to explain symmetric reforms across time. It is plausible that other factors affect their logic of decentralization. This empirical disconnect invites researchers to further develop theories of symmetric decentralization, so far mostly focused on Latin America (Falleti 2005; O'Neill 2003), and to take their institutional effects more seriously.

## 2.6 Conclusion

The dynamics and causes of decentralization are not homogeneous across time. This study theoretically argues that the institutional emergence of regional democracies is a critical juncture that affects the likelihood of decentralization reforms, and *how* and *when* statewide governments decide to decentralize. Theoretically, political actors in the autonomous regional arena have the potential to exercise pressure on the center to obtain more authority and accumulate power. The shift of the power equilibrium to the regional arena is a game-changer for party decentralization strategies, since statewide governments shift their attention, in relative terms, to regional elections, where new resources and electorates are located.

First, the general dynamics of decentralization are revealed by focusing on reform pattern across institutional settings. Federations seem to reach high levels of decentralization before political decentralization. Decentralization in unitary systems, on the contrary, is rather more likely to occur *after* political decentralization, when autonomous regional governments emerge. This core difference indicates that federal systems quickly reach a ceiling of overall decentralization once they are politically decentralized, possibly activating different party mechanisms of decentralization. Second, with an original data-set covering the period 1950-2018, this study investigates party-based explanations in a systematic framework for both symmetric and asymmetric decentralization. The findings show a complex and differentiated picture. While ideological strategies of statewide parties in conjunction with regionalist demands seem to drive asymmetric decentralization reforms before regional democracies emerge, this dynamic disappears afterwards. This is in line with the fact that party politics lose relevance once regional democracy is institutionalized, but the question of which mechanisms start to drive further reforms remains unanswered. Nevertheless, regionalist parties play an important role, depending on institutional context, and are important in understanding asymmetric decentralization.

As Falleti (2005) found, the sequence and type of decentralization matter. For both statewide and regional governments, fiscal decentralization is an important reform, in which ideological proximity is key. Statewide governments' ideological incentives are better predictors of the major political reform and of fiscal decentralization reforms in asymmetric cases. Their importance drastically decreases after regional democracies emerge. In short, party politics is not straightforwardly related to decentralization reforms after the emergence of regional democracies. Regional governments' political influence can be a decisive factor creating cross-pressures between the national and regional arenas. Political decentralization, then, has negative implications for the interests of governing statewide parties; after the major reform, they have to adjust their strategic calculations on subsequent decentralization according to regional governments' bargaining power. Furthermore, the notion that decentralization only implies benefits seems to be a shortcoming in the literature ignoring the burden that comes with policy responsibilities. This study also reveals the previously downplayed role of regionalist parties in the long-run dynamics of decentralization.

The investigation highlights a contrasting logic for symmetric decentralization reforms. The electoral-ideological considerations of governing statewide parties seem to be unrelated to symmetric decentralization reforms, both before and after political decentralization. The dynamics of symmetric decentralization gain speed after political decentralization – indicative of a critical juncture and the beginning of increasing returns through irreversible reforms. However, it is not clear how the mechanisms behind this initial major reform work or to what extent they matter for subsequent decentralization. Overall, symmetric decentralization and decentralization in unitary systems require further theoretical and empirical investigation. One limitation of this study is the exclusive focus on ideological incentives and cabinet politics of statewide parties in both symmetric and asymmetric decentralization. Future research can investigate to what degree such decentralization reforms are driven instead by the electoral performance expectations of the governing statewide party. Additionally, studying exactly how regional governments in the immediate aftermath of political decentralization interact and defend their interests vis-à-vis the statewide government will enable a deeper understanding of the role of regional democracy in the overall dynamics of decentralization. Here, comparative case studies can offer rich insights.

To what degree the emergence of regional democracies and the new role of regional elections is a critical juncture remains a contested issue. As with the Europeanization process (Hooghe and Marks 2018), political decentralization in democracies produces an autonomous regional political arena that reorganizes multi-level political conflicts and reshapes party competition. How statewide governments are challenged by

electoral multi-level cross-pressures to decentralize subsequently is still unknown, and how governments' interests evolve *after the regional critical juncture* remains under-investigated. This is a call to focus on such phenomena while accounting for the symmetry-asymmetry divide which appears strongly relevant to different explanations of decentralization and to the notion that the rise of regional democracy can fundamentally alter multi-level politics.

# THE MEASUREMENT OF ISSUE ATTENTION ACROSS POLITICAL ARENAS: EXPLOITING INCREASING RETURNS WITH OPTIMIZED DICTIONARIES

(co-authored with Leonce Röth and Lea Kaftan)

## Abstract

We present a semi-automated way of sorting text fragments into issues, with the aim of building valid and transferable optimized dictionaries. Text-as-data approaches have facilitated analyses of huge amounts of text, but have not so far fully exploited their potential for application in combination. We discuss the strengths and weaknesses of dictionaries, machine learning, and topic models, and argue that the precise combination of them will yield the best opportunities for valid and comparable issue identification. Developing optimized dictionaries is initially resource-intensive, but optimized dictionaries can be applied to all sorts of texts and will yield increasing returns. We demonstrate our case with a discussion of attention to the issue of territorial politics in leading newspapers compared to the parliaments in Spain (1976–2019) and the UK (1900–2020). Substantially, we demonstrate that parliaments favor technical and competence-oriented issues over conflictual ones, whereas the media’s priorities are the opposite.

### 3.1 Introduction

Attention to issues is an indispensable yet scarce political resource (Bachrach and Baratz 1962; Franck 2019; Jones and Baumgartner 2005; Proksch and Slapin 2015). Issue selection and relative emphasis are necessary prerequisites for actors to frame issues and adopt attitudes towards them (Robertson 1976). Success in political competition rests to a large extent on the ability to direct attention to some issues and divert it from others (Bélanger and Meguid 2008; Budge 2015; Budge and Farlie 1983; Carmines and Stimson 1989; Dolezal et al. 2014; Green-Pedersen 2007; Petrocik 1996; Pogorelis et al. 2005; Riker 1990; van der Brug and Berkhout 2015). Conversely, lack of attention deprives citizens and politicians of the opportunity to be heard and diminishes their prospects for political relevance (Gilens 2001).

Scholars have addressed many aspects of the distribution of attention and, to some extent, its denial. For example, we know a great deal about party competition for issue attention (Carmines and Stimson 1989; de Sio and Weber 2014; de Sio, de Angelis and Emanuele 2017; Green-Pedersen 2007; Green-Pedersen and Mortensen 2015; Spoon, Hobolt and de Vries 2014); the selection and prioritization of issues in the media (Allern 2017; Domingo et al. 2008; Donsbach 2004; Herman and Chomsky 1988; McChesney 2000; Niblock and Machin 2007, p. 191; Staab 1990; Strömbäck, Karlsson and Hopmann 2012, p. 726; Welbers et al. 2016); and, increasingly, how parliamentarians attach attention to issues (Bräuninger and Debus 2009; Green-Pedersen 2019; Proksch and Slapin 2015; Proksch et al. 2019). Those strains of research have typically analyzed attention within only one political arena – such as legislatures, executives, or the media – because one of the biggest challenges for the comparison of issue attention across democratic arenas is how to arrive at valid and comparable estimates.

In this study, we present a semi-automated way of sorting text fragments into issues, with the primary aim of identifying valid and comparable issues in text across political arenas. Recent advances in text-as-data approaches provide a welcome remedy for the limitations of resource-intensive hand-coding, extending the scope of data generation and classification. Applied text-as-data approaches have facilitated analyses of huge amounts of texts, but have not so far fully exploited the potential benefit of applying them in combination. We discuss the strengths and weaknesses of existing methods (dictionaries, machine learning, and topic models) and argue that we need to find the precise combination of them which will yield the best opportunities for valid and comparable issue identification from text. Our approach centers optimized dictionaries for specific political issues and sub-issues. Developing optimized dictionaries is initially

resource-intensive, but once achieved, they can be applied to all sorts of texts and will yield increasing returns.

To demonstrate our case, we assess the issue of attention to territorial politics across parliaments and the media in Spain (1976–2019) and the UK (1900–2019). This is an issue that arguably varies substantially within and across both countries and, furthermore, contains independent sub-issues which can be sorted along one of the most important dimensions within the politics of attention in parliament and the media: conflictual versus technical.

The investigation proceeds as follows. First, we discuss the logic of attention to issues across political arenas. Second, we suggest a procedure for obtaining comparable data on issue attention across political arenas, based on dictionary optimization, combining expert-based keyword selection with machine learning, and topic models. Third, we apply our approach to relative attention in the parliaments and leading newspapers in Spain and the UK, using the example of territorial politics.

### **3.2 The logic of issue attention across political arenas in democracies**

Many normative accounts point to the conclusion that a democracy cannot exist without a media system that rigorously holds people in power to account. The media also play a special role in providing a wide range of informed opinion on the important issues of the day. Most of what people, including politicians, know comes to them “secondhand” from the media (McCombs and Shaw 1972, p. 176; Midtbø et al. 2014; van Aelst and Walgrave 2011; Walgrave 2008). Media should keep citizens informed about the behavior and policies of representatives and keep representatives informed of the preferences of citizens.

Empirically, we know that neither the media nor elected representatives behave as ideal transmitters of information. Both follow their own logic of communication (for the media, see Allern 2017; Domingo et al. 2008; Donsbach 2004; Herman and Chomsky 1988; McChesney 2000; Niblock and Machin 2007, p. 191; Staab 1990; Strömbäck, Karlsson and Hopmann 2012, p. 726; Welbers et al. 2016; for parliaments, Altheide and Snow 1979; Esser and Strömbäck 2014; Mazzoleni and Schulz 1999; Strömbäck 2008; Walgrave and van Aelst 2006), raising questions about agenda-setting, mediatization, and media bias in the reporting of political debates in representative arenas.

Media scholars have shown how news selection is driven by, for example, journal-

istic intentions (Staab 1990), subjective judgments (Herman and Chomsky 1988; McChesney 2000). Commercial incentives impact news selection (Allern 2017) by giving priority to sensational news with lower production costs and higher appeal to target audiences (Niblock and Machin 2007, p. 191; Strömbäck, Karlsson and Hopmann 2012, p. 726). In short, media studies remind us that news selection is a process falling far short of fulfilling its ideal democratic function.<sup>1</sup>

In parliamentary arenas, governments influence issue attention by emphasizing policy problems and their solutions (Green-Pedersen 2019). Politicians react to trigger events by strategically raising public awareness of issues and narrowing the range of possible solutions towards beneficial legislation (Dearing and Rogers 1996, pp. 86–88). Government parties have more resources to shape agendas (Den Hartog and Monroe 2011) and tend to highlight more technical and bill-oriented issues, given their role in initiating legislation (Bräuninger and Debus 2009, p. 814). Meanwhile, since media driven by newsworthiness favor conflictual issues over technical ones, we might expect that they are more likely to report strongly contested parliamentary debates and challenges by political outsiders.

Self-evidently, the dynamics of issue attention in parliaments and the media interact. This interaction has been widely studied under the heading of agenda-setting, informing us of several tendencies: media shape the agenda of politicians more strongly than vice versa (Kleinnijenhuis and Rietberg 1995; Walgrave and van Aelst 2006); negative news generates more political attention than positive news (Thesen 2013); opposition parties are more influenced by media than government parties (Green-Pedersen and Stubager 2010); and the agenda-setting power of media varies across issues (Soroka 2002).

However, most studies are typically single-country studies (Vliegenthart et al. 2016). To our knowledge, there are only four exceptions (van Noije, Kleinnijenhuis and Oegema 2008; Vliegenthart and Montes 2014; Vliegenthart and Walgrave 2011; Vliegenthart et al. 2016). These comparative studies all corroborate the case-study insights that politicians react more to media than the other way around. Furthermore, there are substantial differences in politicians' responsiveness to the media, determined by the government/opposition divide and across political systems (Vliegenthart et al. 2016).

Beyond differences across political systems, it all boils down to a very simple incentive structure for the emphasis and framing of issues. Moving up the delegation chain of political representation means turning voter preferences, their differences, and conflicts, their tendency to sensationalism, into sober technical solutions wrapped in an aura of competence. Accordingly, the more issues refer to technical solutions and

---

<sup>1</sup>See Harcup and O'Neill (2017) for an overview.



expertise, the more beneficial this distribution of attention is to established authorities (governments, ministries, civil servants). The media, and political actors without direct access to the executive, are incentivized to favor just the opposite. A vast literature on newsworthiness indicates that conflict and sensational news sell best and thus logically align with political challengers to feature attention to conflicts and underrepresent attention to its solutions.

Before we ask whether the cases of territorial politics in Spain and the UK align with the incentive structure described, we lay out a path to valid and comparable data on issue attention across political arenas such as the parliament and the media.

### **3.3 Valid and comparable data on issue attention across political arenas**

Researchers have invested substantial resources to come up with measures of issue attention. The vast majority of measures are based on expert surveys or derived from hand-coded texts. These procedures demand respect for the (internally) valid identification of content, because humans are, and will probably remain for some time, the best coders of text. But human codings are subject to at least three failings.

First, expert surveys and hand-coded text are difficult to compare because underlying concepts are typically not synchronized and cannot be adapted *ex post*. For example, an expert survey uses a very different concept of salience from scholars carrying out media or manifesto content analysis (see Chaqués-Bonafont, Palau and Baumgartner 2015; Helbling and Tresch 2011; John et al. 2013). Second, both approaches are hard to replicate and thus difficult to extend to new content. Third, they are very resource-intensive, hence limiting the scope of analysis. For example, the Comparative Agenda Project (CAP) manually codes newspaper content (Baumgartner, Breunig and Grossman 2019; Barberá et al. 2020; Jacobi, van Atteveldt and Welbers 2016), but resource constraints mean only the front pages are coded. Resource constraints also often mean only short periods are available, limiting the ability to test dynamic processes over longer timespans.

Automated approaches to text classification can be a powerful complement, enabling more text to be processed, thus leading to better coverage in time and space; and dictionaries travel easily across different political arenas, in principle increasing comparability.<sup>2</sup> Recent applications of automated text classification to the salience of

---

<sup>2</sup>Automated processes face language barriers but language barriers also apply to hand-coding.

specific policy issues in the media demonstrate the obvious advantage of higher coverage with fewer resources (see Gilardi, Shipan and Wüest 2020; Soroka and Wlezien 2019). The biggest challenge of automation in comparison to hand-coding and expert surveys is to ensure validity. Automation is substantially more likely to produce false positives and false negatives than hand-coding.

What exactly should be identified is a deductive decision, a deliberative process to cut up political communication into meaningful pieces. Typically, concepts such as issues or topics represent such pieces. Issues are composed of sub-issues (see Green-Pedersen 2019, p. 26). For example, we can collect references to the welfare state (overarching issue) and we can disaggregate those references into pieces referring to the domains of retirement, health care, or unemployment, to name just three possible sub-issues. This distinction is particularly important for the representation of political communication because it allows for distinguishing between relative emphases of sub-issues within more general debates. In the following, we describe how automated text analysis can increase the valid identification of issues, a baseline we need to establish in order to measure attention to issues.

### **3.3.1 The measurement of attention to issues in text**

The ideal classification of text into political issues demands the minimization of three sources of error and the optimization of transferability to new sources of text. The three main sources of error are false negatives, false positives, and endogeneity bias. If a method identifies more text fragments than in “reality” contain relevant content, we speak of false positives. If a method does not identify text fragments that in “reality” mention the issue, we speak of false negatives. Endogeneity bias is introduced by the preconceived and/or biased ideas of the researcher. Often, researchers are driven by a specific event or biased perspective on a specific matter, which translates into some methods of text identification more than others. For example, historical bias can be introduced by developing dictionaries with keywords related to more recent events, ignoring past concepts, wordings, or debates. The resulting measurements of issue salience will accordingly reflect the researchers’ bias and will not yield valid results. Finally, the efficiency of methods is influenced by their transferability. Our aim is to arrive at text-identification methods that are easy to apply to different corpora and contexts and allow text classifications to be compared across documents and political arenas.

Before we discuss the abilities of different methods to minimize these three major sources of error and improve transferability, we urge researchers to reflect on the most

important features of their text corpora. Researchers need to define their corpus (universe of text) and the issue(s) of interest, because key features of the corpus and issues have an impact on the succeeding decisions.<sup>3</sup> We discuss the three most frequently used approaches to classifying text and how well they perform the task of minimizing the three sources of error, and outline the properties of transferability.

### 3.3.2 Hand-coding

Hand-coding texts is considered the most internally valid method of identifying issues (Hutter 2014). Hand-coding can handle low-quality text corpora and minimizes false negatives, false positives, and endogeneity biases (van Atteveldt, van der Velden and Boukes 2021). Humans can identify optical character recognition (OCR) or typing errors, humans are best at detecting irony, negations, and metaphors. Humans can also identify endogeneity bias and, ideally, correct for preconceived misspecifications. In short, hand-coding is still the best method of minimizing all three sources of error.

However, although best practice in hand-coding provides codebooks and example texts to explain coding decisions, they are not easy to replicate, adapt, or apply to new corpora because hand-coding is extremely resource-intensive. Even slight adaptations of existing coding decisions or application to new sources are severely hampered by resource constraints, meaning that, typically, specific and limited amounts of texts are coded for specific purposes which are difficult to compare to other projects.

Since hand-coding is extremely costly, researchers have developed alternative methods, most importantly keyword searches with dictionaries (Barberá et al. 2020; Hayes and Weinstein 1990; King, Lam and Roberts 2017; Radford 2021, machine learning (Zhou and Goldberg 2009), and topic models (Blei and Lafferty 2006; Blei, Ng and Jordan 2003; Roberts et al. 2013, 2014). We discuss these in the following sub-sections.

### 3.3.3 Dictionaries

Dictionaries are a set of words signaling the emphasis of a specific issue. Researchers can either use any single word in their dictionaries as a deterministic indicator of whether a text refers to an issue of interest, or they can use Boolean search queries (Welbers, van Atteveldt and Benoit 2017, p. 254). In the context of measuring attention to issues in corpora, researchers can use dictionaries for confirmatory analyses of issues that are described by precise words (Albugh, Sevenans and Soroka 2013). Un-

---

<sup>3</sup>The most important differences between corpora are whether researchers have complete access or only partial access via search engines, and whether the text documents include OCR or typing errors because different text identification approaches are to different degrees affected by this.

like human coders, dictionaries are very susceptible to all three sources of error, but have strong properties of transferability. For example, if words in the dictionary are ambiguous, they will identify too many text fragments that, in reality, do not mention the issue of interest (false positives) (see Barberá et al. 2020; King, Lam and Roberts 2017). If the dictionary does not contain all relevant words for the issue, it will miss relevant text fragments and thus underestimate attention (false negatives). The selection of keywords is also strongly influenced by endogeneity bias because researchers can only select keywords for issues they are aware of. On a more technical note, dictionaries might also be biased if texts have many typing and/or OCR errors, a feature of particular importance when using historic documents.

Dictionaries also have advantages. Their creation and application are not costly, and they can be easily applied to large and to new corpora. A significant advantage of dictionaries is that researchers do not need full access to the corpus for their application. A keyword search can be used to measure attention to issues in corpora provided by corporations such as Nexis or ProQuest or by digital newspaper archives via text sampling based on search queries. It is, furthermore, irrelevant whether issues of interest appear often or rarely in the corpus. Overall and in contrast to hand-coded text, issue attention scores based on dictionary searches are most open to error because dictionaries are prone to all three major sources of bias. However, dictionaries have good properties in terms of replicability, adaption, and transferability.

### **3.3.4 Machine learning**

Like dictionaries, machine-learning classifiers can serve researchers as a tool for measuring attention to issues for confirmatory analysis at a considerably lower cost than hand-coding. But researchers nevertheless have to hand-code a randomly selected and large enough training set of texts, especially if the issue of interest is rare and the corpus is large (Aggarwal 2018; Cieslak and Chawla 2008; see also Mitts 2019). Furthermore, unbalanced training sets and data can lead to poor performance by machine-learning algorithms (Cieslak and Chawla 2008). Most importantly, if researchers want to use machine learning as a tool for the measurement of attention to issues, they must have access to a large random sample of the corpus.

Machine learning can handle OCR or typing errors better than dictionaries, since single words do not lead to deterministic but probabilistic identification of issues. For the same reason, it is less problematic if words that describe an issue are ambiguous. False negatives and positives appear more often than in hand-coding, but should be significantly lower in comparison to dictionaries if the machine learning is appropriately

tuned and if the hand-coded training set is large enough. Machine learning should also reduce endogeneity bias in comparison to dictionaries, because text fragments with related contexts are likely to be identified even if researchers were not aware of them. In terms of our framework, machine learning has advantages in minimization of errors in comparison to dictionaries, but is less well suited in terms of adaptations, replicability, and transferability, although the last three features show substantial improvements over hand-coded text.

### 3.3.5 Topic models

In contrast to the previously discussed methods, topic models are designed for explorative research only. Thus, researchers can use topic models if they are interested in attention to unknown issues in texts. They conceptualize texts as mixtures of topics, and topics as clusters of words that often appear together (Blei and Lafferty 2006; Blei, Ng and Jordan 2003; Roberts et al. 2013, 2014). While some topic models allow topics to be correlated (Roberts et al. 2013, 2014), others do not (Blei and Lafferty 2006; Blei, Ng and Jordan 2003). Besides possible issues with computational power and topic-model selections, they are easy to apply even to large text corpora and have been demonstrated to be cost-efficient.

Topic models perform much better than dictionaries and hand-coding in instances in which issues are described by highly ambiguous words. This is because topic models allow words to be part of different word clusters and, thus, to describe different issues simultaneously although in different combinations. OCR and typing errors are problematic, but not as much as for dictionary analyses, because single words do not dichotomously determine whether a text mentions an issue or not. Moreover, it does not matter how (in)frequently specific issues arise, as long as researchers choose an appropriately large number of topics for their analysis. However, researchers need access to a large random sample of the corpus.

Since topic models are completely unsupervised methods of quantitative text analysis, endogeneity is not a problem as long as researchers pre-process their corpus and select their models with rigor. For the same reason, the logic of false positives and negatives does not apply to topic models. Since the proportions of topics are based on the distribution of words across texts, and since the words appear in each text, it is rather a theoretical question whether derived word clusters make any sense instead of whether they are actually included in that text. Due to the explorative nature of topic models, researchers face general problems of validity attribution, because clear conceptual benchmarks are absent. Furthermore, the comparability of topic-model results

**Table 3.1:** Properties of measurement approaches for issue identification in text corpora

Method	Hand-coding	Dictionaries	Machine Learning	Topic Models	Optimized dictionaries
Sources of bias					
False negatives	Low	High	Medium	-	Low
False positives	Low	High	Medium	-	Low
Endogeneity bias	Low	High	Medium	Low	Low
Transferability					
Transferability	Low	High	Low	Low	High
Other features					
Required access	Full corpus needed	Keyword access sufficient	Large random sample	Large random sample	Keyword access and small random sample
Research interest	Explorative & confirmatory	Confirmatory	Confirmatory	Explorative	Confirmatory
Resource-intensity	High	Low	Medium	Low	Increasing returns
OCR or typing errors	Not problematic	Very problematic	Problematic	Problematic	Problematic

across different corpora is problematic.

Table 3.1 summarizes the benefits and pitfalls of these approaches to the measurement of attention to issues in text corpora. None of them can on its own provide valid, comparable, and transferable measures of issue attention. In the following section, we argue that we can exploit the different strengths of each of these methods to arrive at optimized dictionaries with the best properties in terms of bias reduction, transferability, and investment of resources.

### 3.4 The optimization of dictionaries

Inadequately developed dictionaries will yield largely misleading inferences (Grimmer and Stewart 2013, p. 274).<sup>4</sup> Nonetheless, dictionaries for issue identification have the best properties in terms of transferability because, once developed, they can easily be applied to all types of sources. Investment in good dictionaries thus promises substan-

<sup>4</sup>Equally, sentiment dictionaries – well-elaborated keyword lists to capture tonality in text – can bring up major biases (Rauh 2018; van Atteveldt, van der Velden and Boukes 2021).

tially increasing returns; the major task is to minimize the three sources of error they are subject to.

We argue that the bad properties of dictionaries can be substantially diminished if the complementary strength of other methods is exploited (see also van Atteveldt, van der Velden and Boukes 2021; Rice and Zorn 2021, for similar arguments). We propose to complement the weaknesses of dictionaries with the strength of hand-coding and machine learning, with the ultimate aim of creating optimized dictionaries for the identification of issues in text, where bias and resource investment are low but transferability is high. We organize this procedure in four steps.

**Step 1: Initial dictionary.** Researchers should first carefully define and describe their issues of interest and create an initial dictionary based on existing dictionaries, expert knowledge, and/or secondary sources. Although such a dictionary is more than what Grimmer and Stewart (2013, p. 274) describe as rudimentary, any bias it bears is still unknown.

**Step 2: Reduction of false positives and false negatives.** The second task is to reduce false positives and negatives. This is done by assessing a sample of hand-coded text. While we recommend selecting texts for hand-coding as randomly as possible, researchers can make use of a “guided” random selection to reduce the number of texts that have to be hand-coded. If researchers use texts from different sources or a wide period, their selection of texts for hand-coding should take into account possible variations deriving from different word usage across sources or periods for the issue description, as well as for parts of texts in which the issue is not described. It is advisable to select both texts that contain a word from the dictionary and texts that do not. The hand-coding of a random sample reveals instances of false positives and negatives and guides the first iteration of dictionary adjustment by erasing words which create many false positives and including words which reduce false negatives.

**Step 3: Reduction of endogeneity bias.** In the third step, researchers should assess the degree of endogeneity bias in their dictionary and further reduce false positives and negatives. Researchers should use their hand-coded sample of text from step 2 as a training set for an appropriate machine-learning algorithm. A comparison of the results from the machine-learning algorithm and the dictionary approach will reveal the dictionary’s blind spots and enable further reduction of false negatives and positives. For example, the researcher might have been unaware of a debate that would conceptually fit the definition of the issue but whose key terms are not featured in the dictionary. The machine-learning algorithm would identify such a debate because of its general semantic coherence, whereas the dictionary alone would not. High cor-

respondence of identified texts between machine-learning and dictionary approaches signals low endogeneity bias; low correspondence reveals greater bias. Hand-coding a random sample of those texts identified by machine learning but not by the dictionary guides the second iteration of dictionary adjustment. Now researchers can repeat steps 1 through 3 as often as necessary until they reach satisfactory levels of identification.

We use F1 scores to assess the performance of the dictionary in terms of bias. The F1 score indicates the joint performance of sensitivity (true positive rate) and precision (true negative rate) (Derczynski 2016):

$$F1 = 2 * \left( \frac{Precision * Sensitivity}{Precision + Sensitivity} \right),$$

while

$$Precision = \frac{TP}{TP + FP}$$

and

$$Sensitivity = \frac{TP}{TP + FN}$$

A value of zero indicates only false negatives (FN) and positives (FP), whereas a value of 1 indicates only true positives (TP) and true negatives (TN). F1 scores can be calculated by comparing the hand-coded with the dictionary-based classifications. Comparing the F1 scores across the different iterations of dictionary optimization indicates the progress of the procedure.<sup>5</sup> Once a satisfactory F1 score is achieved, we can turn to an assessment of the key strength of dictionaries, transferability to other text corpora.

**Step 4: Transferability.** A key strength of dictionaries is the simplicity of applying them to other text corpora. The dictionary optimized through steps 1–3 can readily be applied to all types of corpora. The easiest way to assess the appropriateness of using the optimized dictionary on a new corpus is to hand-code another random sample of the new corpus and calculate the F1 scores of the optimized dictionary for the new corpus. This allows researchers to assess the quality of the dictionary predictions in previously unseen parts or new corpora and enables inferences about the comparability of issue identification across different sources.

---

<sup>5</sup>F1scores aim at achieving a balance between precision and sensitivity (typically also called recall). They are easy to interpret. Initially, precision is typically very high with dictionary approaches (it was actually perfect (1.00) with the first version of our dictionary). However, optimizing dictionaries means increasing sensitivity without reducing precision too much (sensitivity was 0.28 in the first version and rose to 0.88, whereas precision dropped from 1.00 to 0.88).



### 3.5 Case study: territorial politics in Spain and the UK

To demonstrate our procedure, we assess attention to territorial politics within parliaments and the media in Spain (1976–2019) and the UK (1900–2019), an issue that arguably varies substantially within and across both countries. In both countries, we had full access to one newspaper (*El País*, 1976–2019, and *The Times*, 1900–2013) and keyword-based access to a second via Nexis (*El Mundo*, 2002–2019, and the *Guardian*, 1985–2017). We also used Spanish and British parliamentary speeches from ParlSpeech (Rauh and Schwalbach 2020) for the periods 1996–2018 and 1989–2019, respectively. Our case study demonstrates that optimized dictionaries can identify fragments of text even without access to the full corpus and that optimized dictionaries travel extremely well to new text corpora.<sup>6</sup>

Following steps 1 and 2 of the optimization procedure, we created separate initial dictionaries for territorial issues in Spain and the UK. These were based on historical and political science research, party manifestos, and homepages of non-governmental organizations. We structured territorial issues into different debates for the period of interest and identified keywords for each of these periods and territorial issues for a single dictionary. Simultaneously, we drew the first round of random newspaper articles for hand-coding and compared the hand-coded with the dictionary-based identification (991 articles from *El País* and 570 from *The Times*; both newspapers with full access). We used the first random draws to check the inter-coder reliability of two different coders (result for Spain: 78.3%).

Two things became clear with the first round of hand-coding. Whereas true positives from the dictionary and hand-coding were highly satisfactory (97% on average), the initial dictionary produced a very high number of false positives. As we had observed that our first random selection of *Times* articles only contained eight territorial politics articles, we drew another random sample of two territorial politics articles per year, identified by dictionary version 1. Furthermore, we observed a high number of OCR-related errors in *The Times*. These errors are not randomly distributed, decreasing over time, and they make the usage of dictionaries problematic.<sup>7</sup> OCR errors rendered the performance of the dictionary for *The Times* much lower than for other British and Spanish sources.

---

<sup>6</sup>See sections SM3.1–SM3.2 of the Supplementary Material for descriptive statistics and complete dictionaries.

<sup>7</sup>As we can show, this also applies to machine-learning algorithms.

Accordingly, given our experiences with steps 1 and 2, we further increased our hand-coded samples for both newspapers. Overall, this increased the hand-coded set for El País to 2,535 articles. We used this hand-coded sample to further adapt the dictionary. As a consequence, for example, false positives were substantially reduced and the proportion of texts concerning territorial issues dropped from 28.7% to 16.8% of all texts in the Spanish newspapers. In the case of The Times, we drew two additional random samples, enlarged our hand-coded sample to 1,368 articles, and simultaneously adjusted the balance in the hand-coded sample to include more texts concerning territorial issues. As in the procedure with El País, we deleted words from the dictionary that led to many false positives and included keywords that led to fewer false negatives.

Following step 3, we applied different machine-learning classifiers<sup>8</sup> based on the hand-coded training sets to identify text with territorial content in the newspapers. The correlation between the optimized dictionary and the machine-learning predictions was considerably higher for El País than The Times, but significant in both cases (compare section SM3.3 of the Supplementary Material). We drew a final random sample of two articles per year where classifications by the optimized dictionary and the machine-learning classifiers diverged. Our analysis of this sample alerted us to our own endogeneity bias. Our initial perspective on territorial politics in Spain was shaped by an implicit institutionalist perspective on authority distribution, and through our process we came to realize the crucial importance of secessionist and violent expressions for the media. Developing the UK dictionary began later, so we were able to incorporate this insight from the beginning. Finally, we adjusted our dictionary to diminish endogeneity bias.

Table 3.2 presents the F1 scores for three different machine-learning approaches and our different dictionary versions. The F1 scores increased from 0.520 to 0.783 in the case of El País and from 0.417 to 0.549 in the case of The Times. Overall, the performance of each final dictionary version is comparable to the performance of the most suitable machine-learning algorithm for each language. The considerably lower F1 scores for The Times might be due to its longer time coverage (114 years compared to 43 years of El País articles) as well as to the OCR errors. However, both machine-learning and dictionary approaches suffer under these conditions. One might expect that dictionary improvement and performance diagnostics are endogenous, since these rely on learning from a hand-coded set of articles that are also used to evaluate the performance of the dictionaries. However, overfitting is largely avoided, since adaptations consider a trade-off between reducing false negatives and false positives and since we

---

<sup>8</sup>Support Vector Machine (SVM), Random Forest, and Naïve Bayes.

**Table 3.2:** F1 scores: performance of dictionaries throughout the steps

	Dictionary version 1	Dictionary version 2	Dictionary version 3	Machine learning	Number of hand- coded documents
El País	0.520	0.770	0.783	0.801	2534
The Times	0.417	0.523	0.549	0.584	1596

Note: Dictionaries were applied to all hand-coded articles. The performance scores of the machine-learning algorithms are based on Leave-One-Out Cross-Validation (LOOCV) without model tuning and with defaults. In the case of El País, we used Random Forest as ML algorithm. In the case of The Times, we used SVM. SVM failed to categorize six Times articles. RF and SVM were the best-performing algorithms for each newspaper. We show F1 scores using all ML algorithms in section SM3.4 of the Supplementary Material.

**Table 3.3:** Transferability

	Dictionary version 1	Dictionary version 2	Dictionary version 3	Machine learning	Number of hand- coded documents
El Mundo	0.10	0.87	0.90	0.55	100
Congreso de los Disputados (Spain)	0.42	0.86	0.89	0.40	100
Guardian	0.82	0.89	0.93	0.60	100
House of Commons (UK)	0.76	0.87	0.88	0.47	100

included further articles after each round of dictionary adaption.

In the last step, we examine the transferability of our dictionaries to other sources of text, a procedure that additionally helps to assess the possibility of overfitting. In our case study, we applied the optimized dictionary to the second newspaper in each country, for which we lacked access to the whole corpus, and to parliamentary speeches. To calculate the F1 scores, we hand-coded a random sample of parliamentary speeches and news articles from the newspaper with limited accessibility. The F1 scores using new sources indicate the transferability of the dictionary. The performance of the optimized dictionary is better than satisfactory for all four external sources (compare Table 3.3), whereas using machine learning or non-optimized dictionaries leads to F1

scores substantially below those achieved using optimized dictionaries.<sup>9</sup>

The performance comparison reveals three key advantages of optimized dictionaries. First, performance for identification is at least as good as other procedures like machine learning. Second, optimized dictionaries work very well for identification in unseen text sources. Third, optimized dictionaries can be applied to measure attention to issues even when the full corpus is not accessible. Interestingly, the good performance of the dictionary for the Guardian shows that dictionaries can be developed with sources including ORC errors, and also indicates the magnitude of bias caused by OCR error.

### **3.5.1 Issue attention: territorial politics in Spain and the UK**

In this sub-section, we discuss the application of our procedure, looking at attention to territorial issues in Spain and the UK. Figure 3.1 shows that territorial politics has long been a dominant issue in Spanish democratic history. From the most recent democratization process in 1976 until 2019, 8% of articles in the two newspapers analyzed include references to territorial politics, increasing to roughly 20% for the domestic politics section. The Congreso de los Diputados devotes much less attention to territorial politics: from 1996 to 2019, 6% of parliamentary interventions and speeches alluded to it. Although the level of attention differs slightly across arenas, Figure 3.1 shows that their developments co-vary over time (correlation of 0.66,  $p < 0.00$ ). Recalling the F1 statistics of roughly 0.90 in the Spanish case and across the newspapers and parliaments allows us to judge the estimates of the emphasis on territorial politics as fairly accurate.

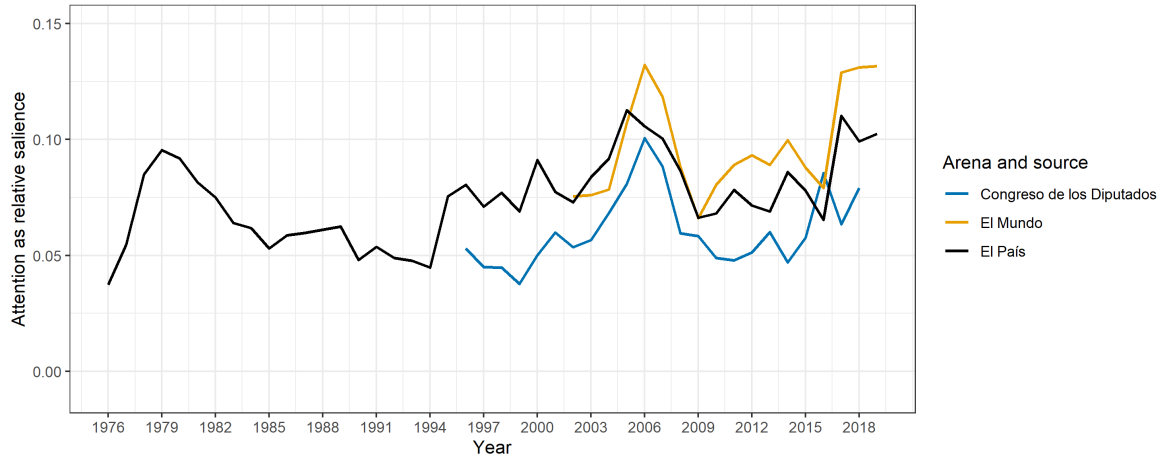
Strong media attention around 1980 reflects the rise of regional democracies when the constitutional architecture of Spain was being set up and fundamental questions of regional autonomies were debated. Shortly after, several decentralization laws were passed to transfer political autonomy or competencies to the Comunidades Autónomas. Such a high degree of saliency was only reached again with the reform of the Catalan statute in 2006, and in 2017 with the Catalan referendum and the declaration of independence by Catalan president Carles Puigdemont. In short, parliamentary and newspaper attention to territorial politics mirrors key developments in struggles for authority in Spain.

