Talking about EU: European Identity and the Publication Office of the European Union

LUCA PARESCHI

luca.pareschi@unibo.it

University of Bologna

Abstract

In this paper we tackle the concept of European identity and the way in which it is communicated by the Publication Office of the European Union. We downloaded the publications of the Publication Office labelled with 'European Identity', and we analysed them through Topic Modelling, eliciting six topics that constitute those documents. We analysed these topics against the different dimensions of European Identity, as described by literature: the civic component of the European identity is partly represented by a topic dealing with the rule of law and National and European identities. The cultural dimension of the European identity characterises analysed documents to a wider extent, as three topics deal with common history, heritage, and culture. In particular, a topic connecting European identity with common cultural values and education is the more used in our sample of documents, and the one which is characterising more and more these documents over time. We argue that this trend signals a particular emphasis by the European Union in identifying European Identity with shared cultural values that can be taught.

1. Introduction

The debate on the concept of European identity intensified over the last 15 years, questioning the drivers that built this shared identity, its relationship with geographical borders and its process of social construction (Caporaso and Kim, 2009; Chacha, 2012; Risse, 2014). In particular, literature still lacks a clear definition (Capello, 2017), but European identity certainly entails two issues: first, it has to be shared among a community; second, this identity is socially constructed (Capello, 2017). Indeed, as a consequence of a shared European identity, a community accepts a similarity that generates solidarity and permits to construct an 'other'. This identity is socially constructed, and it emerges as "the intentional and unintentional consequences of social interactions" (Fligstein *et al.*, 2012: 108). Literature described different drivers that can generate this solidarity: European identity can stem from the acknowledgement of common history and values (Bruter, 2003), or through similar political values, that permits the identification of citizen with a particular institution and its values (Capello, 2017), although national political values can be at odds with European ones (Schilde, 2014). The relationship between European, national, and regional identities is particularly relevant, as citizens may or may not share a sense of European identity, exactly as they can perceive a national or a regional identity. Different scholars tackled the relationship between these territorial (Capello, 2017) identities: we know that these identities are not necessarily in

2.   Method and data

We adopted an inductive approach, as it allows for the emergence of unexpected results, and as it is suited for analysing new phenomena (; Strauss and Corbin, 1990; Eisenhardt and Graebner, 2007). Also, we moved back and forth from data, interpretation and theory in order to better understand our phenomenon (Miles and Huberman, 1984). To tackle our research questions, which is about the communication of European Identity performed by the European Union[1] all the publications identified by the tag *European Identity*.[2] We focused only on publication available on-line (.pdf, .html, .doc), and in English, to avoid analysing languages which are reaching a smaller audience. After removing duplicates, we finally downloaded 75 documents, published between 2003 and 2019, for a total of 475 MB, and 710,093 words. As we are interested in the discursive construction of the idea of European Identity, we firstly analysed these documents through the software *Nvivo* to extract the most relevant parts of the texts. In particular, through an automatic text search, we identified all the occurrences in all the documents of the word *identity:* for each occurrence we automatically extracted the paragraph surrounding that word: overall, we extracted 1,278 paragraphs, accounting for 159,533 words, from 54 of the original 75 documents. For analysing this amount of data, we relied on Topic Modelling (Blei *et al*., 2003), which is a semi-automatic inductive technique for analysing big corpora of texts, based on the Bayesian algorithm Latent Dirilecht Allocation (LDA). As Hannigan *et al*. (2019) point out, Topic Modelling is gaining more and more legitimation in management studies, where it was successfully and recently used to inductively develop classification systems, detect novelties, understand on-line user generated content, analyse cultural dynamics and understand the evolution of frames. In example, Topic Modelling was used to analyse the impact of different frames on the political process: Levy and Franklin (2014) used this technique to detect frames used in the interactions of a public website. Topic Modelling permitted to uncover nuances in meanings deployed, and in the framing processes. Fligstein *et al*. (2017) used Topic Modelling to detect differences in meanings used in communications by the Federal Open Market Committee. Topic Modelling was also used to confront meanings in the actual debate and meanings described by theory; for instance, Ahonen (2015) challenged existing theories on sources of legal tradition in different countries by developing an inductive classification system through this technique.

