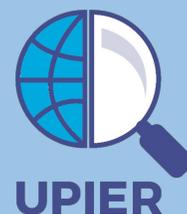# USES OF THE PAST IN INTERNATIONAL ECONOMIC RELATIONS

UPIER

YOU SHALL KNOW A COUNTRY BY THE COMPANY IT KEEPS: EXPLORING SPAIN'S COLLOCATES IN DIE ZEIT, 1946 - 2017

Elisa Garrido Moreno

Foreword

UPIER Working Papers series reflects the work in progress of the researchers associated with the HERA-funded project Uses of the Past in International Economic Relations and of others whose papers directly address UPIER research themes. The papers are peer reviewed by UPIER and associated researchers and seek to advance our understanding of how the past has been constructed and used in international economic relations over the past 200 years.

The views expressed in this working paper, and all errors and omissions, should be regarded as those solely of the authors and are not necessarily the views of the affiliated institutions.

For more information on UPIER visit: www.upier.web.ox.ac.uk

Author

Elisa Garrido Moreno is a Postdoctoral Fellow at University Carlos III Madrid. She holds a PhD in Art Theory and History. Her research interests include also cultural history, visual culture and digital humanities (http://uc3m.academia.edu/ElisaGarrido).

Contact email address: elisa.garrido@uc3m.es

Abstract

In the years of the Eurozone crisis the countries hardest hit by banking distress, rising unemployment and looming crises in sovereign debt, namely, Portugal, Italy, Ireland, Greece and Spain, were often identified in the international press as the PIIGS group – an extension of PIGS – the acronym widely used since the 1990s to identify Southern European "problem" economies. The use of both of these acronyms revealed how an isolated term could serve to condense thinking regarding the economic vulnerabilities and risk factors common to peripheral economies in the Euro area. Both acronyms are examples of country grouping heuristics: cognitive shortcuts which categorize countries into peer groups with (alleged) common characteristics. Such country groupings may be connected to the social construction of ideas, as templates, that enhance the interpretation of economic and political realities and therefore guide the choices and the decisions of actors (whether investors, voters, or policymakers). In this paper, I discuss to what extent the idea of a "European South", as a grouping heuristic, gradually emerged from the perception of Southern countries as correlates on multiple dimensions for a long period of their history. To explore this issue, I draw on recent research in collocation analysis (a sub-field of NLP Natural Language Processing) which uses DiaCollo (Kollokationsanalyse in Diachroner Perspektive, an open-source text-mining software to tackle large, diachronic corpora of German texts). Here, I explore how the statistical association of the word "Spanien" with its primary collocates has evolved over time by studying the entire collection of articles published in the German newspaper, *Die Zeit,* between 1946 and 2017. I find that a strong association of Spain with Southern European countries has been an ongoing characteristic of articles in the press for more than sixty years. Over time, this association has developed different dimensions that emphasize the common political, institutional, social and economic development of Southern European countries.

Introduction

In the years of the Eurozone crisis, European countries hardest hit by banking distress, rising unemployment and looming sovereign debt (Portugal, Italy, Ireland, Greece and Spain) were often identified in the international press as the PIIGS group. This acronym aimed to condense, in a single term, the notion of economic vulnerability and risk factors common to peripheral economies of the Euro area. PIIGS was an extension of PIGS, an earlier acronym for four Southern countries. This had come into widespread use in the press in the mid-1990s, during the harsh negotiations over the terms and conditions for the accession to the EMU (Economic and Monetary Union), at a time when serious concerns about the capacity of countries with a recent history of macroeconomic instability to participate in the single currency were being publicly expressed by policy-makers and analysts. Apparently, the initial of the term - so the story goes - was a Wall Street article published on November 6, 1996. However, the idea of a Southern European country cluster (occasionally including France) with a pejorative connotation – e.g. the popular concept of "Club Med" countries – had been around since the early 1990s and had been used in the context of the process of monetary integration, as well as other EU policy issues (Brazys and Hardiman 2015). In fact, concerns that "Latin monetary views" (with reference to France and other Southern countries) could prevail at the expense of Germany's culture of stability had been expressed by top Bundesbank officials as early as 1988 (James 2012: 235). In the following years, the "battle of ideas" on the Euro brought to light deeper Franco-German differences rooted in divergent historical experiences as well as differences in culture, political economy and economic traditions (Brunnermeier, James and Landau 2016).

The acronym PIGS is an example of country grouping heuristic—a cognitive shortcut based on the categorization of countries into peer groups with (alleged) common characteristics. Country grouping has come to the attention of researchers interested in the social construction of ideas which provide a framework for understanding the world. This literature sees ideas working as templates that enhance the interpretation of economic and political realities, and may possibly influence the choices and decisions made by actors—whether investors, voters or policymakers (Brazys and Hardiman 2015).

Unlike other well-known acronyms invented by business and investment analysts (such as BRICS) the term PIGS has special characteristics that make its symbolic power especially strong. First, it conveys a judgement of value with a clear derogatory bias. As Fourcade (2013) explains: "Who would you rather

put your money on – the BRICs or the PIGS? The terms (which evocate, respectively, a sturdy material and a filthy porcine) are not irrelevant here: we think and feel through language… [T]he economy is always and everywhere a morality play". In fact, the use of this terminology hit a nerve and was often criticized by commentators in the stigmatized countries [1] to the extent that its use was actually discouraged – with little success, apparently – in the international press as well as among business and financial analysts.[2]

Second, PIGS terminology reflects the perception of the South from a northern point of view; reinforcing the idea of a cultural divide in Europe along geographic lines. This version of the "clash of cultures" approach shifts the focus from diverging policies to diverging histories and identities. Indeed, on occasion, the former have been treated as symptoms of the latter. As a former German member of the ECB board wrote recently: "Different economic points of view stem from historical experience and cultural peculiarities", pointing explicitly to the (typically German?) value of individual responsibility as the key determinant of a fundamental divergence with other European countries.[3] Therefore the use of the past in the public discourse on the Eurozone crisis has been gradually drifting towards the realm of morality with all its burden of perceptions and identities – or stereotypes, even on especially unfortunate occasions.[4]

---

[1] In the case of Spain see: F. González Laxe, "Entre los BRIC y los PIGS", *El País*, 30/08/2008: https://elpais.com/diario/2008/08/30/galicia/1220091497_850215.html; S. Pozzi, "El Banco Mundial estigmatiza a los 'PIGS' con una definición errónea", *El País*, 10/06/2010: https://elpais.com/diario/2010/06/10/economia/1276120806_850215.html; J. Soler, "Elogio de los PIIGS", *El País*, 24/06/2012: https://elpais.com/elpais/2012/06/22/opinion/1340367240_720933.html .

[2] In 2010, the Financial Times ("Anything but porcine at BarCap", Feb 5, 2010, https://ftalphaville.ft.com/2010/02/05/142451/anything-but-porcine-at-barcap/) reported the following memo circulated by Barclays Capital, a British bank, to its employees: "Please alert your teams not to use the acronym PIIGS in any written communication. Rather, they should spell out the acronym and say: Portugal, Italy, Ireland, Greece and Spain. Research Production globally have been informed to take out any reference to the acronym in question."

[3] J. Stark, "The historical and cultural differences that divide Europe's union", Financial Times, February 2, 2015: https://www.ft.com/content/e08ec622-ad28-11e4-a5c1-00144feab7de .