Attention to territorial politics in The Times, the Guardian, and the House of Commons is substantially lower across the entire period, although it too varied significantly during the last century (compare Figure 3.2). During the 1910s, territorial politics were

---

<sup>9</sup>See section SM3.5 of the Supplementary Material for a detailed description of the transferability of different dictionary versions and machine-learning algorithms.

**Figure 3.1:** Attention to territorial politics across political arenas in Spain



Note: Parliamentary speeches and media articles; sources: Rauh and Schwalbach 2020; El País (1976–2019); El Mundo (2002–2019). Average saliences: El País 7.4%; El Mundo 9.7%; Congreso de los Diputados 6.1%. Overall Pearson’s correlation between media and parliament arena: 0.66 ( $p < 0.00$ ).

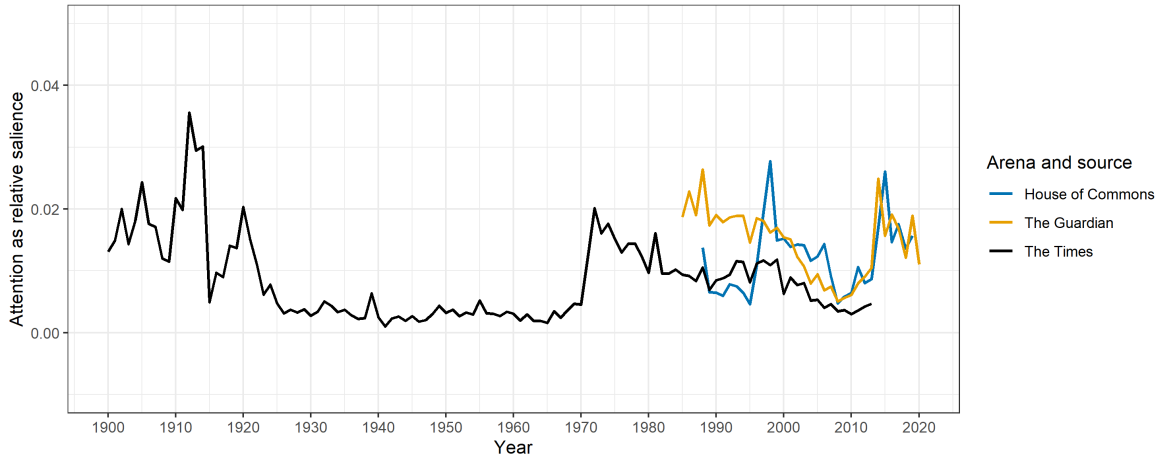
at the center of public debate and attention,<sup>10</sup> due especially to the Scottish Home Rule Bill of 1913 and the Irish War of Independence in 1916. However, the Government of Ireland Act of 1920 and the Anglo-Irish Treaty of 1921 capped a turbulent decade of state-building, and were followed by nearly half a century of non-salient territorial politics.

The 1970s saw a revival in territorial politics with reinvigorated struggles over authority in Scotland, Wales, and particularly in Northern Ireland (the “Troubles”). The Times and the Guardian began to put more emphasis on territorial politics, with around 1.3% of articles addressing the issue. Nevertheless, attention to territorial politics in both newspapers decreased until around 2014 when the Scottish independence movement gained more traction.

On average, 1.2% of parliamentary speeches in the House of Commons since 1985 have referred to territorial issues. Between 1995 and 1998, and again between 2013 and 2019, attention to territorial politics was considerably greater, due in the earlier period to the Good Friday Agreement (ending the Northern Irish conflict) and the devolution process which culminated in the establishment of Welsh and Scottish parliaments.

<sup>10</sup>Issues that would have increased salience were those such as colonialism in general, the Boer war, trade with Latin America, which were substantially more salient than, for example, Irish independence, but we included domestic territorial issues. We should also keep in mind that newspapers in the early 20th century were very different, focusing less on news and more on reporting stock tables, marriages, housing and other markets, death notices, etc.

**Figure 3.2:** Attention to territorial politics across political arenas in the UK



Note: Parliamentary speeches and media articles; sources: Rauh and Schwalbach 2020; The Times (1900–2013); Guardian (1985–2020). Average saliences: Guardian 0.8%; The Times 1.5%; House of Commons 1.2%. Overall Pearson’s correlation between media and parliament arena: 0.45 ( $p = 0.01$ ).

### 3.5.2 Attention to sub-issues

Researchers might be interested not only in measuring attention to pre-defined issues of interest, but also in measuring the distribution of emphasis across sub-issues within them. The optimized dictionary is ideal for identifying the universe of text which can be further subdivided using approaches such as topic models.<sup>11</sup> To have a pre-selected corpus of articles or text fragments identified by optimized dictionaries provides a substantial advantage for the application of topic models. We make use of the selection of news articles and parliamentary speeches by the optimized dictionary to explore and identify sub-issues with Structural Topic Models (STM) (Roberts et al. 2013, 2014). We selected STM since we expect topics to be correlated and because we expect topic distributions across documents to be dependent on the newspaper. STM, in contrast to other topic models, allows topics to be correlated and the distribution of topics to be influenced by external factors (Blei and Lafferty 2006; Roberts et al. 2014). We assess sentences<sup>12</sup> as the meaningful text entity in articles and speeches, because news reporting and parliamentary interventions often only make references to territorial politics within an issue bundle in each text unit. Moreover, we expect to find the

<sup>11</sup> Alternatively, and more time-consumingly, researchers might develop optimized dictionaries for every sub-issue.

<sup>12</sup> Debate continues on whether to study whole texts, paragraphs, sentences, or windows of words preceding and succeeding dictionary keywords (see Hutter 2014 for a similar discussion on content analysis). We excluded interventions by the parliamentary chair, which can bias attention dynamics due to their protocolary nature. When sentence structure was not available in the data due to OCR, e.g. in The Times, we extract paragraph entities.

most interesting differences between democratic arenas such as the media and the parliament in emphasis across sub-issues. From the perspective of strategic incentives for parliamentarians and the media, we would expect the media to overemphasize conflictual sub-issues, such as violent expressions of separatist ambitions, more than parliaments, whereas politicians should overemphasize fiscal or administrative aspects of territorial politics.

To find the optimal number of sub-issues/topics in the STM of each country, we conducted models allowing for 5 to 50 topics ( $K$ ). We then inspected the models by observing the performance diagnostics parameters: semantic coherence, exclusivity, lower bound, held-out likelihood, and residuals. In the Spanish case, the models  $K = 9$  and  $K = 11$  performed best for exclusivity and semantic coherence, while the rest of the parameters behaved relatively similarly in this  $K$  topics range.<sup>13</sup> Following the same procedure, we selected a British model with  $K = 19$  over a model with  $K = 11$ .

The STM predictions indicate the proportion of the text units – that is, sentences – belonging to one of the  $K$  topics. We qualitatively assess the topics and decide based on correlations and substantial content whether to aggregate the topic and their proportions for each observation. In our case, the aggregated topics are conceptualized as sub-issues of territorial politics. It might also be feasible to exclude remaining false-positive texts by recognizing residual topics via STMs. In Spain, four topics were related to Basque separatism, two to Catalan separatism, two to political autonomy in general, and one to fiscal and administrative authority. In the UK, two topics were related to administrative and fiscal autonomy, two to political autonomy, one to Scottish separatism, and four to Irish separatism (for example, by focusing on violence, political accords, or important politicians and religious leaders). Two other topics were related to cultural aspects such as television programs or theatrical productions about the Irish troubles. Another eight topics were miscellaneous, either focusing on territorial politics in other regions of the world or collecting words that were related to speeches in parliament (“hear, hear!”, “Cheers”) and OCR errors (stocks and market news, unreadable “words”).<sup>14</sup>

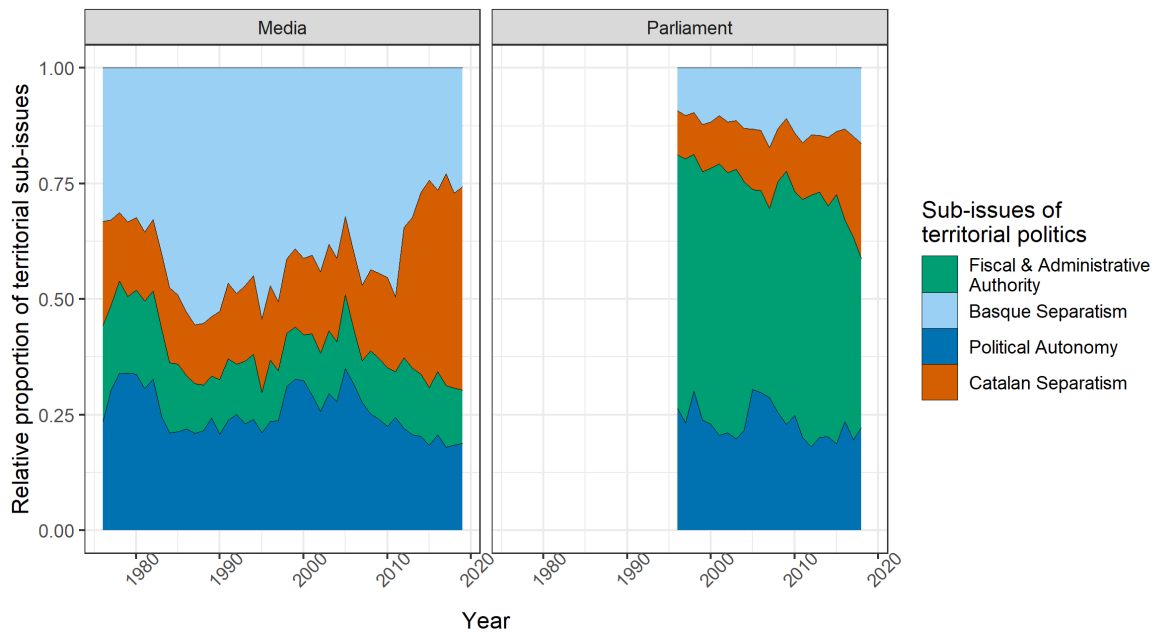
We plot the relative emphasis of the grouped sub-issues for Spain in Figure 3.3. In newspapers, the sub-issue of Basque separatism is the most prominent across almost the whole period, taking up more than half the attention between 1980 and 2010. This

<sup>13</sup>Within both  $K = 9$  and  $K = 11$ , the exclusivity and semantic coherence scores of the topics were alike. The research group then discussed qualitatively both models based on expert knowledge on territorial politics and face validity of the results to select the substantially more meaningful model, in this case,  $K = 11$ . Overall, models with fewer topics, i.e. lower  $K$ , reduce the complexity of the topics. See section SM3.6 and SM3.8 of the Supplementary Material for more details.

<sup>14</sup>See section SM3.7 of the Supplementary Material for Figures with topics without aggregation.

is in line with the idea of the newsworthiness of conflictual issues, due especially to ETA-related violence and counterviolence. In contrast, references to Basque separatism were relatively less frequent in parliament, where fiscal and administrative sub-issues received substantially more attention. Furthermore, attention to political autonomy follows similar patterns in newspapers and parliament. Newspapers depict the importance of this sub-issue preceding and surrounding the constitutional discussions in the late 1970s just before the emergence of autonomous political democracies in the regions. Attention to political autonomy rose again as challenges to the status quo escalated around the turn of the millennium (Orte and Wilson 2009). Similarly, Catalan separatism gains most attention in news outlets, with the street mobilizations of 2011 onwards (Liñeira and Cetrà 2015), and enters the parliamentary arena both later and with less intensity. This is congruent with the two most widespread expectations of attention politics. First, the media are the predominant agenda-setters and, second, the media overemphasize alarming and conflictual issues, thus resulting in the gap in emphasis between the two arenas.

**Figure 3.3:** Attention to territorial sub-issues in Spain



Note: Proportions reflect the yearly aggregation of sub-issues on the bases of sentences in newspaper articles and parliamentary speeches. The sum of substantially relevant topics equals 1. The sub-issues correlate across arenas as follows: Catalan separatism 0.79 ( $p < 0.00$ ); political autonomy 0.53 ( $p < 0.00$ ); fiscal and administrative authority -0.15 ( $p = 0.50$ ); Basque separatism -0.25 ( $p = 0.25$ ).

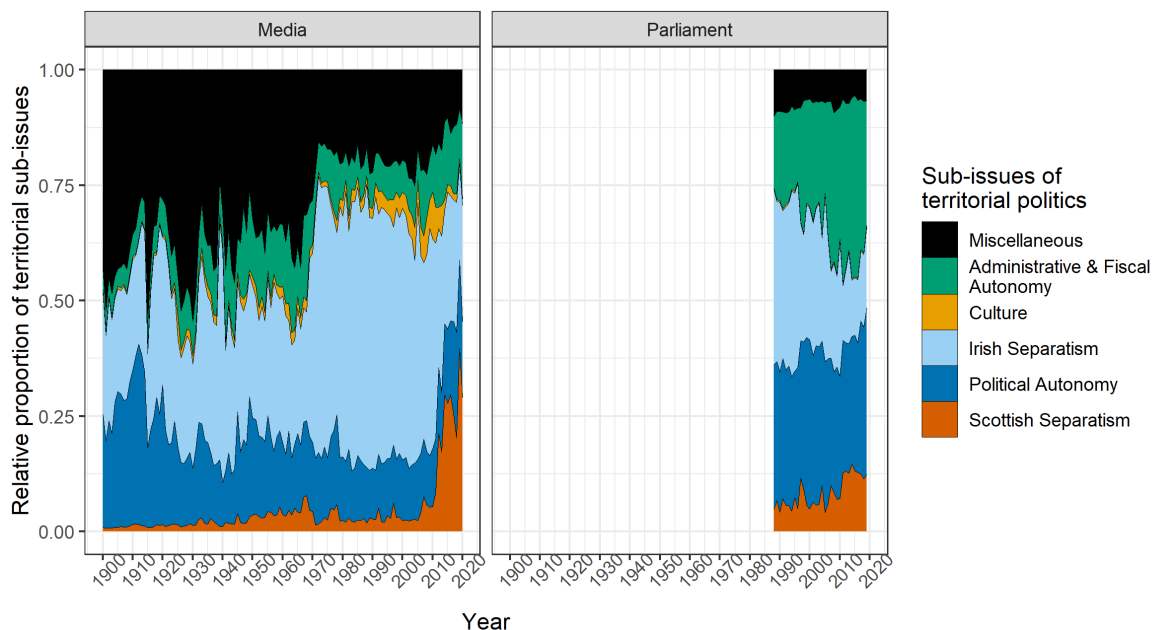
Figure 3.4 displays attention to territorial sub-issues in the UK media and the House of Commons. OCR errors prevented many sub-issues in the Times from being



described with STM. However, the Figure clearly shows that political autonomy and Irish separatism dominated territorial politics articles in The Times until the 1960s. From the 1940s to the 1970s, administrative and fiscal autonomy also played a major role, mirroring debate on relationships between an independent Ireland, Northern Ireland, and the UK, and particularly on the status of Northern Ireland. While attention to political autonomy gradually decreased after the war of independence, attention to Irish separatism continued to increase until the implementation of the Good Friday Agreement, when protest and violence in Northern Ireland dropped. Subsequently, in the light of devolution and Scottish separatism, administrative and fiscal autonomy have become more prominent in Guardian and Times articles on territorial debates.

We see a different picture in the House of Commons, where attention to more technical sub-issues of territorial politics, such as administrative and fiscal autonomy or political autonomy, has been much greater. Surprisingly, Scottish separatism has not played a large role, while the focus on Irish separatism has decreased significantly in the last couple of years.

**Figure 3.4:** Attention to territorial sub-issues in the UK



Note: Proportions reflect the yearly aggregation of sub-issues on the bases of sentences in newspaper articles and parliamentary speeches. The sum of substantially relevant topics equals 1.

Figure 3.4 usefully identifies historical and OCR-related bias within the British sources. The miscellaneous category is substantially reduced by a datapoint from the Guardian entering the model (1980). Within the two phases before and after access to

the Guardian data, there is hardly any substantial difference over time. That means there is little historical bias in the attention scores, but a strong OCR-related bias across time. The miscellaneous category would be further reduced by using only Guardian estimates and would then be comparable to the size of that category in parliament, where OCR-related errors are absent.

The analysis of sub-issues demonstrates that optimized dictionaries can be a good starting point for topic models because they help to delineate the universe of relevant text. Of course, numerous inferences could be drawn from such a fine-grained analysis of attention to territorial issues in both countries. The most important, however, stands out very clearly. Newspapers and parliaments emphasize sub-issues of territorial politics very differently. In line with our initial framework, politicians in parliaments prefer technical and solution-oriented issues to demonstrate competence and suppress internal divisions. Newspapers are more driven by newsworthiness, which leads to a substantially stronger emphasis on conflict. These findings might not be altogether surprising – but we aim not for surprise, but at a valid measurement of the magnitude of differences in attention across sources of text and political arenas.

## 3.6 Conclusion

In this study, we present a semi-automated way of sorting text fragments into issues with the primary aim of identifying valid and comparable issues in text across political arenas. Recent advances in text-as-data approaches enable us to extend the scope of data generation and classification beyond resource-intensive hand-coding. Yet none of the existing methods combines the minimization of all three major sources of error (false positives, false negatives, and endogeneity bias) with transferability to new sources of text. We set out to achieve the precise combination of methods which will yield the best results.

We optimize expert-based dictionaries via machine learning, and demonstrate that they minimize error while simultaneously outperforming all other methods in terms of transferability. Developing optimized dictionaries is initially resource-intensive, but once achieved, they can be applied to all sorts of texts, and thus are characterized by increasing returns. Validated by randomly selected and hand-coded text, we show that optimized dictionaries minimize false positives and false negatives substantially (F1 scores around 0.9 across sources) and, astonishingly, perform equally well on new and even unseen text corpora, thus offering the potential to generate issue attention scores that are comparable across political arenas, time, and space, and will greatly serve the

accumulation of knowledge in the domain of attention politics.

We demonstrate our case with a discussion of attention to territorial politics in the leading newspapers vis-à-vis the parliaments in Spain (1976–2019) and the UK (1900–2020). We show that our optimized dictionaries validly identify the content of text – a necessary condition for drawing substantial inferences on the distribution of attention based on the classification of text. We compare the emphasis placed on territorial issues (and their sub-issues) across media and parliaments and confirm what we theoretically discussed as the different incentive structures of media and parliamentary arenas of political communication.

Politicians in parliaments, in particular members of the executive, have a strong incentive to emphasize solution- and competence-oriented issues, whereas the media, driven by concerns of newsworthiness, has greater incentive to emphasize conflictual issues. Based on the distribution of sub-issues in territorial politics this incentive structure is highly visible in both the UK and Spain. Typically, more than 50% of relevant newspaper articles conveyed information on violent territorial conflict and strident political contestation such as Basque separatism or the Irish troubles, whereas the same issues rarely exceed 20% of parliamentary attention. In contrast, territorial debates in parliaments typically focus on competence-oriented issues such as the distribution of administrative authority across levels. Administrative and fiscal issues account typically for more than 40% of parliamentary attention in the domain of territorial debates and less than 10% in the media.

Overall, we make the case that researchers should abandon applying text-classification approaches in isolation and start to exploit the complementarity of their different strengths. Optimized dictionaries are a promising avenue for such an endeavor, because they can validly identify issues within texts and have unique properties in terms of transferability. Once an optimized dictionary has been developed an (almost) infinite number of applications are possible, leading to comparable measures of attention, one of the scarcest resources in social interactions.

## MEDIATED PARTY POSITIONS FROM NEWSPAPERS

(co-authored with Lea Kaftan and Leonce Röth)

### **Abstract**

Party positions directly communicated by politicians – in election manifestos or in parliamentary debates, for instance – do not usually flow unfiltered to citizens. The media plays an important role by altering the salience of parties’ political issues and in this way affecting the transmission of party position to voters, whose picture of parties is thus a *mediated* one. This possible disconnect in the voter–party relation is fundamental for representation studies. We propose to focus on *mediated* party placements as portrayed by journalists in newspapers. To do so, we develop a methodological procedure to obtain party positions for fine-grained policy (sub-)issues from newspaper texts in an automated manner with sentiment analysis. We focus on the case of territorial politics and its sub-issues in Spain (1976–2019), and make use of the news corpora of El País and El Mundo. Our estimation measure converges to a satisfying degree with expert judgments and manifestos, in accordance with our argument that the media alters party self-placements due to its own logic of communication. By comparing positions across newspapers and with parliamentary speeches, we find that an aggregation of 20 sentences combined with topic certainty information for sub-issues of interest is ideal to arrive at valid party positions. We describe the shifting positions on decentralization of two statewide and two regionalist parties over time in Spain, and apply our measurement to a model of party policy shift and regional voter transition. We conclude by discussing the research implications for a new agenda of mediated party positions.

## 4.1 Introduction

Investigating political parties and their positions is fundamental for the study of democratic representation and policy-making. In comparative politics, the analysis of party positions expressed by parties or politicians themselves in electoral manifestos, legislative speeches and tweets is well advanced, especially in quantitative research based on text (Benoit and Laver 2006; Laver, Benoit and Garry 2003; Sältzer 2020; Slapin and Proksch 2008; Volkens et al. 2020). However, indirect assessments of party positions produced by external sources – e.g. by expert surveys or ideological estimations based on Wikipedia entries (Bakker et al. 2021; Herrmann and Döring 2021) – are rarer. In contrast, estimations based on mass media communication have experienced less success in political science. Still, manually annotated content analysis remains very popular (Kriesi et al. 2012). It is indisputable that political communication about party stances is less about self-placements than being placed by the media. Media party placements can therefore complement our understanding of party politics.

The lacuna of quantitative text analysis of mass media and party positions is striking (see Dumdum and Bankston 2021), although there are a few studies which tackle this link, mostly applying manual coding (see Adams, Weschle and Wlezien 2021; Baumann and Gross 2016; Helbling and Tresch 2011; Ruedin 2013; Schwarzbözl, Fatke and Hutter 2020). This gap is not only methodologically curious, but has substantial consequences for political communication, since citizens are more likely to inform themselves on daily politics through the media than directly from parliament (Banducci, Giebler and Kritzing 2017). The indirect reception of party positions is important for a sense of representation and gives the media a powerful intermediate position in communication between representatives and the represented – the voter–party link (Costello, Thomassen and Rosema 2012; Costello et al. 2021). Surprisingly, researchers have not yet engaged in measuring political actors’ positions, as conveyed by the media, by exploiting the benefits of quantitative text analysis and the availability of constant news reporting on both major and niche issues in non-electoral periods. Needless to say, the mass media does not always offer a truthful representation of party positions: rather it *assigns* a position to a party. This is what we call *mediated* party positions.

To study mediated party positions, we use a new procedure that combines topic models and sentiment analysis. We chose sentiment analysis instead of word-scaling due to the traveling capacity of sentiment across different text-generation contexts, such as different news outlets. Using optimized dictionaries (see chapter 3), we first select sentences from newspaper articles that mention a policy issue of interest and

parties that we are interested in. We then implement topic models, which provide us with the portion of a text that belongs to a certain topic – topic prevalence – and thus functions as a certainty measurement for the appearance of a specific issue or sub-issue of our policy dimension of interest. We use sentiment to measure the position on the issue parties are assigned in each sentence, and weight these with our certainty measurement. Using different convergence comparisons and validity checks, we show that this measurement procedure allows us to capture valid mediated party positions. Our study demonstrates that newspapers offer very good ballpark estimates of party positions.

The empirical application centers on territorial politics in Spain and the analysis is based on the full newspaper corpora of center-left *El País* (1976–2019) and center-right *El Mundo* (2002–2019). Territorial politics is a salient policy issue in Spain. Party positions on this issue have typically been measured on a very aggregate level (Basile 2016; Mazzoleni 2009; Toubreau 2017; Toubreau and Wagner 2015), despite the fact that they contain a great deal of variation on many meaningful territorial sub-issues (Alonso 2012; Meguid 2009; O’Neill 2005; Verge 2013). For example, territorial stances can vary across regions – some may want more autonomy for Catalunya, but not necessarily for Galicia. And positions can vary across sub-issues of autonomy – some support politico-administrative, but not necessarily fiscal autonomy. Territorial politics provides an interesting training ground for assessing party positions on an issue that is especially affected by sub-issue diversity.

We compare our aggregated issue estimations with established expert survey and manifesto data (Bakker et al. 2021; Volkens et al. 2020). An analysis across newspaper outlets using our approach reveals that at least 20 sentences are necessary to estimate valid party positions within a sub-issue. We also conduct a convergence analysis with estimations by our own approach applied to Spanish parliamentary speeches, to see how direct and indirect communication are related. A validation with hand-coding of media sentences tests the performance of our approach. We find interesting commonalities and differences between the media and parliamentary arenas of political communication. First, media transmission of policy positions has clearly different priorities in terms of emphasis – highlighting conflictual, “newsworthy” issues whereas technical issues are more prominent in parliament (see chapter 3 in this dissertation; van der Pas and Vliegenthart 2016). Second, party positions in both arenas converge, but not perfectly – in line with theory. Furthermore, we find mixed evidence of systematic ideological bias induced by specific newspapers. Lastly, using mediated party positions on territorial sub-issues in Spain, we show an application to a popular dilemma in political science (Adams 2012), namely the sequential question of regional voter support

transition and party policy shifts: who reacts to whom? Such analyses demonstrate the potential applicability and the future agenda that mediated party positions enable. Party positions from media data can complement existing approaches and help improve our understanding of political representation and political communication in democratic systems.

## 4.2 Party positions and newspapers

Political parties are a core characteristic of democratic systems, and the understanding of their positions on relevant issues is fundamental for the study of political communication, representation and policy-making. The party politics literature has demonstrated that party competition is all about connecting position to issues and playing with their emphasis (Budge and Farlie 1983; de Sio and Weber 2014; de Sio, de Angelis and Emanuele 2017; Röth 2017). Historically, the measurement of party positions relies either on expert judgments or on the manual analysis of text documents (Keman 2007, p. 77). Digitization has revolutionized access to text data and spurred a rapid increase of techniques for text-as-data approaches (Grimmer and Stewart 2013; Klüver 2009; Lowe et al. 2011; Roberts, Stewart and Airolidi 2016; Spirling 2016), increasingly also applied to the identification of party positions (Laver, Benoit and Garry 2003; Slapin and Proksch 2008; see applications by Arnold, Hug and Schulz 2009; Carter et al. 2018; Guntermann 2016; Kriesi et al. 2012; Lehmann and Zobel 2018; Ruedin 2013; Ruedin and Morales 2019).

These approaches come with widespread advantages, but also restrictions. Expert surveys and social media data have restricted temporal reach, whereas approaches based on documents such as manifestos provide long time series, but temporal snapshots with limited precision on specific issues and for selected parties only (Helbling and Tresch 2011). Press releases published by parties can function as an alternative means of extracting party positions between elections, but are determined by partisan strategic concerns about which topics to address when (van der Velden, Schumacher and Vis 2017). Using legislative speeches is one way to arrive at positional estimates with a denser temporal coverage that allows us to go back in time and obtain positions on issues that parties are often forced to discuss, assuming protocols are available and usable (see e.g. Lauderdale and Herzog 2016; Proksch et al. 2019). All these approaches focus on data of the political communication of elites, whether it be politicians writing manifestos, MPs speaking in parliament, or party politics researchers in universities assessing positions based on their knowledge.

Most citizens neither participate directly nor receive direct communication from

political and academic elites, but typically infer party positions on issues via all sorts of media (McCombs and Shaw 1972, p. 176; Midtbø et al. 2014; Newton and Brynin 2001). Newspapers provide a helpful source of information for party positions over virtually every time span envisaged (Baumann, Debus and Gross 2021). Surprisingly, though, despite the fact that the mass media constitutes the most important arena for public debate (Bennett et al. 2006; Ferree et al. 2002), party positions based on newspaper data have hardly ever been assessed with automated text as data approaches. Some encouraging findings from earlier studies on party positions elicited from media data support the idea of developing an automated process to derive party positions from media data.

Media data have been successfully used to manually measure valid party positions and/or issue salience (Baumann and Gross 2016; Helbling and Tresch 2011; Kriesi et al. 2012). Positions expressed in print media tend to converge with inferences from party manifestos or expert placements (Helbling and Tresch 2011). The Comparative Agenda Project has also greatly contributed to the study of newspaper coverage on political parties through manual content analysis of cover pages (Baumgartner and Chaqués-Bonafont 2015).

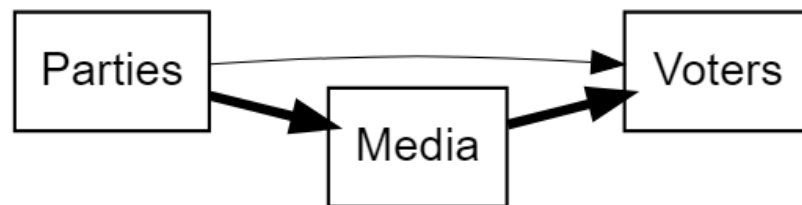
Nonetheless, there is also reason not to take at face value party placements in newspaper data as one might with direct communication. Both media and political elites follow their own logics of communication (Allern 2017; Bräuninger and Debus 2009; Domingo et al. 2008; Green-Pedersen 2019; Staab 1990) and try to influence each other: politicians by influencing the public agenda, and newspapers by *mediatizing* political debates. Political parties aim to set the agenda and influence newspaper coverage of certain issues according to opportunity structures of issue salience and ownership (Guinaudeau and Palau 2016). However, when conveying party stances, media may introduce many sorts of biases (Banducci, Giebler and Kritzing 2017; see Davenport 2009) absent from directly communicated speeches or original text-documents.

Given their news selection is driven by their commercial incentive of *newsworthiness*, newspapers are biased towards covering broken promises rather than pledges fulfilled, conflictual issues rather than consensual ones (Müller 2020; Niblock and Machin 2007, p. 191; Strömbäck, Karlsson and Hopmann 2012, p. 726; van der Pas and Vliegenhart 2016). Newspapers are not neutral or disinterested, but favor some political actors over others, reflected in their tonality, their portrayal of political agents or the relative emphasis of faults and virtues (Baumgartner and Chaqués-Bonafont 2015; Puglisi 2011). Furthermore, journalists follow their own agenda, driven by subjective judg-



ments, ideological leanings and address books (Donsbach 2004; Herman and Chomsky 1988; McChesney 2000; Mügge 2016; Staab 1990). Hence, it is undoubtedly clear that party positions derived from media data are not necessarily the precise ones which political actors intend to convey. Consequently, democratic theory clearly suggests that the media does indeed have an active role in the transmission of political communication (Curran 2011; Graber 2003). As the theoretical model in Figure 4.1 depicts, the mediated transmission channel of party positions to voters is probably stronger than a direct communication channel between parties and voters, e.g. through electoral rallies or official party communication.<sup>1</sup>

**Figure 4.1:** Simple top-down model of party position transmission: parties, media, and voters



Note that citizens receive party position signals from the media which are thus *mediated* (see Newton and Brynin 2001), leading to a divergence between *true* and *media-conveyed* positions (Schwarzbözl, Fatke and Hutter 2020). According to the mediated party positions model, newspapers afford the prevalent debates in society more space, highlighting conflictual and negative issues, interfering in parties’ issue emphasis strategies, and so on (see Baumann, Debus and Gross 2021; de Sio and Weber 2014; Rovny 2012). The study of positional divergences and ideological re-/de-/alignment dynamics can therefore benefit from including a media perspective that may potentially help explain democratic representation gaps (see Brooks and Manza 1997; Costello et al. 2021). Moreover, the degree of influence the media exerts on the portrayal of party stances is of fundamental interest and can only be systematically addressed by obtaining valid placements of political actors in the media. Accordingly, we propose a procedure using sentiment text analysis to gain party positional information on the basis of newspaper articles.

<sup>1</sup>Although we do not discuss it, parties also react to citizens’ electoral support or positional shifts (see Abou-Chadi and Stoetzer 2020), which they are probably informed about by the media.

### 4.3 Measuring positions in newspapers with sentiment analysis

In recent decades, quantitative text-as-data methods have been developed to scale text and measure political preferences in an efficient and cost-saving manner. Unsupervised word-scaling methods like *wordfish* (Slapin and Proksch 2008) and correspondence analysis (Schonhardt-Bailey 2008) use text-dimensionality reduction techniques distinguishing language use to situate political texts in the natural latent space of the data corpus selected, usually a policy-dimensional space. Manifestos are usually couched in terms of left–right positioning, whereas parliamentary debates portray the government–opposition divide, since parties’ wordings are driven by bill debates and the government’s performance. Methodological innovations deal with issues of time, context and environment in both approaches. For example, the *wordshoal* algorithm controls for parliamentary debate specificities, and conducts a wordfish estimation of MPs’ speeches within each debate to control for parties’ language use in different contexts (Lauderdale and Herzog 2016). A right-wing party could more often mention “decrease” in a debate about taxation, but emphasize “increase” in a debate on security spending. Similarly, canonical correspondence analysis (CCA) emulates the wordshoal correction by harmonizing the text wording for position scaling based on environmental factors, e.g. the source of the data (see Sältzer 2020).<sup>2</sup>

Although text-dimension reduction based on semantics works well for electoral manifestos and parliamentary speeches, we argue that this is not the case for indirect placements of parties on low-salience issues, as in textual media data. Thus, these word-sensitive approaches use larger texts within each document of a corpus and compare documents generated in similar contexts and periods sharing a similar communication strategy. Although exploiting different media sources is advantageous to cover time periods of interest and to grasp party positions on specific sub-issues, the variety of sources creates two fundamental challenges for word-scaling.

First, indirect party placement by the media within its news reporting introduces wording selectively and reflects the intentions and motivations of the media mouthpiece rather than those of the political actors themselves (see Davenport 2009; Gentzkow and Shapiro 2010; for an overview, see Hamborg, Donnay and Gipp 2019). A right-leaning outlet is more prone to write about public debt, while a left-leaning outlet writes about public investment – essentially the same issue. When both outlets talk about different political actors, their ideologically driven wording can cancel out party

---

<sup>2</sup>Section SM4.1 in the Supplementary Material describes these approaches in more detail.

positions, rendering environmental or contextual diversity harmonization techniques inadequate to erase the preferred terminology of a particular medium. By comparison, politicians’ direct communication offers either more information on one occasion (one manifesto or one major speech)<sup>3</sup> or many shorter signals from a source that is genuinely the same across time (press releases or tweets).

Second, since we are interested in sub-issues, which tend to be fragmented and often fine-grained, the amount of text of interest is very limited. Newspapers only rarely speak about the position of a party on a specific policy issue, in comparison to the bulk of their coverage. With small amounts of text for each issue position estimation, semantic word-based scaling is sub-optimal, and selective wording by the source very easily shifts the estimation. Such a *mediated* distortion is serious, because when “a party is rather small, its position in the coverage of an issue is likely to be reported only if it is an issue owner” (Schwarzbözl, Fatke and Hutter 2020, p. 803).

As a suitable alternative, we propose to use aggregated signals from sentiment analysis, a straightforward method that has gained popularity in recent years. Taking the tone of text into account improves our understanding of political phenomena as portrayed in the media (Young and Soroka 2012). Sentiment analysis uses deterministic counts of positively and negatively connoted words in text documents based on a dictionary. A key feature of sentiment analysis is that the reference point across texts is fixed, so that sentiments towards a phenomenon can be compared (Rudkowsky et al. 2018; see Castanho Silva and Proksch 2021)<sup>4</sup>. Unlike word-scaling, sentiment analysis is less prone to biases from different sources and data-generation processes. In sum, sentiment analysis more readily travels across heterogeneous communication patterns, due to its universality, and suits indirect party placements better. In the following, we discuss how to tackle some of the measuring challenges for obtaining party positions on sub-issues that sentiment analysis presents.

---

<sup>3</sup>The shorter the texts, the more problematic word-scaling methods become, as they are sensitive to word frequencies after pre-processing (Greene et al. 2016), the so-called “word inclusion criteria” (Proksch and Slapin 2009, pp. 331ff.). This challenge intensifies if short texts are independent of each other and generated using context-dependent language, as in mass media.

<sup>4</sup>Although the true benchmark changes across texts due to topics or journalistic influence (see Rauh 2018), the overall measure is comparable across texts. We deal with exactly this issue, in that we integrate aggregation and certainty optimization. For example, Proksch et al. (2019) elegantly apply sentiment analysis to MPs’ speeches on the Irish budget debate (a fixed benchmark) to reflect the government–opposition divide.