Technically, Topic Modelling analyses corpora of texts, and automatically detects words that co-occur in the same source more often that it would happen by chance. Then the algorithm places together co-occurring words in containers of meaningful words named 'topics' (Mohr and Bogdanov, 2013). This automatic stage is followed by an inductive one, where the researcher has to inductively understand the meaning of each topic, based on the most important words constituting each topic, and on the actual coding of the original documents, which is an available output of the algorithm. Documents pertaining to a corpus are therefore conceptualised as mix of topics, and each text is described by its topics' composition (Antons *et al*., 2016). This technique has several important features: first, it permits to deal with corpora too big to be qualitatively analysed; second, induction is a key strength of the technique, which

---

[1] Available online at https://publications.europa.eu/en/home (last accessed: April 11, 2020).

[2] Interestingly the path to find the subject *European Identity* is Social Question > culture and religion > culture > cultural difference > cultural identity > European Identity.

is only partially automated; third, the technique recognises that words can have different meanings in different contexts; therefore, the same word can appear in different topics (DiMaggio *et al.*, 2013).

Before analysing our 1,278 paragraphs we cleaned data by removing stop-words, which are non-meaningful words, such as pronouns, articles and adverbs. After removing stop-words, our corpus consisted of 76,740 words. Inputs required by the algorithm are, of course, the texts to be analysed, together with the number of topics the researcher wants the algorithm to elicit. This is a very sensitive and important step in the analysis, at it strongly impacts the results. Literature is not proposing a conclusive method for defining the 'right' number of topics: some scholars rely on quantitative methods for defining this number (i.e., Mimno *et al.*, 2011), but other studies highlight how topic models that are better according to quantitative measures are not perceived as better by human judgment (Chang *et al.*, 2009). DiMaggio *et al.* (2013) point out that there is no 'right model', as models are like lenses through which analysing reality, so the better model is the one that permits more insights.

We complemented both quantitative and qualitative methods for understanding the appropriate number of topics. We relied on two software for performing this task: Mallet, an open source software which is the state of the art for performing Topic Modelling (McCallum, 2002; Hannigan *et al.*, 2019), and MITAO, a Python based user friendly interface developed at the University of Bologna, that relies on the NLTK and Gensim libraries (Heibi *et al.*, 2019). In particular, we used MITAO for two tasks: first, we used it to clean data, removing stop-words, and other 'noise' that characterised sources.[3] Second, we used MITAO, following an established procedure, to produce models with 5, 10, 15, 20, and 25 topics. We analysed those topics qualitatively and quantitatively, as MITAO provides two measures for evaluating the quality of a topic model: *perplexity score,* and *coherence.* Perplexity is used in information theory to evaluate how well a statistical model describes a dataset, with lower perplexity denoting a better probabilistic model (Zhao *et al.*, 2015). Coherence depends on the pair-wise word similarity scores of the words in a same topic and can be described as the semantic interpretability of the top terms usually used to describe discovered topics (O'Callagan *et al.*, 2015). Perplexity and coherence are calculated in different ways; therefore, they do not always agree on the 'right' number of topics. Moreover, as we said, both coherence and perplexity have results that can differ from a qualitative analysis. In any case, qualitative analysis, perplexity score, and coherence, pointed at the fact that, with our sample, the best models were the ones containing from 5 to 10 topics. Therefore, we used Mallet to produce Topic Models with 4, 5, 6, 7, 8, 9, and 10 topics, and we qualitatively analysed these new models, focusing on interpretability, analytical power, and validity (Hannigan *et al.*, 2019). Finally, we selected the 6 topic model. For this final stage, we used Mallet as it is more legitimised in the field.