[4] In March 2017, the Dutch minister of Finance Jeroen Dijsselbloem, then head of the Eurogroup (the group of Eurozone's ministries of finance), came under crossfire from Southern European governments and EU authorities for declaring in an interview to the German newspaper Frankfurter Allgemeine Zeitung: "Ich kann nicht mein ganzes Geld für Schnaps und Frauen ausgeben und anschließend Sie um Ihre Unterstützung bitten. Dieses Prinzip gilt auf persönlicher, lokaler, nationaler und eben auch auf europäischer Ebene" – a statement interpreted in the international press as an indirect but clear criticism to Greece and other Southern European countries. In his defence, he claimed that he merely wanted to illustrate a moral principle. https://www.faz.net/aktuell/wirtschaft/wirtschaftspolitik/nach-aeusserung-in-der-f-a-z-ruecktritt-von-eurogruppenchef-dijsselbloem-gefordert-14937323.html

In this paper, I aim to discuss to what extent the idea of a "European South" grouping (beyond its obvious geographical connotation) can be regarded as a social construct that gradually emerged from history. The intuition I'm pursuing here is, basically, that the likelihood of the idea of a country cluster emerging will be higher, and its heuristic power will be stronger, where countries are perceived as correlates on multiple dimensions, whether geographical, geo-political, political, institutional, or cultural, for a long period of their history. To explore this issue, I draw on recent research advances in NLP (Natural Language Processing). More precisely, I adopt a text mining approach based on collocations: pairs of words that co-occur at high frequency in a meaningful context. In the framework of European integration and EU policy-making, the most critical voices on Europe's South came from Germany. Accordingly, I apply this methodology to a text corpus composed of the entire collection of articles published in the German magazine, *Die Zeit,* from 1946 to 2017. The strategy is to explore diachronically which word pairs (collocates) typically occurred together with Spain – the S in PIGS – in order to assess whether, when and why Spain was systematically associated with other European countries in a way that could convey to German readers the idea of a "Southern club". To paraphrase the famous statement by English linguist J.R. Firth ("You shall know a word by the company it keeps"). It is plausible to assume that a systematic association of certain countries in the news might act over time as a heuristic device. In turn, this could be based either on availability (judging the likelihood of an event based on how easily an example or a case comes to mind), representativeness (judging an event based on how strongly it resembles other events stored in memory) or affect (making judgments about an event based on sentiment from past experiences. See: Kahneman and Tversky 1972; Tverski and Kahneman 1974; Kahneman and Frederick 2002)[5].

My main findings suggest that the strong association of Spain with other Southern European countries was an ongoing characteristic of German news features for the entire second half of the 20th century and the first decades of the 21st century. Over time, this association developed in different ways, by emphasizing the common political, institutional, social and economic development of Southern European countries. This may have encouraged German readers to perceive the European South as a single entity – a group of countries with common characteristics that made them fit together. Such "commonalities" may have contributed to the emergence of the notion of PIGS as an "economic problem" area that arose in the 1990s and was revived in the post-2008 crisis. At the same time, this study shows

---

[5] These heuristic devices are used to explain investors' decision in the literature on behavioral finance.

that the association of Ireland with the European South emerged well before the outbreak of the recent crisis.

The paper is structured as follows. In the first section I briefly discuss some aspects of the extensive literature on the North-South cultural divide in Europe. Sections II and III describe the text mining tools, the text sources and the methodology of country collocation analysis. Section IV present the results. Section C presents the conclusion.

I.   The North-South divide in Europe: religion, culture or ideology?

According to Elliot (2000), the roots of the political and social differences between northern and southern European countries began in the late sixteenth century during the conflict between the Protestant North and the Catholic South. In fact, the role of religion as a key factor in the analysis of economic and social development dates back to the analysis of Max Weber (1930 [1905]), who argued that Protestants were more inclined to succeed in business than Catholics thanks to their superior work ethic, thrift and social capital (see also Barro and McCleary 2003). Although economists as a rule tend to reject cultural explanations of comparative economic performances, the post-2008 Eurozone crisis has revived the interest in Weber's approach. Some analysts emphasize cultural and social characteristics that seem to have influenced the development of the economy and empowered the dominance of the Protestant part of Europe. Such factors could include: the secularization and consequent freedom of the economy from religious control, differences in education (self-education by reading the Bible autonomously), the consequences of the Catholic Counter-Reformation and the importance of the Atlantic (slave) trade in creating an autonomous business class that would demand modernizing institutional reform (Young 2009; Sheremeta and Vernon 2017). More generally, the notion that cultural and social differences between the "two Europes" might help to explain systematic differences in economic and policy outcomes – for instance, in the characteristics and performance of welfare states – has been gaining traction in the policy debate in Europe and elsewhere.[6]

---

[6] See for example T. Sanandaji, "The America left's two-Europes problem", 2012, http://www.aei.org/publication/the-american-lefts-two-europes-problem/

The assumption that European countries emphasize different cultural values, and can therefore be grouped into relatively homogenous "cultural clusters", that are rooted in the past, retains significant traction in the social sciences. In Hofstede and Minkov's comparative graphs, which rate cultural values such as 'power distance', 'uncertainty avoidance', 'long-term orientation', and so on, Spain's graph matches that of other Southern European countries, and is sharply distinguished from Germany's (Hofstede et al. 2010; Figure 1). Likewise, in the GLOBE Project (https://globeproject.com) -- a global research initiative that explores the connections between national cultural values, societal effectiveness and organizational leadership – finds that Italy, Spain and Portugal, together with France, French-speaking Switzerland and Israel, are strongly correlated along a number of the identified dimensions of societal culture (some of which are derived from Hofstede's modeling), supporting the view of a Latin Europe cluster (House et al 2004; see Figure 2).

Figure 1: PIGS in Hofstede's dimensions on national culture
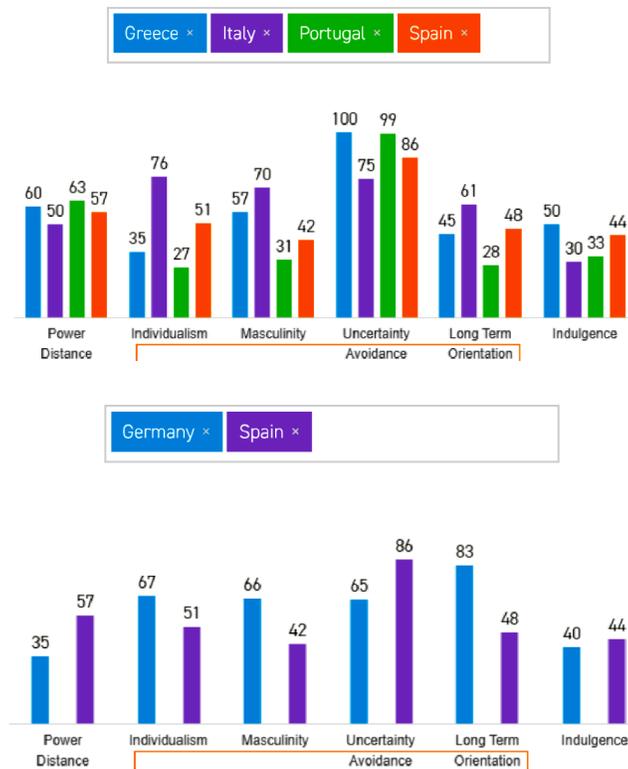
Source: Hofstede Insights (https://www.hofstede-insights.com/product/compare-countries/)