### 4.3.1 Information load

As indicated above, newspaper and media signals in general offer more fragmented policy signals than speeches or party manifestos, particularly as the media is not constantly referring to parties and issues, but rather reports on these in small segments diluted across news. Accordingly, we discuss questions of aggregation of these signals and inclusion of surrounding sentences as contextual information load.

#### 4.3.1.1 Aggregation

In the world of bag-of-words, carefully choosing the unit of analysis and the level of aggregation is key. Sentiment within sentences or text segments can generate biased signals if connoted words do not refer to the phenomena of interest or if negation is frequent. Without embedding models or direction-sensitive transformers, this cannot be solved at the local level (Liu, Jia and Vosoughi 2021; Rodríguez and Spirling 2021; Rudkowsky et al. 2018). However, in party politics research we are interested in global positions – “Instead of seeking to classify any individual document, most social science literature that has hand- (or computer-) coded text is primarily interested in broad characterizations about the whole set of documents” (Hopkins and King 2010, p. 243; see also van Attevelde, Kleinnijenhuis and Ruigrok 2008, p. 437). Aggregation of text for specific time periods is therefore required in order to acquire unbiased party positions from newspaper text (see Ceron, Curini and Iacus 2015; O’Connor et al. 2010).<sup>5</sup>

Hence, sentences on the sub-issue of interest that also report on political actors seem to be the most appropriate unit of analysis for our endeavor; they are then aggregated to capture valid sentiments and positions. The established sentiment measure for counts is the natural logarithm ratio of the positive and the negative words, adding +0.5 to each count because division by and the logarithm of zero are not defined mathematically (see Lowe et al. 2011).<sup>6</sup> The logarithm reflects the marginal decrease in tonality caused by repetitive notions (Weber–Fechner law) and can be noted as follows for party  $p$  in time period  $t$ :

---

<sup>5</sup>When using manually annotated party positions, as in the case of MARPOR/CMP’s quasi-sentences (Volkens et al. 2020), local estimations at the lowest level of analysis can be accurate, since each assessment has a higher certainty probability and “noise” is reduced through human coding (see Röth 2017).

<sup>6</sup>Additionally, a quantitative approach with local sentence estimation would be biased towards zero, since sentences without sentiment-loaded wording appear frequently. This bias does not occur with previously annotated text data, since sentences are pre-filtered by humans based on content-load.

$$sentiment_{p,t} = \log\left(\frac{pos + 0.5}{neg + 0.5}\right)$$

Although aggregation is a helpful tool to measure political actors' positions in news texts, some difficulties remain. There is still uncertainty, both conceptual and empirical, as to the optimal aggregation period. Further, if even humans have difficulties "reading" and identifying sentiment and positions, automated approaches are likely to amplify this challenge. For example, sentiment dictionaries are agnostic about what a "neutral" position can mean, especially with unbalanced wording lists for negativity and positivity (Rauh 2018). Furthermore, indirect communication of party placements will increase uncertainty compared to unmediated communication by politicians. Journalists inflate uncertainty by rapidly changing narrative contexts, through comparisons, and by using sarcasm or injecting subjective evaluations. Unlike directly communicated party positions, we cannot take the estimated sentiment of any media statement at face value. Aggregation of more sources and indirect party mentions is, then, the primary option to improve measurement and reduce the weight of misclassified statements, but it is not the only one.

#### **4.3.1.2 Context**

We explained that a reduced selection of sentences related to a specific political phenomenon (sub-issue) mentioning the political actor is the optimal unit of analysis to minimize random noise. However, often short text passages such as sentences can be either uninformative, because there is no sentiment signal (more usual for shorter sentences), or are noisy by chance. Therefore, including the preceding and succeeding reported sentences in media articles can help to capture contextual information related to parties and the specific political phenomenon of interest, even though they lack clear reference to the target subject-matter. Imagine, for example, the following example: "It appears to be the worst, most mediocre fiscal proposal of all time. Therefore, the Socialist party rejected the bill vehemently". Within article news, we include the sentences which precede and follow our sentences of interest, as long as they do not mention other political actors.

### **4.3.2 Certainty**

#### **4.3.2.1 Issue certainty**

We know from party research that the emphasis of issues on the agenda interacts with party positions in order to transmit policy signals. In content analysis, human

annotators usually assign a text to a single category – e.g. topic – without coding the certainty with which it belongs to this and only this category. In quantitative text analysis, however, texts are frequently allocated multiple memberships, and certainty and proportional measurements are a common by-product. As a consequence, media sentences on different political phenomena can be assigned to different issues and, by default, to sub-issues to different degrees. Topic modeling allows us to infer what fragment of a document belongs to a specific topic (Blei and Lafferty 2007; Roberts, Stewart and Airolidi 2016). If a sentence reporting on a political party, for example, relates to two policy issues, its positional signal towards each of the issues should not be weighted as much as if only one issue is being raised. We can make use of topic prevalence to give greater weight to text passages carrying more informative load relating to the issue of interest, in order to estimate a position on that issue.

Text passages can be highly fragmented or belong to a residual topic, which would be uninformative for our purposes.<sup>7</sup> We adapt Lowe et al.’s (2011) scaling formula and introduce a condition whereby we make use only of sentences with a topic prevalence with absolute majority (prevalence > 0.5). We adjust each sentence’s positive and negative counts by the prevalence ( $\sim$  weights). Accordingly, the positive–negative rate variation is influenced by topic content certainty for the issue or sub-issue of interest. In other words, for every sentence and only for every sentence with *topic prevalence* > 0.5, our weighted sentiment position is

$$sentiment_{p,y,s-i} = \log\left(\frac{(\sum_{s=1}^n pos \times topic\ prevalence) + 0.5}{(\sum_{s=1}^n neg \times topic\ prevalence) + 0.5}\right)$$

where the sentiment position of the party  $p$  in period  $t$  within a sub-issue  $s-i$  (or issue, more generally) is the logarithmic transformation of the ratio of positive and negative counts, plus 0.5 for each count. The count of positive and negative mentions is weighted at the sentence level  $s$  by the sub-issue weight *topic prevalence* above 0.5 and then aggregated by period for all party sub-issue sentences.<sup>8</sup>

#### 4.3.2.2 Direction

Since sentiment analysis measures the global tone of a sentence, it is ignorant of the order of words. Within a bag-of-words framework, "The commission supported the

---

<sup>7</sup>Challenges of certainty are not exclusive to text analysis. Expert judgments are often based on uncertain assessments and on different understandings of issues among experts (see Castanho Silva and Littvay 2019).

<sup>8</sup>Simultaneously with the development of our work, Diaf and Fritsche (2021) introduced a methodological approach inspired by the same intuition of utilizing topic modeling for scaling.

fiscal reform, whereas Party X did not” would portray the same sentiment as ”The commission did not support the fiscal reform, whereas Party X did.” Although negation is a conundrum in bag-of-words analyses, in the aggregate such misleading cases can be canceled out and deliberately accepted. Still, controlling for wording associated with the political issue direction is easier. If we are measuring preferences or positions on the welfare state, sentiment regarding state intervention/investment, welfare expansion, and social programs is the naturally flipped version of sentiment regarding liberalization, privatization, and retrenchment. In that case, sentiment signals should be re-scaled for comparability.

#### 4.3.2.3 Media source harmonization

One last challenge of working with newspaper data and indirect party placements is source and journalistic subjectivity. The literature has systematically shown that media sources have inherent ideological bias when reporting on political phenomena (Banducci, Giebler and Kritzing 2017; Davenport 2009; Mugge 2016; Staab 1990). This is problematic for party position estimation, since it can be driven by (a) the sentiment journalists associate with certain parties due to general ideological affinity, independent of the policy issue, or (b) the sentiment journalists associate with specific policy issues due to political convictions. Accounting for overall newspaper sentiment towards the specific policy (sub-)issue and setting this as the centered benchmark of sentiment can help to reduce journalistic influences. Thus, the deviation of the party position from the newspaper sentiment benchmark harmonizes source biases.

**Table 4.1:** Sentiment measurement adaptations for party positions in newspapers

<b>Information load</b>	
Aggregation	Measure of global sentiment through text aggregation for optimal time frames
Context	Include sentences preceding and succeeding the text passage relevant to the political phenomenon and the political actor
<b>Certainty</b>	
Issue certainty	Weight sentiment counts at the sentence level by topic prevalence of the issue of interest
Direction	Re-scale sentiment at the sentence level based on wording direction of the policy issue captured
Media source harmonization	Deviation from the overall sentiment of the newspaper in non-party sentences

On the whole, then, sentiment analysis is a promising approach to obtaining automated placements of political parties on specific issues as portrayed in the media. In

the following, we introduce our case study, territorial politics in Spain, and describe the data-generation process of acquiring policy-specific sentences from newspaper media.

## 4.4 Case study

To put the measurement of partisan placements in the media to the test, we apply our method to territorial politics in Spain.<sup>9</sup> We have three reasons for this choice. First, Spain is a country where the media is truly politicized, so politics matter (Baumgartner and Chaqués-Bonafont 2015). This is why we can expect party positions to be communicated. Second, territorial politics is highly salient in Spain, but it consists of various fine-grained sub-issues which are not easily grouped in an overall territorial issue. For example, parties may embrace some substantive sub-issues of decentralization (fiscal, political, or administrative),<sup>10</sup> but reject others. Third, parties may pursue different policies and objectives in different regions (Alonso 2012; Bednar 2004; Meguid 2009; O’Neill 2003, 2005; Röth and Kaiser 2019; Röth et al. 2016). This sub-issue diversity provides a hard test of whether media data is able to distinguish specific stances of parties within broader political issues such as territorial politics. Furthermore, to capture this sub-issue diversity would be a major complement to existing measures, since territorial politics is typically dealt with on a very abstract level in conventional party politics research (Bakker et al. 2021, codebook, p. 23; Benoit and Laver 2006; Volkens et al. 2020, codebook, p. 13).

This diversity also exists empirically. For example, the political parties *Partido Popular* (PP) and *Partido Socialista Obrero Español* (PSOE) differ in the ways in which they have positioned themselves towards Basque separatism, especially regarding the PSOE government’s peace negotiations with the terrorist group ETA in the 2000s. They also adopted obviously different positions on the 2006 Catalan statute of autonomy; by taking the reform to the constitutional court, the PP clearly demonstrated its opposition (Basta 2017, p. 59; Toubreau 2017, p. 90). However, they have taken very similar stances on the decentralization of political autonomy to slow-track regions like Valencia and Andalusia in the 1980s and 1990s – the *café para todos* strategy (Toubreau 2017, p. 87).

We are the first to apply automated positional measurement of political parties

---

<sup>9</sup>Deriving from the Comparative Agenda Project, another case study has focused on coverage and general tone of news covering territorial politics in Spain (Baumgartner and Chaqués-Bonafont 2015). We extend their study in three ways. First, their analysis is based only on front pages. Second, their coding is entirely manual, while we use a quantitative automated approach. Third, we drill down through territorial politics to its most meaningful sub-issues.

<sup>10</sup>See Hooghe et al. (2016) for definitions of those sub-issues.



using full corpus newspapers (El País and El Mundo) over a long time period. Our selected case is nonetheless partly restricted due to data availability. Before we turn to the analysis of positions using media data, we roughly describe the data-generation process and the selection of sentences used for the location of parties' positions on the territorial politics (sub-)issues.

## 4.5 Data generation

We use all sentences mentioning territorial politics that were published in the two most important statewide newspapers: the left-leaning El País (1976–2019) and the right-leaning El Mundo (2002–2019). The data from El País was web-scraped and the data from El Mundo was collected with LexisNexis. To select sentences with a reference to territorial politics, we use an optimized dictionary for territorial politics in Spain, based on hand-coded validation and machine learning improvements (see dictionary, as described in chapter 3, in section SM3.2 in the Supplementary Material). The dictionary performs extremely well with F1 scores of 0.78–0.89 in newspapers and 0.90 in parliamentary speeches, and outperforms exclusive machine learning predictions.

We identified 572,831 territorial sentences in both newspapers. We then used topic models to identify sub-issues of territorial politics, driven by our theoretical motivation to grasp sub-issues as well. Topic models are an unsupervised method of clustering words into latent topics with the assumption that topics reveal sub-topics when applied to sentences about a specific issue. Because we expect topics to be correlated, and because we expect topic distributions across documents to depend on the newspaper concerned, we use Structural Topic Models (STMs), which allow topics to be correlated and the distribution of topics to be influenced by external factors (Blei and Lafferty 2007; Roberts et al. 2014; Roberts, Stewart and Airolidi 2016).

After optimizing meta-parameters for STMs with  $K = 2$  to  $K = 50$  (exclusivity, semantic coherence, held-out likelihood, residuals and lower bounds), we selected the STM with  $K = 15$  as the best-fit model by also accounting for face validity of the topics (see section SM4.6 of the Supplementary Material). One topic was a residual one containing general government news, and one topic mixed conceptually different territorial sub-issues, which could complicate party position estimation. The number of 13 topics still represents high fragmentation, so we reviewed the literature on territorial politics and read extracts from the topics identified in order to aggregate them in a substantially meaningful manner beyond purely semantic logic. We arrived at four sub-issues of territorial politics for the Spanish case: Basque Separatism, Catalan

Separatism, Fiscal-Administrative Authority, and Political Autonomy.

We created short party dictionaries consisting of the party labels and the prime ministers' and party presidents' names from 1976 to 2020 to identify the parties within our territorial politics sentences. We focus our analysis on two statewide parties, PP and PSOE, and two regionalist parties, *Partido Nacionalista Vasco* (PNV) and *Convergència i Unió* (CiU). For the regionalist parties lacking presence in national government, we also included parliamentary party leaders. To avoid mixed positions, we selected those territorial politics sentences exclusively mentioning one party ( $N = 93,319$ ; this translates into 16.3% of all territorial politics sentences in the media data).<sup>11</sup> Lastly, we used an automatized Spanish translation of the Lexicoder sentiment dictionary to count positive and negative connoted words (Daku, Soroka and Young 2015; Proksch et al. 2019).<sup>12</sup>

In the following, we put our estimation under empirical scrutiny with comparisons (1) with manifestos and expert judgments, (2) between newspaper outlet sources, and (3) between political arenas, to evaluate convergence of mediated party positions with positions derived with the same sentiment approach from parliamentary speeches.

## 4.6 Estimation and comparison analyses

### 4.6.1 Convergence of positions with manifestos and expert surveys

First, we compare our approach of measuring mediated party positions with established methods and data in political science: the Chapel Hill Expert Survey (CHES) (Bakker et al. 2021), covering the years 2006, 2010, 2014 and 2019, and MARPOR/CMP (Volkens et al. 2020), with estimates for Spanish general elections since 1977. For the analyses based on sentiment, we collapse all territorial politics sentences in both newspapers to the year level as an appropriate time unit for aggregation. We began to explore a higher-level time unit for aggregation, which proved suitable for estimation, as we argue below. We use the CHES *regions* item, reflecting the "position on political

---

<sup>11</sup>We also tested long party lists including all members of the party boards and group spokespersons. However, since these politicians are mentioned less often, we capture only around 15% more party sentences, while at the same time increasing the level of uncertainty due to the heterogeneity of the positions represented by different regional groups.

<sup>12</sup>Like Proksch et al. (2019), we used Google to make an alternative Spanish translation of a second dictionary, the German SentiWS dictionary (Remus, Quasthoff and Heyer 2010). All territorial sentences both for positive and negatives counts achieved correlations of  $r = 0.605$  ( $p < 0.000$ ) and  $r = 0.651$  ( $p < 0.000$ ) respectively, with the chosen Spanish Lexicoder dictionary measures.

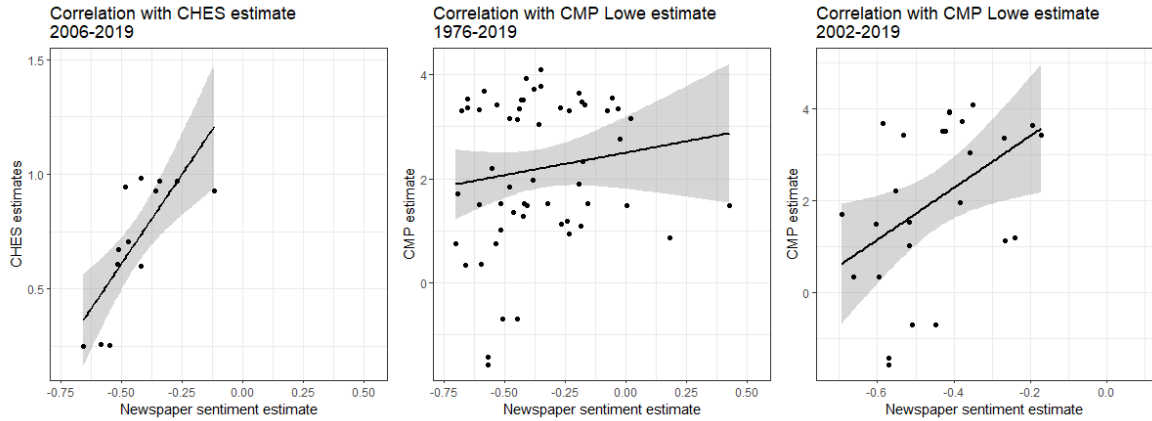
decentralization to regions/localities” from strongly in favor to strongly opposed, with values re-scaled to 0–1. In the case of MARPOR/CMP, items per301 and per302 inform on party ”support for decentralization” of political and/or economic power and on ”general opposition to political decision-making at lower political levels” – i.e. support for decentralization and centralization, respectively. We use the Lowe et al. (2011) transformation to set support and opposition in relation.

Figure 4.2 shows the correlation between our simple newspaper sentiment estimate of party positions and the data from two external sources. First, we see that our *mediated* estimates strongly correlate with the CHES positions at 0.76 ( $p < 0.00$ ; .95 CI: 0.42, 0.91). The correlation with the manifesto data is still positive, but substantially lower at 0.14 and not statistically significant ( $p = 0.30$ ; .95 CI: -0.12, 0.39). As already highlighted, sentiment signals are noisier and more uncertain with less text data (see Figure SM4.18 of section SM4.4 in the Supplementary Material). We therefore test one last correlation with a sub-set of manifestos from the period from 2002, for which we have texts from both newspapers. The average number of sentences per year rises from 544 to 938, a substantial increase of 72.4%. The right panel of Figure 4.2 shows the subset correlation of 0.453 ( $p = 0.02$ ; .95 CI: 0.0887, 0.7109). Although both estimations converge to a higher degree, estimates based on fewer sentences still remain less accurate compared to expert judgments or manifestos.<sup>13</sup>

Two aspects are worth mentioning. First, newspaper sentiment-based party locations on territorial politics in Spain seem to converge to some extent with positions derived from expert surveys and annotated manifesto content, as long as we have *enough* news reporting on parties and the issue of interest. Second, we should be aware that mediated party positions do not fully reflect – nor should they, according to expectations – positions assigned by experts or self-placements by political parties. Mediated party positions should be a complement to traditional estimates and not a substitute. Although we could have expected less convergence, due to mediatization and media bias, the comparison shows relatively high correlation between sources, probably due to the high level of political dimension aggregation. Hence, the comparison between different arenas – media, electoral, expert-based – cannot fully ascertain whether a newspaper sentiment approach is adequate to obtain (mediated) party positions. In

<sup>13</sup>Figure SM4.19 of section SM4.4 in the Supplementary Material portrays different correlation tests with different sentiment analysis adjustments: normal mean sentiment, bootstrapped sentiment, mean based on the most recent year and last two years, wording direction control, and context inclusion. We see that, overall and for the aggregated territorial issue, context inclusion and direction control do not substantially improve correlation with expert survey and manifesto estimates. Including more information from the most recent year and the last two years increases the level of convergence, since more sentences are accounted for. This last dynamic is in accordance with our findings so far.

**Figure 4.2:** Party positions from newspaper sentiment estimates, expert survey and electoral manifestos



Note: confidence bands at the 0.95 level. Point estimates of sentiment are based on the mean of 200 bootstrapped samples by omitting one sentence each time.

order to compare our approach with word-scaling, we conduct a wordshoal and CCA analysis, showing that our sentiment measure outperforms word-scaling methods in capturing mediated party positions (see Figure SM4.20 of section SM4.5 in the Supplementary Material). We bear in mind, nevertheless, that this is only a convergence analysis, and external validity cannot be fully achieved between self-placements and mediated placements.

#### 4.6.2 Newspaper sources: comparison, biases and (un-)certainty

In this section we compare position estimates between newspaper outlets on the territorial sub-issues we identified: Basque Separatism, Catalan Separatism, Fiscal-Administrative Authority, and Political Autonomy. If our sentiment-based newspaper estimates are a faithful representation of mediated *reality*, we should expect different sources to converge in their assessment of party positions. Here, we can better compare the performance of different adjustments of sentiment text measures and obtain valid results using different newspaper sources. Furthermore, we can try to answer the question of *how much text is necessary for deriving accurate party positions*. The comparison is restricted to the period from 2002, for which we have available data for El País and El Mundo.

We compare different simple transformations of sentiment measurement at the sentence level which are then aggregated (see Table 4.1). For direction control, we flip the sentiment count of positive and negative words in sentences identified by the territorial politics dictionary with (re-)centralization wording, for example "nación única",

”lengua común” or ”españolista”. (Re-)centralization wording applies only to 7.3% of all territorial party sentences, since the Spanish debate is generally framed as ”less decentralization”. For inclusion of context, we add the sentences preceding and succeeding the relevant sentence within a news article.

In order to account for topic content certainty within each of the four territorial sub-issues, we deploy the STM prevalence values determining the sub-issue classification. Remember that STM helps us identify what share of the sentence belongs to our conceptual territorial policy sub-issue of interest. The sum of sub-issue topics’ prevalence ( $k$ ) that predominated within each sentence determines the single category assigned to the text passage. Dealing with four territorial sub-issues, any with a prevalence of only  $> 0.25$  in a sentence can predominate by simple majority. We also include the topic content certainty measure by weighting sentences with prevalence values above 0.5 and aggregate them per year. For media source harmonization, we calculate the difference between the party-sentence-conveyed sentiment and the source’s overall sentiment on the relevant sub-issue.

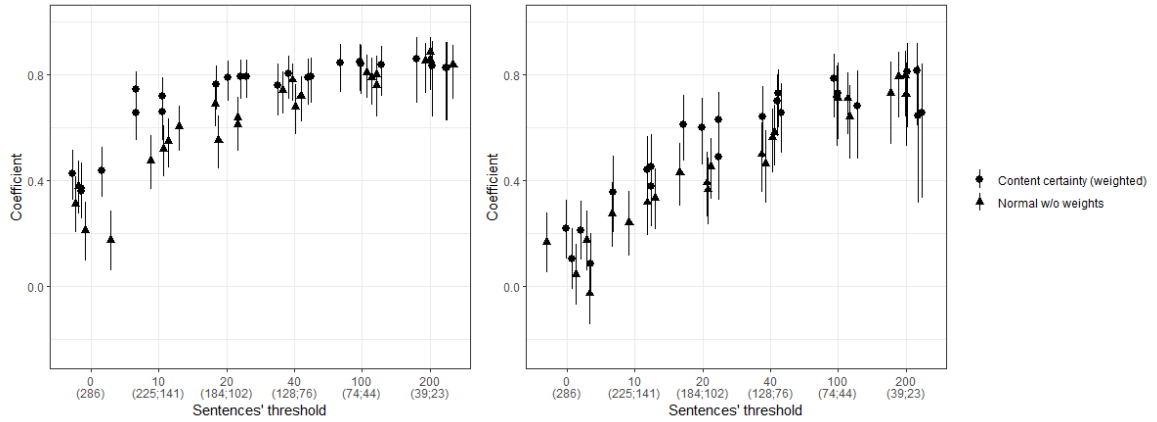
Figure 4.3 shows the correlation of yearly party positions for territorial sub-issues in Spain since 2002 between estimates by El País and El Mundo. The correlations are separated by different thresholds of minimum number of sentences by estimate and newspaper. The left panel depicts how the 286 position estimates ( $\sim 18$  years  $\times$  4 parties  $\times$  4 policy sub-issues) significantly correlate at around 0.40. With a higher threshold of at least 10 sentences by estimation and newspaper, we lose information on some time points, but the convergence of both newspapers increases to 0.60 for a total of 141 party positions. A threshold of at least 20 sentences seems to fit the trade-off best, reaching convergence values of 0.80, since the relationship is asymptotic. Higher thresholds minimally improve the convergence in the aggregation, but many party-position estimates get lost.

Different measurement adaptations – direction control, context inclusion, and both together — produce no systematic improvement in newspaper estimation convergence compared to a straightforward global sentiment analysis. Yet, topic content certainty substantially improves newspaper convergence of estimates across almost every threshold level. Excluding sentences with low issue topic load optimally reduces uncertainty in sentiment analysis. We see that for very few sentences or for many sentences, content certainty does not make a big difference. The best range of sentences (10 to 20) indicates the balance between minimum aggregation and reaching certainty.

Finally, the right panel presents the correlations between newspapers after harmonizing for overall newspaper sentiment on the sub-issue level. Surprisingly, convergence

**Figure 4.3:** Convergence of party positions by newspaper outlet with different measurements and minimum sentence thresholds

Left panel: w/o newspaper adjustment. Right panel: w/ media source harmonization



levels are systematically lower than without newspaper source harmonization. It could be that non-party reporting in newspaper articles does not actually reflect the ideology of that outlet. Additionally, it is not clear how overall newspaper sentiment relates to news about parties. It is also plausible that an outlet fiercely against Catalan Separatism attributes more negative wording to certain parties supportive of the Catalanist cause.

In short, we show that our sentiment approach for measuring party positions on fine-grained policy sub-issues is consistent across newspaper sources. The most important text adaptation for improving internal convergence is topic content certainty. Topic content certainty informs sentiment analysis about the most informative sentences on the issue of interest. We could determine that ideally a minimum of 20 sentences is necessary in order to obtain valid party positions from newspaper data. We believe that this threshold resembles the intuition that a human being with no knowledge at all of a country's party politics would need a similar amount of information to make an informed guess as to a party position.

### 4.6.3 Text-analytical comparison with parliamentary speeches

To compare mediated party positions on territorial sub-issues with an external source, we make use of parliamentary debates in the Congreso de los Diputados. Parliamentary speeches are helpful to infer party positions on overarching issues such as the left-right

dimension (Lauderdale and Herzog 2016; Rheault and Cochrane 2020). Although parliamentary speeches overall tend to reflect the government–opposition divide, due to the parliamentary logic of communication and legislative dynamics (Proksch et al. 2019), we argue that narrowing parliamentary speeches down to sentences on substantial policy (sub-)issues is suitable for extracting information on sub-issue-specific party positions.

We follow the same procedure as with the newspaper data to generate the parliamentary territorial corpus, exploiting the encompassing ParlSpeech dataset (Rauh and Schwalbach 2020) which covers parliamentary debates in the Congreso from 1996 to 2018. Overall, territorial politics receives less attention in the parliamentary arena than in the mass media. While 7.9% of all news articles mention territorial politics in one way or another, only 6.7% of speeches by Spanish MPs pay attention to the territorial issue. We acquire our selection of 37,685 territorial sentences by party MPs using document-feature matrices with the optimized STM based on newspaper sentences.

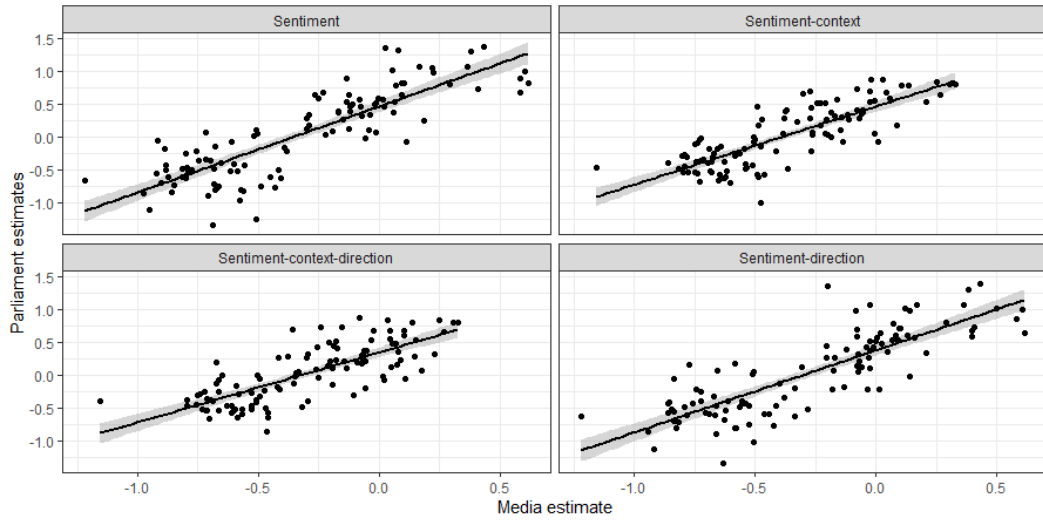
Figure 4.4 presents the correlations between yearly newspaper and parliamentary party positions on the four territorial sub-issues based on sentiment analysis for the PSOE, PP, CiU and PNV. The analysis makes use of the topic content certainty weighting for all measurement comparisons. The correlations confirm the indication from the previous newspaper source comparison, namely that topic content certainty seems to be a good-fit corrector, whereas other adaptations do not systematically differ in their performance. Even the straightforward sentiment estimates without measurement adaptations shows the highest correlation (0.844) compared to the adjusted versions of context inclusion (0.809), direction control (0.827), and both context inclusion and direction control (0.765). All correlations are positive and statistically significant at  $p\text{-value} < 0.000$ .

The correlation patterns with parliamentary speeches reveal a high level of convergence with our newspaper sentiment-based measure of party position. We compare territorial sub-issue positions and are able to test measurement adaptations in different sources from two different political arenas, parliament and mass media. The comparison uncovers a high degree of external validity even for the measurement of fine-grained, often low-salience issue positions.

## 4.7 Qualitative validation

In order to confirm confidence in our measurement approach, we conducted a qualitative, hand-coded validation of party positions taken by the statewide parties PP and

**Figure 4.4:** Correlation between sentiment-based positions in mediated estimates and parliament estimates for territorial sub-issues



Notes: correlation lines with 0.95 confidence bands. Positions estimated with at least 30 sentences input.

PSOE and the regionalist parties PNV (Basque) and CiU (Catalan).<sup>14</sup> We selected two samples of articles with different guided random procedures for each of the sub-issues, covering a total of  $N = 919$  party-territorial sentences (see section SM4.9 in the Supplementary Material for more details; both samples remain comparable). The parties were coded on a 1–5 scale with negative, negative conditional, neutral, positive conditional, and positive attitudes towards more decentralization within each sub-issue. The qualitative validation is intended to resemble the automatized measurement as far as possible. For example, we measure hand-coded positions at the sentence level, where we assume that annotators are not always fully certain about each position assignment. We then obtain the mean position for the period of interest.

The qualitative validation of the two samples uncovers sobering correlations of  $r = 0.37$  and  $r = 0.28$  for the topic content certainty weighted estimate, but neither correlations are statistically significant (at  $p\text{-value} < 0.10$ ). Inferences should be drawn cautiously, since we only consider 30 estimated positions out of the 919 aggregated sentences. Additionally, the hand-coding of the first sample revealed that 20% of the sentences portrayed party stances ambiguous to the human eye; and 41% of the sentences were coded as problematic for a text-as-data approach.<sup>15</sup> The uncertainty

<sup>14</sup>Although internal validation would have been a necessary condition previous to external convergence analysis, we invert the sequence of analysis in order to derive a general estimate of optimal aggregation level of text driven by the overall external convergence.

<sup>15</sup>Hand-coded sample A included a dummy where the coder indicated whether the text segment would generate problems concerning a sentiment measure based on bag-of-words – e.g. long sentences with

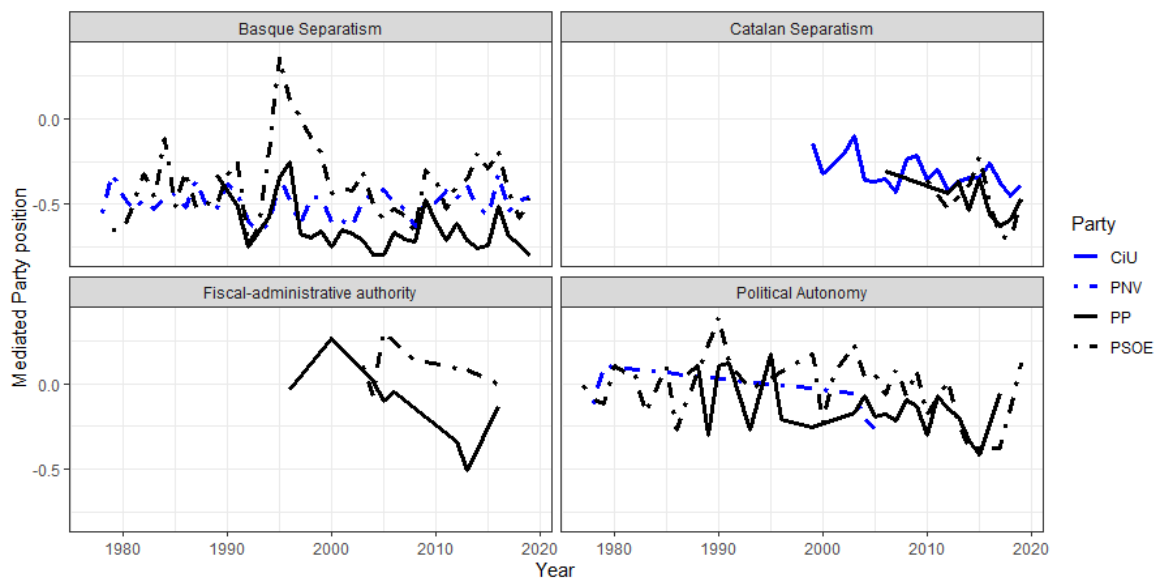


levels revealed by reading party-territorial news reporting emphasize the importance of text aggregation to obtain valid signals by reducing random noise. The validation of more positions might improve our confidence in this assessment.

## 4.8 Position estimates and application

Having validated our estimates of Spanish party positions on territorial sub-issues, we can turn to inspecting the dynamics of positioning and shifting across time. Figure 4.5 illustrates the mediated positions of the PP, PSOE, CiU and PNV on the sub-issues of Basque Separatism, Catalan Separatism, Political Autonomy and Fiscal-Administrative Authority over the decades using the content-weighted measure. Note that positions are not available for every party for every year, in compliance with our conservative threshold of at least 30 sentences by estimation. For example, we have very few positional signals for the PNV on Political Autonomy between 1980 and 2005. Conceptually, this *interrupted* portrayal of party positions is a more truthful representation of media reality: if we do not have enough news reporting on party stances on (sub-)issues, it is difficult or even impossible to derive a party position. In other words, we only observe party positions which we can confidently measure.

**Figure 4.5:** Mediated party positions on territorial sub-issues in Spain



Note: Sentiment-based estimates based on at least 30 sentences input.

---

many different sentiment signals referring to other topics or issues. Topic content certainty would help to correct this.

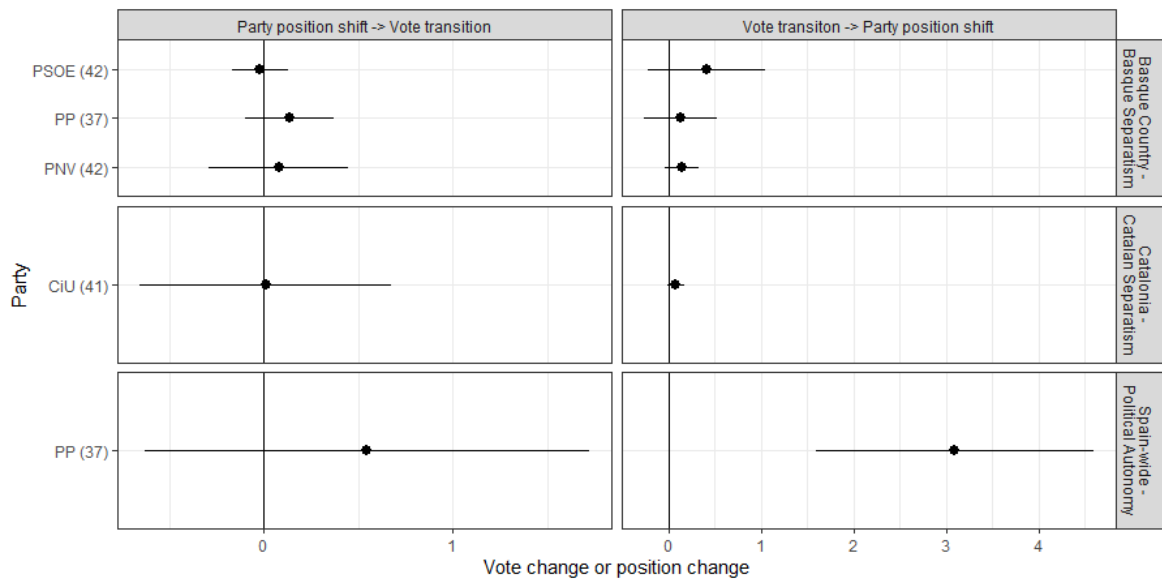
Empirically, the resulting estimates can partly be taken at face value, at least in part. Overall, it is very plausible that Basque Separatism is the sub-issue with most opposing placements, due to the association in Basque sovereignty debates with the terrorist organization ETA. Similarly, PSOE has held a more accommodating position across time than PP. Conversely, party positions on Catalan Separatism are almost indistinguishable. Taking the polarization of this issue into account, especially during the *procés* period since 2012, it is not persuasive that in specific years the PSOE has a more pro-decentralization position than Catalan regionalist CiU. We believe that enhancing news input with more sources and including regional press can improve our estimation.