3.   Findings

For interpreting the six Topic Model that we selected, we relied on the three main outputs of the algorithm: (i) a *list of the most important words for defining each topic* (see Table 3.1), which was the most important tool for making sense of elicited topics; (ii) a *source per topic matrix,* detailing the topic composition of each paragraph; and (iii) a *word per topic matrix,* detailing

---

[3] This 'noise' was a consequence of the transformation of all sources in .txt files.

the coding of each word in the original texts. The first two outputs were used to interpret all the topics, the last one for fine-tuning our interpretation, where needed. Table 3.1 presents the 50 most important words for each topic, together with an inductively developed label. In the following part of this paragraph we describe the topics and how we interpreted them.

Topic 1 - National versus European Identity and the rule of law: according to our interpretation, this topic discusses European identity in its relationship with national identities. If we analyse the most important words defining this topic, we find several terms that refer to *national*[4] *state* and *national identity*: *czech*, *british*, *national*, *britain*, *polish*, *french*, *english*, *czechs*, *britishness*, *poland*, *france*, *spanish*, *spain*. These words are used in a context where the rule of law is discussed, in particular focusing on the most important *principles* defining the *states.* Indeed, we find words as *constitutional*, *constitution*, *state*, *law*, *principles*, *republic*, *government*, *member*, *states*, *parliament*, and *treaty*. The paragraphs mostly coded at this topic, in example, discuss the relationships between national (constitutional) identities and the European one in reference to the Treaty of Lisbon, or deal with those constitutional rules and principles which are specific to a nation (i.e., those that are protected in the national legal order) and/or in the European one.

Topic 2 - Identity and Political Integration: the second topic tackles identity in the process of European integration. Among the most important words we find *european*, *europe*, *identity*, *integration*, and *national*. We deal with an *integration* that must be *political*, at the level of *nations* as *Italy*, but also *civic*, involving the *identification* of *citizen* with the *union*. Paragraphs that are constituted by this topic in a higher percentage, indeed, discuss the compatibility of regional, national and European identities in different countries, assessing that regional and national identities are compatible with a feeling of closeness to the EU and may even, at least to a certain extent, reinforce such a feeling.

Topic 3 - Common cultural history: the third topic tackles identity as a product of a common history. The focus on history is, of course, identified by words such as *history*, *historical*, *century*, but also by words such as *modern*, *time*, *present*, *past*, *contemporary*. Indeed, the most important clue for understanding the meaning of this topic is the presence of the words *greek*, and *greece*, which point at the idea of Europe as directly coming from the place where democracy was born. The common history has therefore a strong *cultural* layer, but also a *civic*, and a *social* one. These layers are fundamental to define the development of the *society* in *state* and *nation*, in Europe. The more important paragraphs for this topic take modern Greece as the example of an ethnic identity which is deeply rooted in the long history of western civilisation from the beginning. As it is the case for the whole Europe, the issue for modern Greece is that the long cultural heritage is not automatically transformed into a shared political identity, without a strong nation state.

Topic 4 - Common cultural values and education: Topic 4 defines *european identity* with *reference* to *common cultural values*, common *heritage*, and common *culture* in *europe*. Cultural values are a funding core of the European Union, as they are shared among citizens. Related to this point there are two issues, that both affect *cities* and *citizens*: on the one hand, the issue of *education*, to transmit and preserve cultural values. And on the other hand, the issue of *policies*, that is defined through *projects*, *activities*, *institutions*, *programme*, and *participation*. All these activities are meant to secure a *democratic future* for the European Union. Paragraphs mostly coded at

---

[4] We use *italic* do denote words among the most important ones for the topic we are describing.

this topic present programmes that aim at promoting a sense of European active citizenship and community by supporting universities and schools involved in young and adult education and sport activities. This education should specifically focus on fundamental values of the EU, such as democracy, participation in democratic decision making, tolerance, and common understanding. Topic 3 and Topic 4 share a focus on culture, but while the former emphasises culture as the product of the history, the latter emphasises shared values as the basis for a common future.