Figure 2: GLOBE's Latin Europe cluster
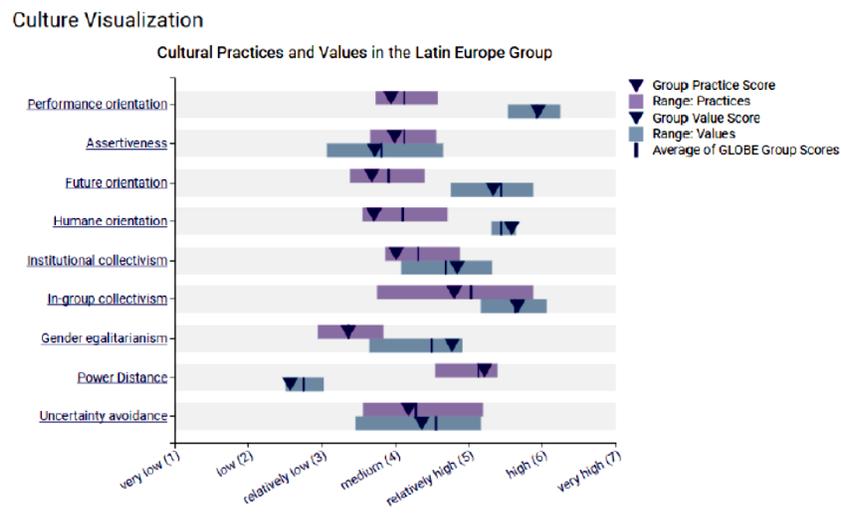
Source: The Global Project (https://globeproject.com/results/clusters/latin-europe)

If, in the social sciences, "national cultures" are treated as objective, measurable entities, other approaches emphasize their ideological nature. The influential constructivist school, inspired by Gellner (1983) and Anderson (1983), identified fundamental changes in the political and socio-economic environment of 18th and 19th century Europe, when, as they see it, identity politics became the principal instrument used by the state to secure its legitimacy. By extension, the desire to belong to a unified cultural entity works as a tool for exclusion, based on making clear distinctions between oneself and others, in-group and out-group (Habermas 1998). For Anderson, nations historically developed as *imagined communities*: fictive projections motivated by the concern to be identified as belonging to the same nation, where identity is reinforced in opposition to others, whether along the polarities of North-South, center-periphery, ours-theirs, or rational-irrational, and so on.

Following a similar thread, Dainotto (2007) offers a historical analysis of the evolution of ideas about Europe that continue to define culture, politics, and identity today. He sees the origins of the North-versus-South divide as an ideological relocation within Europe of a West-versus-East divide, which led to a representation of Southern national cultures as irrational and rude in stark comparison to those of the rational, civic-minded nations depicted as the cradle of "true" European values. Dainotto also suggests that the creation of a borderless Europe – a process officially set in motion on March 25, 1995, with the entry into force of the Schengen Treaty for seven members of the European Union (Germany,

France, Belgium, Netherlands, Luxembourg, Spain and Portugal), soon followed by the rest of the EU-15 countries by the end of 1997 – was not only an economic issue but a social phenomenon that, by enhancing the interaction between so different cultures, exacerbated the perception of the profound differences between the North and the South.

One way or another, the European past – either reflected in measurable cultural dimensions, or projected through ideological representations – is unlikely to work as a ground for a common identity. At the Copenhagen European Summit of December 1973, the Heads of State or Government of the nine Member States of the enlarged European Community (Germany, Belgium, France, Italy, Luxemburg, the Netherlands, and the recently added Denmark, Ireland and United Kingdom) affirmed their determination to introduce the concept of European identity into their common foreign relations.[7] Since then, however, EU cultural policy has shifted from the initial ambition to build an identity based on "unity", emphasizing common heritage, traditions and values to a minimalist approach based on "diversity". The latest ambiguous variant of this is the concept of "unity in diversity", thus displacing the construction of a common ground from the past to the future – a "community of destiny" (Sassatelli 2002). This suggests that "the historical-geographical and socio-political characteristics of Europe as a continent [that] have generated a 'multiple identity area' of overlapping territorial and historical spaces at local, regional and national territorial level" (Walkenhorst 2008) make the building of Europe as a new "imagined community" a daunting task. In fact, the recent economic crisis and the ensuing harsh political controversies exacerbated a sentiment of deeply rooted diversities that seems to challenge the idea of a common destiny. As suggested by Alfred Korzybski, the founder of general semantics, "the map is not the territory" – and much less so in Europe in the last decade.

II.   Tools and sources: using DiaCollo to explore *Die Zeit*

In this paper, I aim to discuss the degree to which the association of Spain with other Southern European countries can be thought of as a social construct that emerged from history. The intuition is basically that the probability of a country cluster emerging will be higher, and its heuristic power will be stronger where

---

[7] See "Declaration on European Identity", *Bulletin of the European Communities*. No 12, Luxembourg: Office for official publications of the European Communities (December 1973): 118-122.

countries are associated on multiple dimensions (geographical, geo-political, political, institutional, cultural) for a long period of their history. To explore this issue, I draw on recent research advances in NLP (Natural Language Processing), a scientific method combining the fields of artificial intelligence and linguistics. This approach is increasingly used by scholars as data sharing becomes increasingly important in modern scientific research. DiaCollo-Kollokationsanalyse in Diachroner Perspektive (Collocation analysis in diachronic perspective: https://www.clarin.eu/showcase/diacollo) is an open-source text-mining software hosted by CLARIN-D (https://www.clarin-d.net/en/), the German partner of CLARIN ("Common Language Resources and Technology Infrastructure": https://www.clarin.eu/), a Europe-wide research infrastructure that promotes the sharing, use and sustainability of digital language data for research in humanities and social sciences. Text mining is a subfield of data mining, which is, in common with NLP, a subfield of artificial intelligence. This interdisciplinary field combines machine learning, statistics and computational linguistics. Text mining technology permits the categorization, extraction and sentiment analysis of large text corpora.[8] Collocation analysis is one of many tasks in statistical NLP. It consists of a target term (*collocant*) and its *collocates* – i.e. words associated to it, either grammatically or symbolically, as a means of identifying "some conventional way of saying things" (Manning and Schütze 1999).

DiaCollo allows researchers to measure and visualize the strength of typical co-occurrences between a *collocant* and its *collocates* in a diachronic perspective. Unlike most conventional NLP tools, which implicitly assume corpus homogeneity (even when they have temporal extension), DiaCollo's can accommodate temporal heterogeneity (i.e. it can measure variations in the strength of co-occurrences over time) thus helping researchers to understand the historical context from which they derive. For this reason, it can be used to analyze semantic shifts, for example, the evolution of word usage, or discourse trends, which reflect changes in the discourse environment (Jurish 2015). DiaCollo gives researchers the possibility to see these semantic shifts by tracing pair words: words which appear together at a certain period in time. In this way, changes in meaning will be indexed by changes in word combinations that may reveal broader political or cultural changes as well. Over time, the relevance of some connections will fade, while other connections gain salience.

In order to be able to track the change of a word and its collocates over a long period of time, large corpora with detailed bibliographic metadata (including occurrence dates) must be available. The

---

[8] The term *corpus* refers to a collection of data to be processed and we use the plural *corpora* to speak about more than one corpus.

most important historical corpus archived and searchable through DiaCollo is the *Deutsches Textarchiv* (http://www.deutschestextarchiv.de/), a large collection of German texts published in different fields (the sciences, the arts, and newspapers) between 1650 and 1900. Moreover, DiaCollo can be used also to explore more recent newspaper corpora such as the *Berliner Zeitung* and *Tagespsiegel* for the period from the mid-1990s to 2005) and the entire collection of articles published in the *Die Zeit* (1946-2017). Since the latter is the only corpus covering the entire period, from WW2 to the recent crisis, this is the source used in this paper.