The sentiment-based estimation of yearly positions on Spanish territorial sub-issues can serve to address key dilemmas in the decentralization literature. We use our estimations to answer the question: does region-specific electoral support positively impact party-positional shifts on decentralization in this region, or does more party support for region-specific decentralization increase regional voter support?

We make use of the RD|SED and RED datasets (Kaiser et al. 2021 $a,b$ ) to obtain average regional electoral support in the statewide and regional arena; this offers an optimal signal of support for political parties when systematic historical polls are not available. For example, we investigate whether an increase in voter support in Catalonia led parties to shift towards a position more accommodating of Catalan Separatism, or if the relationship is inverse. Voter transition is the change in vote share and party-position shift is the change on the territorial sub-issue, both at the year level. We conduct party- and sub-issue-specific Vector Auto-regressive (VAR) models to account for time-series endogeneity and specify both current and one-year-lagged voter transition and party-position shift (respective  $Y$  have two time-lags). We re-scale positions just for illustration; therefore positional shift size can only be interpreted in relative terms (full models in section SM4.10 in the Supplementary Material). Analyses are only available for cases where we have enough information over the decades.

Results in Figure 4.6 show that neither parties nor voters were reactive to each other’s changing dynamics in Catalonia and Basque Country. The non-existent relationship might be driven by the polarization of the sub-issues in both regions, with large regionalist mobilization. In Basque Country especially, we would not expect parties to be more accommodating of Basque Separatism and open to negotiating with Basque terrorist actors due to electoral support in that region. However, the statewide party PP shifted towards supporting the decentralization of authority overall when they gained statewide electoral support. This simple application, based on news-media-

**Figure 4.6:** Voter transition and party-position shifts in Spanish territorial politics – VAR models



Note: observations ( $N$ ) per VAR-estimation in parentheses. Standard deviation based on 0.95 CI. Lagged and control variables not shown.

assessed sentiment, to territorial sub-issues shows the potential for expanding established applications in party politics to fine-grained dimensions of issue-competition.

## 4.9 Towards an agenda of mediated party positions

The study of party positions is enriching for our understanding of democratic representation. As we have argued, self-portrayed positions by parties and representatives do not reach citizens unaffected. Most people absorb party positions through the media. The process can lead to a *disconnect* between representatives and the represented, since the media intervenes in the salience and selection of sub-issues, as well as in the pure portrayal of party positions. Media interference also involves journalists' ideological leanings, prioritization of conflictual and negative news, and more space available in mass media for established parties. Therefore, studying *mediated* party positions can be helpful in bridging the party–voter congruence link. *Mediated* party positions might help us to better understand many of the political communication dynamics that exist in democratic delegation.

However, most research on party positions focuses on quantitative approaches analyzing original texts by politicians or surveys by experts. The literature centered around party positions as conveyed in the media instead uses manual annotation tech-

niques. To fill this gap, we propose an automatized text approach to capture *mediated* party positions on fine-grained (sub-)issues with sentiment analysis. Sentiment is more suitable for analyzing text passages from different news outlets that strongly impose their own wording patterns. Compared to word-scaling, sentiment is more universal and can travel more easily across different wording contexts, and can grasp indirectly communicated party positions originating from distinctive text-generation processes. To tackle the issue of benchmarking for sentiment analysis, we include a measure of topic certainty derived from topic models to buttress our estimation from measuring sentiments towards the sub-issue of interest. An optimal combination of topic content certainty and sentiment signal aggregation allows us to arrive at valid yearly *mediated* party positions.

We apply the methodological procedure to the case of territorial politics in Spain using the news outlets El País and El Mundo (1976–2019). Spanish territorial politics is a suitable case for applying our approach, since territorial sub-issues reveal a diversity of party positions with interchanging levels of salience over time. First, we find middle–high convergence levels on the overall territorial dimension between our *mediated* position estimation and measures by expert surveys and manifestos. To evaluate the text approach on sub-issues, a between-newspapers comparison identifies that topic certainty together with an aggregation of at least 20 sentences is an optimal configuration to arrive at converging party positions in newspapers. Similarly, we find high convergence with parliamentary speeches across territorial sub-issues.

Finally, the investigation describes the position shifts of the PSOE, PP, PNV and CiU across the sub-issues Basque Separatism, Catalan Separatism, Political Autonomy and Fiscal-Administrative Authority. We show that, to a large extent, face-valid mediated party positions can be efficiently extracted from newspaper data with sentiment analysis. We make use of the analysis’ product and implement a straightforward VAR model of regional voter transition and party policy shifts to demonstrate the potential applicability of mediated party positions. We find that neither parties nor voters in Basque Country or Catalonia were reactive to each other concerning Basque and Catalan Separatism, respectively. We only find that PP was overall more pro-Political Autonomy when its electoral support increased statewide in Spain.

The potential of sentiment analysis for mediated party positions is very promising. Certain aspects remain open for future research, which can improve the procedure to acquire mediated party positions. For example, we only focus on one-party-exclusive sentences. Sentences mentioning several parties are excluded; it is not clear whether these can, for example, portray different party positions, e.g. polarized or rather more

consensual ones. Furthermore, our investigation uses a bag-of-words framework that can benefit greatly from word-embedding analysis to refine the estimations. We believe that future research can expand on this procedure to improve the measurement of mediated party positions. The analysis of mediated party positions can generate fruitful insights for representation studies and for party politics scholarship in general.

## SUPPLEMENTARY MATERIAL

## SM2 The Long-run Dynamics of Decentralization: Regional Democracy and Statewide Governments' Strategies, 1950-2018

### SM2.1 Overview of theory

**Table SM2.1:** Summary of theoretical hypotheses and arguments' references

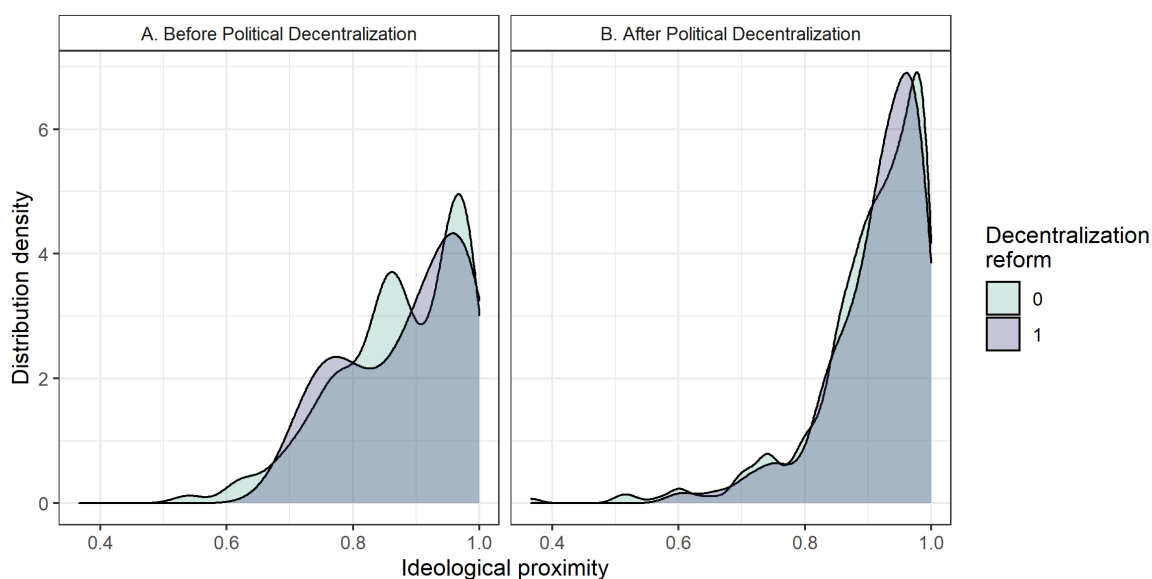
N°	Hypothesis	Direct argument	Indirect argument
<b>H1</b>	After political decentralization, decentralization becomes more likely	Toubeau (2017), among other arguments of regionalization	Falletti's (2005; 2013) sequential and institutional logic; see also Pierson (2000)
<b>H2.1</b>	Electoral, partisan and ideological incentives of statewide parties explain decentralization dynamics	Masseti and Schakel (2013); Röth and Kaiser (2019)	O'Neill (2003, 2005); Meguid (2009, 2015)
<b>H2.2 and H2.3</b>	-After political decentralization, statewide parties' electoral and ideological preferences in statewide elections do not explain subsequent decentralization reforms, compared to before -After political decentralization, statewide parties' electoral and ideological preferences in regional elections explain subsequent decentralization reforms	<i>New theoretical argument derived from the historical institutionalist argument (H1) combined with party based mechanisms of decentralization (H2.1)</i>	-

## SM2.2 Summary and descriptive statistics

**Table SM2.2:** Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Ideological proximity (statewide election)	2817	0.897	0.092	0.366	0.857	0.968	1
Ideological proximity (regional election)	1575	0.886	0.089	0.481	0.833	0.956	1
RAI <sub>t-1</sub>	2791	15.265	8.005	0	10	23	28
Regional democracy	2870	0.725	0.447	0	0	1	1
Electoral democracy - polyarchy (VDEM)	2866	0.801	0.141	0.244	0.796	0.881	0.924
EU	2870	0.495	0.5	0	0	1	1
Years of main cabinet party (log)	2862	1.302	1.039	0	0	2.079	3.761
Reform passed - dummy	2818	0.057	0.233	0	0	0	1
Decentralization reform	2732	0.057	0.232	0	0	0	1
Autonomist parties	2863	4.857	12.858	0	0	0	88.231
Secessionist parties	2863	3.203	8.929	0	0	0	52.608
Regions Nr	2869	6.815	12.364	1	1	9	81
Fiscal reform	2732	0.018	0.131	0	0	0	1
After First Reform	2870	0.55	0.498	0	0	1	1

**Figure SM2.1:** Statewide Ideological Proximity, Political Decentralization and Reforms



## SM2.3 Data Aggregation and Fundamental Decisions

A link to the whole data, the selected aggregated data as well as the whole code for aggregation and modelling is available in case of request or publication.

### SM2.3.1 Regional and electoral aggregation

In order to weight the regions affected by symmetric decentralization and accounting for their electoral vulnerability for the statewide government (Alonso 2012), I create a relative weighted measure for each region based on their relative seat allocation in the national parliament. If this information is missing, the relative registered voters are taken. In the case of Mexican elections in the 1960s and 1970s with both information values missing, I assign all regions the same relative weight. The same occurs for the Danish regional elections, the results of which are collected at a lower sub-national tier than the statewide elections; here, I assign all regions the same relative weight. In the case of the Italian region Trentino-Alto Adige, regional elections took place separately in Trentino and Adige. For comparison with statewide elections, their respective electoral results were merged with 50 % weights.

Ideological *Center of Gravity* measures are largely calculated based on seat distributions for Turkey and Switzerland, since vote share information in the electoral results is not available. This can lead to more extreme values implying very low or very high shares of support for either the governmental or opposition parties.

### SM2.3.2 Electoral results structure and missing elections

For the period of study, two sets of regional elections are missing, for Nunavut (Canada) in 1997 and for the Faroe Islands before 1970. In other cases, not all of the most recent regional elections have been collected, especially in countries with non-coordinated regional elections, e.g. Spain and Germany. In these cases, year observations at the region level are only considered with at least 80 % of the regional coverage measured by their relative importance weights, which only applies for few cases after 2011. The regional elections coverage ends in Canada in 2011, in Italy in 2016 and in Austria in 2018. In addition, certain elections present idiosyncratic issues: at statewide elections in Greenland before 1979 only independent candidates ran for office; for these cases we lack ideological information and ergo *Centres of Gravity*. For the 2003 Serbia election, electoral disaggregated results were not available, and ideological measurements are missing. The same is the case for the exceptional 1994 statewide election in Italy (information missing for the 1995 observation).



Most of the Mexican ideological estimations are party family imputations (see datasets codebooks), since CMP data for older Mexican elections is missing (Volkens et al. 2020). Note, also, that Mexican election results before the 2000 only consider the election of plurality MPs, which presents a more contrasted electoral picture with more intense government strongholds that would the proportional system results depict. This also affects the ideological representation of the regions. Further specific design and coding decisions can be found in the coding scripts in the online supplementary material.

## SM2.4 Description of cases and reforms

### SM2.4.1 Cases of region/region groups by decentralization reform type

**Symmetric decentralization cases (N = 20):** Australia States, Austria Länder, Belgium Provinces/ Provinces, Canada Provinces, Denmark Regioner/ Amtskommuner, France Rég./ Départements, Germany Länder, Italy Province/Regioni ord., Mexico Estados, Nicaragua Departamentos, Norway Fylker, Portugal Distritos, Serbia Okruzi-districts/ Republika Srbija, Spain CC. AA. I, Spain CC. AA. II, Spain Provincias, Sweden Län, Switzerland Cantons, Turkey Iller/ Kalkinma ajanslari, the United Kingdom (English) Counties.

**Asymmetric decentralization cases (N = 39):** Australia Australian Capital Territory, Australia Northern Territory, Belgium Brussels/ Rég. Capitale, Belgium Comm. française/ Rég. wallone, Belgium Vlaamse Gemeenschap, Bosnia and Herzegovina Federacija, Bosnia and Herzegovina Republika Srpska, Canada Northwest Territories, Canada Nunavut, Canada Quebec, Canada Yukon, Spain Cantabria, Denmark Faroe Islands, Denmark Greenland, France Corse, Italy Friuli-Venezia Giulia, Italy Sardegna, Italy Sicilia, Italy Trentino-Alto Adige/ Südtirol, Italy Valle d'Aosta, Mexico California Baja Sur, Mexico Distrito Federal, Mexico Quintana Roo, Nicaragua Managua-Distrito Nacional, Nicaragua Región Autónoma del Norte, Nicaragua Región Autónoma del Sur, Portugal Açores, Portugal Madeira, Serbia Vojvodina, Spain Andalucía, Spain Catalunya/ Cataluña, Spain Ceuta, Spain Comunidad Foral de Navarra/ Nafarroa, Spain Euskadi/ Pais Vasco, Spain Galiza/ Galicia, Spain Melilla, the United Kingdom Northern Ireland, the United Kingdom Scotland, the United Kingdom Wales.

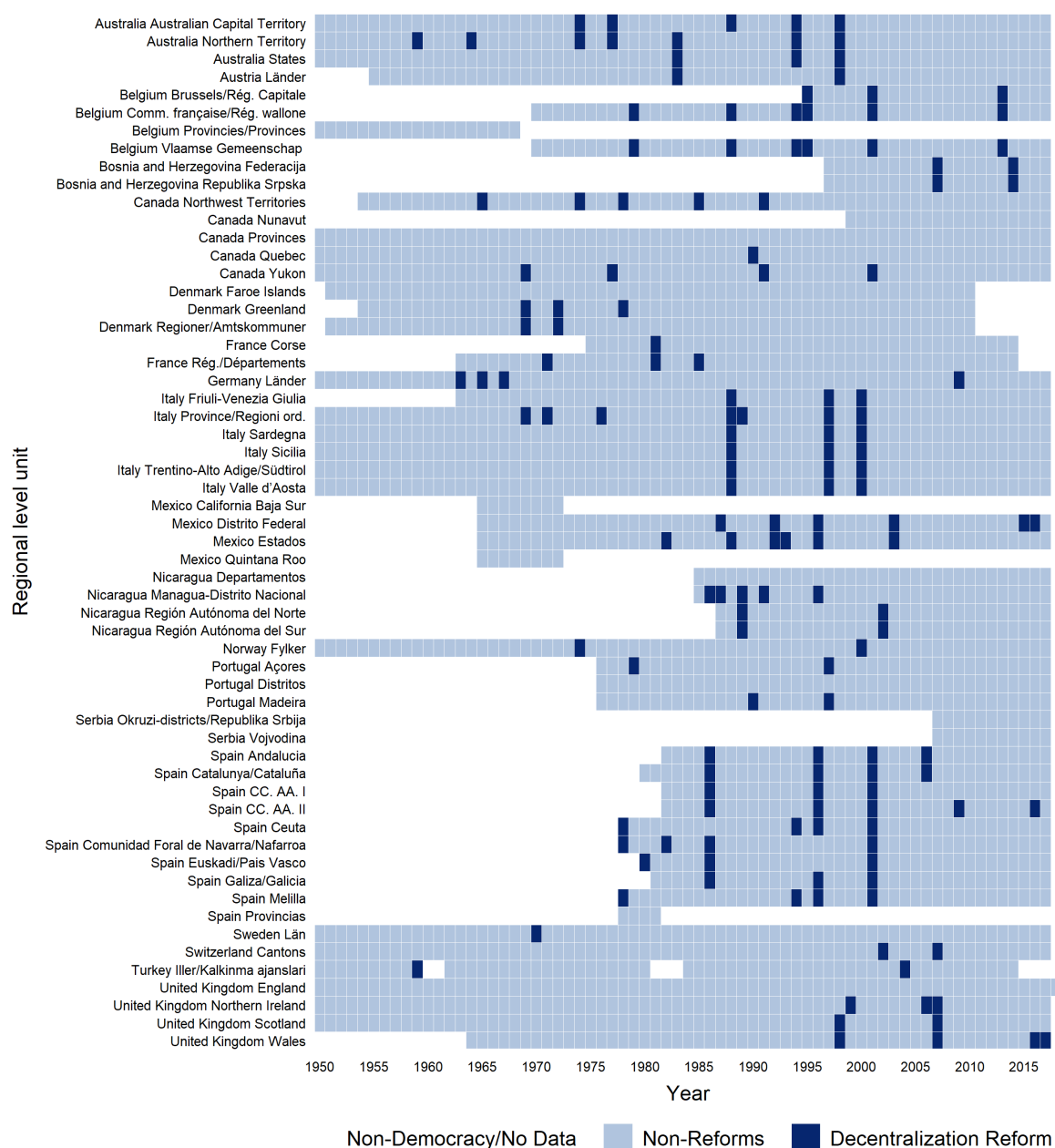
#### **SM2.4.2 Political decentralization: year of emergence of regional democracy**

**No regional democracy during the coverage period (N = 10):** Nicaragua Departamentos, Portugal Distritos, Spain Provincias, Turkey Iller/ Kalkınma ajansları, the United Kingdom (English) Counties, Mexico Quintana Roo, Mexico California Baja Sur, Spain Cantabria (one year), Belgium Provincies/Provinces.

**Regional democracy during the whole coverage period (N = 13):** Australia States, Austria Länder (coverage starts 1955 - before occupied by allies), Canada Provinces & Canada Quebec, Germany Länder, Italy Sardegna, Italy Sicilia, Italy Trentino-Alto Adige/ Südtirol, Italy Valle d'Aosta, Serbia Vojvodina, Sweden Län, Switzerland Cantons, Portugal Açores and Portugal Madeira (start 1976).

**Regional democracy emerges in between the coverage period (N = 36):** Canada Yukon (1978), Canada Northwest Territories (1979), Canada Nunavut (1999), Mexico Distrito Federal (1988), Mexico Estados (1989), Nicaragua Región Autónoma del Norte, Nicaragua Región Autónoma del Norte & Nicaragua Managua-Distrito Nacional (1990), Australia Northern Territory (1974), Australia Australian Capital Territory (1989), Spain Comunidad Foral de Navarra/Nafarroa (1979), Spain Andalucía (1982), Spain Cantabria, Spain CC. AA. I & Spain CC. AA. II (1983), Spain Catalunya/Cataluña & Spain Euskadi/Pais Vasco (1980), Spain Galiza/Galicia (1981), Spain Melilla & Spain Ceuta (1995), Italy Friuli-Venezia Giulia (1963), Italy Province/ Regioni ord. (1970), France Corse (1982), France Rég./Départements (1986), Belgium Brussels/Rég. Capitale (1989), Belgium Comm. française/Rég. wallonne & Belgium Vlaamse Gemeenschap (1995), Denmark Regioner/Amtskommuner & Denmark Greenland (1970), Serbia Okruzi-districts/ Republika Srbija (exception: before 2006), Norway Fylker (1974), the United Kingdom Northern Ireland (exception: until 1971, 2000-2003, 2007-2018), the United Kingdom Scotland and the United Kingdom Wales (1999), Bosnia and Herzegovina Federacija & Bosnia and Herzegovina Republika Srpska (exception: before 1998 and after 2015).

**Figure SM2.2:** Decentralization Reforms across Time and Cases



Country selection including symmetric and asymmetric decentralization region(s) units.  
Source: RAI (Hooghe et al. 2016).

## SM2.5 Full models and interaction figures

**Table SM2.3:** Ideological Incentives and Decentralization Degree *Before* and *After* Political Decentralization

	DV: Decentralization Reform (0 — 1)			
	Asymmetric Reforms		Symmetric Reforms	
	(1)	(2)	(3)	(4)
Ideological proximity	1.489 (1.127) [0.187]	2.847 (2.185) [0.193]	−0.360 (2.685) [0.894]	−5.478 (3.676) [0.137]
Regional democracy		2.773 (2.366) [0.242]		−6.197 (4.171) [0.138]
RAI <sub>t-1</sub>	−0.030 (0.016) [0.066]	0.031 (0.027) [0.250]	0.037 (0.023) [0.109]	0.097 (0.070) [0.166]
Electoral democracy - polyarchy (VDEM)	−0.151 (0.891) [0.866]	0.439 (0.934) [0.639]	−1.411 (1.367) [0.303]	−1.523 (1.578) [0.335]
EU	0.259 (0.252) [0.305]	0.332 (0.260) [0.203]	0.926 (0.395) [0.019]	0.953 (0.404) [0.019]
Years of main cabinet party (log)	−0.128 (0.098) [0.191]	−0.126 (0.097) [0.198]	0.344 (0.163) [0.035]	0.294 (0.162) [0.071]
Reform passed - dummy	−1.127 (0.604) [0.063]	−1.145 (0.605) [0.059]	0.038 (0.764) [0.961]	−0.014 (0.769) [0.986]
Regional democracy x Ideological proximity		−1.963 (2.624) [0.455]		8.447 (4.837) [0.081]
Regional democracy x RAI <sub>t-1</sub>		−0.104 (0.038) [0.007]		−0.095 (0.076) [0.211]
Regions Nr			0.014 (0.013) [0.313]	0.020 (0.015) [0.191]
Constant	−3.347 (1.201) [0.006]	−5.354 (2.195) [0.015]	−3.463 (2.608) [0.185]	0.473 (3.041) [0.877]
Random effects	✓	✓	✓	✓
N	1,673	1,673	1,002	1,002
Regional level units	39	39	20	20
Countries	13	13	18	18
N	1,673	1,673	1,003	1,003
Log Likelihood	−402.266	−398.117	−167.089	−164.855
AIC	820.533	818.233	352.178	353.710
BIC	863.912	877.879	396.375	412.639

Standard deviation in brackets; p-values in square brackets.

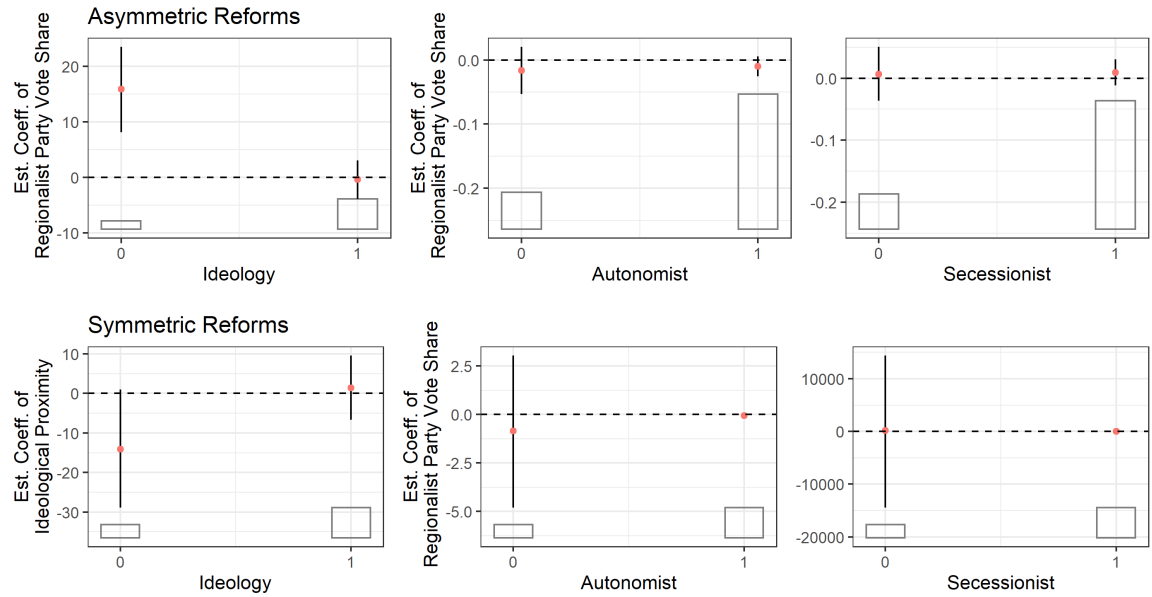
**Table SM2.4:** Influence of Regionalist Parties Across Time

	DV: Decentralization Reform (0 — 1)			
	Asymmetric Reforms		Symmetric Reforms	
	(5)	(6)	(7)	(8)
Ideological proximity	4.948 (1.873) [0.009]	15.921 (4.632) [0.001]	−2.094 (4.088) [0.609]	−14.123 (9.090) [0.121]
Autonomist parties	−0.008 (0.008) [0.294]	−0.016 (0.023) [0.487]	−0.048 (0.075) [0.519]	−0.791 (2.350) [0.737]
Secessionist parties	0.003 (0.010) [0.770]	0.006 (0.026) [0.806]	0.256 (0.108) [0.018]	−17.265 (264.397) [0.948]
Regional democracy	−0.642 (0.485) [0.186]	14.281 (4.738) [0.003]	−0.240 (0.815) [0.769]	−13.586 (8.259) [0.100]
RAI <sub>t-1</sub>	0.008 (0.030) [0.781]	−0.001 (0.034) [0.988]	0.052 (0.053) [0.327]	0.053 (0.054) [0.325]
Regions Nr			0.045 (0.031) [0.142]	0.035 (0.031) [0.267]
Electoral democracy - polyarchy (VDEM)	0.612 (3.030) [0.840]	−0.477 (2.945) [0.872]	3.406 (4.749) [0.474]	1.965 (4.727) [0.678]
EU	0.926 (0.331) [0.006]	1.242 (0.381) [0.002]	1.471 (0.758) [0.053]	1.623 (0.787) [0.040]
Years of main cabinet party (log)	−0.178 (0.119) [0.136]	−0.124 (0.123) [0.315]	0.339 (0.236) [0.153]	0.314 (0.237) [0.187]
Reform passed - dummy	−24.584 (76,983.820) [1.000]	−17.400 (1,977.498) [0.993]	−0.728 (1.106) [0.511]	−0.805 (1.107) [0.467]
Regional democracy x Ideological proximity		−16.305 (5.108) [0.002]		15.668 (9.938) [0.115]
Regional democracy x Autonomist parties		0.006 (0.025) [0.796]		0.740 (2.346) [0.753]
Regional democracy x Secessionist parties		0.003 (0.030) [0.931]		17.501 (264.397) [0.948]
Constant	−7.622 (3.177) [0.017]	−16.900 (5.050) [0.001]	−6.625 (5.852) [0.258]	4.661 (8.909) [0.601]
Random effects	✓	✓	✓	✓
N	1,090	1,090	489	489
Regional level units	25	25	11	11
Countries	7	7	9	9
Log Likelihood	−247.498	−240.596	−90.037	−88.163
AIC	516.997	509.191	204.075	206.325
BIC	571.930	579.106	254.383	269.211

Standard deviation in brackets; p-values in square brackets.

**Figure SM2.3:** Effect of Autonomist and Secessionist Parties' Electoral Support Before and After Political Decentralization

Models 6 and 8.



Note: Error bars reflect .90 level of confidence intervals.

**Table SM2.5:** Multi-Level Incentives *After* Political Decentralization

	DV: Decentralization Reform (0 — 1)	
	Asymmetric Reforms	Symmetric Reforms
	(9)	(10)
Ideological proximity statewide election	1.945 (2.460) [0.430]	4.066 (4.885) [0.406]
Ideological proximity regional election	−0.084 (1.685) [0.961]	−0.393 (3.344) [0.907]
RAI <sub>t-1</sub>	−0.131 (0.038) [0.001]	0.0004 (0.041) [0.992]
Regions Nr		−0.029 (0.037) [0.426]
Electoral democracy - polyarchy (VDEM)	17.433 (4.857) [0.0004]	−5.037 (5.684) [0.376]
EU	0.148 (0.358) [0.679]	0.834 (0.537) [0.121]
Years of main cabinet party (log)	−0.022 (0.141) [0.878]	0.162 (0.202) [0.423]
Reform passed - dummy	−21.147 (228.978) [0.927]	−0.477 (1.076) [0.658]
Constant	−16.908 (4.661) [0.0003]	−2.387 (6.316) [0.706]
Random effects	✓	✓
N	950	578
Regional level units	25	12
Countries	7	12
Log Likelihood	−202.161	−107.416
AIC	422.322	234.832
BIC	466.030	278.428

Note: Standard deviation in brackets; p-values in square brackets. Country sample with political decentralization. Sample excludes regional elections in Mexico and the asymmetric regions of Portugal, Nicaragua, Serbia and Bosnia and Herzegovina.

## SM2.6 Fiscal Decentralization Reforms

**Table SM2.6:** Main models adapted to only predict fiscal decentralization

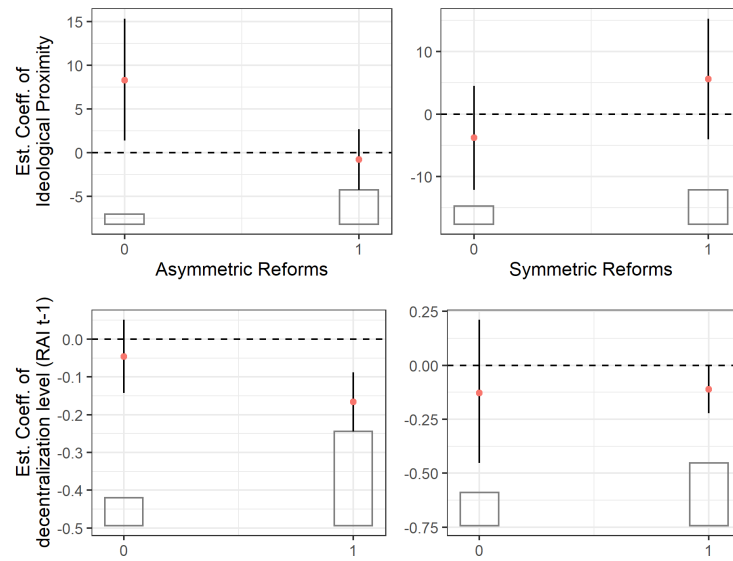
	DV: Fiscal decentralization reforms (0 — 1)					
	(11)	(12)	(13)	(14)	(15)	(16)
Ideological proximity	8.276	−3.671	16.227	−345.106	4.571	8.170
statewide election	8.276	−3.671	16.227	−345.106	4.571	8.170
	(4.201)	(5.019)	(6.638)	(448.717)	(4.805)	(9.685)
	[0.049]	[0.465]	[0.015]	[0.442]	[0.342]	[0.399]
Ideological proximity					−3.930	3.088
regional election					(3.349)	(7.065)
					[0.241]	[0.663]
RAI <sub>t-1</sub>	−0.046	−0.131	−0.135	−0.170	−0.254	−0.192
	(0.059)	(0.203)	(0.062)	(0.118)	(0.088)	(0.087)
	[0.440]	[0.521]	[0.029]	[0.150]	[0.005]	[0.029]
Regional democracy	10.013	−5.635	12.107	−261.381		
	(4.422)	(6.490)	(7.184)	(333.673)		
	[0.024]	[0.386]	[0.092]	[0.434]		
Autonomist parties			−0.006	4.222		
			(0.024)	(5.810)		
			[0.785]	[0.468]		
Secessionist parties			0.002	1.019		
			(0.029)	(2,696.358)		
			[0.932]	[1.000]		
Ideological proximity x	−9.046	9.266	−12.912	350.557		
Regional democracy	(4.851)	(7.454)	(7.770)	(448.467)		
	[0.063]	[0.214]	[0.097]	[0.435]		
Regional democracy x	−0.120	0.018				
RAI <sub>t-1</sub>	(0.076)	(0.216)				
	[0.114]	[0.935]				
Regional democracy x			−0.006	−4.184		
Autonomist parties			(0.031)	(5.800)		
			[0.836]	[0.471]		
Regional democracy x			−0.014	−0.889		
Secessionist parties			(0.039)	(2,696.358)		
			[0.711]	[1.000]		
<i>Controls</i>						
Electoral democracy - polyarchy (VDEM)	1.918	−2.465	20.745	0.592	44.448	−2.873
	(1.883)	(2.852)	(9.509)	(16.679)	(18.058)	(13.320)
	[0.309]	[0.388]	[0.030]	[0.972]	[0.014]	[0.830]
EU	0.911	0.149	1.922	0.666	0.163	−0.176
	(0.435)	(0.704)	(0.760)	(1.545)	(0.797)	(0.919)
	[0.037]	[0.833]	[0.012]	[0.667]	[0.838]	[0.848]
Years of main cabinet party (log)	−0.499	−0.603	−0.571	−0.830	−2.460	−1.257
	(0.185)	(0.312)	(0.250)	(0.583)	(0.857)	(0.553)
	[0.007]	[0.054]	[0.023]	[0.155]	[0.005]	[0.023]
Reform passed - dummy	−23.678	1.246	−25.888	1.402	−24.361	1.636
	(95,245.790)	(1.097)	(724.083)	(1.312)	(207,072.200)	(1.201)
	[1.000]	[0.256]	[0.972]	[0.286]	[1.000]	[0.174]
Regions Nr		0.011		−0.040		−0.146
		(0.026)		(0.081)		(0.074)
		[0.674]		[0.617]		[0.049]
Constant	−11.993	0.705	−35.260	255.123	−37.941	−5.929
	(4.312)	(3.942)	(10.270)	(342.758)	(15.836)	(14.627)
	[0.006]	[0.859]	[0.001]	[0.457]	[0.017]	[0.686]
Random effects	✓	✓	✓	✓	✓	✓
N	1,673	1,003	1,090	489	950	578
Regional level units	39	20	25	11	25	12
Countries	13	18	7	9	7	12
Log Likelihood	−149.608	−59.582	−87.384	−24.138	−47.248	−30.843
AIC	321.215	143.163	202.767	78.276	112.495	81.687
BIC	380.861	202.092	272.682	141.161	156.203	125.282

Standard deviation in brackets; p-values in square brackets. Models labels refer to the tables and main models of decentralization reforms in the paper.



**Figure SM2.4:** Effect of Ideological Incentives, Decentralization Degree and Political Decentralization on Fiscal Authority Transfers

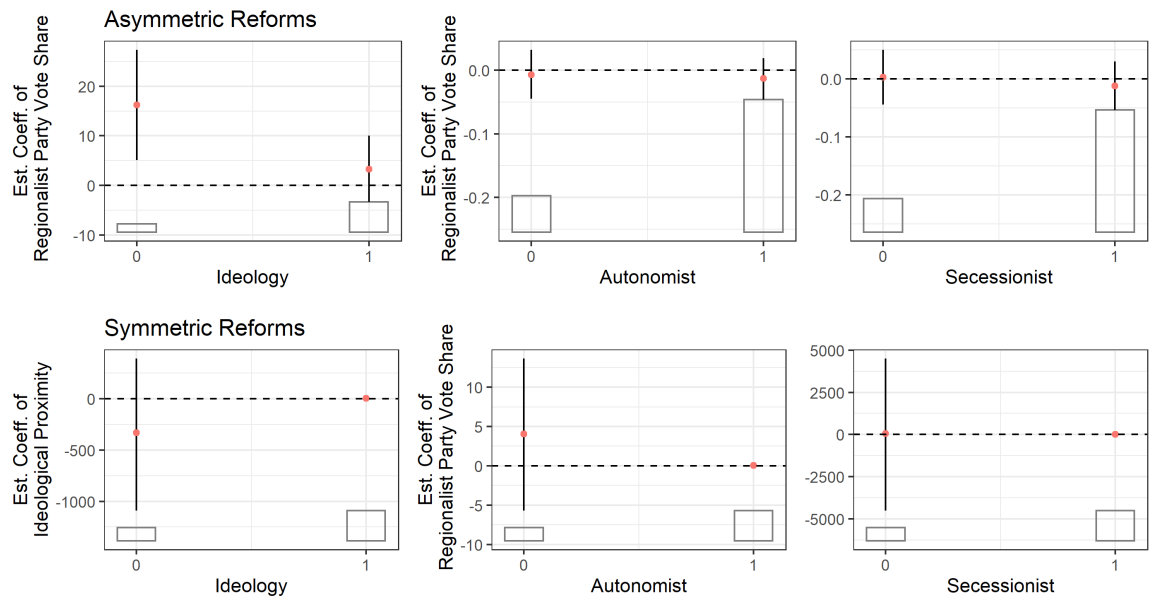
Models 11 and 12.