Topic 5 – European, German, Austrian identities, and the war: Topic 5 tackles European identity as emerging from the wars of the previous century (*first*, *post*). In particular, this topic is used by paragraphs emphasising the relationships between Austrian and German identities with European identity: this is the reason why several words connected to Germany and Austria characterise this topic (*german*, *national*, *austrian*, *nation*, *germany*, *austria*, *germans*). National identities are analysed and described in their different layers and also in connection with relevant political moments of the previous century. WWII, in example, was a milestone for the end of Nationalism, which is a threat to the Union. But also the fall of the Berlin Wall was relevant for the *political* equilibrium of the previous century. We find traces of these discourses in the words *west*, *east*, *eastern*, *western*.

Topic 6 - Borders and minorities: the focus of this topic regards *border* and *borders* of the European *identity*. Of course we deal with physical borders, but in this topic borders are also defining how European *identity* is interacting with *different communities'* defining identities: *people* pertaining to *different social groups* might have a stronger or weaker *group identity*, that can interact with European identity to a *different* extent, This is the reason why we find several words pointing at *minority groups* that, although very different among them, share the fact that *group identity* is important. We find words as *women*, *gender*, *young*, but also *polish*, *italian*, *slovenian*. The issue, here, is how the fact of belonging to minorities interacts with European identity. Paragraphs mostly coded at this topic, in example, analyse how age and gender impact identity formation, or how borders affect group identity, or focus on migrant women and their identity.

Topic	1	2	3	4	5	6
Label	**National versus European Identity and the rule of Law**	**Identity and Political Integration**	**Common cultural history**	**Common cultural values and education**	**European, German, Austrian identities, and the war**	**Borders and minorities**
50 most important words per topic	Identity	european	identity	european	german	border
	reference	national	political	identity	national	reference
	Czech	identity	historical	reference	austrian	social
	British	europe	ethnic	europe	identity	identity
	constitutional	integration	greek	culture	nation	different
	national	identities	state	values	political	communities
	State	people	nation	cultural	germany	people
	Law	political	cultural	level	west	group
	Britain	between	history	common	east	borders
	movement	reference	development	sense	european	between
	university	identification	society	cities	europe	language
	Case	civic	modern	citizens	austria	narratives
	Art	media	time	union	eastern	work
	Polish	respondents	greece	education	nationalism	research
	Power	common	fact	heritage	between	groups
	fundamental	analysis	concept	history	class	italian
	French	research	process	citizenship	germans	construction
	constitution	important	collective	role	formation	relations
	principles	union	community	diversity	war	local
	republic	citizens	question	policies	public	economic
	Idea	countries	important	shared	western	interviewees
	Rights	survey	between	belonging	social	place
	English	nation	social	programme	party	women
	Czechs	italian	present	understanding	post	polish
	government	cultural	past	project	first	experience
	London	results	contemporary	participation	regional	cross
	britishness	economic	first	policy	components	living
	population	positive	because	democratic	historical	differences
	Press	discourse	civic	future	empire	ways
	minorities	elements	context	making	middle	informants
	States	italy	issue	society	collective	minority
	member	feel	point	building	great	space
	immigrants	spanish	world	activities	unification	individual
	Century	sense	discourse	institutions	attitudes	slovenian
	Legal	general	different	potential	orientations	young
	Poland	being	nations	transnational	past	side
	France	ethnic	image	human	framework	change
	spanish	country	long	rights	state	events
	respect	findings	century	local	central	gender
	Article	representations	report	active	habsburg	position
	instance	main	identities	strengthening	system	belonging
	principle	discourses	made	mobility	orientation	members

Studies	public	elements	develop	liberal	study
influence	project	terms	promote	anti	personal
Spain	attitudes	unity	civil	building	experiences
Set	relationship	concepts	support	cultural	people
parliament	perceived	model	make	enlargement	life
Treaty	parties	being	impact	major	home
Cent	lay	dominant	based	construction	food
discussion	issues	relationship	cultures	world	context

Table 3.1. 50 most important words per topic.