Die Zeit is a liberal-oriented weekly newspaper published in Hamburg. It has a wide national circulation. This has remained stable, at roughly 450, 000 to 500, 000 issues per week, from the 1980s until the present. It was one of the first German newspapers to be licensed by the British after the War and it played an important role in the resurrection of democracy in West Germany (Humphreys 1996: 82). *Die Zeit* publishes dossiers and essays, as well as long, detailed articles, and multi-authored reports where different points of view are expressed on a single topic. Unlike the other widely-read German magazine, *Der Spiegel*, whose style is based on investigative research, anonymous writing and provocative opinion pieces, *Die Zeit* privileges the analysis of individual authors with a clearly identifiable signature in order to convey the balance between political camps (https://www.goethe.de/ins/gb/en/kul/mag/20704049.html make this reference clearer and move to footnote). For this reason, it is seen to be one of the most influential newspapers to shape Germany's public opinion and social communication. Among its editors and publishers we find distinguished intellectuals and political personalities. Marion Dönhoff (1909-2002), one of the most important intellectuals of post-war Germany, joined the magazine as political editor in 1946, becoming editor-in-chief in 1968 and publisher in 1972 – a role she kept until her death in 2002, aged 92. A bestselling author and an advocate of the reconciliation between East and West, she contributed to set the moral standards for the coexistence of the peoples in a united Europe (Harpprecht 2008). Another influential personality in *Die Zeit* was Helmut Schmidt, a distinguished leader of the SPD (the Social-Democratic Party) who served as Chancellor of the Centre-Left government coalition from 1974 to 1982. A pioneer of international economic cooperation, he became Editor in 1983 and collaborated with the newspaper until his death in 2015 (Soell 2014).

III.   Method: country collocation analysis

Country grouping heuristics are based on the premise that within-cluster countries are associated with each other more intensely than they are with countries outside the cluster. Here, I use the DiaCollo software to measure the time-varying statistical association between Spain, as a "collocant" country, and its country collocates. It is important to stress that DiaCollo uses only content words as potential collocates; co-occurrences with high-frequency non-content items (e.g. articles) are automatically filtered out (Jurish 2018). In DiaCollo the query can be specified for different "profiles"; a profile is defined on the base of the break units into which the text corpus is segmented. The "collocations" profile searches for collocant-collocate pairs within sentence boundaries, using a fixed-width moving "context window" of a maximum of 5 adjacent words; a collocation pair is included in the results only if it passes a minimum frequency threshold (set at 5 co-occurrences by default). However, this "native co-occurence" approach tends to miss out sparse data – i.e. infrequent terms for which very few potential collocates can be identified. In order to include a broader range of candidate collocates, DiaCollo includes also a "TDF" (Term x document frequency) profile, which allows users to search for co-occurrences at "document" level; by default, documents are identified by paragraph boundaries.

To measure the strength of co-occurrences, DiaCollo makes it possible to use different scoring functions, which include absolute raw- and log-frequency and normalized raw- and log-frequency per million tokens. The most widely used score, however, is the log Dice coefficient, a "lexicographer-friendly association score" that allows meaningful comparisons between corpora of a very different size (i.e. number of tokens included) (Rychlý 2008: 6-9). It is defined as:

$$\text{score}_{\text{ld}}(w_1, w_2) = 14 + \log_2 \frac{2(f_{12} + \varepsilon)}{(f_1 + \varepsilon) + (f_2 + \varepsilon)}$$

where w1 and w2 are (respectively) the collocant (query) and collocate tuples (i.e. sequences of content words); f1 and f2 are (respectively) the total independent number of co-occurrences for the query term (w1,*) and the collocate term (*,w2); f12 is the number of co-occurrences for the collocation-pair (w1,w2); and ε is an error.

For each query DiaCollo gives results for each time-slice specified: the total number of co-occurrences (for any word included in the corpus), the f1, f2 and f12 co-occurrence statistics (see above), the selected score for the top k collocates and their attributes (POS, i.e. Part-of-Speech: noun, proper name, adjective, verb). Results can be displayed by using different formats of online visualization – e.g., as an HTML table or a chart (both with hyperlinks to underlying corpus hits), a tab-separated plain text, bubbles or clouds (in which collocates are displayed as labelled circles or text labels, respectively, whose size and color represent the corresponding association score).

In the example below (Figure 3), DiaCollo is instructed to return collocations (PROFILE) for the target term "Spanien" (QUERY) and its 20 strongest collocates (KBEST) on the base of the log Dice score (SCORE) aggregated at intervals of ten years (SLICE) for the entire period covered by the text corpus (the DATE(S) box is left blank). Figure 4 shows results for the first time-slice (1940 = 1946-49) in HTML format; KWIC (Keyword in context) boxes are the hyperlinks to the underlying text hits; an example for the collocate word "Beziehung" (relationship) is presented in figure 5. Figure 6 shows the same results in Cloud and Bubble formats, in which the strongest associations (ranked along the side bar) are in red and the weakest ones in blue.

Figure 3: DiaCollo interface

Figure 4: Results in HTML Table format

N	f1	f2	f12	score	label	lemma	pos		
15657298	2717	2717	18	6.7621	1940	Spanien	NE	KWIC	
15657298	2717	983	11	6.6061	1940	Franco	NE	KWIC	
15657298	2717	766	7	6.0412	1940	Portugal	NE	KWIC	
15657298	2717	6543	16	5.8232	1940	Beziehung	NN	KWIC	
15657298	2717	3130	10	5.8084	1940	Marshall-Plan	NN	KWIC	
15657298	2717	18658	33	5.6608	1940	Frankreich	NE	KWIC	
15657298	2717	669	5	5.5966	1940	Abbruch	NN	KWIC	
15657298	2717	2048	7	5.5891	1940	spanisch	ADJA	KWIC	
15657298	2717	5144	11	5.5189	1940	Uno	NN	KWIC	
15657298	2717	2471	7	5.4664	1940	Argentinien	NE	KWIC	
15657298	2717	1065	5	5.4370	1940	Guthaben	NN	KWIC	
15657298	2717	2181	5	5.0640	1940	Beteiligung	NN	KWIC	
15657298	2717	2487	5	4.9765	1940	diplomatisch	ADJA	KWIC	
15657298	2717	3260	5	4.7767	1940	arabisch	ADJA	KWIC	
15657298	2717	3543	5	4.7100	1940	Griechenland	NE	KWIC	
15657298	2717	8050	8	4.6057	1940	Italien	NE	KWIC	
15657298	2717	11421	9	4.3826	1940	erklären	VVFIN	KWIC	
15657298	2717	6814	6	4.3666	1940	Großbritannien	NE	KWIC	
15657298	2717	16176	10	4.1164	1940	USA	NE	KWIC	
15657298	2717	49406	26	4.0308	1940	Land	NN	KWIC	

Figure 5: Hyperlinks to text hits from KWIC boxes

Figure 6: Results in Cloud and Bubble formats
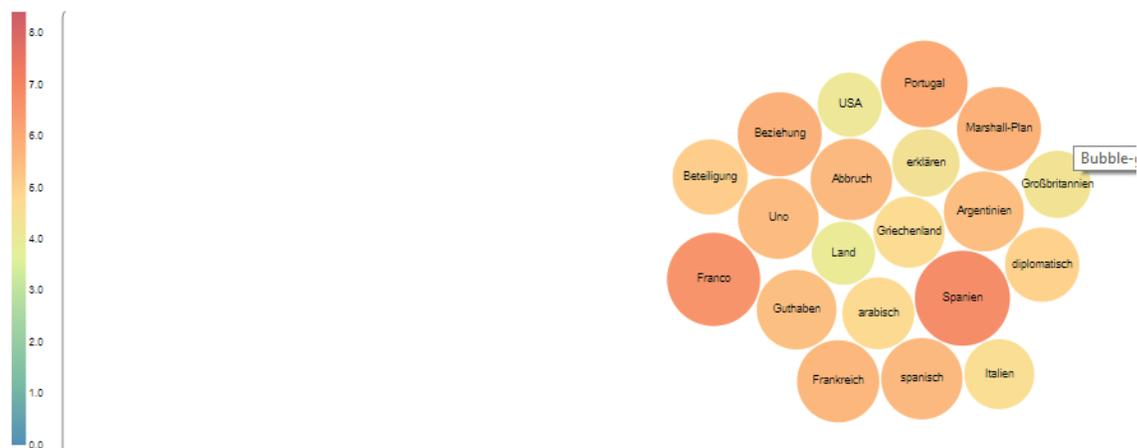
IV.   Results: Spain's collocates over time