Note: Error bars reflect .90 level of confidence intervals.

**Figure SM2.5:** Effect of Autonomist and Secessionist Parties' Electoral Support Before and After Political Decentralization on Fiscal Authority Transfers

Models 13 and 14.



Note: Error bars reflect .90 level of confidence intervals.

## SM2.7 Test of First Reform as Hypothetical Critical Juncture

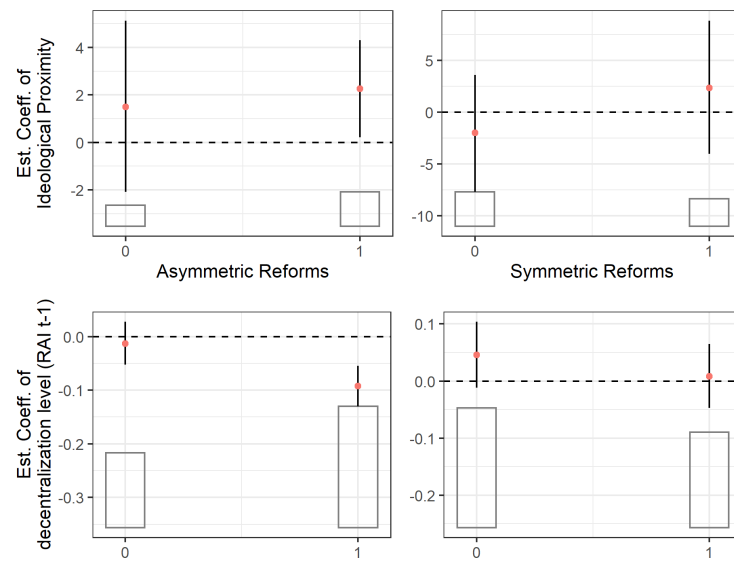
**Table SM2.7:** Ideological Proximity, the First Reform and Posterior Decentralization

	DV: Decentralization Reform (0 — 1)	
	Asymmetric Reforms	Symmetric Reforms
	(15)	(16)
Ideological proximity	1.492 (2.183) [0.495]	−1.925 (3.457) [0.578]
After First Reform	1.302 (2.314) [0.574]	−2.797 (4.371) [0.523]
RAI <sub>t-1</sub>	−0.013 (0.025) [0.607]	0.045 (0.035) [0.201]
Electoral democracy - polyarchy (VDEM)	−0.943 (0.865) [0.277]	−1.564 (1.443) [0.279]
EU	0.339 (0.243) [0.163]	0.746 (0.473) [0.115]
Years of main cabinet party (log)	−0.151 (0.099) [0.127]	0.295 (0.163) [0.071]
Reform passed - dummy	−1.351 (0.602) [0.025]	−0.077 (0.773) [0.921]
Regions Nr		0.002 (0.016) [0.874]
Ideological proximity x After First Reform	0.772 (2.605) [0.768]	4.264 (4.985) [0.393]
After First Reform RAI <sub>t-1</sub>	−0.079 (0.033) [0.017]	−0.037 (0.047) [0.424]
Constant	−3.378 (2.153) [0.117]	−1.944 (3.123) [0.534]
Random effects	✓	✓
N	1,673	1,003
Regional level units	39	20
Countries	13	18
Log Likelihood	−392.152	−165.726
AIC	806.305	355.452
BIC	865.951	414.381

Standard deviation in brackets; p-values in square brackets.

**Figure SM2.6:** Effect of Ideological Incentives and Decentralization Degree Before and After the First Reform

Models 15 and 16.



Note: Error bars reflect .90 level of confidence intervals.

## SM3 The Measurement of Issue Attention across Political Arenas: Exploiting Increasing Returns with Optimized Dictionaries

### SM3.1 Data description

**Table SM3.8:** Territorial articles distribution

<b>Spain</b>	<b>El País</b>	<b>El Mundo</b>	<b>Congreso de los Diputados</b>
Source	Website archive scrape & Lexis- Nexis	LexisNexis	ParlSpeech V2 (Rauh and Schwalbach 2020)
Coverage period	1976-2019	2002-2019	1996-2018
Documents: articles or speeches	1,972,504	735,792	262,276
- territorial	145,778 (7.4%)	70,477 (9.6%)	15,695 (5.98%)
Territorial sentences (total)	377,618	195,213	37,434

<b>United Kingdom</b>	<b>The Times</b>	<b>The Guardian</b>	<b>House of Commons</b>
Source	Times Archive	Digital LexisNexis	ParlSpeech V2 (Rauh and Schwalbach 2020)
Coverage period	1900-2013	1985-2020	1988-2019
Documents: articles or speeches	7,293,823	2,571,009	1,956,223
- territorial	56,982 (0.8%)	34,421 (1.34%)	23,629 (1.21%)
Territorial sentences (total)	102,911	79,877	73,716

## SM3.2 Optimized dictionaries

All dictionaries were applied ignoring lower or upper cases.

### Spain

**List 1:** First expert knowledge-based territorial politics dictionary for Spain – version 1

Preautonomía, preautonom\*, preautonómico, preautómica, vía rápida, pactos autonómicos, autonomía regional, regionalism\* espa\*, LOAPA, loapa, loap\*, vi\* rapid\*, vía rápida”, independentismo, regional, plurinacional, plurilingüistic\*, uniformación, nación única, pluricult\*, pluricultural, derechos históricos, derech\* historic\*, espa\* federal, federalizar espa\*, plurinac\*, descentra\*, regionalist\*, reforma regional, competencias regionales, poder regional, negociaciones regionales, negociaciones con las comunidades autónomas, reforma de las comunidades, competencias de las comunidades, reforma de competencias fiscales, reforma de competencias regionales, reforma de competencias locales.

**List 2:** First optimized territorial politics dictionary for Spain – version 2

Note: categorization in bold only for orientation

**Recentralization keywords:** loap\*, nación única, lengua oficial del estado, imposición lingüística, castellano como lengua vehicular, desafíos rupturistas, unidad de españa, gal, lengu\* vehicul\*.

**Decentralization keywords:** preautonom\*, pactos autonómicos, autonomía regional, regionalism\* espa\*, independentismo, plurinacional, plurilingüistic\*, pluricult\*, pluricultural, derechos históricos, espa\* federal, federalizar espa\*, plurinac\*, descentra\*, regionalist\*, reforma regional, competencias regionales, poder regional, negociaciones con las comunidades autónomas, reforma de las comunidades, competencias de las comunidades, protagonismo de las comunidades autónomas, estado de las autonomías, estado autonómico, equiparación competencial, distribución de competencias, traspaso de competencias, transferencia de competencias, organización territorial, traspasos a las comunidades autónomas, historia autonómica, modelo autonómico, solidaridad interterritorial, identidades de nuestras nacionalidades y regiones, estructura territorial, pactos locales autonómicos, marco estatuario, pacto de ajuria enea, españa de las autonomías, modelo de financiación autonómico, acuerdos autonómicos de 1992, concierto económico, pluralidad de españa, sistema autonómico, descentralización política, descentralización fiscal, gestión descentralizada, administraciones territoriales, conferencia general de cooperación autonómica, autonomía de las nacionalidades, autonomía de las regiones, lengua cooficial, lengua común, pluralidad lingüística, pluralidad cultural, derecho a elegir el idioma, bilingüismo equilibrado, españa autonómica, cohesión territorial, marco competencial, marco autonómico, soberanía regional, desafíos territoriales, diálogo autonómico, derechos forales, estatuto de sau, declaración de barcelona, nación sin estado, espíritu de ermua, catalanismo, catalanista, estatutos de autonomía, estatuto de autonomía,

financiación autonómica, autogobierno, independentistas, independentista, eta, secesionist\*, reconocer a las autonomías, referéndum de autonomía, demandas autonómicas, demandas regionales, rupturista\*, exigencias autonóm\*, reconocimiento autonómico, solidaridad autonómica, diversidad ling\*, fomento autonóm\*, presupuest\* de las autonom\*, etarra, terra lliure, reivindicación territorial, reivindicación autonómica, soberanía autonómica, nacionalidades, devolución, financiar a las autonomías, autodeterminacionist\*, normalización ling\*, estatuari\*, abertzal\*, lizarr\*, andalucist\*, territorialidad, antiespañol\*, autonomismo, autonomist\*, aragonesista, antiautonomista, proetarra, federalización, vasquist\*, catalanidad, cosoberanía, transferencia a las autonomías, soberanista, soberanism\*, descentralizador, alta inspección.

**List 3:** Last optimized territorial politics dictionary for Spain – version 3

**Note:** categorization in bold only for orientation

**Decentralization keywords:** secesionist\*, reconocer a las autonomías, referéndum de autonomía, demanda\* autonómica\*, demanda\* regional\*, exigencias autonóm\*, reconocimiento autonómico, solidaridad autonómica, diversidad ling\*, fomento\* autonóm\*, presupuest\* de las autonom\*, terra lliure, reivindicaciones territoriales, reivindicación territorial, reivindicación\* autonómica\*, soberanía autonómica, financiar a las autonomías, autodeterminacionist\*, autogobierno, estatutari\*, abertzal\*, nlizarr\*, territorialidad, antiespañol\*, autonomismo, autonomist\*, aragonesista\*, proetarra\*, federalización, vasquist\*, catalanidad, cosoberanía, transferencia\* a las autonomías, soberanista\*, soberanism\*, preautonóm\*, preautonom\*, pacto\* autonómico\*, autonomía regional, regionalism\* espa\*, independentismo\*, plurilingüistic\*, pluricult\*, pluricultural\*, derecho histórico, derechos históricos, espa\* federal, federalizar espa\*, plurinac\*, descentra\*, nregionalist\*, reforma\* regional\*, competencia\* regional\*, poder\* regional\*, negociaciones con las comunidades autónomas, reforma\* de las comunidades, competencia\* de la\* comunidad\*, protagonismo de las comunidades autónomas, estado de las autonomías, estado autonómico, equiparación competencial, distribución de competencias, traspaso de competencias, transferencia de competencias, organización territorial, traspasos a las comunidades autónomas, historia autonómica, nmodelo\* autonómico\*, solidaridad interterritorial, identidades de nuestras nacionalidades y regiones, estructura territorial, pacto\* local\* autonómic\*, marco\* estatutario\*, pacto de ajuria, españa de las autonomías, modelo de financiación autonómico, acuerdos autonómicos de 1992, concierto económico, pluralidad de españa, sistema\* autonómico\*, gestión descentralizada, administraciones territoriales, conferencia general de cooperación autonómica, autonomía de las nacionalidades, nautonomía de las regiones, lengua cooficial, pluralidad lingüística, pluralidad cultural, derecho a elegir el idioma, bilingüismo equilibrado, españa autonómica, cohesión territorial, marco competencial, cohesión autonómica, marco\* autonómico\*, soberanía\* regional\*, desafío\* territorial\*, diálogo\* autonómico\*, derecho\* foral\*, estatuto de sau, declaración de barcelona, nación sin estado, espíritu de ermua, catalanismo\*, catalanista\*, estatutos de autonomía, nestatuto de

autonomía, financiación autonómica, independentistas, independentista, plurilingüismo, estatut\* autonómic\*, estatutos autonómicos, reforma\* estatutaria\*, reforma\* de los estatutos, reforma\* del estatuto, reformar el estatuto, estatut d'autonomia, plan ibarretxe, nacionalismo catalán, nacionalismo vasco, competencias territoriales, procesos autonómicos, acceso a la autonomía, estatuto de cataluña, estatuto catalán, estatuto del país vasco, estatuto vasco, nestatuto valenciano, estatuto de valencia, estatuto de galicia, estatuto gallego, estatuto de andalucía, estatuto andaluz, estatuto de madrid, estatuto madrileño, estatuto de murcia, estatuto murciano, reintegración y mejoramiento del régimen foral de navarra, estatuto de extremadura, estatuto extremeño, estatuto de la rioja, estatuto riojano, estatuto asturiano, estatuto de asturias, estatuto de aragón, estatuto aragonés, estatuto canario, estatuto de islas canarias, nestatuto de las islas canarias, estatuto de canarias, estauto de cantabria, estatuto cántabro, estauto de castill\*, estatuto de castilla y león, estatuto balear, estatuto de las islas baleares, estatuto de islas baleares, estatuto de baleares, estatuto de ceuta, estatuto de melilla, competencia\* autonómica\*, mejoramiento del fuero, conferencia de presidentes de las comunidades autónomas, autonomía catalana, autonomía vasca, autonomía andaluza, autonomía aragonesa, nautonomía riojana, autonomía valenciana, autonomía gallega, autonomía murciana, autonomía balear, autonomía madrileña, autonomía extremeña, autonomía asturiana, autonomía canaria, autonomía cántabra, autonomía castellana, procés, independencia de catalunya, independencia de cataluña, independencia catalana, antiespañolism\*, estatuto de catalunya, estatuto de euskadi.

**(Re-)Centralization keywords:** nación única, lengua oficial del estado, imposición lingüística, castellano como lengua vehicular, alta inspección, lengua común, unidad de españa, lengu\* vehicul\*, normalización ling\*, antiautonomista, loapa, desafíos rupturistas, anticatalanist\*, antiregionalist\*, antiindependentist\*, españolism\*, andalucism\*.

**Territorial terrorism keywords:** eta, gal, etarra.

## United Kingdom

**List 4:** First expert knowledge-based territorial politics dictionary for the United Kingdom – version 1

confederalism, devolution, federalism, secession, self-government, separation, unionism, unionist, anti-treaty, anti-Treatyite, IRA, Orangeman, Orangemen, pro-treaty, Treatyite, Unionists, Anglo-Irish Treaty, Bloody Friday, devolved powers, Easter Rising, federal system, home rule, independent Ireland, Ireland's independence, Irish Bill, Irish independence, Irish nationalism, Irish problem, Irish question, Irish troubles, Irish Troubles, Kilbrandon Report, powers delegated, republican Ireland, Scotland Act, Scotland Bill, Scottish Assembly, Scottish independence, tax-varying powers, Ulster Covenant, Welsh disestablishment, Welsh independence, Crowther Commission, Kilbrandon Commission, Orange order, Scottish Parliament, separate aspirations, Smith Commission, Speaker's Conference, United Ireland, Independence referendum, Ireland act, British union, Welsh devolution, Cymru Fydd, Welsh

Board, Welsh Acts, Welsh affairs, Welsh Office, Lord Crowther, Lord Kilbrandon, Welsh Assembly, Richard Commission, Welsh Government, Silk Commission, Administrative devolution, Devolution Referendum, Scottish Parliament, Reserved Powers Model, Scottish Board, independence of Scotland, independence of Wales, Irish Citizen Army, Irish Republican Army, Irish Republican Brotherhood, Scottish Covenant Association, Stone of Destiny, Council of Wales, Welsh Assembly Government, Wales Act, disestablishment in Wales, Independence of Ireland, Irish Parliamentary Party, Joint Ministerial Committees, Regional Assemblies Bill, Welsh Assembly, Yorkshire parliament, Northern assembly campaign, Royal Commission on the Constitution, Campaign for a Northern Assembly, Regional Assemblies Preparation Bill, Regional economic planning board, Regional economic planning council, Your Region, Your Choice, Secretary of State for Wales, Secretary of State for Scotland, Council for Wales and Monmouthshire, A Voice for Wales, Government of Wales Act 1998, Government of Wales Act 2006, Powers for a Purpose in 2015, Tax Collection and Management Act.

**List 5:** First optimized territorial politics dictionary for the United Kingdom – version 2

H Block, H-Block, IRA , UAC , UDA , UFF , UVF , a Parliament in Northern Ireland, Air Passenger Duty, All Wales Convention, All-Wales Convention, Anglo-Irish Agreement, Anglo-Irish Treaty, Anti-treaty, Anti-treatyite, assemblies for Scotland, assemblies for Wales, assembly for England, assembly for Scotland, assembly for Wales, Belfast Agreement, Blanket protest, Bloody Friday, Bloody Sunday, Border poll , Campaign for a Northern Assembly, Catholic areas, Commission on the Powers and Electoral Arrangements of the National Assembly for Wales, Confederalism, Constitutional convention, Crowther Commission, Cymru Fydd, Devolution, Devolved powers, Easter Rising, Fair Employment Act, Fair Employment Agency of Northern Ireland, Fair Employment Northern Ireland Act, Free Derry, Fresh Start Agreement, Good Friday Agreement, Hillsborough Agreement, Hillsborough Castle Agreement, Home rule, I. R. A., I.R.A., Independence of Ireland, Independence of Scotland, Independence of Wales, Independence referendum, Independent Ireland, Independent Scotland, Independent Wales, International Body on Arms Decommissioning, Ireland Act, Ireland's independence, Irish bill, Irish Citizen Army, Irish Free State Act, Irish Free State Consequential Provisions Act, Irish independence, Irish National Liberation Army, Irish nationalism, Irish nationalist, Irish Parliamentary Party, Irish problem, Irish question, Irish Republican Army, Irish Republican Brotherhood, Irish troubles, Irish unity, Irish Volunteers, Irish War of Independence, Joint Ministerial Committees, Kilbrandon Report, Long Kesh, Maze Prison, Mitchell principles, Mitchell report, Northern Assembly Campaign, Northern Ireland Act, Northern Ireland Constitution, Northern Ireland's Fair Employment Agency, Orange order, Orangeman, Orangemen, Parliament for Northern Ireland, peace in Northern Ireland, peace in Ulster, Powers for a Purpose, Pro-treaty, Republican Ireland, Reserved Powers Model, Richard Commission, Rome rule, Royal Commission on the Constitution, Scotland Act, Scotland Bill, Scotland's independence, Scottish Covenant Association, Scottish independence, Scottish local government, Scottish nationalism, Scottish referendum,



Separate aspirations, Silk Commission, Smith Commission, Special category status, St Andrews Agreement, Status of Northern Ireland, Status of Scotland, Status of Wales, Stone of Destiny, Stormont House Agreement, Sunningdale Agreement, Suspensory Act, Tax Collection and Management Act, Tax Collection and Management Wales Act, Tax-varying powers, Treatyite, U. A. C., U. D. A., U. F. F., U. V. F., U.A.C., U.D.A., U.F.F., U.V.F., Ulster Army Council, Ulster Covenant, Ulster crisis, Ulster Defence Association, Ulster Freedom Fighters, Ulster unionism, Ulster Volunteer Force, Ulster Workers' Council strike, Ulster's Solemn League and Covenant, Unified Ireland, UWF strike, Voice for Wales, Wales act, Wales bill, Wales referendum, Wales' independence, Welsh devolution, Welsh Government, Welsh independence, Welsh nationalism, Welsh nationalist, Welsh referendum, Welsh taxes, West Lothian question, Yorkshire parliament, Your region, your choice.

**List 6:** Last optimized territorial politics dictionary for the United Kingdom – version 3

Note: categorization in bold only for orientation

**(Northern) England and Cornwall keywords:** Campaign for a Northern Assembly, Cornish assembly, assembly for England, Northern Assembly Campaign, Yorkshire parliament, Your region, your choice

**Northern Ireland (and Ireland) keywords:** Irish born loyalists, Irish Free State Act, Irish Free State Agreement, Irish Free State Consequential Provisions Act, Irish independence, Irish War of Independence, Loyalists in Southern Ireland, Parliament for Southern Ireland, H Block, H-Block, IRA , UAC , UDA , UFF , UVF , a Parliament in Northern Ireland, Anglo-Irish Agreement, Anglo-Irish Treaty, Belfast Agreement, blanket protest, Bloody Friday, Bloody Sunday, border poll , Catholic areas, constitutional convention, direct rule, Easter Rising, Fair Employment Act, Fair Employment Agency of Northern Ireland, Fair Employment Northern Ireland Act, Free Derry, Fresh Start Agreement, Good Friday Agreement, Hillsborough Agreement, Hillsborough Castle Agreement, I. R. A., I.R.A., Independence of Ireland, Independent Ireland, International Body on Arms Decommissioning, Ireland Act, Ireland Bill, Ireland's independence, Irish bill, Irish Citizen Army, Irish National Liberation Army, Irish Parliamentary Party, Irish Republican Army, Irish Republican Brotherhood, Irish troubles, Irish Volunteers, Kilbrandon Report, Long Kesh, Maze Prison, Mitchell principles, Mitchell report, Northern Ireland Act, Northern Ireland Constitution, Northern Ireland's Fair Employment Agency, Orange Order, orangemen, Parliament for Northern Ireland, peace in Northern Ireland, peace in Ulster, pro-treaty, republican Ireland, Rome rule, Special category status, St Andrews Agreement, Status of Northern Ireland, Stormont House Agreement, Sunningdale Agreement, Suspensory Act, Treatyite, U. A. C., U. D. A., U. F. F., U. V. F., U.A.C., U.D.A., U.F.F., U.V.F., Ulster Army Council, Ulster Covenant, Ulster crisis, Ulster Defence Association, Ulster Freedom Fighters, Ulster Volunteer Force, Ulster Workers' Council strike, Ulster's Solemn League and Covenant, Unified Ireland, UWF strike, anti-treaty, anti-treatyite, Irish loyalist, Irish nationalism, Irish nationalist, Irish problem, Irish ques-

tion, Irish unity, Loyalists in Ireland, Parliament for Ireland, peace in Ireland, Ulster loyalist, Ulster unionism.

**Scotland keywords:** Air Passenger Duty, assembly for Scotland, Crowther Commission, dispute resolution process, Independence of Scotland, Independent Scotland, powers to Scotland, Scotland Act, Scotland Bill, Scotland's independence, Scottish Covenant Association, Scottish independence, Scottish nationalism, Scottish referendum, Status of Scotland, Stone of Destiny, tax-varying powers, Wales Bill, West Lothian question.

**Scotland, Wales and England keywords:** assemblies for England, assemblies for Scotland, assemblies for Wales.

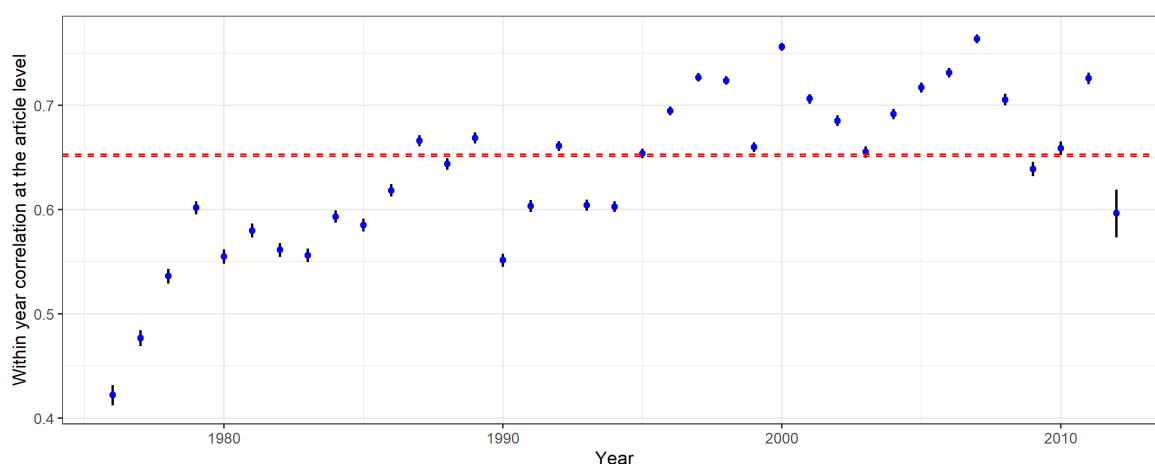
**United Kingdom keywords:** act of the Union, Barnett floor, Barnett formula, devolve powers, devolved government, devolved institutions, devolved powers, federal Britain, federal constitution, federal UK, federal United Kingdom, fiscal powers, home rule, Independence referendum, Parliaments for Ireland, Parliaments for Scotland, Parliaments for Wales, Reserved Powers Model, Royal Commission on the Constitution, Separate aspirations, Smith Commission.

**Wales keywords:** All Wales Convention, All-Wales Convention, Commission on the Powers and Electoral Arrangements of the National Assembly for Wales, Cymru Fydd, devolution in Wales, Holtham Commission, Independence of Wales, Independent Wales, Powers for a Purpose, powers to Wales, Richard Commission, Silk Commission, Status of Wales, Tax Collection and Management Act, Tax Collection and Management Wales Act, Voice for Wales, Wales Act, Wales' devolution, Wales referendum, Wales' independence, Welsh devolution, Welsh independence, Welsh nationalism, Welsh nationalist, Welsh referendum, Welsh taxes.

### SM3.3 Territorial issue identification using machine learning and optimized dictionaries

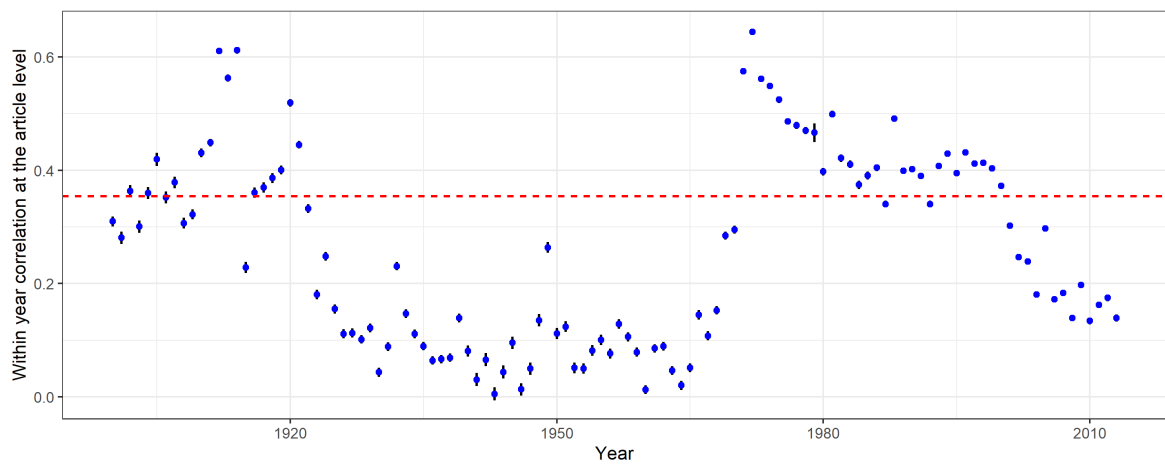
We test how much the identification of territorial issues based on our optimized dictionaries correlates with an identification based on the best-performing machine-learning algorithm at the level of articles. Thus, we compare how much machine-learning algorithms and optimized dictionaries coincide in their decision whether a single article contains territorial issues. Three aspects are worth mentioning: 1) When territorial politics is more salient, both methods converge more strongly; 2) Both methods correlated less in earlier periods both in the UK (The Times) and in Spain (El País). This might be related to historical biases of both measures; 3) Correlations are overall lower in the UK (The Times). We explain this with the high prevalence of OCR errors in The Times. Salience measures based on daily data correlate substantially more than measures based on the article level.

**Figure SM3.7:** Correlation of territorial issue identification within articles in Spain, El País 1976-2012



Note: Pearson's correlations across years, unit of analysis: article (El País);  $N = 1,555,553$ , training set  $N = 2,535$ . Average correlation: 0.65 (0.95-CI: 0.65; 0.65); Horizontal red lines show 0.95-confidence intervals for overall correlation of both methods across time; Vertical black lines show 0.95-confidence intervals for correlations within each year. Machine-learning algorithm used: Random Forest. Average salience across whole period: 6.1% (ML), 7.1% (optimized dictionaries).

**Figure SM3.8:** Article correlation between methods' predictions in the UK case, The Times 1900-2013



Note: Pearson's correlations across years, unit of analysis: article (The Times);  $N = 7,292,227$ , training set  $N = 1,368$ . Average correlation: 0.35 (0.95-CI: 0.35; 0.35); Horizontal red lines show 0.95-confidence intervals for overall correlation of both methods across time; Vertical black lines show 0.95-confidence intervals for correlations within each year. Machine-learning algorithm used: Support Vector Machine. Average salience across whole period: 1.1% (ML), 0.8% (optimized dictionaries).

### SM3.4 Dictionary performance compared to the performance of machine-learning algorithms

Note: All performance tests are conducted with the whole period under investigation. However, if we only test performance on exclusively overlapping periods (for Spain 2002-2018 and for UK 1985-2013), both dictionary and machine learning prediction performance increase slightly.

**Table SM3.9:** Territorial politics prediction confusion matrix and performance parameters of different dictionary versions and different ML algorithms in the Spanish case, El País (N = 2,535)

Prediction method	TN	FN	FP	TP	Sensitivity	Precision	Accuracy	F1
Dictionary 1	1991	236	116	191	0.45	0.62	0.86	0.52
Dictionary 2	1964	70	143	357	0.84	0.71	0.92	0.77
Dictionary 3	1962	59	145	368	0.86	0.72	0.92	0.78
SVM	2018	125	89	302	0.71	0.77	0.92	0.74
Random forest	2052	105	55	322	0.75	0.85	0.94	0.80
Naïve Bayes	1832	54	275	373	0.87	0.58	0.87	0.69

**Table SM3.10:** Territorial politics prediction confusion matrix and performance parameters of different dictionary versions and different ML algorithms in the UK case, The Times (N = 1,368)

Prediction method	TN	FN	FP	TP	Sensitivity	Precision	Accuracy	F1
Dictionary 1	883	39	486	188	0.83	0.28	0.67	0.42
Dictionary 2	1203	88	166	139	0.61	0.46	0.84	0.52
Dictionary 3	1242	93	127	134	0.59	0.51	0.86	0.55
SVM	1302	106	61	121	0.53	0.66	0.89	0.59
Random forest	1356	164	13	63	0.28	0.83	0.89	0.42
Naïve Bayes	557	18	812	209	0.92	0.20	0.48	0.33

### SM3.5 Transferability of optimized dictionary to new sources and political arenas

#### Spain

**Table SM3.11:** Transferability of an optimized dictionary of territorial politics to a new media source, Spain (El Mundo)

Prediction method	TN	FN	FP	TP	Sensitivity	Precision	Accuracy	F1
Dictionary 1	45	52	0	3	0.06	1.00	0.48	0.10
Dictionary 2	43	11	2	44	0.80	0.96	0.87	0.87
Dictionary 3	43	8	2	47	0.86	0.96	0.90	0.90
Random forest	28	28	17	27	0.49	0.61	0.55	0.55

Note: Prediction confusion matrix and performance parameters of different dictionary versions and different ML algorithms for transferring a territorial politics optimized dictionary to a new newspaper outlet (El Mundo, N = 100) based on El País (N = 2,535).

**Table SM3.12:** Transferability of an optimized dictionary of territorial politics to a different arena, Spain (Congreso de los Diputados)

Prediction method	TN	FN	FP	TP	Sensitivity	Precision	Accuracy	F1
Dictionary 1	51	36	0	13	0.27	1.00	0.64	0.42
Dictionary 2	47	9	4	40	0.82	0.91	0.87	0.86
Dictionary 3	45	5	6	44	0.90	0.88	0.89	0.89
Random forest	20	29	31	20	0.41	0.39	0.40	0.40

Note: Prediction confusion matrix and performance parameters of different dictionary version and different ML algorithms for transferring a territorial politics optimized dictionary to a new political arena (Spanish parliament Congreso de los Diputados, N = 100) based on El País (N = 2,535).

## United Kingdom

**Table SM3.13:** Transferability of an optimized dictionary of territorial politics to a different media source, UK (the Guardian)

Prediction method	TN	FN	FP	TP	Sensitivity	Precision	Accuracy	F1
Dictionary 1	44	8	9	39	0.83	0.81	0.83	0.82
Dictionary 2	48	5	5	42	0.89	0.89	0.90	0.89
Dictionary 3	48	2	5	45	0.96	0.90	0.93	0.93
SVM	53	27	0	20	0.43	1.00	0.73	0.60

Note: Prediction confusion matrix and performance parameters of different dictionary versions and SVM for transferring a territorial politics optimized dictionary to a new newspaper outlet (the Guardian, N = 100) based on The Times (N = 1,368).

**Table SM3.14:** Transferability of an optimized dictionary of territorial politics to a different arena, UK (House of Commons)

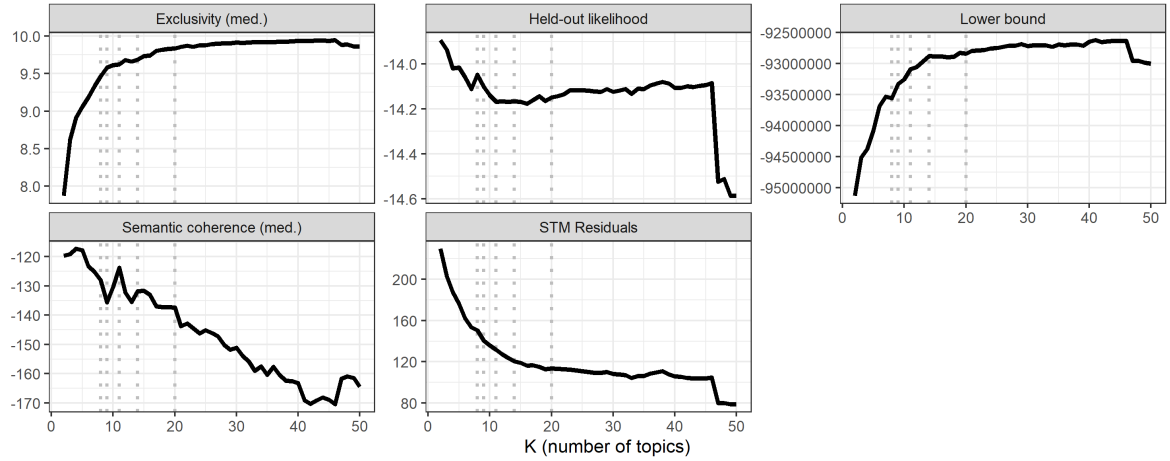
Prediction method	TN	FN	FP	TP	Sensitivity	Precision	Accuracy	F1
Dictionary 1	49	12	8	31	0.72	0.79	0.80	0.76
Dictionary 2	49	4	8	39	0.91	0.83	0.88	0.87
Dictionary 3	48	2	9	41	0.95	0.82	0.89	0.88
SVM	54	29	3	14	0.33	0.82	0.68	0.47

Note: Prediction confusion matrix and performance parameters of different dictionary versions and SVM for transferring a territorial politics optimized dictionary to a new political arena (British parliament House of Commons, N = 100) based on The Times (N = 1,368).

## SM3.6 STM selection

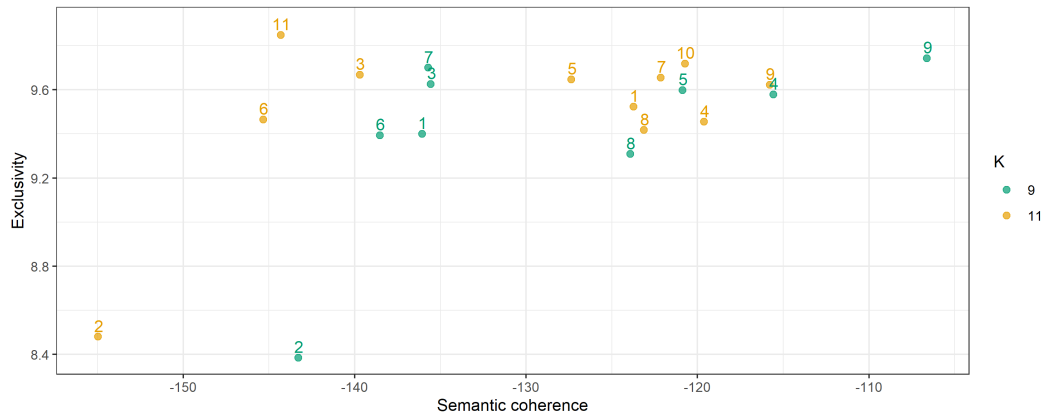
### Spain

**Figure SM3.9:** STM optimization parameters for territorial sentences in newspapers and parliament, Spain



Note: K range from 2 to 50; specification: spectral initialization without covariates. Grey vertical lines mark visual local optima.