4.   Discussion

The topics that we elicited differ according to the way they are used. Figure 4.1 highlights the average use of each topic in the sample and the percentage of times when each topic is the most important one for a textual source. Let's focus first on the average use of the topics, or average prevalence, which tells us the importance of each topic for the whole corpus analysed. As topics are constituted by the same amount of words, this information provides us with the answer to the question 'how often are used words pertaining to a certain topic, on average?' Topic 4 - Common cultural values and education, which constitutes almost a quarter of the published documents, is the most important one, signalling how for the EU the cultural values are the most important issues to deal with, when dealing with European identity. In particular, the stress on education, which is pinpointed by Topic 4, signals how the EU is forward looking when dealing with this issue. Not surprisingly, the second most used topic is Topic 2, Identity and Political Integration. The fact that 19% of the analysed documents are composed by this topic reflects the importance of European political Integration for the building of European identity. The other four topics account for a percentage between 16% and 12.7%, with the least important one being Topic 5¸ European, German, Austrian identities, and the war.
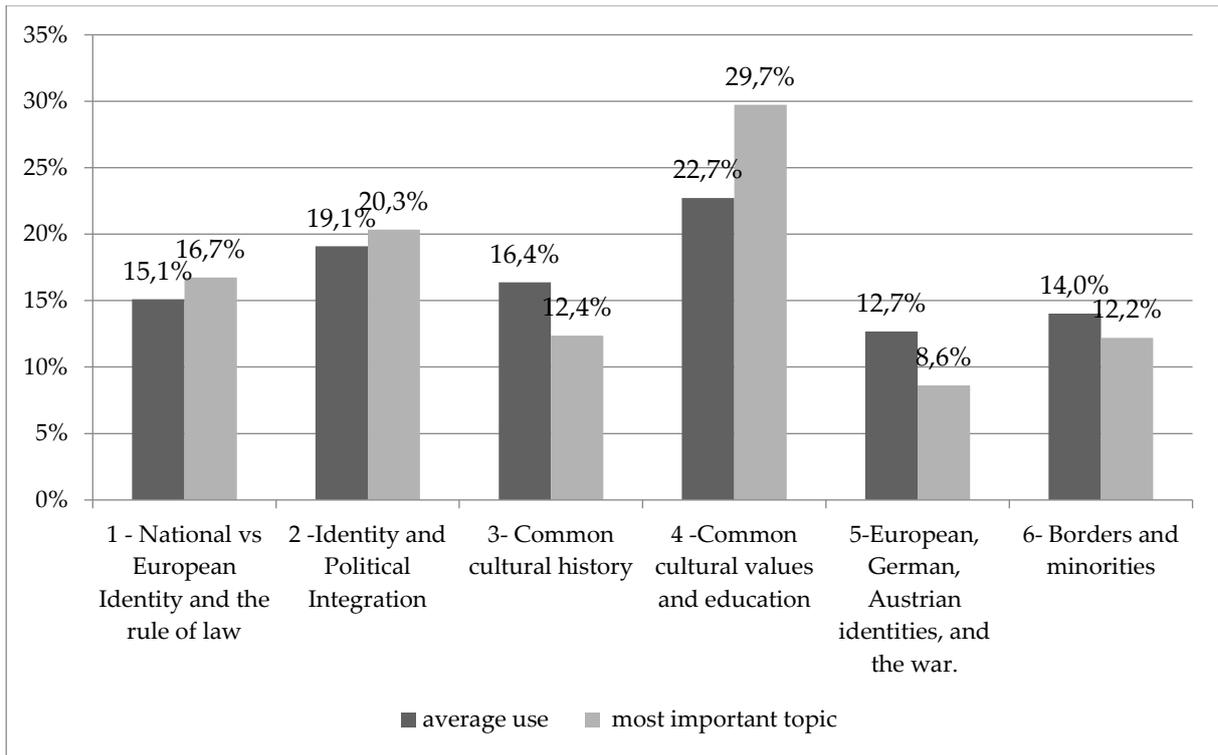
Figure 4.1. Average use of topics in the sample and percentage of times where a topic is the most important one for a textual source.