A first relevant finding is how "visible" Spain was as a news-worthy subject in *Die Zeit* compared to other foreign countries. For this purpose, I use the DiaCollo profile "unigrams". Results are presented in Figure 7. The rate of presence is calculated as the number of times the name of each country appears each year, divided by the total number of unigrams (i.e. content words) included in the corpus in the same year. Since the total number of times "Spanien" appears is not an appropriate measure of its visibilty, it needs to be normalised. For reasons possibly related to the number and size of published articles, there has been a considerable increase in the number of unigrams in recent years. The top graph shows that before 2010 Spain was featured less in the German compared to the USA, the Soviet Union or other European countries such as France, Britain or Italy. Mild spikes can be observed in the early 1960s and the mid 1970s; these were associated, respectively, with the rise of Spanish emigration to Germany and the fall of the Franco regime and the transition to democracy. Measured on this scale, the relevance acquired by Spain after 2010 as a consequence of its role in the Eurozone crisis was definitely unprecedented.

Figure 7: Rate of presence of foreign countries in *Die Zeit*, 1946-2017
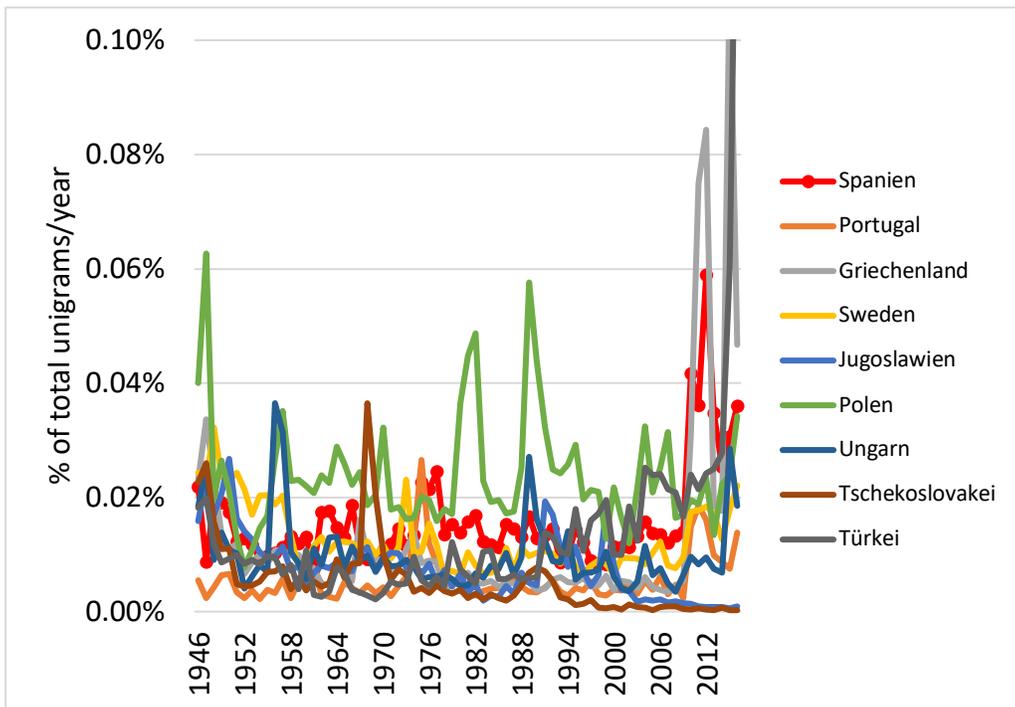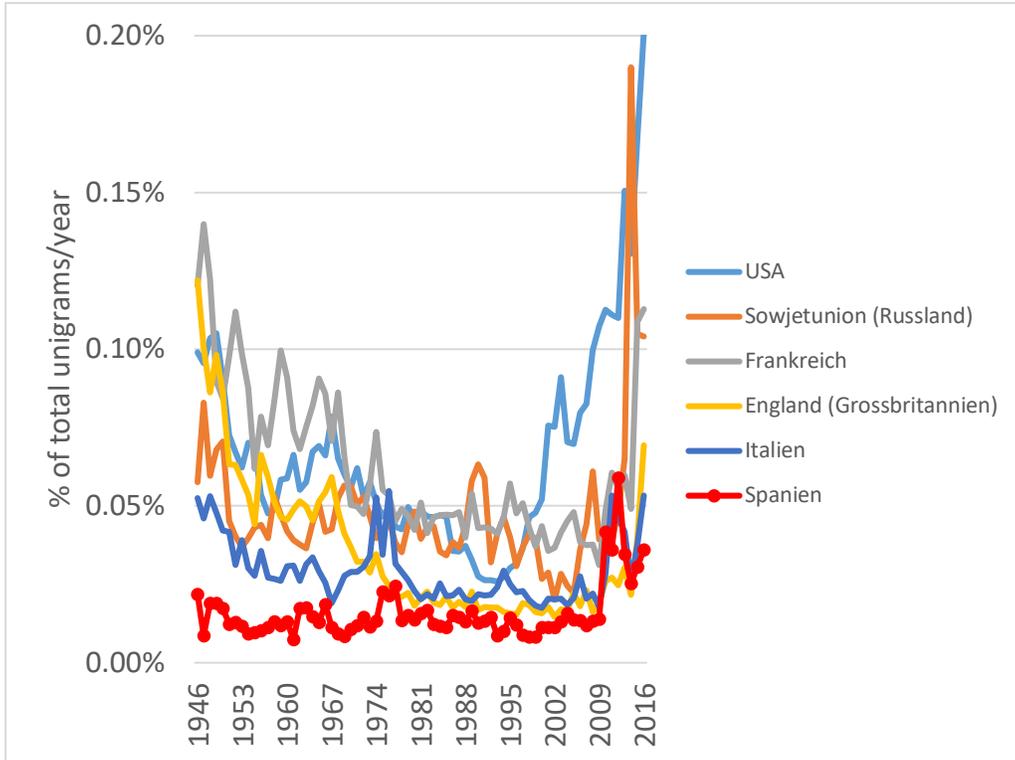
However, compared to other European peripheral countries, the presence of Spain in the German news was not negligible. Given Germany's geopolitics as well as the traditional eastward

projection of German foreign political and economic relationships, Poland was the most "visible" country in this group. However, Spain was "seen" in the news by readers of *Die Zeit* much more than any other peripheral country in the North, East or South of Europe.

As a second step, by using the DiaCollo profiles "collocations" and "term-document matrix", we can assess which words were more strongly associated with Spain in *Die Zeit*. Tables 1 and 2 report Spain's top 20 collocates by decade from the 1940s onwards ranked on the base of the log Dice score, while Figure 8 shows annual scores for the four Southern European countries. It emerges clearly that for *Die Zeit's* readers Spain was often portrayed in close association with at least one of the other Southern European countries: Portugal, Italy and Greece – the PIG group. Until the late 1990s, the strongest association was with Portugal, which can be explained not only by the geographical contiguity of the two countries, but, more importantly, by their parallel political development: both remained under autocratic regimes until the mid-1970s; both experienced the transition to democracy in the late 1970s; and, both applied for accession to the European Economic Community in 1986. This is suggested by the presence of the words "*beitritt*" (accession) and "*gemeinschaft*" (community) among the top collocates in the 1980s. Political developments could also explain the strong association of Spain with Greece and Italy in the 1960s and 70s – namely, in the case of Greece, the establishment of a military regime (1967-74), the return to democracy and its accession to the EEC in 1981; in the case of Italy, the strong electoral success of the communist party in the 1970s and early 1980s, which led to the rise of a common political strategy (jointly with French and Spanish communists) under the flag of Eurocommunism.