**Figure SM3.10:** Comparison of exclusivity and semantic coherence of STMs with  $K = 9$  and 11, Spain

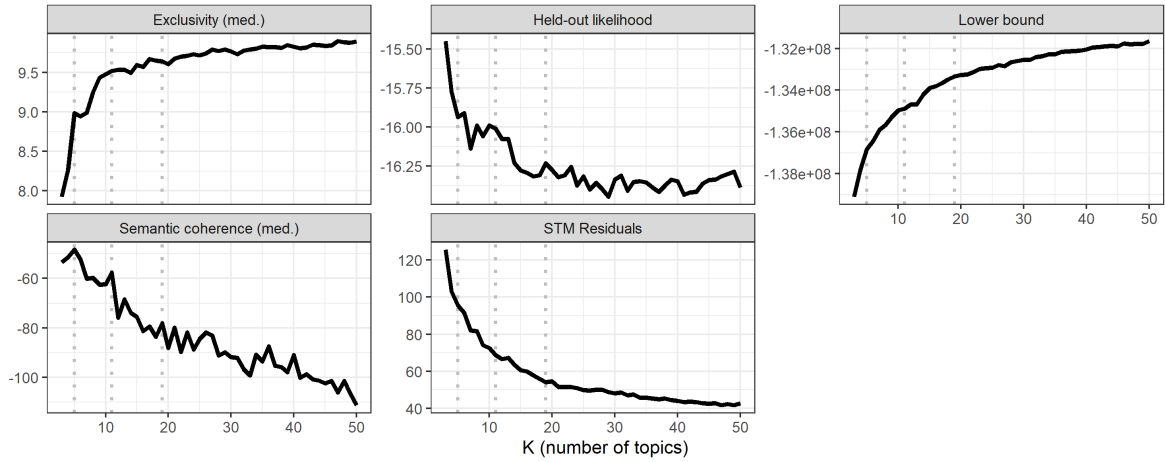


Note: Model comparison of topics with Ks 9 and 11; Model selected according to (1) local optima across the range of topics, and (2) discussions in the research group. We based our decision on the distribution of words within topics, exemplarity texts for each topic and topic correlations. We selected the model with  $K = 11$ .



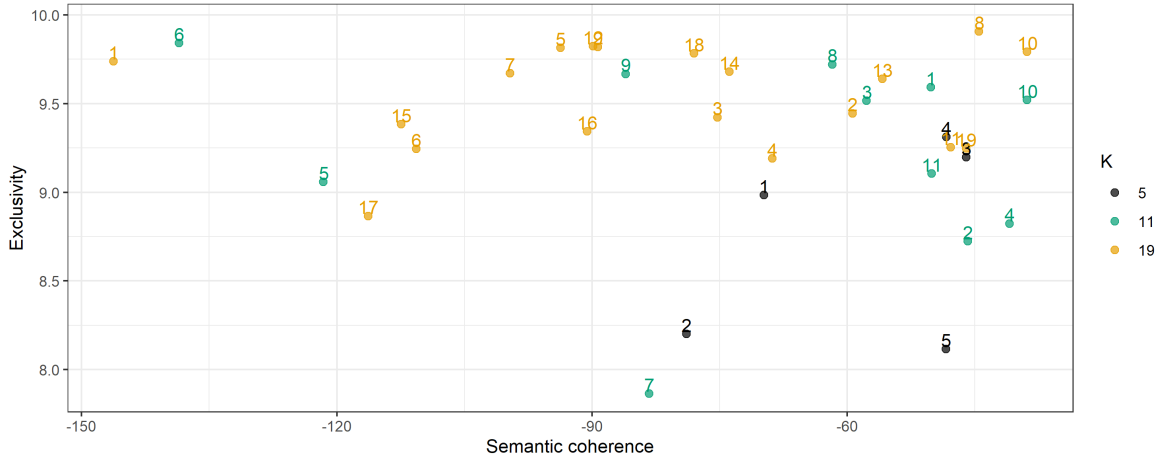
## United Kingdom

**Figure SM3.11:** STM optimization parameters for territorial sentences in newspapers and parliament, UK



Note: K range from 2 to 50; specification: spectral initialization without covariates. Grey vertical lines mark visual local optima. In the case of The Times, we used paragraphs instead of sentences due to issues with sentence recognition because of OCR errors.

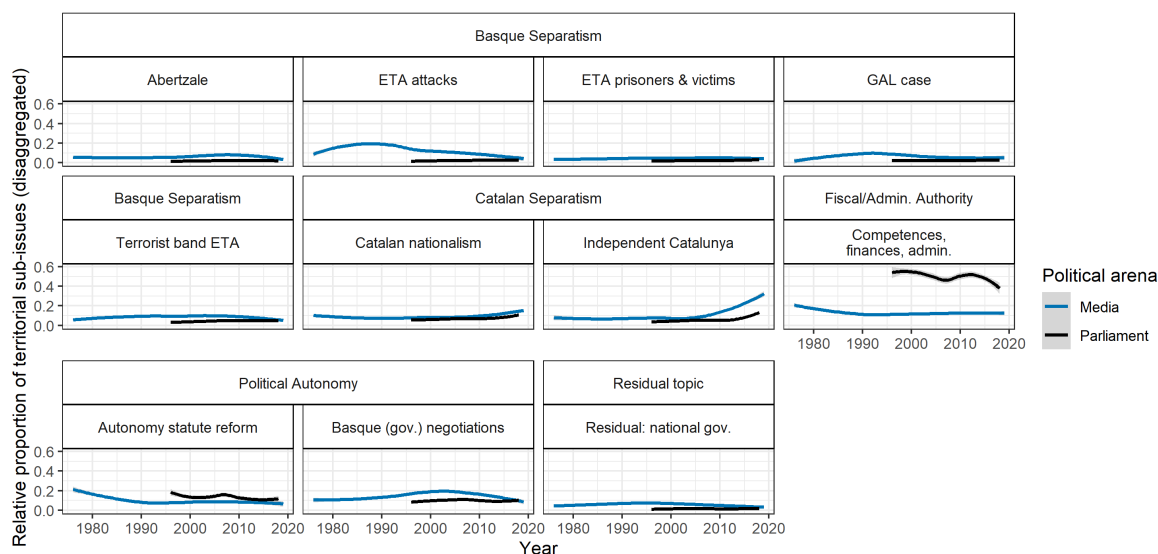
**Figure SM3.12:** Comparison of exclusivity and semantic coherence of STMs with  $K = 5$ , 11, and 19, UK



Note: Model comparison of topics with  $K$ s 5, 11 and 19; Model selected according (1) local optima across the range of topics, and (2) discussions in the research group. We based our decision on the distribution of words within topics, exemplarity texts for each topic and topic correlations. We selected the model with  $K = 19$ .

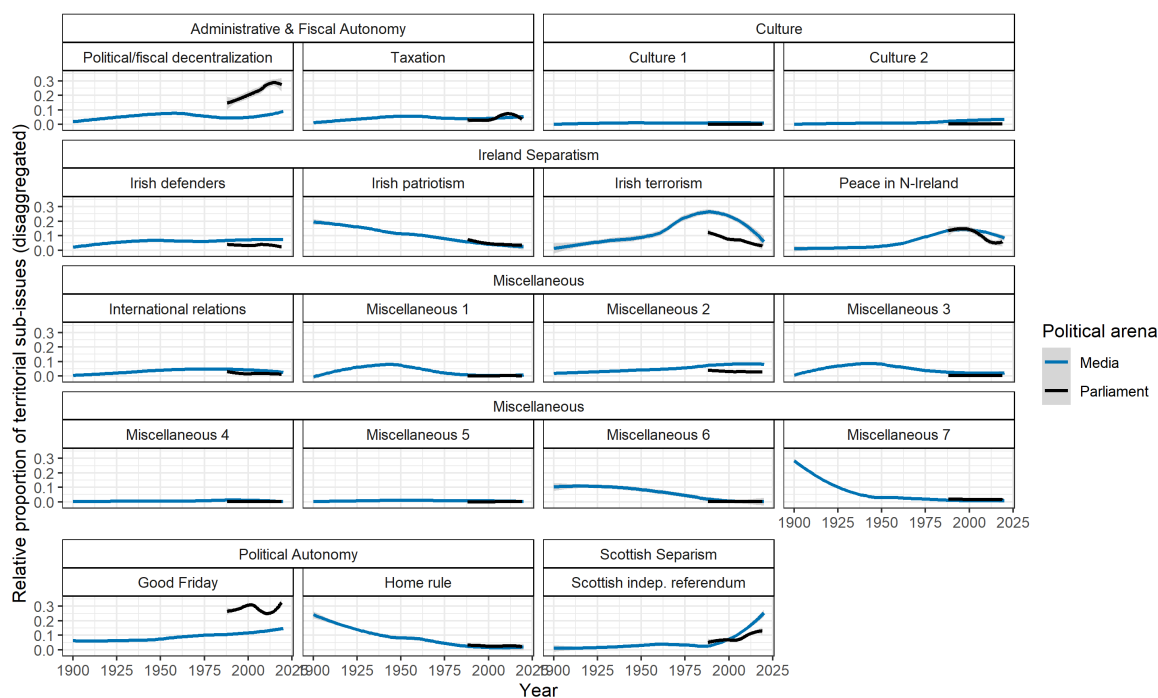
## SM3.7 Territorial topics without aggregation across time and arenas

**Figure SM3.13:** Territorial topics without aggregation, Spain 1976-2019



Note: Mean prevalence of each topic in each year, based on an STM with  $K = 11$ .

**Figure SM3.14:** Territorial topics without aggregation, UK 1900-2019



Note: Mean prevalence of each topic in each year, based on an STM with  $K = 19$ .

## SM3.8 Representative texts of the topics without aggregation

### List 7: Representative sentences for 11 territorial topics, Spain

[1: Terrorist attacks band ETA] “En ella dice que oyo una voz que le llamaba pronunciando dos veces su apellido, lo que le sobresalto porque, por haber recibido varias amenazas de muerte, creyo que se trataba de algun activista de ETA que iba a atentar contra su vida, pero, que al ver que quien le llamaba era un joven muy pulcro y [...]”

[2: Catalan separatism] “Entenent que la ficció te aixopluc fins i tot quan l'autor parteix de la pròpia vida, ja que en realitat; indica Amat, el que fan molts autors es partir d'una veritat personal, emocional i sentida i embolicar-la amb mentides, es a dir amb ficció; Primera Persona proposa la presència d'escriptors [...]”

[3: GAL case] “El caso del secuestro de Larretxea fue desgajado del sumario principal de los GAL y se encuentra pendiente de instruccion en el Juzgado Central numero 1 de la Audiencia Nacional hasta que el fiscal interponga la correspondiente querella, El asunto se encuentra pendiente de resolucion del fiscal, qui [...]”

[4: Abertzale] “APOYO LAS LISTAS QUE EL PP NO PUDO PARAR Electoral Mendi (AEN) Alegikoalde Azkertiar Abertzalea (Atea-Alegia) Anueko Indarra (AI) Bagoaz ( Zestoa Belauntzako Sustraiak ( Belauntza Berriozar Baietz (BB) Branka ( Hondarribia Erreil Bizirik ( Errezil Herriarengatik Irun Herria Izustarri Maeztuko Auker [...]”

[5: Residual national government] “ÁBALOS MECO, Jose Luis ACEDO PENCO, Pedro AGIRRETXEA URRESTI, Joseba Andoni AGUIAR RODRÍGUEZ, Ernesto AGUIRRE RODRÍGUEZ, Ramon ALBA GOVELI, Nayua Miriam ALBA MULLOR, Maria Dolores ALBADALEJO MARTÍNEZ, Joaquin ALCONCHEL GONZAGA, Miriam ALLI MARTÍNEZ, Ínigo Jesus ALONSO ARANEGUI, Alfonso ALONSO CANTOR [...]”

[6: Independent Catalunya] “¿El líder del PSC y candidato a presidir la Generalitat, Pere Navarro, acuso ayer al líder de ICV-EUiA, Joan Herrera, de contribuir a hacer crecer la deriva independentista del presidente de la Generalitat y candidato a la reelección, Artur Mas (Ci En un mitin en Manresa ante 300 personas, Navarro [...]”

[7: Basque (government) negotiations] “Ahora bien, pese a dejar clara la voluntad del Pacto de propiciar ese final dialogado de la violencia, Ardanza dejó claro que para llegar a ese punto ETA debe dar muestras inequívocas de querer abandonar la violencia porque lo contrario equivaldría a provocar fracaso y frustración:”

[8: ETA attacks] “APOYO Un completo arsenal etarra Entre el material incautado a ETA figuran 180 kilogramos de nitrato amónico, 15 litros de nitrometano, un subfusil MAT con dos cargadores, una pistola Browning con dos cargadores, un revolver Smith Wesson calibre 38 con munición, varios 'tupper' para la confección de b [...]”

[9: Comptences, finances and administration] “Son de especial relevancia: 1) incluir en la Constitucion, como sugirio el Consejo de Estado, mencion expresa a las comunidades autonomas; 2) regular el Senado como Camara que represente eficazmente a los territorios tanto por su composicion como por sus funciones; 3) reconocer las singularidades y [...]”

[10: Autonomy statute reform] “Los representantes de UCD y PSOE han acordado, tras una reunion conjunta, solicitar que en el orden del dia de la asamblea que el pleno del Consejo General de Castilla y Leon celebrara el sabado en Avila se incluya una peticion al Gobierno para que convoque la Asamblea de Parlamentarios y Diputacion [...]”

[11: ETA prisoners and victims] “Al igual que Bolinaga fue puesto en libertad con el pretexto de que sufría una enfermedad terminal, al igual que decenas de presos han salido a la calle por la aplicacion de la sentencia de Estrasburgo y al igual que otros muchos disfrutaban de permisos y beneficios penitenciarios por la via Nanclares [...]”

Note: Representative documents for each topic in the STM selected for Spain. For space reasons, we chose only the first 300 characters of each document.

#### **List 8: Representative paragraphs for 19 territorial politics, UK**

[1: Miscellaneous 1] “A Aberavon E 50,025 V 35,963 (71.9%) John Morris (Lab) 25,650 Ron McConville (LD) 4,079 Peter Harper (C) 2,835 Phil Cockwell (PC) 2,088 Peter David (Ref) 970 Captain Beany (Beanus) 341 Lab hold Maj 21,571 Swing 2.7% from LD to Lab 1992: Lab 26,877; C 5,567; LD 4,999; PC 1,919; Real Bean 707 Aberdeen [...]”

[2: Home rule] “A- meeting of the Opposition peers was held yesterday at the House of -Lords to consider the contentious Bills which the Government are sending up shortly. There were about 40 present, including Lord Lansdowne, the Duke of Devonshire, Lord Midletont, Lord Salisbury, Lord Camperdown, Lord Kenyon, Lor [...]”

[3: International relations] “Decades of discord 1951 Iran nationalises precursor of BP, the Anglo-Iranian Oil Company, triggering a dispute with Britain 1953 The Prime Minister, Mohammed Mossadeq, deposed in a coup with British and US backing 1980 Britain closes its embassy in Tehran after the Islamic revolution 1989 The Irania [...]”

[4: Miscellaneous 2] “Remember Kia-Ora Remember Kia- Ora Remember Kia-Ora Remember ftsa-OrRi be 'iaa Remem her n wMnber EKia-Ora RemeJAR . EW'cmember Kia-Ora 1 I5 l vAr Kia-Ora Remcmber Kiaa Fmember Kia-Ora Remem ber Kia-Ora Remember Kia-Ora Remember Kia-Ora Remember Kia- Ora Remember Kia-Ora Remember Kia-Ora Remember Kia-Ora Remember [...]”

[5: Culture 1] “ mmmmMfWftmi swasw'w\$?mmmm!8MjmmMm!m Theatres ADELPHI

0844 412 4651 loveneverdies.com 'ANDREW LLOYD WEBBER AT HIS MUSICAL BEST'  
Times LOVE NEVER DIES Mon-Sat 730pm, Wed Sat 2.30pm ALDWYCH THEATRE  
0844 847 1714 DIRTY DANCING THE CLASSIC STORY ON STAGE Mon-Thur 730, Fri  
5 8.30pm, Sat 3 h 730 [...]"

[6: Miscellaneous 3] "Home Away P W D L F A W D L F A GD Pts 1 Walsall 22 9 1 1  
21 3 5 5 1 14 10 22 48 2 Swindon 22 7 1 3 16 9 6 3 2 14 9 12 43 3MKDons 22 7 1 3 19 13 6  
2 3 20 16 10 42 4 Lincoln City 22 6 2 3 23 14 7 0 4 20 14 15 41 5 Wycombe 22 6 4 1 15 7 5  
1 5 11 12 7 38 6 Peterborough 22 5 2 4 25 21 6 3 2 [...]"

[7: Taxation] "JwÔMAN AVIATION SERVICES CO It t \*Ji J+\* -A \* H- Uff tMa  
tYUTOK UTKII CO. CUOQ Excellent Career Challenge Attractive Tax Free Salary Other  
Benefits Oman Aviation Services Company SAOG is a growth orientated public company in  
the Aviation industry based in the Sultanate of Oman The company s a [...]"

[8: Good Friday] "There are very few references to the border at all in the Belfast agree-  
ment, but where there are references, they do not in any way suggest that this decision cannot  
take place."

[9: Scottish indep. referendum] "Here's the agenda for the day. 10am: Conference opens  
with announcement of the results of the deputy leadership election. 10.15am: Welcome  
address by Elizabeth Grant, provost of Perth and Kinross council. 10.30am: Debates on the  
independence referendum, the minimum wage, social justice, cycling and [...]"

[10: Miscellaneous 4] "Japan Growth 294.20 313.701 - 1.90 ... Japanindex 81.65 87.14 -  
0.21 0.09 Japan Smlr Cos 37.28 39.79 -0 .60 ... international High Growth Funds Asian 57.24  
61.09 + 0.63 0.15 Hong Kong Gwlh 99.13 105.80 +0.50 0.78 Spore fiMlysn Gth 78.41 83.68  
-0 .13 ... Tiger Index 208.30 2223W +0.10 0.22 INVESCO F [...]"

[11: Irish patriotism] "The relief whihh a settlement would bring to right-minded people  
in America would be only less acute than t.hat which it woould bring to the Irish and the  
British. The three peoples have very strong ties of blood, culture and sympat hy, which have  
not been severed during the last few tragic years of [...]"

[12: Political/fiscal decentralization] "Gentleman agree with the recommendations of the  
final Holtham report, published today, which calls for an immediate Barnett floor to protect  
Wales from further convergence, the implementation of transition mechanisms towards a  
needs-based formula, and a place at the table for the Welsh Government in [...]"

[13: Irish terrorism] "Nine regular soldiers were injured in two attacks by the Provisional  
IRA near the Irish border late on Saturday and early yesterday. None was seriously hurt and  
only one was kept in hospital. The attacks occurred in the same area where a corporal aged  
30 was killed last Thursday by a landmine laid b [...]"

[14: Irish defenders] "O'Rahilly was a born rebel: a self-described anarchist whose grand-

father had died while storming Dublin's General Post Office during the 1916 Easter Rising."

[15: Culture 2] "Thandie Newton stars in The Chronicles of Riddick (Sky Movies 4, 8pm) 7.00PM 7.30 8.00 8.30 9.00 9.30 10.00 See Choice (F) available on Freeview (HD) High Definition MAIN CHANNELS SKY ONE The Simpsons Three back-to-back episodes of the cartoon: The Regina Monologues; Special Edna; and Goo Goo Ga/' Pa [...]"

[16: Miscellaneous 5] "B HI niyn / Wi " 2O' i 14 Locker CO A 142 ... 3.5 ... 49 3B ML Hogs 39 ... 2.7 3.5 . 67 26 MS Inll 26 ... 12.0 5.7 179 129 Mang Bronzet 141 + 1 3.5 11.0 539 412 McKecliniet 428 ... 4.3 15.B 120 71'=MeBQiltt 82 6.0 13.2 120 101 Metalr 106 ... 4.4 17.6 133 99 MdlsKt 131 ... II I? 589 518 Molins [...]"

[17: Miscellaneous 6] "J.D. YAXMULO.CO nntrollm ,theetir of nt TADING r IT d the d 1INMY (c) N AsiCuTv 190 Ute and to the Matter ot CEItXIAItD cltm HEY 1 2 d nr tomand Apostle. Q i.ond n ckrentYEC. aOrerof the ead of adbe, date th1t (df july196 pruodr Scto d. sob-scutIo 11(b) or the above mentiod ct.or\*equiry rD th t inh [...]"

[18: Peace in N-Ireland] "The collapse of Northern Ireland's political institutions moved closer yesterday after David Trimble, the Ulster Unionist Party leader, said that he was withdrawing his ministers from the Stormont Executive in protest at the IRA's continued refusal to decommission. He said he was pressing ahead afte[...]"

[19: Miscellaneous 7] "The Chancellor of the Exchequer had said that almost the whole of the provisions to be found in this Bill were taken from the report to which his right hon. friend the member for Wimbledon referred and that all they objected to was the machinery That seemed to him to be begging the question. On a su[...]"

Note: Representative documents for each topic in the STM selected for the UK. For space reasons, we chose only the first 300 characters of each document.

## SM4 Mediated party positions from newspapers

### SM4.1 Established word-scaling methods in political science

In the last decades, quantitative text-as-data methods have been developed to scale text and measure political preferences in an efficient and cost-saving manner. One early implemented supervised technique using quantitative text-as-data is *wordscores* (Laver, Benoit and Garry 2003). *wordscores* compare to what degree virgin texts are similar to anchor texts with known and exogenously given positions. However, it is difficult to choose consensual anchor positions, especially since they do not travel across text sources due to the sensitive changing language. To tackle these issues, unsupervised techniques without benchmark texts, like *wordfish* (Slapin and Proksch 2008; see implementations by Klüver 2009; Wratil and Hobolt 2019)<sup>1</sup> or correspondence analysis (Schonhardt-Bailey 2008), appear more adequate. Based on item-response theory, the wordfish algorithm calculates the probability that a certain word co-occurs in a certain text considering a Poisson distribution and a Naïve Bayes condition. The algorithm also controls length of the documents and the overall frequency of words, in order to find optimal discriminatory words. In a similar fashion, correspondence analysis implements a parametric dimensional space reduction procedure similar to principal component analysis (Lowe 2016). Like any dimension reduction approach, the latent scaling space depends on the data generation, in this case corpus selection. Manifestos *usually* conceive left-right positioning and parliamentary debates rather the government-opposition divide, since wordings are driven by bill proposal and criticisms.

Methodological innovations deal with issues of time, context and environment in both approaches. The *wordshoal* algorithm, e.g., controls for parliamentary debates specificities and conducts a wordfish estimation of MPs’ speeches within each debate (Lauderdale and Herzog 2016). The idea is to control parties’ language use in different contexts and periods. For example, a right-wing party could more often mention “decrease” in a debate about taxation but emphasize “increase” in a debate on security investments. Similarly, canonical correspondence analysis (CCA) emulates the wordshoal correction by harmonizing the text escalation based on environmental factors, e.g. the source of the data (see Sältzer 2020).

---

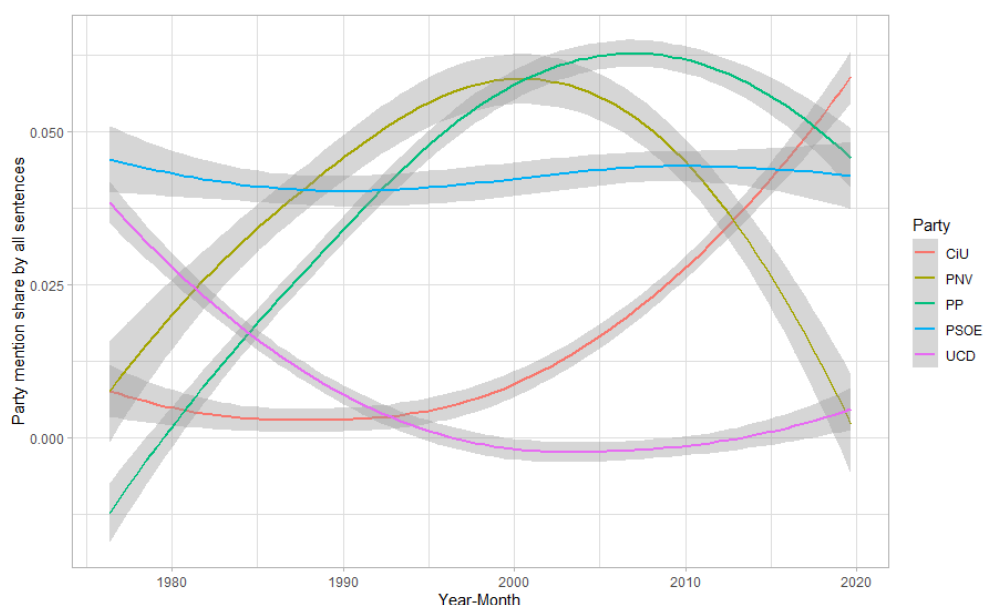
<sup>1</sup>In this same vein, part-of-speech tagging, and word embeddings/transformers are novel methodologies to approach text-as data from a more linguistic and morphological perspective. Rheault and Cochrane (2020) develop a similar approach as wordfish but based on party embeddings with a neural network estimation.

## SM4.2 Regional newspaper readership in Spain

Since we are interested in territorial issues that normally are of greater interest for peripheral regions, the selection of four newspapers based in state-wide capitals can raise concerns. However, the study of territorial politics is often focused on decisions by statewide actors in the *center* (Verge 2013; Toubreau 2017). The role of regionalist actors in statewide governments territorial strategies should not be underestimated either (Field 2016). Further, the selected national press does not comprise a substantial regional bias in terms of readership. The Spanish newspapers El Mundo and El País have confronted challenges in their regional expansion. In the 1980s, only 0.8% of the Catalan population read El País, mostly students and young readers. By 1996, only 10% of the Basque readers were informed by statewide newspapers: 2.6% read El País and 6.4% El Mundo. In other words, 9 in every 10 Basque readers read a regional newspaper (Moureau 2004). The regional readership problems have been explicitly tackled by El País since the 1980s, for example, with regional outlets and editorial teams as well as the purchase of small newspapers in Basque provinces where the traditional outlet El Correo is not well established. In 2004, El País was selling 60% of its diaries outside of Madrid (Moureau 2004). However, by 2017, El País was read in Madrid three times more than in Catalonia and the Balearic Islands (AIMC 2018).

## SM4.3 Data description: party sentences over time

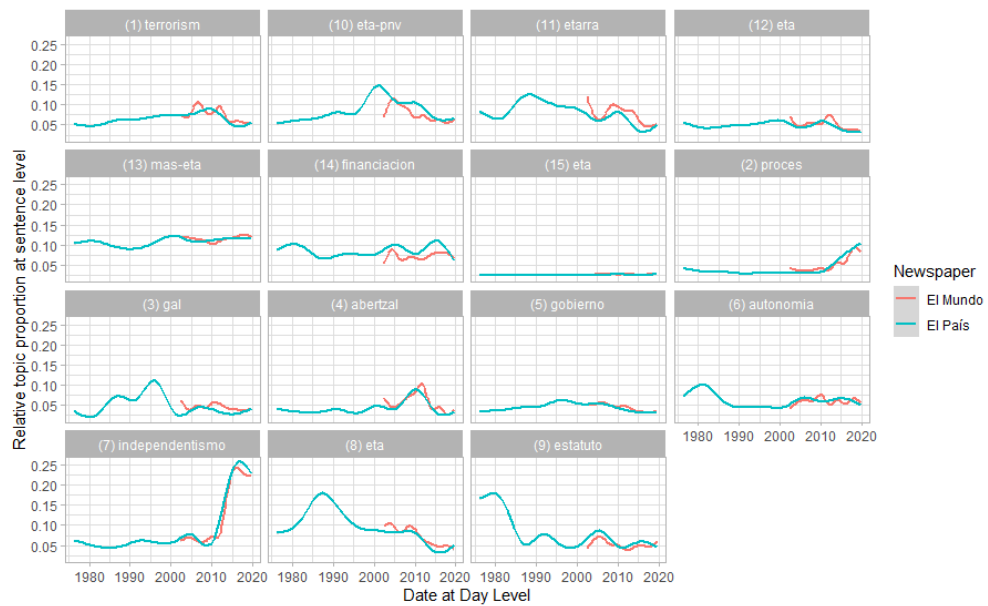
**Figure SM4.15:** Share of party sentences within territorial sentences in newspapers, 1976-2019



Note: Prominence of parties in El Mundo and El País in 572.831 territorial sentences and based on party and party leaders dictionary.

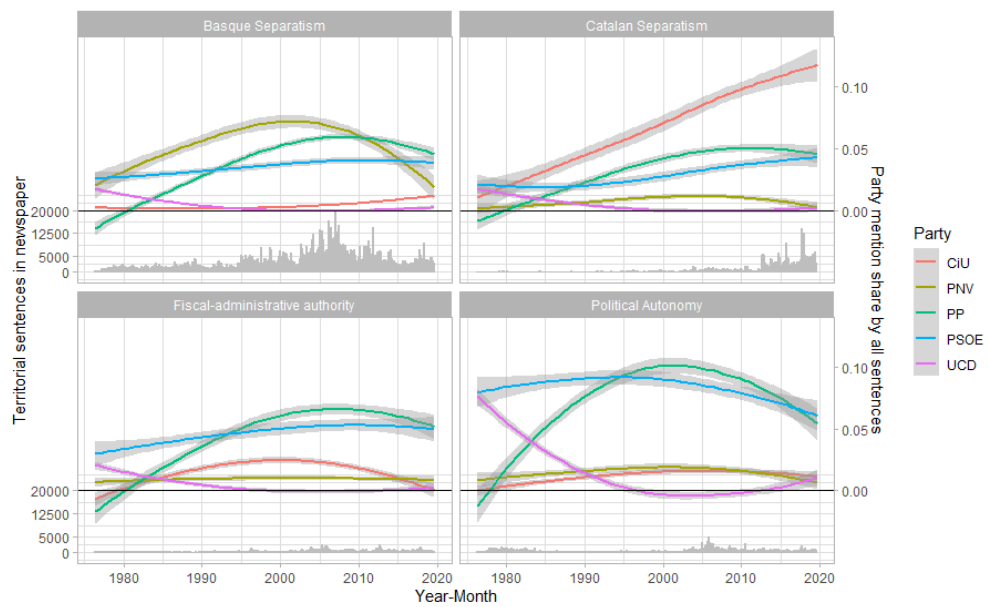


**Figure SM4.16:** Sub-issues by newspaper over time (predicted topic prevalences), 1976-2019



Note: Based on 572.831 territorial sentences in El Mundo and El País, estimation without estimation

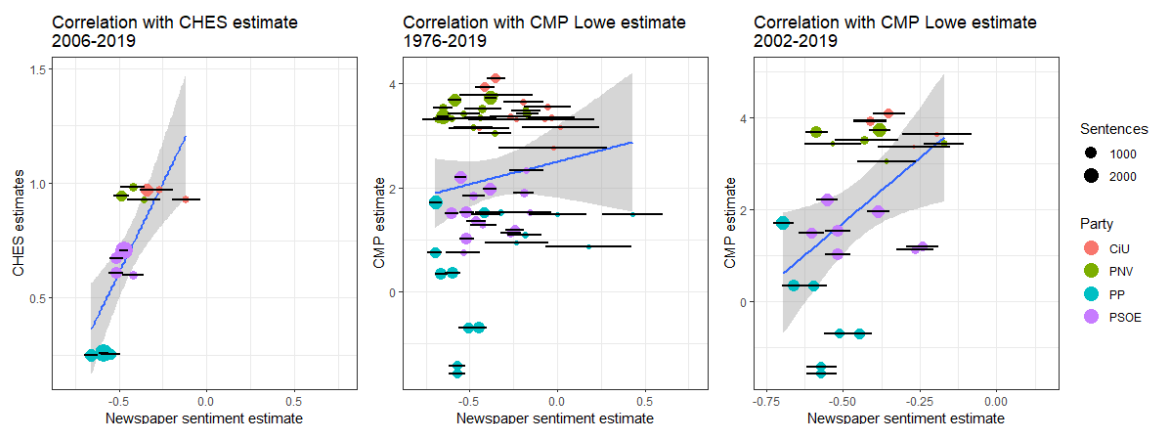
**Figure SM4.17:** Share of party sentences within territorial sentences in newspapers by topic sub-issue



Note: Prominence of parties in El Mundo and El País in 572.831 territorial sentences by territorial sub-issue and based on party and party leaders dictionary.

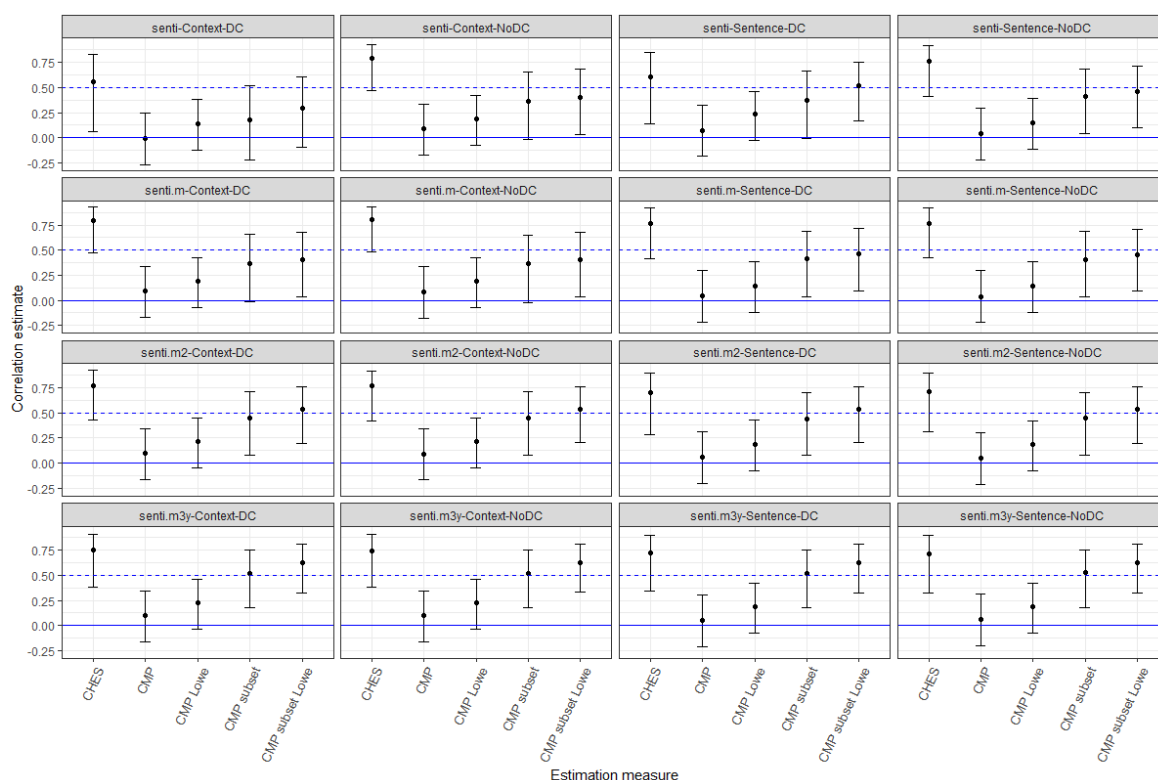
## SM4.4 Correlation of aggregated position with manifestos and expert surveys

**Figure SM4.18:** Mediated party positions, expert survey and electoral manifestos



Note: confidence bands at the 0.95 level. Horizontal bars of sentiment estimates reflect standard deviation of 200 bootstrapped samples by leaving each time one sentence out.