**Table 1. Profile "collocations": Spain's newsworthy top-20 collocates by decades**

1946-49		1950-59		1960-69		1970-79	
6,762	*Spanien*	7,096	Italien	7,789	Portugal	7,785	Portugal
6,606	Franco	7,054	Portugal	7,386	Italien	7,692	Griechenland
6,041	Portugal	5,920	Schweiz	6,966	Griechenland	7,035	Italien
5,823	Beziehung	5,896	Frankreich	6,431	Jugoslawien	6,750	Jugoslawien
5,808	Marshall-Plan	5,676	Schweden	6,242	Schweiz	6,307	Frankreich
5,661	Frankreich	5,515	*Spanien*	5,836	Österreich	6,071	Türkei
5,597	Abbruch	5,441	Griechenland	5,732	Franco	5,908	Franco
5,589	*spanisch*	5,281	Österreich	5,672	Frankreich	5,453	Schweiz
5,519	Uno	4,831	Holland	5,636	Ahlers	5,424	Österreich
5,466	Argentinien	4,759	Jugoslawien	5,188	Türkei	5,315	*spanisch*
5,437	Guthaben	4,656	*spanisch*	5,094	*Spanien*	5,207	England
5,064	Beteiligung	4,150	England	5,090	Madrid	4,601	Madrid
4,977	diplomatisch	4,009	Marokko	4,775	England	4,509	Schweden
4,777	arabisch	3,997	Türkei	4,565	*spanisch*	4,492	Großbritannien
4,710	Griechenland	3,479	Ungarn	4,065	Großbritannien	4,425	Marokko
4,606	Italien	3,474	Belgien	4,039	Land	4,421	baskisch
4,383	erklären	3,364	Verteidigung	3,982	Gibraltar	4,256	Carrillo
4,367	Großbritannien	3,263	Großbritannien	3,952	Belgien	4,239	demokratisch
4,116	USA	3,229	USA	3,935	Argentinien	4,121	Demokratie
4,031	Land	3,118	Polen	3,582	Tourist	4,048	Arbeiterpartei
1980-89		**1990-99**		**2000-09**		**2010-17**	
8,419	Portugal	8,331	Portugal	7,912	Italien	8,292	Italien
7,226	Italien	7,908	Italien	7,695	Portugal	7,720	Portugal
7,214	Griechenland	7,303	Griechenland	7,279	Frankreich	6,937	Frankreich
6,430	Frankreich	6,975	Frankreich	7,017	Griechenland	6,776	Griechenland
6,359	Beitritt	6,499	Großbritannien	6,664	Großbritannien	6,671	Madrid
5,564	Großbritannien	6,439	Irland	6,430	Irland	6,518	Europameister
5,395	*spanisch*	6,280	Belgien	6,148	England	6,477	Irland
5,353	Franco	5,529	England	5,673	Polen	6,273	Titelverteidiger
5,142	Österreich	4,337	Deutschland	5,590	Belgien	6,166	England
4,975	Gemeinschaft	4,152	Franco	5,476	José	5,909	Nadal
4,924	Belgien	4,042	González	5,446	Zapatero	5,846	Rafael
4,917	Jugoslawien	3,877	Türkei	5,406	Aznar	5,759	Lesen
4,813	*Spanien*	3,857	*Spanien*	5,235	Österreich	5,729	Großbritannien
4,743	Schweiz	3,780	Österreich	5,086	Schweden	5,700	Niederlande
4,734	Eg	3,742	Schweden	4,730	Marokko	5,645	Weltmeister
4,733	England	3,691	Holland	4,660	Deutschland	5,637	Rajoy
4,733	Madrid	3,654	Europa	4,626	Türkei	5,593	welt-
4,504	Eg-Beitritt	3,641	Dänemark	4,576	Dänemark	5,590	Belgien
4,477	Irland	3,578	Land	4,507	Baskenland	5,464	Thema
4,315	Weltmeisterschaf	3,448	*spanisch*	4,442	*Spanien*	5,458	Mariano

Table 2. Profile "Term-Document Matrix": Spain's newsworthy top-20 collocates by decades

1946-49		1950-59		1960-69		1970-79	
14,000	*Spanien*	*14,000*	*Spanien*	*14,000*	*Spanien*	*14,000*	*Spanien*
12,088	*spanisch*	*11,743*	*spanisch*	*11,603*	*spanisch*	*11,487*	*spanisch*
11,860	Franco	11,115	Franco	11,131	Franco	11,033	**Portugal**
11,231	Madrid	10,752	Madrid	10,882	**Portugal**	10,959	Madrid
10,459	**Portugal**	10,679	**Portugal**	10,845	Madrid	10,888	**Griechenland**
10,019	Bürgerkrieg	*10,560*	*Spanier*	10,659	**Italien**	10,805	Franco
10,016	*Spanier*	10,450	**Italien**	*10,441*	*Spanier*	*10,340*	*Spanier*
9,872	Uno	9,957	**Griechenland**	10,235	**Griechenland**	10,291	**Italien**
9,700	**Italien**	9,944	Schweiz	9,771	**Jugoslawien**	10,088	**Jugoslawien**
9,693	iberisch	9,799	Bürgerkrieg	9,770	Schweiz	9,771	Juan
9,678	Juan	9,776	Belgien	9,750	Österreich	9,764	Frankreich
9,675	Don	9,666	Holland	9,703	Bürgerkrieg	9,642	Carlos
9,608	diplomatisch	9,536	Schweden	9,478	Barcelona	9,485	Marokko
9,515	Barcelona	9,477	Frankreich	9,421	Gibraltar	9,406	Österreich
9,464	Botschafter	9,465	**Türkei**	9,317	**Türkei**	9,398	**Türkei**
9,463	Frankreich	9,458	**Jugoslawien**	9,268	Frankreich	9,390	Bürgerkrieg
9,451	Argentinien	9,387	Österreich	9,164	Belgien	9,333	Barcelona
9,353	NordAfrika	9,351	Dänemark	9,126	Del	9,160	baskisch
9,333	Beziehung	9,215	Marokko	9,045	Dänemark	9,117	Schweiz
9,215	Marshall-Plan	9,195	Gibraltar	9,037	Marokko	9,018	König
1980-89		1990-99		2000-09		2010-17	
14,000	*Spanien*	*14,000*	*Spanien*	*14,000*	*Spanien*	*14,000*	*Spanien*
11,675	*spanisch*	11,296	**Portugal**	*11,110*	*spanisch*	11,418	**Italien**
11,365	**Portugal**	*11,113*	*spanisch*	10,884	**Italien**	11,012	Madrid
10,935	Madrid	10,955	**Italien**	10,659	**Portugal**	*10,883*	*spanisch*
10,505	*Spanier*	10,412	**Griechenland**	10,476	Madrid	10,877	**Portugal**
10,485	**Griechenland**	10,266	Madrid	10,432	Frankreich	10,381	Frankreich
10,429	**Italien**	10,230	Frankreich	10,128	**Griechenland**	10,163	**Griechenland**
10,231	Franco	*9,920*	*Spanier*	*10,068*	*Spanier*	9,885	**Irland**
9,994	Felipe	9,886	González	10,043	Zapatero	9,878	Barcelona
9,895	Frankreich	9,822	Großbritannien	10,034	Aznar	9,853	Europameister
9,790	Beitritt	9,814	Belgien	9,915	Großbritannien	9,662	Rajoy
9,492	iberisch	9,797	Felipe	9,825	José	9,644	Titelverteidiger
9,443	Gemeinschaft	9,751	**Irland**	9,614	**Irland**	9,584	England
9,439	Barcelona	9,440	Barcelona	9,530	Belgien	*9,563*	*Spanier*
9,418	Eg	9,366	Franco	9,396	Eta	9,478	Staatsanleihe
9,390	Sozialist	9,342	Aznar	9,347	England	9,406	Weltmeister
9,289	González	9,177	England	9,293	baskisch	9,345	Niederlande
9,271	Großbritannien	9,170	Sevilla	9,249	Franco	9,293	Mariano
9,269	Belgien	9,044	Dänemark	9,180	Luis	9,282	Real
9,256	Jugoslawien	9,041	Sozialist	9,131	Schweden	9,223	Halbfinale

**Figure 8: Annual log Dice score for PIIG countries**
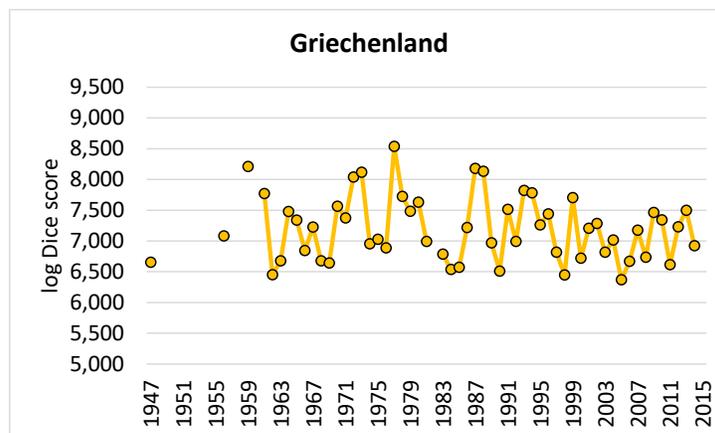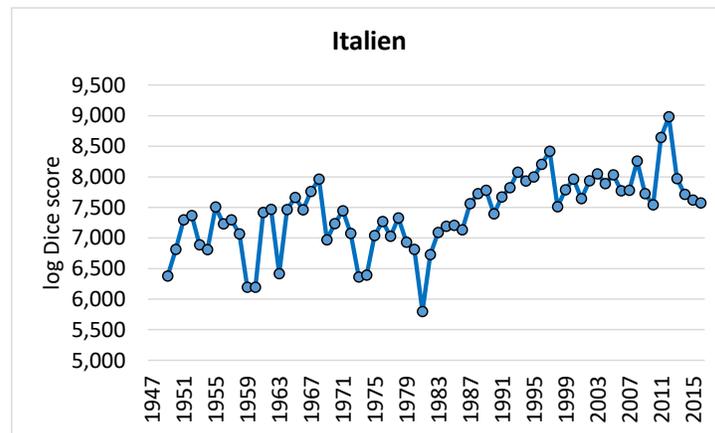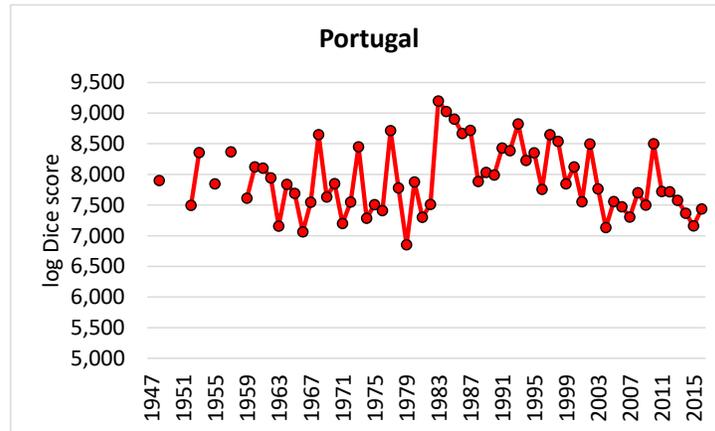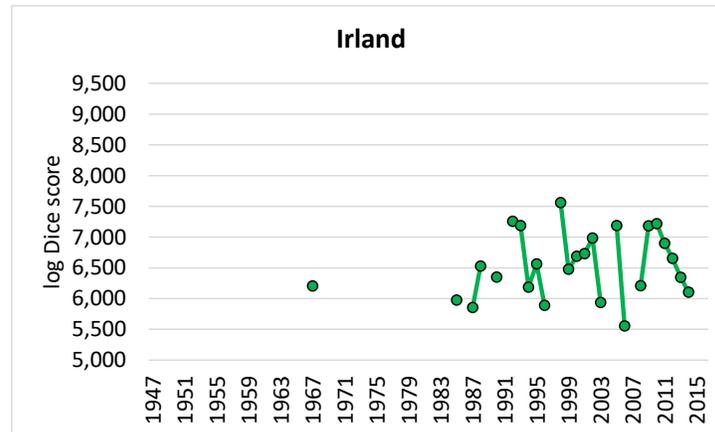
However, Italy's very high score and the rising scores of Greece, Turkey and Yugoslavia after the 1950s suggests two additional channels of association. The first is related to the rise of German tourism abroad during this period. While, in the 1950s, the preferred destinations of German travellers were neighboring countries, such as Austria and Switzerland, from the 1960s Spain, Italy and other southern destinations gained increasing prominence (Wohlmann 1993). Thanks to the expansion of organized tourism as well as travel and hotel infrastructure, Spain and other southern European countries quickly turned themselves from a "blank spot" (*weißer Fleck*) to an increasingly popular holiday destination for West Germans (Kopper 2007). The second factor is the mass recruitment of foreign workers through bilateral agreements between West Germany and Italy (1955), Spain and Greece (1960), Turkey (1961), Portugal (1964) and Yugoslavia (1968). These worker migrants were known as *Gastarbeiter* ('guest workers'), an expression that precisely defined their status within the German society. From the point of view of most of the German population, they were useful visitors to fill temporary gaps in the German labour market; they were invited to work under temporary contracts and expected to return home afterwards, in spite of their fundamental contribution to the German 'economic miracle' (Höhne et al 2014; Hübschmann 2015). Whilst tourism and emigration enhanced West Germans' awareness of the cultural diversity in the South, it may simultaneously have enhanced the perception of the South of Europe as a unified whole.

While links based on tourism and emigration seem to lose relevance after the 1970s, Ireland emerges as a leading collocate in the 1990s. This coincides chronologically with the expansion of the EU budget for its cohesion policy (the so-called "structural funds"), whose main beneficiaries were in fact Portugal, Greece, Ireland, most Spanish regions and the Southern Italian regions (Figure 9). The common characteristic emphasized here was their relative economic underdevelopment – technically defined as a GDP per capita below 75 per cent of the EU average. On the other hand, Spain is increasingly associated

with leading European countries (France, Britain, Germany), which may be related to its increasing relevance in EU policy-making – i.e. in the reform of member countries' voting weights in the Council of the EU, which were the focus of the harsh negotiations leading to the Treaty of Nice (2001) and Lisbon (2007).