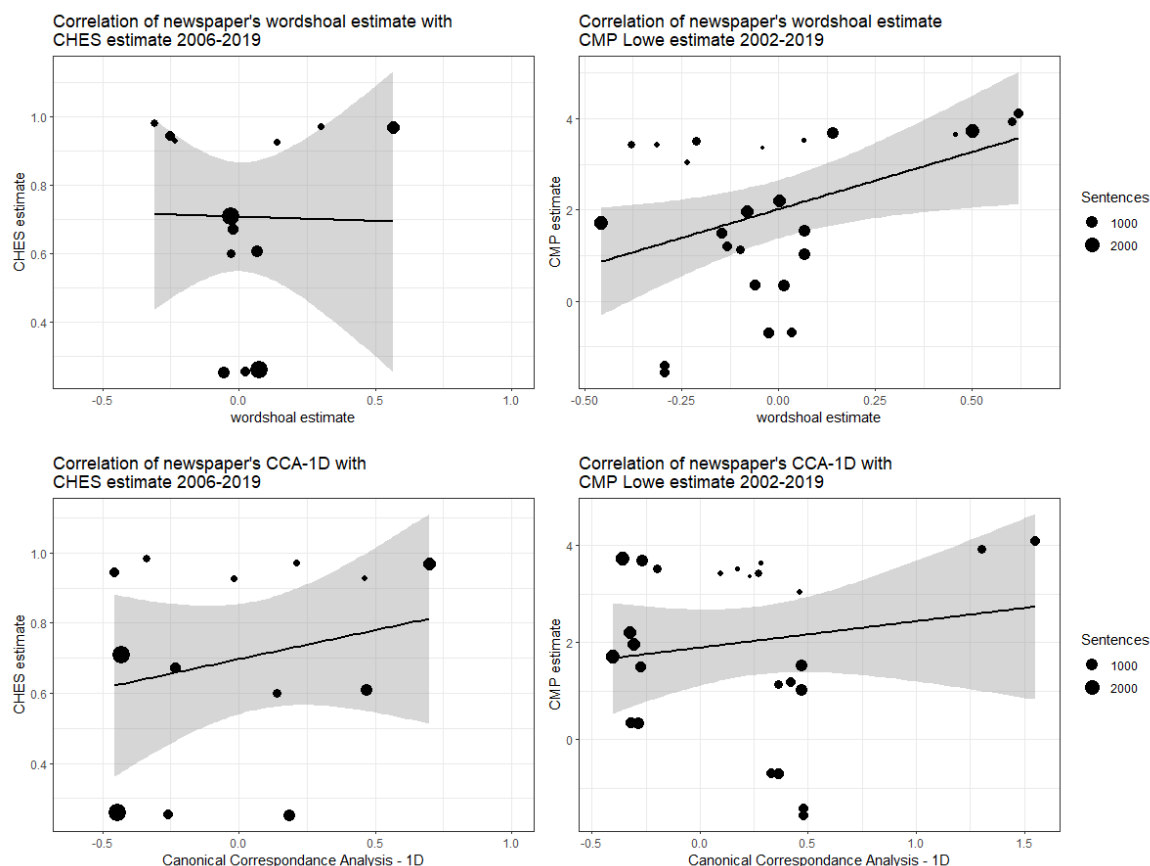
**Figure SM4.19:** Party positions of newspaper sentiment estimates, expert surveys, and electoral manifestos with different adjustments



Note: confidence bands at the 0.95 level. Sentiment estimate is the mean of 200 bootstrapped samples by leaving each time one sentence out.

## SM4.5 Manifestos, expert surveys, media positions based on word-scaling

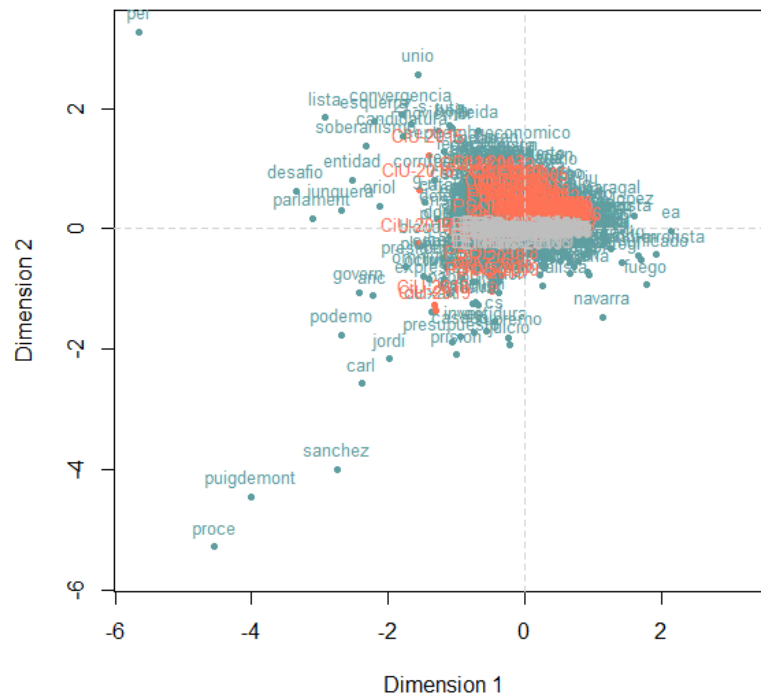
**Figure SM4.20:** Single territorial issue position estimates by wordshoal and CCA compared with CHES and MARPOR/CMP



Note: confidence bands at the 0.95 level. Sentiment estimate is the mean of 200 bootstrapped samples by leaving each time one sentence out.

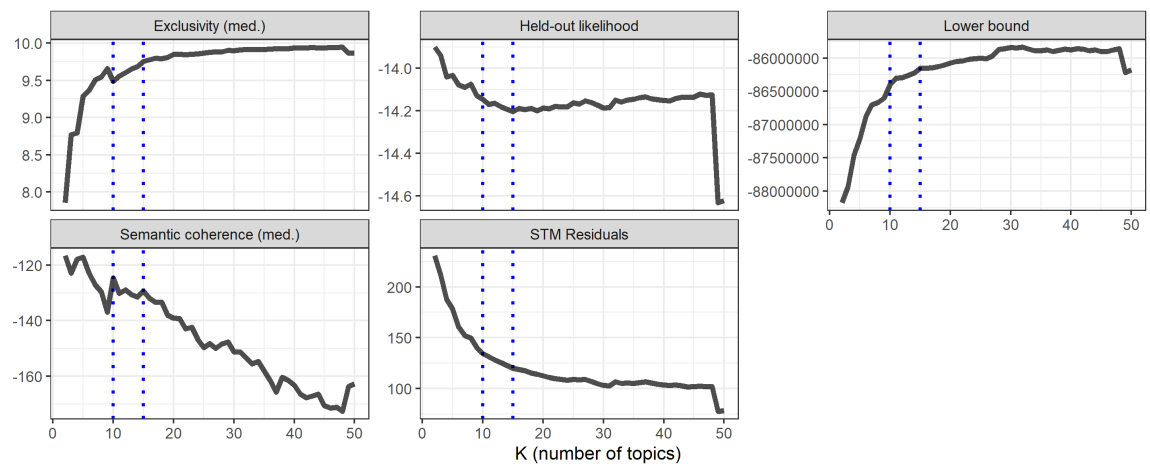
The analyses in Figure SM4.20 are estimated only based on all party sentences aggregated by year. As we argue above, word-scaling is very sensitive to language especially when few observation per text in the observation year are available, as can be seen in the Figure. Furthermore, Figure SM4.21 shows the word-scaling positions in the first and second latent dimensions using CCA. As can be seen, party-year-estimates can not be clearly clustered and mostly overlap (orange labels), and the estimated newspaper positions do not seem to explain any kind of the variation (grey labels). For both estimation, the newspaper outlet at the year level was the grouping variable for language use harmonization.

**Figure SM4.21:** Wording and positions of parties according to CCA



## SM4.6 Model performance analysis for STM

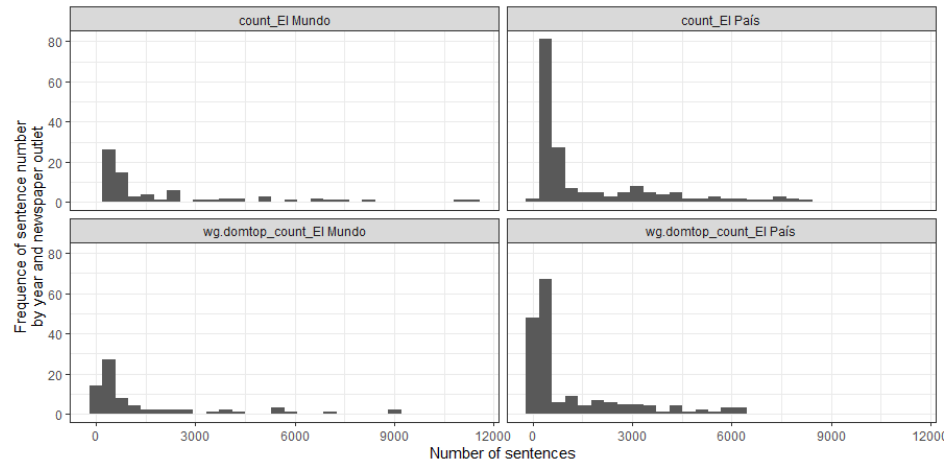
**Figure SM4.22:** STM optimization parameters for territorial sentences in newspapers



Note: K range from 2 to 50; specification: spectral initialization without covariates. Blue vertical lines mark visual local optima.

## SM4.7 Sentences reduction for the increase of certainty

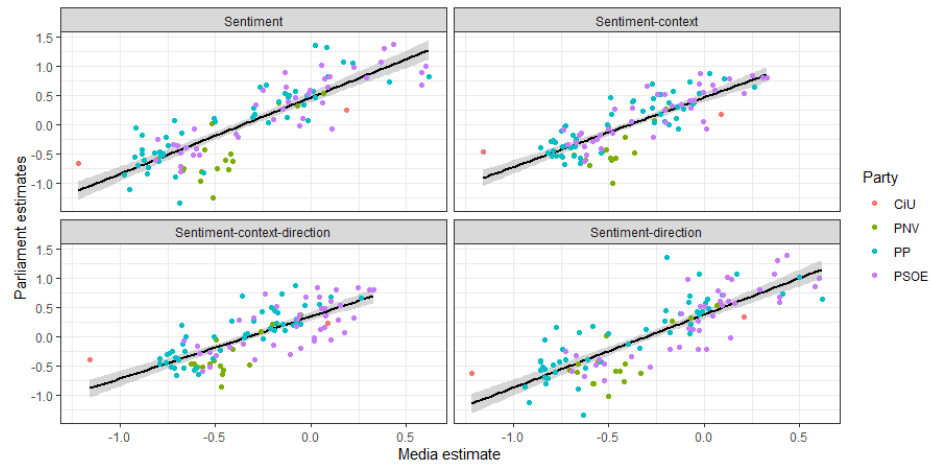
**Figure SM4.23:** Yearly sentences' distribution by condition and newspaper outlet



Note: Upper graphs shows the number of sentences for year estimation without conditioning based on El Mundo and El País. Lower graphs shows the distribution once the topic prevalence  $> 0.5$  condition is applied. This reduces the overall size of text data but, as shown, increases certainty in the positional measurement with sentiment.

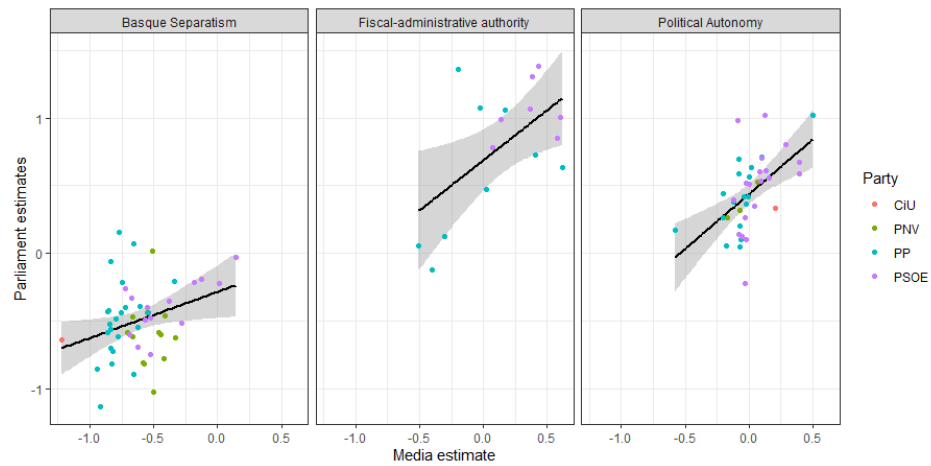
## SM4.8 Convergence with parliament - extension

**Figure SM4.24:** Convergence with parliament by party and measurement



Note: Based measurement of topic content  $> 0.5$  and weighted by prevalence.

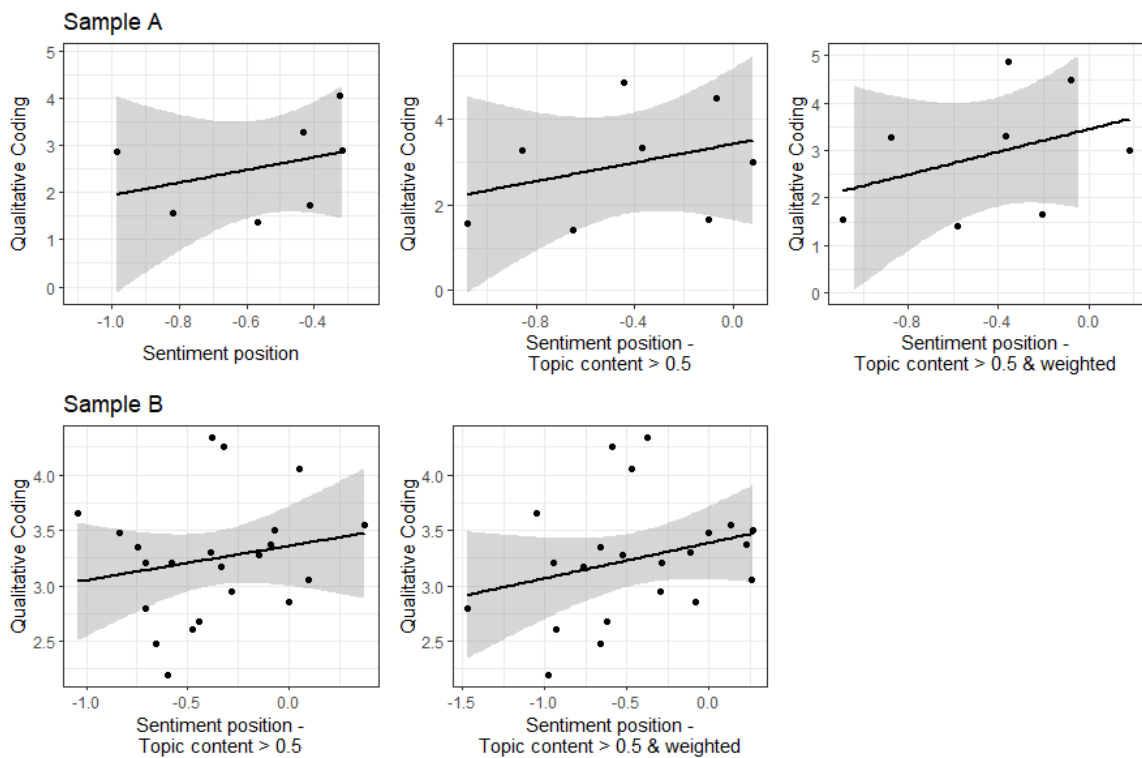
**Figure SM4.25:** Convergence with parliament by territorial sub-issue and party



Note: Based measurement of topic content  $> 0.5$  and weighted by prevalence.

## SM4.9 Qualitative validation - extension

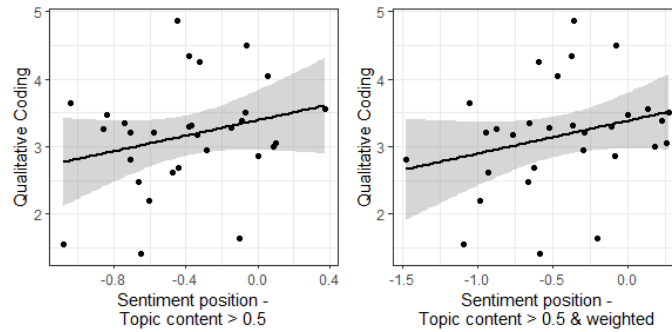
**Figure SM4.26:** Qualitative validation - two samples



Notes: Upper correlations:  $r = 0.54$  ( $p = 0.17$ ; .95 CI:  $-0.27, 0.90$ ) –  $r = 0.33$  ( $p = 0.42$ ; .95 CI:  $-0.49, 0.84$ ) –  $r = 0.37$  ( $p = 0.36$ ; .95 CI:  $-0.45, 0.85$ ). Lower correlations:  $r = 0.20$  ( $p = 0.38$ ; .95 CI:  $-0.24, 0.57$ ) –  $r = 0.28$  ( $p = 0.20$ ; .95 CI:  $-0.16, 0.63$ ). Within both samples, the comparison of manually annotated party positions between newspaper outlets El País and El Mundo (for the weighted subset of  $> 0.5$  prevalence) is strongly correlated. The correlation is  $0.90$  ( $p < 0.00$ ) for Sample A and  $0.58$  ( $p < 0.01$ ) for Sample B. A correlation comparison for the sentiment measure is not feasible due to the rather lower number of observations and our rule of aggregating at least 20 sentences.

The qualitative validation was conducted with two different but comparable sampling selection strategies. Sample A is a random selection of highly salient weeks for each territorial sub-issue, for which we coded the full weeks in June and July 2020 ( $N = 479$ ). Once we updated our understanding of aggregation and how mediated party positions can be better grasped, we changed the qualitative validation strategy. Sample B consist of 22 party-sub-issue year observations based on 440 sentences (20 sentences per observation). The observations were randomly selected for years with at least three parties consisting of a minimum of 20 sentences per territorial sub-issue. Sample B was coded in July and August 2021.

**Figure SM4.27:** Qualitative validation - joint samples



Note:  $r = 0.26$  ( $p = 0.17$ ; .95 CI: -0.12, 0.56) –  $r = 0.27$  ( $p = 0.14$ ; .95 CI: -0.10, 0.58).

## SM4.10 Regional voter transition and party policy shifts

**Table SM4.15:** Party Position Shifts - VAR Models

	DV: Party Position Shifts				
	Basque Country PNV (1)	Basque Country PP (2)	Basque Country PSOE (3)	Catalonia CiU (4)	Spain-wide PP (5)
Vote transition <sub>t</sub>	0.148 (0.124) p = 0.242	0.130 (0.306) p = 0.677	0.414 (0.318) p = 0.203	0.081 (0.096) p = 0.416	3.096 (1.349) p = 0.071
Vote transition <sub>t-1</sub>	0.064 (0.125) p = 0.613	-0.389 (0.282) p = 0.186	-0.313 (0.318) p = 0.332	0.112 (0.098) p = 0.282	-1.506 (0.979) p = 0.185
Party position shift <sub>t-1</sub>	-0.457 (0.155) p = 0.006	-0.370 (0.192) p = 0.070	-0.145 (0.170) p = 0.399	-0.343 (0.239) p = 0.182	-0.894 (0.426) p = 0.090
Party position shift <sub>t-2</sub>	-0.405 (0.148) p = 0.010	-0.455 (0.181) p = 0.022	-0.102 (0.169) p = 0.550	-0.220 (0.230) p = 0.362	-0.367 (0.360) p = 0.356
Constant	-0.022 (0.325) p = 0.946	-0.622 (0.514) p = 0.242	0.019 (0.714) p = 0.979	-0.027 (0.515) p = 0.960	-1.460 (0.852) p = 0.148
N	39	23	38	15	10
R <sup>2</sup>	0.300	0.362	0.091	0.300	0.671
Adjusted R <sup>2</sup>	0.217	0.221	-0.020	0.020	0.408
Residual Std. Error	2.026 (df = 34)	2.434 (df = 18)	4.394 (df = 33)	1.807 (df = 10)	1.815 (df = 5)
F Statistic	3.640** (df = 4; 34)	2.557* (df = 4; 18)	0.823 (df = 4; 33)	1.072 (df = 4; 10)	2.552 (df = 4; 5)



**Table SM4.16:** Regional Voter Transitions - VAR Models

	DV: Regional Voter Transitions				
	Basque Country PNV (1)	Basque Country PP (2)	Basque Country PSOE (3)	Catalonia CiU (4)	Spain-wide PP (5)
Vote transition <sub>t-1</sub>	0.030 (0.177) p = 0.867	0.447 (0.277) p = 0.125	0.590 (0.179) p = 0.003	-0.133 (0.298) p = 0.665	0.362 (0.448) p = 0.451
Vote transition <sub>t-2</sub>	-0.036 (0.178) p = 0.841	-0.190 (0.256) p = 0.469	-0.310 (0.179) p = 0.093	-0.294 (0.307) p = 0.360	-0.584 (1.637) p = 0.734
Party position shift <sub>t</sub>	0.078 (0.222) p = 0.729	0.134 (0.174) p = 0.452	-0.020 (0.096) p = 0.835	0.010 (0.744) p = 0.990	0.545 (0.561) p = 0.370
Party position shift <sub>t-1</sub>	-0.152 (0.211) p = 0.476	-0.028 (0.164) p = 0.868	0.050 (0.095) p = 0.605	-0.171 (0.715) p = 0.817	0.318 (0.566) p = 0.595
Constant	-0.042 (0.464) p = 0.929	-0.301 (0.465) p = 0.526	0.079 (0.402) p = 0.846	-1.483 (1.606) p = 0.378	-0.185 (1.098) p = 0.873
N	39	23	38	15	11
R <sup>2</sup>	0.025	0.141	0.256	0.101	0.352
Adjusted R <sup>2</sup>	-0.089	-0.050	0.166	-0.259	-0.080
Residual Std. Error	2.891 (df = 34)	2.205 (df = 18)	2.471 (df = 33)	5.628 (df = 10)	3.037 (df = 6)
F Statistic	0.222 (df = 4; 34)	0.736 (df = 4; 18)	2.844** (df = 4; 33)	0.281 (df = 4; 10)	0.815 (df = 4; 6)

## SM5 Statistical Software

- Benoit, Kenneth, Kohei Watanabe, Wang Hayan, Paul Nulty, Adam Obeng, Stefan Müller and Akitaka Matsuo. 2018. `quanteda`: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. doi: 10.21105/joss.00774, URL: <https://quanteda.io>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Stefan Müller, Patrick O. Perry, Benjamin Lauderdale and William Lowe. 2020. `quanteda.textmodels`: Scaling Models and Classifiers for Textual Data. R package version 0.9.1. <https://CRAN.R-project.org/package=quanteda.textmodels>
- Croissant, Yves and Millo Giovanni. 2018. *Panel Data Econometrics with R: the plm package*. Wiley.
- de Leeuw, Jan and Patrick Mair. 2009. Simple and Canonical Correspondence Analysis Using the R Package `anacor`. *Journal of Statistical Software*, 31(5), 1-18. URL <http://www.jstatsoft.org/v31/i05/>.
- Grün, Bettina and Kurt Hornik. 2011. `topicmodels`: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1-30. doi:10.18637/jss.v040.i13.
- Hlavac, Marek. 2018. `stargazer`: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. <https://CRAN.R-project.org/package=stargazer>
- Kuhn, Max and Davis Vaughan. 2020. `yardstick`: Tidy Characterizations of Model Performance. R package version 0.0.6. <https://CRAN.R-project.org/package=yardstick>
- Lauderdale, Benjamin. 2020. `wordshoal`: `quanteda` textmodel extensions for the `wordshoal` textmodel. R package version 0.4. <http://github.com/kbenoit/wordshoal>
- Liu, Licheng and Yiqing Xu. 2019. `panelView`: Visualizing Panel Data. R package version 1.1.2. <https://CRAN.R-project.org/package=panelView>
- Liaw, Andy and Wiener, Matthew. 2002. Classification and Regression by `randomForest`. *R News* 2(3), 18–22.
- Pedersen, Thomas Lin. 2019. `patchwork`: The Composer of Plots. R package version 1.0.0. <https://CRAN.R-project.org/package=patchwork>
- Roberts, Margaret, Brandon M. Stewart and Dustin Tingley. 2019. `stm`: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91(2), 1-40. doi: 10.18637/jss.v091.i02.

- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer Verlag New York.
- Wickham et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Yan, Yachen. 2016. *MLmetrics: Machine Learning Evaluation Metrics*. R package version 1.1.1. <https://CRAN.R-project.org/package=MLmetrics>

## BIBLIOGRAPHY

- Aalen, Lovise and Ragnhild Louise Muriaas. 2017. “Power Calculations and Political Decentralisation in African Post-Conflict States.” *International Political Science Review* 38(1):56–69.
- Abou-Chadi, Tarik and Lukas F. Stoetzer. 2020. “How Parties React to Voter Transitions.” *American Political Science Review* 114(3):940–945.
- Adams, James. 2012. “Causes and Electoral Consequences of Party Policy Shifts in Multi-party Elections: Theoretical Results and Empirical Evidence.” *Annual Review of Political Science* 15(1):401–419.
- Adams, James, Simon Weschle and Christopher Wlezien. 2021. “Elite Interactions and Voters’ Perceptions of Parties’ Policy Positions.” *American Journal of Political Science* 65(1):101–114.
- Aggarwal, Charu. 2018. *Machine Learning for Text*. New York: Springer.
- AIMC. 2018. “Segundo informe anual - Marco General de los Medios en España; Oficina de Justificación de la Difusión (OJD) y Asociación para la Investigación de Medios de Comunicación (AIMC).” *El País*. *EL PAÍS lidera la prensa generalista con 1.069.000 lectores* June 27, 2018.
- Albugh, Quinn, Julie Sevenans and Stuart Soroka. 2013. *Lexicoder Topic Dictionaries, June 2013 versions*. Montreal: McGill University.
- Alesina, Alberto, Reza Baqir and William Easterly. 1999. “Public Goods and Ethnic Divisions.” *Quarterly Journal of Economics* 114(4):1243–1284.
- Allern, Sigurd. 2017. “Journalistic and Commercial News Values;” *Nordicom Review* 23(1-2):137–152.

- Alonso, Sonia. 2012. *Challenging the State: Devolution and the Battle for Partisan Credibility*. Oxford: Oxford University Press.
- Altheide, David and Robert Snow. 1979. *Media Logic*. Beverly Hills: Sage.
- Amat, Francesc and Albert Falcó-Gimeno. 2014. “The Legislative Dynamics of Political Decentralization in Parliamentary Democracies.” *Comparative Political Studies* 47(6):820–850.
- Ansell, Ben W. and Johannes Lindvall. 2020. *Inward Conquest: The Political Origins of Modern Public Services*. Cambridge: Cambridge University Press.
- Arnold, Christine, Simon Hug and Tobias Schulz. 2009. Cross-Validating Measurement Techniques of Party Positioning.
- Bachrach, Peter and Morton Baratz. 1962. “Two Faces of Power.” *The American Political Science Review* 56(4):947–952.
- Bakker, Ryan, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Vachudova. 2021. “Chapel Hill Expert Survey (CHES) Europe 1999-2019 trend file.”.
- Banducci, Susan, Heiko Giebler and Sylvia Kritzinger. 2017. “Knowing more from less: How the information environment increases knowledge of party positions.” *British Journal of Political Science* 47(3):571–588.
- Barberá, Pablo, Amber Boydstun, Suzanna Linn, Ryan McMahon and Jonathan Nagler. 2020. “Automated Text Classification of News Articles: A Practical Guide.” *Political Analysis (online first)* pp. 1–24.
- Bardhan, Pranab and Dilip Mookherjee. 2006. Decentralization, Corruption and Government Accountability. In *International Handbook on the Economics of Corruption*, ed. Susan Rose-Ackerman. Cheltenham: Edward Elgar chapter 6.
- Basile, Linda. 2016. “Measuring party strategies on the territorial dimension: a method based on content analysis.” *Regional & Federal Studies* 26(1):1–23.
- Basta, Karlos. 2017. “The State between Minority and Majority Nationalism: Decentralization, Symbolic Recognition and Secessionist Crises in Spain and Canada.” *Publius: The Journal of Federalism* 48(1):51–75.
- Baumann, Markus, Marc Debus and Martin Gross. 2021. “Strategic issue emphasis in parties’ election campaign statements.” *Party Politics* 27(3):515–527.

- Baumann, Markus and Martin Gross. 2016. "Where is my party? Introducing new data sets on ideological cohesion and ambiguity of party positions in media coverage." *Arbeitspapiere/Mannheimer Zentrum für Europäische Sozialforschung; Working paper* 167.
- Baumgartner, Frank and Laura Chaqués-Bonafont. 2015. "All News is Bad News: Newspaper Coverage of Political Parties in Spain." *Political Communication* 32:268–291.
- Baumgartner, Fran, Christian Breunig and Emiliano Grossman. 2019. *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press.
- Bawn, Kathleen. 1993. "The Logic of Institutional Preferences: German Electoral Law as a Social Choice Outcome." *American Journal of Political Science* 37(4):965–989.
- Bednar, Jenna. 2004. "Authority Migration in Federations: A Framework for Analysis." *PS: Political Science Politics* 37(3):403–408.
- Bell, Andrew and Kelvyn Jones. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3(1):133–153.
- Bennett, W. Lance, Victor W. Pickard, David P. Iozzi, Carl L. Schroeder, Taso Lagos and C. Evans Caswell. 2006. "Managing the Public Sphere: Journalistic Construction of the Great Globalization Debate." *Journal of Communication* 54(3):437–455.
- Benoit, Kenneth and Michael Laver. 2006. *Party Policy in Modern Democracies*. London: Routledge.
- Blei, David, Andre Ng and Michael Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.
- Blei, David and John Lafferty. 2006. "Correlated Topic Models." *Advances in Neural Information Processing Systems* 18:147–154.
- Blei, David and John Lafferty. 2007. "A correlated topic model of Science." *The Annals of Applied Statistics* 1(1):17–35.
- Brancati, Dawn. 2006. "Decentralization: Fueling the Fire or Dampening the Flames of Ethnic Conflict and Secessionism?" *International Organization* 60(03):651–685.
- Brancati, Dawn. 2008. "The Origins and Strengths of Regional Parties." *British Journal of Political Science* 38(1):135–159.
- Brooks, Clem and Jeff Manza. 1997. "The Social and Ideological Bases of Middle-Class Political Realignment in the United States, 1972 to 1992." *American Sociological Review* 62(2):191–208.

- Broschek, Jörg, Bettina Petersohn and Simon Toubeau. 2018. "Territorial Politics and Institutional Change: A Comparative-Historical Analysis." *Publius: The Journal of Federalism* 48(1):1–25.
- Bräuninger, Thomas and Marc Debus. 2009. "Legislative Agenda-Setting in Parliamentary Democracies." *European Journal of Political Research* 48:804–839.
- Budge, Ian. 2015. "Issue Emphases, Saliency Theory and Issue Ownership: A Historical and Conceptual Analysis." *West European Politics* 38(4):761–777.
- Budge, Ian and Dennis Farlie. 1983. *Explaining and Predicting Elections: Issue Effects and Party Strategies in Twenty-three Democracies*. London/Boston: Allen Unwin.
- Bélanger, Éric and Bonnie M. Meguid. 2008. "Issue Salience, Issue Ownership, and Issue-Based Vote Choice." *Electoral Studies* 27(3):477–491.
- Capoccia, Giovanni and R Daniel Kelemen. 2007. "The Study of Critical Junctures: Theory, Narrative, and Counterfactuals in Historical Institutionalism." *World Politics* 59(3):341–369.
- Caramani, Daniele. 2004. *The Nationalization of Politics. The Formation of National Electorates and Party Systems in Western Europe*. Cambridge University Press.
- Caramani, Daniele. 2015. *The Europeanization of Politics: The Formation of a European Electorate and Party System in Historical Perspective*. Cambridge, UK: Cambridge University Press.
- Carmines, Edward and James Stimson. 1989. *Issue Evolution: Race and the Transformation of American Politics*. New Jersey: Princeton University Press.
- Carter, Neil, Robert Ladrech, Conor Little and Vasiliki Tsagkroni. 2018. "Political parties and climate policy: A new approach to measuring parties' climate policy preferences." *Party Politics* 24(6):731–742.
- Castanho Silva, Bruno and Levente Littvay. 2019. "Comparative Research is Harder Than We Thought: Regional Differences in Experts' Understanding of Electoral Integrity Questions." *Political Analysis* 27(4):599–604.
- Castanho Silva, Bruno and Sven-Oliver Proksch. 2021. "Politicians unleashed? Political communication on Twitter and in parliament in Western Europe." *Political Science Research and Methods* p. 1–17.
- Ceron, Andrea, Luigi Curini and Stefano M. Iacus. 2015. "Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy." *Social Science Computer Review* 33(1):3–20.

- Chandra, Kanchan. 2001. "Ethnic Bargains, Group Instability, and Social Choice Theory." *Politics & Society* 29(3):337–362.
- Chaqués-Bonafont, Laura, Anna Palau and Frank Baumgartner. 2015. *Agenda Dynamics in Spain*. London: Palgrave Macmillan.
- Cieslak, David and Nitesh Chawla. 2008. Learning Decision Trees for Unbalanced Data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Berlin: Springer.
- Clark, Tom S. and Drew A. Linzer. 2015. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3(2):399–408.
- Collignon, Sofia. 2019. "Governments, Decentralisation, and the Risk of Electoral Defeat." *West European Politics* 42(1):173–200.
- Convery, Alan and Thomas Carl Lundberg. 2017. "Decentralization and the Centre Right in the UK and Spain: Central Power and Regional Responsibility." *Territory, Politics, Governance* 5(4):388–405.
- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I Lindberg, Jan Teorell, David Altman, Michael Bernhard, M Steven Fish, Adam Glynn, Allen Hicken et al. 2020. "V-Dem dataset v10. Varieties of Democracy (V-dem) project."
- Costello, Rory, Dimitar Toshkov, Barend Bos and André Krouwel. 2021. "Congruence between voters and parties: The role of party-level issue salience." *European Journal of Political Research* 60(1):92–113.
- Costello, Rory, Jacques Thomassen and Martin Rosema. 2012. "European Parliament Elections and Political Representation: Policy Congruence between Voters and Parties." *West European Politics* 35(6):1226–1248.
- Curran, James. 2011. *Media and democracy*. Taylor Francis.
- Daku, Marc, Stuart Soroka and Lori Young. 2015. *Lexicoder, version 3.0*. Available at [www.lexicoder.com](http://www.lexicoder.com).
- Dardanelli, Paolo, John Kincaid, Alan Fenna, André Kaiser, André Lecours and Ajay Kumar Singh. 2019a. "Conceptualizing, Measuring, and Theorizing Dynamic De/Centralization in Federations." *Publius: The Journal of Federalism* 49(1):1–29.
- Dardanelli, Paolo, John Kincaid, Alan Fenna, André Kaiser, André Lecours, Ajay Kumar Singh, Sean Mueller and Stephan Vogel. 2019b. "Dynamic De/Centralization in Federations: Comparative Conclusions." *Publius: The Journal of Federalism* 49(1):194–219.



- Davenport, Christian. 2009. *Media Bias, Perspective, and State Repression: The Black Panther Party*. Cambridge Studies in Contentious Politics Cambridge University Press.
- de Sio, Lorenzo, Andrea de Angelis and Vincenzo Emanuele. 2017. "Issue Yield and Party Strategy in Multiparty Competition." *Comparative Political Studies* 51(9):1208–1238.
- de Sio, Lorenzo and Till Weber. 2014. "Issue Yield: A Model of Party Strategy in Multidimensional Space." *American Political Science Review* 108(4):870–885.
- Dearing, James and Everett Rogers. 1996. *Agenda Setting. Communication Concepts 6*. Thousand Oaks: Sage Publications.
- Deferre, Gaston. 1981. "Décentralisation : Ce Que Sera La Nouvelle France." *Le Nouvel Observateur* 18 July 1981.
- Den Hartog, Chris and Nathan Monroe. 2011. *Agenda Setting in the US Senate: Costly Consideration and Majority Party Advantage*. Cambridge: Cambridge University Press.
- Derczynski, Leon. 2016. "Complementarity, F-score, and NLP Evaluation." *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16)* 16:261–266.
- Deschouwer, Kris. 2009. Towards a Regionalization of Statewide Electoral Trends in Decentralized States? The Cases of Belgium and Spain. In *Territorial party politics in Western Europe*, ed. Wilfried Swenden and Bart Maddens. Basingstokes: Palgrave Macmillan pp. 31–46.
- Deutsch, Karl W. 1953. "The Growth of Nations: Some Recurrent Patterns of Political and Social Integration." *World Politics* 5(2):168–195.
- Diaf, Sami and Ulrich Fritsche. 2021. "Topic Scaling: A Joint Document Scaling – Topic Model Approach To Learn Time-Specific Topics."
- Dolezal, Martin, Laurenz Ennser-Jedenastik, Wolfgang C. Müller and Anna K. Winkler. 2014. "How Parties Compete for Votes: A Test of Saliency Theory." *European Journal of Political Research* 53(1):57–76.
- Domingo, David, Thorsten Quandt, Ari Heinonen, Steve Paulussen, Jane Singer and Marina Vujnovic. 2008. "Participatory Journalism Practices in the Media and Beyond." *Journalism Practice* 2(3):326–342.
- Donsbach, Wolfgang. 2004. "Psychology of News Decisions: Factors behind Journalists' Professional Behavior." *Journalism* 5(2):131–157.
- Dumdum, Omar O. and Levi Bankston. 2021. "The Interplay of Actors in Political Communication: The State of the Subfield." *Political Communication* 0(0):1–14.