Fig. 9 Beneficiaries of EU Structural Funds 1989-1999
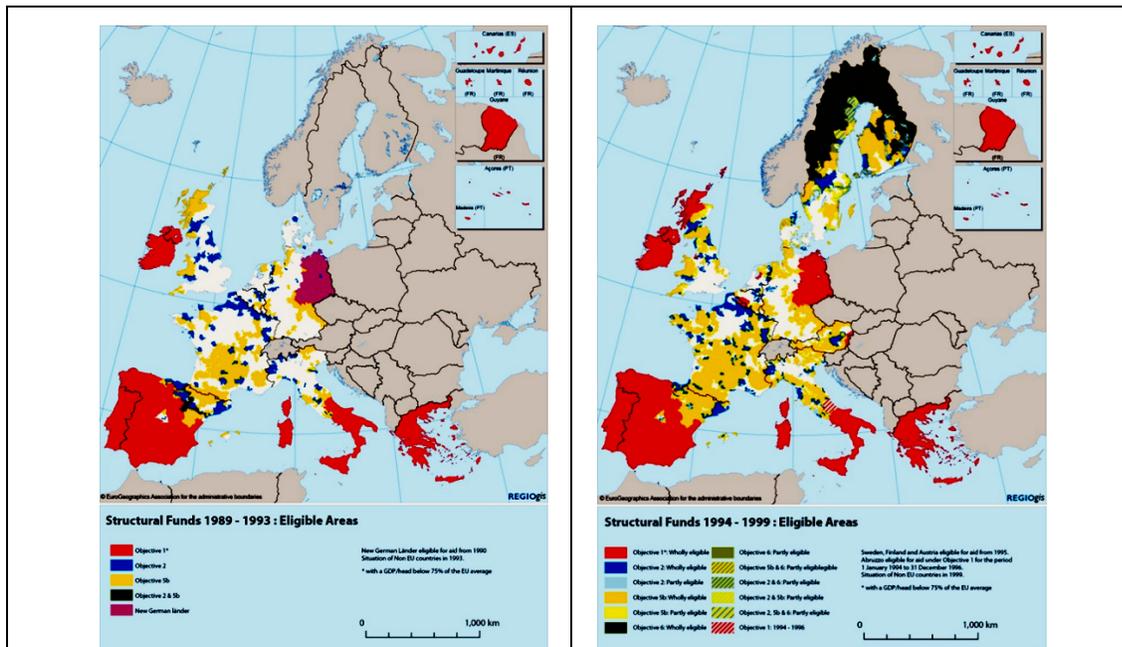
Conclusions and future research

This analysis of diachronic country collocations suggests that the strong association of Spain with Southern European countries was a permanent characteristic of German news for the entire second half of the 20th century and the first decades of the 21st. Over time, this association worked along different dimensions, emphasizing common political, institutional, social and economic development of Southern European countries. This may have led German readers to see the European South as a whole – a group of countries with common characteristics that made them fit together. These "commonalities" might

have provided a fertile ground for the emergence of the notion of PIGS as "economic problem" countries that arose in the 1990s and was revived in the post-2008 crisis. At the same time, it shows that the association of Ireland with the European South emerged well before the outbreak of the recent crisis.

NLP and text mining are research tools with considerable potential, notwithstanding their limitations. In any case, findings must be understood as incomplete and partial results that will always be susceptible to change. This holds for the findings presented in this paper, where the information obtained from analyzing the German newspaper, *Die Zeit*, indicates how the perception of northern European readers of the European South evolved over time. The statistical association of countries (collocations) can be useful way of visualizing ideas and concepts, but obviously contextual information and complementary research will be needed to develop this analysis further.

Bibliographical references

Anderson B. (1983). *Imagined Communities* (London: Verso)

Barro R.J., McCleary R. (2003). "Religion and economic growth across countries", American Sociological Review, 65 (5): 760-781

Brazys S., Hardiman N. (2015). "From 'Tiger' to 'PIIGS': Ireland and the use of heuristics in comparative political economy", *European Journal of Political Research*, 54: 23-42

Brunnermeier M., James H., Landau J-P. (2016). *The Euro and the Battle of Ideas* (Princeton NJ: Princeton University Press)

Dainotto R.M. (2007). *Europe (in Theory)* (Durham NC: Duke University Press).

Elliott J. (2000). *Europe Divided: 1559 – 1598* (London: Wiley-Blackwell).

Fourcade M. (2013). "The material and symbolic construction of the BRICs", *Review of International Political Economy* 20 (2): 256–267

Gellner E. (1983). *Nations and Nationalism* (Oxford: Blackwell).

Habermas J. (1998). *The Inclusion of the Other* (London: Polity Press).

Harpprecht K. (2008), *Die Gräfin: Marion Dönhoff, eine Biographie* (Reinbek bei Hamburg: Rowohlt Verlag, 2008).

Hofstede G., Hofstede G.J., Minkov M. (2010). *Cultures and Organizations. Software of the Mind* (New York: McGraw Hill)

Höhne J., Linden B., Seils E., Wiebel A. (2014), "Die Gastarbeiter. Geschichte und aktuelle soziale Lage", WSI Report, n. 16: https://www.boeckler.de/pdf/p_wsi_report_16_2014.pdf

House R. et al (eds.) (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies* (Thousand Oaks CA: Sage Publishing)

Hübschmann Z. (2015). "Migrant Integration Programs: The Case of Germany", *Global Migration Research Paper*, No. 11 (2015)

Humphreys P. (1996). *Mass Media and Media Policy in Western Europe* (Manchester University Press, 1996), 82

James H. (2012). *Making the European Monetary Union* (Cambridge MA: Harvard University Press)

Jurish B. (2015). "DiaCollo: on the trail of diachronic collocations", Proceedings of the CLARIN Annual Conference 2015. Book of Abstracts (Wroclaw), 28–31.

Jurish B. (2018). ""Diachronic Collocations, Genre, and DiaCollo", in R. J. Whitt (ed.), *Diachronic Corpora, Genre, and Language Change* (Amsterdam: John Benjamins)

Kahneman D., Frederick S. (2002). "Representativeness revisited: Attribute substitution in intuitive judgment", in T. Gilovich, D. Griffin, and D. Kahneman (eds.), *Heuristics of intuitive judgment: Extensions and applications* (Cambridge: Cambridge University Press), pp. 49-81.

Kahneman D., Tversky A. (1972). "Subjective probability: A judgment of representativeness", *Cognitive Psychology*, 3, 430-454

Kopper C. (2007). "Die Reise als Ware. Die Bedeutung der Pauschalreise für den westdeutschen Massentourismus nach 1945". *Zeithistorische Forschungen/Studies in Contemporary History*, 4, pp. 61-83

Manning C., Schütze H. (1999), *Foundations of Statistical Natural Language Processing*, (Cambridge MA: MIT Press).

Rychlý P. (2008). "A lexicographer-friendly association score", in P. Sojka, A. Horák (eds.), Proceedings of Recent Advances in Slavonic Natural Language Processing (Karlova Studánka: RASLAN)

Sassatelli M. (2002) "Imagined Europe. The Shaping of a European Cultural Identity through EU Cultural Policy", *European Journal of Social Theory*, 5(4): 435–451

Sheremeta R.M., Smith V.L. (2017). "The Impact of the Reformation on the Economic Development of Western Europe". Available at SSRN: https://ssrn.com/abstract=3192660 or http://dx.doi.org/10.2139/ssrn.3192660

Soell H. (2014), *Helmut Schmidt: Pioneer of International Economic and Financial Cooperation* (New York: Springer)

Tversky A., Kahneman D. (1974). "Judgment under uncertainty: Heuristics and biases", *Science* (New Series), 185, 1124-1131

Walkenhorst H. (2008). "Constructing the European identity - trap or gap? European integration between community-building and path-dependency", Limerick Papers in Politics and Public Administration, No. 1, (2008): 10.

Weber M. (1930 [1905]). *The Protestant Ethic and the Spirit of Capitalism*, (New York: Routledge)

Wohlmann R. (1993), "Entwicklung des Tourismus 1954-1991", in H. Hahn, H.J. Kagelmann (eds.), *Tourismuspsychologie und Tourismussoziologie* (München 1993)

Young C. (2009), "Religion and Economic Growth in Western Europe: 1500-2000", (Paper presented at the annual meeting of the American Sociological Association, San Francisco).