- Döring, Holger and Hanna Schwander. 2015. "Revisiting the Left Cabinet Share: How to Measure the Partisan Profile of Governments in Welfare State research." *Journal of European Social Policy* 25(2):175–193.
- Eaton, Kent. 2012. *Decentralization and Federalism*. Routledge chapter Chapter 3.
- Elazar, Daniel J. 1987. *Exploring Federalism*. Tuscaloosa: University of Alabama Press.
- Esser, Frank and Jesper Strömbäck, eds. 2014. *Mediatization of Politics*. London: Palgrave Macmillan.
- Faguet, Jean-Paul. 2014. "Decentralization and Governance." *World Development* 53:2–13.
- Falleti, Tulia G. 2005. "A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective." *American Political Science Review* 99(3):327–346.
- Falleti, Tulia G. 2013. Decentralization in time: A process-tracing approach to federal dynamics of change. In *Federal Dynamics*, ed. Arthur Benz and Jörg Broschek. Oxford University Press pp. 140–166.
- Ferree, Myra Marx, William Anthony Gamson, Dieter Rucht and Jürgen Gerhards. 2002. *Shaping abortion discourse: Democracy and the public sphere in Germany and the United States*. Cambridge University Press.
- Field, Bonnie. 2009. "Minority Government and Legislative Politics in a Multilevel State: Spain under Zapatero." *South European Society and Politics* 14(4):417–434.
- Field, Bonnie N. 2016. *Why Minority Governments Work. Multilevel Territorial Politics in Spain*. Palgrave Macmillan UK.
- Franck, George. 2019. "The Economy of Attention." *Journal of Sociology* 55(1):8–19.
- Freidenberg, Flavia and Julieta Cao. 2014. *Territorio y Poder: Nuevos actores y competencia política en los sistemas de partidos multinivel en América Latina*. Salamanca: Ediciones Universidad de Salamanca.
- Gentzkow, Matthew and Jesse M Shapiro. 2010. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78(1):35–71.
- Gilardi, Fabrizio, Charles R. Shipan and Bruno Wüest. 2020. "Policy Diffusion: The Issue-Definition State." *American Journal of Political Science* 65(1):21–35.
- Gilens, Martin. 2001. "Political Ignorance and Collective Policy Preferences." *American Political Science Review* 95(2):379–396.
- Graber, D. 2003. "The media and democracy: Beyond myths and stereotypes." *Annual Review of Political Science* 6 (1) pp. 139–160.

- Green-Pedersen, Christoffer and Peter B. Mortensen. 2015. "Avoidance and Engagement: Issue Competition in Multiparty Systems." *Political Studies* 63(4):747–764.
- Green-Pedersen, Christoffer and Rune Stubager. 2010. "The Political Conditionality of Mass Media Influence: When Do Parties Follow Mass Media Attention?" *British Journal of Political Science* 40(3):663–677.
- Green-Pedersen, Christoffer. 2007. "The Growing Importance of Issue Competition: The Changing Nature of Party Competition in Western Europe." *Political Studies* 55(3):607–628.
- Green-Pedersen, Christoffer. 2019. *The Reshaping of West European Party Politics: Agenda-Setting and Party Competition in Comparative Perspective*. Oxford: Oxford University Press.
- Greene, Zac, Andrea Ceron, Gijs Schumacher and Zoltan Fazekas. 2016. "The Nuts and Bolts of Automated Text Analysis. Comparing Different Document Pre-Processing Techniques in Four Countries." *Working Paper OSF Preprint*.
- Greif, Avner and David D Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98(4):633–652.
- Grimmer, Justin and Bramdon Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.
- Guinaudeau, Isabelle and Anna M. Palau. 2016. "A matter of conflict: How events and parties shape the news coverage of EU affairs." *European Union Politics* 17(4):593–615.
- Guntermann, Eric. 2016. Assessing the Influence of Parties on Support for Regional Nationalism in Spain PhD thesis Université de Montréal.
- Hamborg, Felix, Karsten Donnay and Bela Gipp. 2019. "Automated identification of media bias in news articles: an interdisciplinary literature review." *International Journal on Digital Libraries* 20(4):391–415.
- Harbers, Imke. 2010. "Decentralization and the Development of Nationalized Party Systems in New Democracies: Evidence from Latin America." *Comparative Political Studies* 43(5):606–627.
- Harcup, Tony and Deirdre O'Neill. 2017. "What Is News?" *Journalism Studies* 18:12.
- Hayes, Philip J. and Steven P. Weinstein. 1990. "Construe-TIS: A System for Content-Based Indexing of a Database of News Stories." *Association for the Advancement of Artificial Intelligence Proceedings* 90:49–64.

- Helbling, Marc and Anke Tresch. 2011. "Measuring Party Positions and Issue Salience from Media Coverage: Discussing and Cross-validating New Indicators." *Electoral Studies* 30(1):174–183.
- Herman, Edward and Noam Chomsky. 1988. *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.
- Hernández Rodríguez, Rogelio. 2008. *El centro dividido. La nueva autonomía de los gobernadores*. Mexico D.F.: El Colegio de México.
- Herrmann, Michael and Holger Döring. 2021. "Party Positions from Wikipedia Classifications of Party Ideology." *Political Analysis* p. 1–20.
- Hooghe, Liesbet and Gary Marks. 2018. "Cleavage theory meets Europe's crises: Lipset, Rokkan, and the transnational cleavage." *Journal of European Public Policy* 25(1):109–135.
- Hooghe, Liesbet, Gary Marks, Arjan H. Schakel, Sandra Chapman Osterkatz, Sara Niedzwiecki and Sarah Shair-Rosenfield. 2016. *Measuring Regional Authority: A Post-functional Theory of Governance, Volume I*. Oxford: Oxford University Press.
- Hopkin, Jonathan. 2009. "Party Matters: Devolution and Party Politics in Britain and Spain." *Party Politics* 15(2):179–198.
- Hopkins, Daniel J. and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54(1):229–247.
- Hutter, Swen. 2014. Protest Event Analysis and Its Offspring. In *Methodological Practices in Social Movement Research*, ed. Donatella della Porta. Oxford: Oxford University Press.
- Jacobi, Carina, Wouter van Atteveldt and Kasper Welbers. 2016. "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling." *Digital Journalism* 4(1):89–106.
- John, Peter, Anthony Bertelli, Will Jennings and Shaun Bevan. 2013. *Policy Agendas in British Politics*. Basingstoke: Palgrave Macmillan.
- Jones, Bryan D. and Frank Baumgartner. 2005. *The Politics of Attention: How Government Prioritizes Problems*. Chicago: University of Chicago Press.
- Jurado, Ignacio and Sandra León. 2021. "Economic Globalization and Decentralization: A Centrifugal or Centripetal Relationship?" *Governance* 34(3):665–686.
- Kaiser, André. 1997. "Types of Democracy: From Classical to New Institutionalism." *Journal of Theoretical Politics* 9(4):419–444.

- Kaiser, André, Leonce Röth, Lea Kaftan and Daniel Saldivia Gonzatti. 2021a. *Regional Elections Data-Set (RED)*. Version 2021a. University of Cologne.
- Kaiser, André, Leonce Röth, Lea Kaftan and Daniel Saldivia Gonzatti. 2021b. *Regionally Disaggregated Statewide Elections Data-Set (RD—SED)*. Version 2021a. University of Cologne.
- Keating, Michael. 1983. “Decentralization in Mitterand’s France.” *Public Administration* 61(3):237–252.
- Keil, Soeren and Paul Anderson. 2018. Decentralization as a Tool for Conflict Resolution. In *Handbook of Territorial Politics*, ed. Klaus Detterbeck and Eva Hepburn. Cheltenham: Edward Elgar pp. 89–104.
- Keman, Hans. 2007. “Experts and manifestos: Different sources – Same results for comparative research?” *Electoral Studies* 26(1):76–89.
- King, Gary and Langche Zeng. 2001. “Logistic Regression in Rare Events Data.” *Political Analysis* 9(2):137–163.
- King, Gary, Patrick Lam and Margaret Roberts. 2017. “Computer-Assisted Keyword and Document Set Discovery from Unstructured Text.” *American Journal of Political Science* 61(4):971–988.
- Kleinnijenhuis, Jan and Ewald M. Rietberg. 1995. “Parties, Media, the Public and the Economy: Patterns of Societal Agenda-setting.” *European Journal of Political Research* 28:95–118.
- Klesner, Joseph L. 2005. “Electoral Competition and the New Party System in Mexico.” *Latin American Politics and Society* 47(2):103–142.
- Klüver, Heike. 2009. “Measuring Interest Group Influence Using Quantitative Text Analysis.” *European Union Politics* 10(4):535–549.
- Kriesi, Hanspeter, Edgar Grande, Martin Dolezal, Marc Helbling, Dominic Höglinger, Swen Hutter and Bruno Wüest. 2012. *Political Conflict in Western Europe*. Cambridge: Cambridge University Press.
- Lauderdale, Benjamin E. and Alexander Herzog. 2016. “Measuring Political Positions from Legislative Speech.” *Political Analysis* 24(3):374–394.
- Laver, Michael, Kenneth Benoit and John Garry. 2003. “Extracting Policy Positions from Political Texts Using Words as Data.” *American Political Science Review* 97(2):311–331.
- Lehmann, Pola and Malisa Zobel. 2018. “Positions and saliency of immigration in party manifestos: A novel dataset using crowd coding.” *European Journal of Political Research* 57:1056–1083.

- Levi, Margaret. 2004. An Analytic Narrative Approach to Puzzles and Problems. In *Problems and Methods in the Study of Politics*, ed. Ian Shapiro, Roger Smith and Tarek Masoud. Cambridge: Cambridge University Press pp. 201–226.
- Levy, Jonah. 1999. *Tocqueville's Revenge. State, Society, and Economy in Contemporary France*. Cambridge, MA: Harvard University Press.
- León, Sandra. 2006. “The Political Economy of Fiscal Decentralization: Bringing Politics to the Study of Intergovernmental Transfers.” *CEACS: Tesis doctorales* .
- León, Sandra, Ignacio Jurado and Amuitz Garmendia Madariaga. 2018. “Passing the buck? Responsibility Attribution and Cognitive Bias in Multilevel Democracies.” *West European Politics* 41(3):660–682.
- Lipset, Seymour Martin and Stein Rokkan. 1967. *Party systems and voter alignments: Cross-national perspectives*. Vol. 7 Free Press.
- Liu, Ruibo, Chenyan Jia and Soroush Vosoughi. 2021. “A transformer-based framework for neutralizing and reversing the political polarity of news articles.” *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW1):1–26.
- Liñeira, Robert and Daniel Cetrà. 2015. “The Independence Case in Comparative Perspective.” *Political Quarterly* 86(2):257–264.
- Lowe, W. 2016. “Scaling things we can count.” *Working Paper* .
- Lowe, Will, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. “Scaling Policy Preferences from Coded Political Texts.” *Legislative Studies Quarterly* 36(1):123–155.
- Marks, Gary and Liesbet Hooghe. 2000. “Optimality and Authority: A Critique of Neoclassical Theory.” *JCMS: Journal of Common Market Studies* 38(5):795–816.
- Masseti, Emanuele and Arjan H. Schakel. 2013. “Ideology matters: Why Decentralisation has a Differentiated Effect on Regionalist Parties’ Fortunes in Western Democracies.” *European Journal of Political Research* 52(6):797–821.
- Masseti, Emanuele and Arjan H Schakel. 2016. “Between Autonomy and Secession: Decentralization and Regionalist Party Ideological Radicalism.” *Party Politics* 22(1):59–79.
- Masseti, Emanuele and Arjan H Schakel. 2017. “Decentralisation Reforms and Regionalist Parties’ Strength: Accommodation, Empowerment or Both?” *Political Studies* 65(2):432–451.
- Mazzoleni, Gianpietro and Winfried Schulz. 1999. ““Mediatization” of Politics: A Challenge for Democracy.” *Political Communication* 16(3):247–261.

- Mazzoleni, Martino. 2009. "The Saliency of Regionalization in Party Systems: A Comparative Analysis of Regional Decentralization in Party Manifestos." *Party Politics* 15(2):199–218.
- McChesney, Robert W. 2000. "The Political Economy of Communication and the Future of the Field." *Media, Culture Society* 22(1):109–116.
- McCombs, Maxwell and Donald L. Shaw. 1972. "The Agenda-setting Function of Mass Media." *Public Opinion Quarterly* 36(2):176–187.
- Meguid, Bonnie M. 2005. "Competition between Unequals: The Role of Mainstream Party Strategy in Niche Party Success." *American Political Science Review* 99(3):347–359.
- Meguid, Bonnie M. 2008. *Party Competition between Unequals: Strategies and Electoral Fortunes in Western Europe*. Cambridge: Cambridge University Press.
- Meguid, Bonnie M. 2009. "Institutional Change as Strategy: The Role of Decentralization in Party Competition." *APSA Annual Meeting, Toronto* .
- Meguid, Bonnie M. 2015. "Multi-Level Elections and Party Fortunes: The Electoral Impact of Decentralization in Western Europe." *Comparative Politics* 47(4):379–398.
- Midtbø, Tor, Stefaan Walgrave, Peter van Aelst and Dag A. Christensen. 2014. Do the Media Set the Agenda of Parliament or Is It the Other Way Around? Agenda Interactions between MPs and Mass Media. In *Representing the People: A Survey among Members of Statewide and Sub-state Parliaments*., ed. Kris Deschouwer and Sam Depauw. Oxford: Oxford University Press p. 188–209.
- Mitts, Tamar. 2019. "From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West." *American Political Science Review* 113(1):173–194.
- Moureau, Marie-Christine. 2004. "La sociedad PRISA y su estrategia de expansión periodística en las Comunidades Autónomas." *Centros y Periferias* pp. 285–299.
- Mügge, Liza. 2016. "Bridging the qualitative-quantitative divide in comparative migration studies: newspaper data and political ethnography in mixed method research." *Comparative Migration Studies* 4(17).
- Müller, Stefan. 2020. "Media Coverage of Campaign Promises Through the Electoral Cycle." *Political Communication (online first)* .
- Newton, Kenneth and Malcolm Brynin. 2001. "The National Press and Party Voting in the UK." *Political Studies* 49(2):265–285.
- Niblock, Sarah and David Machin. 2007. "News Values for Consumer Groups: The Case of Independent Radio News, London, UK." *Journalism* 8(2):184–204.

- Niedzwiecki, Sara, Sandra Chapman Osterkatz, Liesbet Hooghe and Gary Marks. 2018. "The RAI travels to Latin America: Measuring Regional Authority under Regime Change." *Regional & Federal Studies* 31(2):235–260.
- Oates, Wallace E. 1999. "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37(3):1120–1149.
- O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R Routledge and Noah A Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.
- O'Neill, Kathleen. 2003. "Decentralization as an Electoral Strategy." *Comparative Political Studies* 36(9):1068–1091.
- O'Neill, Kathleen. 2005. *Decentralizing the State: Elections, Parties, and Local Power in the Andes*. Cambridge: Cambridge University Press.
- Orte, Andre and Alex Wilson. 2009. "Multi-level Coalitions and Statute Reform in Spain." *Regional Federal Studies* 19(3):415–436.
- Pallarés, Francesc and Michael Keating. 2003. "Multi-Level Electoral Competition: Regional Elections and Party Systems in Spain." *European Urban and Regional Studies* 10(3):239–255.
- País, El. 2006. "'Zapatero ha cumplido su compromiso'; Interview with Pasqual Maragall by Enric Company." *El País* 2006-01-23 .
- Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40(3):825–850.
- Pierson, Paul. 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94(2):251–267.
- Pogorelis, Robertas, Bart Maddens, Wilfried Swenden and Elodie Fabre. 2005. "Issue salience in regional and national party manifestos in the UK." *West European Politics* 28(5):992–1014.
- Proksch, Sven-Oliver and Johathan B. Slapin. 2015. *The Politics of Parliamentary Debate: Parties, Rebels and Representation*. Cambridge: Cambridge University Press.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2009. "How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany." *German Politics* 18(3):323–344.
- Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle and Stuart Sorokoa. 2019. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* 44(1):97–131.



- Puglisi, Riccardo. 2011. “Being the New York Times: the political behaviour of a newspaper.” *The BE Journal of Economic Analysis Policy* 11:1.
- Radford, Benjamin J. 2021. “Automated Dictionary Generation for Political Event Coding.” *Political Science Research and Methods* 9(1):157–171.
- Rauh, Christian. 2018. “Validating a Sentiment Dictionary for German Political Language—A Workbench Note.” *Journal of Information Technology Politics* 15(4):319–343.
- Rauh, Christian and Jan Schwalbach. 2020. “The ParlSpeech V2 Data Set: Full-text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies.” *Harvard Dataset, V2* 1.
- Remus, Robert, Uwe Quasthoff and Gerhard Heyer. 2010. SentiWS-A Publicly Available German-language Resource for Sentiment Analysis. In *LREC*.
- Rheault, Ludovic and Christopher Cochrane. 2020. “Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora.” *Political Analysis* 28(1):112–133.
- Rice, Douglas R. and Christopher Zorn. 2021. “Corpus-based Dictionaries for Sentiment Analysis of Specialized Vocabularies.” *Political Science Research and Methods* 9(1):20–35.
- Riker, William H. 1980. “Implications from the Disequilibrium of Majority Rule for the Study of Institutions.” *American Political Science Review* 74(2):432–446.
- Riker, William H. 1990. Heresthic and Rhetoric in the Spatial Model. In *Advances in the Spatial Theory of Voting*, ed. James Enelow and Melvin Hinich. Cambridge: Cambridge University Press pp. 46–65.
- Roberts, Margaret, Brandon M. Stewart, Dustin Tingley and Edoardo M. Airolidi. 2013. “The Structural Topic Model and Applied Social Science.” *Conference paper: Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*.
- Roberts, Margaret, Brandon Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. “Structural Topic Models for Open-ended Survey Responses.” *American Journal of Political Science* 58(4):1064–1082.
- Roberts, Margaret, Brandon Stewart and Edoardo Airolidi. 2016. “A Model of Text for Experimentation in the Social Sciences.” *Journal of the American Statistical Association* 111(515):988–1003.
- Robertson, David. 1976. *A Theory of Party Competition*. New York: Wiley.

- Rodríguez, Pedro and Arthur Spirling. 2021. "Word Embeddings. What works, what doesn't, and how to tell the difference for applied research." *Journal of Politics* forthcoming.
- Rokkan, Stein and Peter Flora. 2000. *Staat, Nation und Demokratie in Europa: die Theorie Stein Rokkans aus seinen gesammelten Werken*. Vol. 1473 of *Suhrkamp Taschenbuch Wissenschaft* Suhrkamp.
- Rovny, Jan. 2012. "Who emphasizes and who blurs? Party strategies in multidimensional competition." *European Union Politics* 13(2):269–292.
- Rudkowsky, Elena, Martin Haselmayer, Matthias Wastian, Marcelo Jenny, Štefan Emrich and Micheael Sedlmair. 2018. "More than Bags of Words: Sentiment Analysis with Word Embeddings." *Communication Methods and Measures* 12(2-3):140–157.
- Ruedin, Didier. 2013. "Obtaining Party Positions on Immigration in Switzerland: Comparing Different Methods." *Swiss Political Science Review* 19(1):84–105.
- Ruedin, Didier and Laura Morales. 2019. "Estimating party positions on immigration: Assessing the reliability and validity of different methods." *Party Politics* 25(3):303–314.
- Röth, Leonce. 2017. "Equivalence presupposes validity. Toward comparable party positions on the market dimension." *EPSA Conference Paper*.
- Röth, Leonce and André Kaiser. 2019. "Why accommodate minorities asymmetrically? A theory of ideological authority insulation." *European Journal of Political Research* 58(2):557–581.
- Röth, Leonce, André Kaiser, Çağan Varol and Uğur Sadioğlu. 2016. "Centre-Periphery Conflict and Ideological Distance in Turkey." *Swiss Political Science Review* 22(4):585–607.
- Schakel, Arjan H. 2018. Denationalization of Elections? Tracing the Developments in the Conceptualization and Measurement of the Nationalization of the Vote. In *Handbook of Territorial Politics*, ed. Klaus Detterbeck and Eva Hepburn. Cheltenham: Edward Elgar pp. 107–119.
- Schakel, Arjan H. and Charlie Jeffery. 2013. "Are Regional Elections really 'Second-Order' Elections?" *Regional Studies* 47(3):323–341.
- Schonhardt-Bailey, Cheryl. 2008. "The Congressional Debate on Partial-Birth Abortion: Constitutional Gravitas and Moral Passion." *British Journal of Political Science* 38(3):383–410.
- Schwarzbözl, Tobias, Matthias Fatke and Swen Hutter. 2020. "How party–issue linkages vary between election manifestos and media debates." *West European Politics* 43(4):795–818.

- Shair-Rosenfield, Sarah, Arjan H. Schakel, Sara Niedzwiecki, Gary Marks, Liesbet Hooghe and Sandra Chapman-Osterkatz. 2021. "Language Difference and Regional Authority." *Regional & Federal Studies* 31(1):73–97.
- Shepsle, Kenneth A. 2006. Old Questions and News Answers about Institutions: The Riker Objection Revisited. In *The Oxford Handbook of Political Economy*, ed. Barry R. Weingast and Donald A. Wittman. Oxford: Oxford University Press pp. 1031–1049.
- Slapin, Jonathan and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.
- Sorens, Jason. 2009. "The Partisan Logic of Decentralization in Europe." *Regional & Federal Studies* 19(2):255–272.
- Sorens, Jason. 2018. "The Effects of Decentralization on Minority Territorialist Party Positions on Self-Government and Vote Shares, 1950–2017." *APSA Annual Meeting*.
- Soroka, Stuart. 2002. *Agenda-Setting Dynamics in Canada*. Vancouver: University of British Columbia Press.
- Soroka, Stuart and Christopher Wlezien. 2019. "Tracking the Coverage of Public Policy in Mass Media." *Policy Studies Journal* 47:471–491.
- Spirling, A. 2016. "Democratization and Linguistic Complexity: The Effect of Franchise Extension on Parliamentary Discourse 1832–1915." *The Journal of Politics* 78(1):120–136.
- Spoon, Jae-Jae, Sara B. Hobolt and Catherine E. de Vries. 2014. "Going green: Explaining issue competition on the environment." *European Journal of Political Research* 53:363–380.
- Staab, Joachim F. 1990. "The Role of News Factors in News Selection: A Theoretical Reconsideration." *European Journal of Communication* 5(4):423–443.
- Stepan, Alfred C. 1999. "Federalism and Democracy: Beyond the US Model." *Journal of Democracy* 10(4):19–34.
- Stewart, Charles and Barry R. Weingast. 1992. "Stacking the Senate, Changing the Nation: Republican Rotten Boroughs and American Political Development in the Late 19th Century." *Studies in American Political Development* 6:223–371.
- Strömbäck, Jasper. 2008. "Four Phases of Mediatization: An Analysis of the Mediatization of Politics." *The International Journal of Press/Politics* 13(3):228–246.
- Strömbäck, Jasper, Michael Karlsson and David Hopmann. 2012. "Determinants of News Content: Comparing Journalists' Perceptions of the Normative and Actual Impact of Different Event Properties When Deciding What's News." *Journalism Studies* 13:718–728.

- Swenden, Wilfried and Bart Maddens, eds. 2009. *Territorial Party Politics in Western Europe*. Basingstokes: Palgrave Macmillan.
- Sältzer, Marius. 2020. "Finding the bird's wings: Dimensions of factional conflict on Twitter." *Party Politics* 0(0).
- Thesen, Gunnar. 2013. "When Good News Is Scarce and Bad News Is Good." *European Journal of Political Research* 52:364–389.
- Thorlakson, Lori. 2006. Party Systems in Multi-level Contexts. In *Devolution and Electoral Politics*, ed. Dan Hough and Charlie Jeffery. Manchester: Manchester University Press pp. 37–52.
- Thorlakson, Lori. 2009. "Patterns of Party Integration, Influence and Autonomy in Seven Federations." *Party Politics* 15(2):157–177.
- Toubeau, Simon. 2017. "Restructuring the State: Mainstream Responses to Regional Nationalism." *Publius: The Journal of Federalism* 48(1):76–101.
- Toubeau, Simon and Emanuele Massetti. 2013. "The Party Politics of Territorial Reforms in Europe." *West European Politics* 36(2):297–316.
- Toubeau, Simon and Markus Wagner. 2015. "Explaining party positions on decentralization." *European Journal of Political Research* 55:340–357.
- Toubeau, Simon and Markus Wagner. 2016. "Party Competition over Decentralisation: The Influence of Ideology and Electoral Incentives on Issue Emphasis." *European Journal of Political Research* 55(2):340–357.
- Treisman, Daniel. 2007. *The Architecture of Government: Rethinking Political Decentralization*. Cambridge: Cambridge University Press.
- Tsebelis, George. 1990. *Nested Games: Rational Choice in Comparative Politics*. Oakland: University of California Press.
- van Aelst, Peter and Stefaan Walgrave. 2011. "Minimal or Massive? The Political Agenda-Setting Power of the Mass Media According to Different Methods." *The International Journal of Press/Politics* 16(3):295–313.
- van Atteveltdt, Wouter, Jan Kleinnijenhuis and Nel Ruigrok. 2008. "Parsing, Semantic Networks, and Political Authority Using Syntactic Analysis to Extract Semantic Relations from Dutch Newspaper Articles." *Political Analysis* 16(4):428–446.
- van Atteveltdt, Wouter, Mariken van der Velden and Mark Boukes. 2021. "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms." *Communication Methods and Measures* 15(2):121–140.

- van Biezen, Ingrid and Jonathan Hopkin. 2006. Party Organization in Multi-level Contexts. In *Devolution and Electoral Politics*, ed. Dan Hough and Charlie Jeffery. Manchester: Manchester University Press pp. 14–36.
- van der Brug, Wouter and Joost Berkhout. 2015. “The Effect of Associative Issue Ownership on Parties’ Presence in the News Media.” *West European Politics* 38(4):869–887.
- van der Pas, D. and R. Vliegenthart. 2016. “Do Media Respond to Party Conflict? Debates of European Integration in British, Dutch and German Party Manifestos and Newspapers 1987–2006.” *Political Studies* 64(2):260–278.
- van der Velden, M., G. Schumacher and B. Vis. 2017. “Living in the Past or Living in the Future? Analyzing Parties’ Platform Change In Between Elections, The Netherlands 1997–2014.” *Political Communication* 35(3):393–412.
- van Noije, Lonneke, Jan Kleinnijenhuis and Dirk Oegema. 2008. “Loss of Parliamentary Control Due to Mediatization and Europeanization: A Longitudinal and Cross-Sectional Analysis of Agenda Building in the United Kingdom and the Netherlands.” *British Journal of Political Science* 38(3):455–478.
- Verge, Tània. 2013. “Party Strategies on Territorial Reform: State-wide Parties and the State of Autonomies in Spain.” *West European Politics* 36(2):317–337.
- Vliegenthart, Rens and Noemi M. Montes. 2014. “How Political and Media System Characteristics Moderate Interactions between Newspapers and Parliaments: Economic Crisis Attention in Spain and the Netherlands.” *International Journal of Press/Politics* 19(3):318–339.
- Vliegenthart, Rens and Stefaan Walgrave. 2011. “When the Media Matter for Politics: Partisan Moderators of the Mass Media’s Agenda-Setting Influence on Parliament in Belgium.” *Party Politics* 17(3):321–342.
- Vliegenthart, Rens, Stefaan Walgrave, Frank R. Baumgartner, Breunig Christian Bevan, Shaun, Sylvain Brouard, Laura Chaqués-Bonafon, Emiliano Grossman, Will Jennings, Peter B. Mortensen, Anna M Palau, Sciarini Pascal and Anke Tresch. 2016. “Do the Media Set the Parliamentary Agenda? A Comparative Study in Seven Countries.” *European Journal of Political Research* 55(2):283–301.
- Volgens, Andrea, Tobias Burst, Werner Krause, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Regel, Bernhard Weßels and Lisa Zehnter. 2020. “The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR) Version 2020.” *Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)*.
- Walgrave, Stefaan. 2008. “Again, the Almighty Mass Media? The Media’s Political Agenda-Setting Power According to Politicians and Journalists in Belgium.” *Political Communication* 25(4):445–459.

- Walgrave, Stefaan and Peter van Aelst. 2006. "The Contingency of the Mass Media's Political Agenda Setting Power: Toward a Preliminary Theory." *Journal of Communication* 56:88–109.
- Welbers, Kasper, Wouter van Atteveldt, Jan Kleinnijenhuis, Nel Ruigrok and Joep Schaper. 2016. "News selection criteria in the digital age: Professional norms versus online audience metrics." *Journalism* 17(8):1037–1053.
- Welbers, Kasper, Wouter van Atteveldt and Kenneth Benoit. 2017. "Text Analysis in R." *Communication Methods and Measures* 11(4):245–265.
- Wrátil, Christopher and Sara B. Hobolt. 2019. "Public deliberations in the Council of the European Union: Introducing and validating DICEU." *European Union Politics* 20(3):511–531.
- Young, Lori and Stuart Soroka. 2012. "Affective news: The automated coding of sentiment in political texts." *Political Communication* 29(2):205–231.
- Zhou, Xiaojin and Andrew B. Goldberg. 2009. "Introduction to Semi-Supervised Learning." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 3(1):1–130.
- Zuber, Christina Isabel. 2011. "Understanding the Multinational Game: Toward a Theory of Asymmetrical Federalism." *Comparative Political Studies* 44(5):546–571.
- Zuber, Christina Isabel and Edina Szöcsik. 2021. "Does Decentralization Turn Minority Parties into Secessionists? Insights from Eastern and Western Europe." *West European Politics* 44(1):825–851.

## **Eidesstattliche Erklärung**

### **nach § 8 Abs. 3 der Promotionsordnung vom 17.02.2015**

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich (zutreffendes unterstreichen) geholfen:

Weitere Personen, neben den ggf. in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Ort, Datum

Unterschrift

# Daniel Saldivia Gonzatti

Research Fellow at the Center for Civil Society Research, WZB Berlin  
Reichpietschufer 50, D-10785 Berlin - Room D 025  
+49 30 25491 182 | daniel.saldivia-gonzatti@wzb.eu | Website | github.com/DaSalGon

## CURRENT POSITION

---

**Research Fellow** 2021 – present  
*Center for Civil Society Research, WZB Berlin Social Science Center* Berlin, GER

- Project: Political Protest and Radicalization, WZB Protest Monitoring & MOTRA  
PIs: Edgar Grande, Swen Hutter
- Research Areas: Protest, Radicalization, Public Discourses, Content and Text Analysis, Political Cleavages

## EDUCATION

---

**PhD Fellowship & Doctoral Studies in Comparative Politics** 2017 – 2021 (planned)  
*Cologne Center for Comparative Politics & Cologne Graduate School, University of Cologne* Cologne, GER

- Supervision: André Kaiser, Sven-Oliver Proksch
- Dissertation Title: Statewide Parties' Perspectives in the Long-Term Dynamics of Decentralization
- Research Topics: Comparative Political Institutions, Party Politics, Quantitative Text Analysis, Territorial Politics, Parliaments in Presidentialism

**Master of Arts in Political Science; Minor: Energy Economics** 2014 – 2017  
*University of Cologne* Cologne, GER

- Master Thesis: Comparing Electoral Behaviour in Direct Democratic Decisions and Elections of Representatives. Supervision: André Kaiser

**Research Residence Abroad** 2016  
*Universidad Nacional de San Martín* Buenos Aires, ARG

- DAAD-Research Grant of the Global South Studies Center – Cologne
- Project Area: Social Movements and Struggles for Participation

**Bachelor of Arts in Political Science, Economics, and Economic History** 2011 – 2014  
*RWTH Aachen University* Aachen, GER

- Bachelor Thesis: Local Autonomy and Centralisation in Venezuela

**Erasmus Semester Abroad** 2013  
*Université Jean Moulin, Lyon 3* Lyon, FRA

- ERASMUS Scholarship – Fields of Study: Politics, History, and Economics

## EXPERIENCE

---

**Research Assistant** 2018 – 2020  
*Cologne Center for Comparative Politics, University of Cologne* Cologne, GER

- DFG Project „Decentralization and Electoral Geographies II“ (KA 1741/10-2)  
Project Direction: André Kaiser, Leonce Röth  
Research Areas: Research Conceptualization, Text-As-Data Analysis, Data Management
- Lecturer of *The Political System of Germany (BA Level)* with Lea Kaftan
- Bachelor Theses Supervision in Comparative Politics



## Student Assistant

2015 – 2017

*Cologne Center for Comparative Politics, University of Cologne*

*Cologne, GER*

- DFG Project „Decentralization and Electoral Geographies“ (KA 1741/10-1)  
Project direction: André Kaiser, Leonce Röth

*Chair of European and Multilevel Politics (Interim), University of Cologne*

- Research and Data Collection for the Project *Parlamentarisation Process of the Parlasur* (2017).  
PI: Dr. Aleksandra Maatsch
- Bachelor tutor of the lecture *The Political System of the EU*. Lecturer: Aleksandra Maatsch (2016)

## University Intern

2013

*Central Bank of Venezuela - BCV*

*Caracas, VEN*

- Internship Unit: Department for Macroeconomic Analysis
- Report and Presentation Preparations on Topics of Fiscal and Monetary Economics

## TEACHING

---

<b>Análisis de Texto con R</b>   <i>Laboratorio Nacional de Políticas Públicas – CIDE, Mexico DF</i>	10/2020
• Text Analysis with R, Methods Workshop - in Spanish, Online	
<b>Introduction to R (Workshop)</b>   <i>University of Cologne</i>	10/2020
• Software Workshop at the Cologne Center for Comparative Politics, UoC	
• Introductory Workshop for Students of MA Political Science – Online	
<b>Academic Supervision of Bachelor Theses</b>   <i>University of Cologne</i>	04/2018 – 02/2021
• Subject: Comparative Politics	
<b>The Political System of Germany (BA Level)</b>   <i>University of Cologne</i>	10/2018 – 03/2019
• Lecturer; together with Lea Kaftan	
<b>Europa weiterdenken in NRW (High School Level)</b>   <i>Friedrich Ebert Stiftung</i>	01/2018 – 04/2018
• Project Teacher in NRW High Schools; FES-NRW	
<b>The Political System of the EU (BA Level)</b>   <i>University of Cologne</i>	04/2016 – 09/2016
• Lecture Tutor; Lecturer: Aleksandra Maatsch	

## WORKING PAPERS

---

### Mediated Party Positions from Newspapers

*Co-authored with Lea Kaftan and Leonce Röth*

**Statewide and Regionalist Parties' Perspectives in the Long-Term Dynamics of Decentralization**  
(Original Title) — *Submitted*

**Introducing Topic Models and Sentiment Analysis for the Identification of Salience and Party Positions in Territorial Debates** (Original Title) — *Under Review*

*Co-authored with Lea Kaftan and Leonce Röth*

**Legislative Debates in Presidential Systems. Government-Opposition Dynamics in MPs' Speeches in Latin America**

*Co-authored with Jan Schwalbach*

## DATA

---

**Regionally disaggregated state-wide elections data set - (RD|SED). Version 2021a.**

**University of Cologne**

*Co-authored with Leonce Röth, André Kaiser and Lea Kaftan*

Forthcoming

**Regional elections data set - (RED). Version 2021a. University of Cologne**

*Co-authored with Leonce Röth, André Kaiser and Lea Kaftan*

Forthcoming

## CONFERENCE CONTRIBUTIONS

---

**EPSA, ECPR, APSA, Puebla-BUAP Government & Methods Seminar**

2020

**CES Madrid, ECPR Wrocław, Zurich Text as Data Workshop**

2019

**GraPa Düsseldorf**

2018

## GRANTS

---

**DAAD-Research Grant for Doctoral Stay Abroad**

*Department of Comparative Politics, University of Bergen (Cancelled due to Covid-19)*

2020

**Cologne Graduate School PhD Scholarship**

*University of Cologne*

2017 – 2021

**DAAD-Research Grant for Stay Abroad (MA Level)**

*Program: Global South Studies Center, University of Cologne*

*Host: Universidad Nacional de San Martín, Buenos Aires*

2016

**Erasmus Scholarship for Bachelor Studies Abroad**

*Host: Université Jean Moulin, Lyon 3, France*

2016

## METHODOLOGICAL FORMATION (SELECTION)

---

**Quantitative Text Analysis**

*Sven-Oliver Proksch (University of Cologne), University of Cologne*

2017/2018

**Topic Models**

*Molly Roberts (UC-Los Angeles), University of Cologne*

2017/2018

**Collecting and Analyzing Social Media Data**

*Pablo Barberá (UC-Southern California), University of Cologne*

2017/2018

**Multilevel Analysis**

*Ian Brunton-Smith (University of Surrey), GESIS Cologne*

2018

**Experiments and Causal Inference**

*Christopher Wratil (University College London), University of Cologne*

2018

**Web Survey Design**

*Mick Couper (University of Michigan), GESIS Cologne*

2018

**Panel Data Analysis**

*Andrew Li (Central European University), ECPR Summer School*

2018

**Web Scraping with R**

*Simon Munzert (Hertie School), University of Cologne*

2018

**Comparative Historical Analysis**

*Marcus Kreuzer (Vilanova University), University of Cologne*

2019

**Machine Learning for the Social Sciences**

*Bruno Castanho Silva (University of Cologne), GESIS Cologne*

2020

## From NLP to CSS: Practical Guide to Transformers

*Christopher Kamm (TU Darmstadt), Moritz Laurer (CEPS)*  
*& Elliott Ash (ETH Zürich), International Conference on CSS*

2021

### SKILLS

---

**Languages:** Spanish (Native), German (Fluent), English (Fluent), French (Intermediate - B1)

**Software:** MS-Office (Advanced Knowledge), R (Advanced Knowledge), Stata (Advanced Knowledge),  
L<sup>A</sup>T<sub>E</sub>X (Advanced Knowledge)



Updated September 2021