Let's now focus on the second histogram of Figure 4.1. This histogram tells us how often a topic is the most important one for a textual source. By comparing the first and the second histogram of Figure 4.1, topic by topic, we can better understand how different topics are used. Indeed, if the second histogram is taller than the first one, the percentage of times when a topic is the most important one is higher than the average use of that topic. This feature denotes topics that strongly characterise texts where they are used. Conversely, topics with an average prevalence higher than the percentage of times when they are the most important ones are topics used to fill up texts characterised by other topics (Ferri *et al.*, 2018). In our case, discourses about the European identity are mainly built by Topic 1, Topic 2, and Topic 4. Topic 4 - Common cultural values and education, is definitely a characterising topic, for the texts where it is used, as it is the most important topic in about 30% of sources, while its average use is 22.7%. A similar situation occurs in the cases of Topic 1 – National versus European Identity and the rule of law, and Topic 2 – Identity and Political Integration. Conversely, the other three topics tend to fill up other discourses, more than being at the centre of the discursive space: common cultural history, the legacy of the war, and the issues affecting borders and minorities seem to complement discourses created around most characterising topics.

		second topic					
		1 - National vs European Identity and the rule of law	2 - Identity and Political Integration	3- Common cultural history	4 - Common cultural values and education	5- European, German, Austrian identities, and the war	6- Borders and minorities
most important topic	1 - National vs European Identity and the rule of law		21.5%	36.9%	16.8%	13.6%	11.2%
	2 - Identity and Political Integration	14.2%		41.2%	17.7%	16.9%	10.0%
	3- Common cultural history	18.4%	32.3%		25.9%	10.1%	13.3%
	4 -Common cultural values and education	27.6%	26.3%	22.9%		5.8%	17.4%
	5- European, German, Austrian identities, and the war	14.5%	45.5%	30.9%	8.2%		0.9%
	6- Borders and minorities	14.7%	30.1%	22.4%	16.7%	16.0%	

Table 4.1. Associations between first and second topic.

We further explored this line of thought by analysing the associations between topics. In particular, for each topic, we calculated the associations between the most important topic for a textual source, and the second most important one. The results of this analysis are presented in Table 4.1. Each row contains data regarding sources where the topic in that row is the most important one. Then, columns provide data regarding how often a given topic is second most important one. In example, take the first row, representing Topic 1 – *National versus European Identity* and the rule of law: if we consider all the textual sources where Topic 1 is the most important one, in 36.6% of these sources the second most important topic is Topic 3 – Common cultural history. This information highlights that discourses around the relationship between national and European identity are mainly constructed with reference to common cultural history. The topics least used to complement National versus European Identity are Topic 5- European, German, Austrian identities, and the war, and Topic 6 – Borders and minorities. Lets' now focus on Topic 4, Common cultural values and education, as this is the most important topic both for average prevalence and for the number of times when it is the most important for a paragraph. Paragraphs characterised by discourses about cultural value are often complemented by Topic 1, about national and European Identity, Topic 2, political integration, and Topic 3, common cultural history. Rarely Topic 4 co-occurs with discourses about the war. The other characterising topic is the number 2, about political integration.

Discourses built on this topic are very often (41.2% of times) complemented by common cultural history (Topic 3). Conversely, when the common cultural history is the main focus of a source dealing with European identity, often the second topic is number 2. Topic 5 and Topic 6, dealing respectively with conflicts of the previous century, and minority and borders, are the least used and the ones that more rarely are the most important for a textual source. Both Topic 5 and Topic 6, when they are the more important for a textual source, are associated with Topic 2, regarding political integration, and Topic 3, about common cultural history.
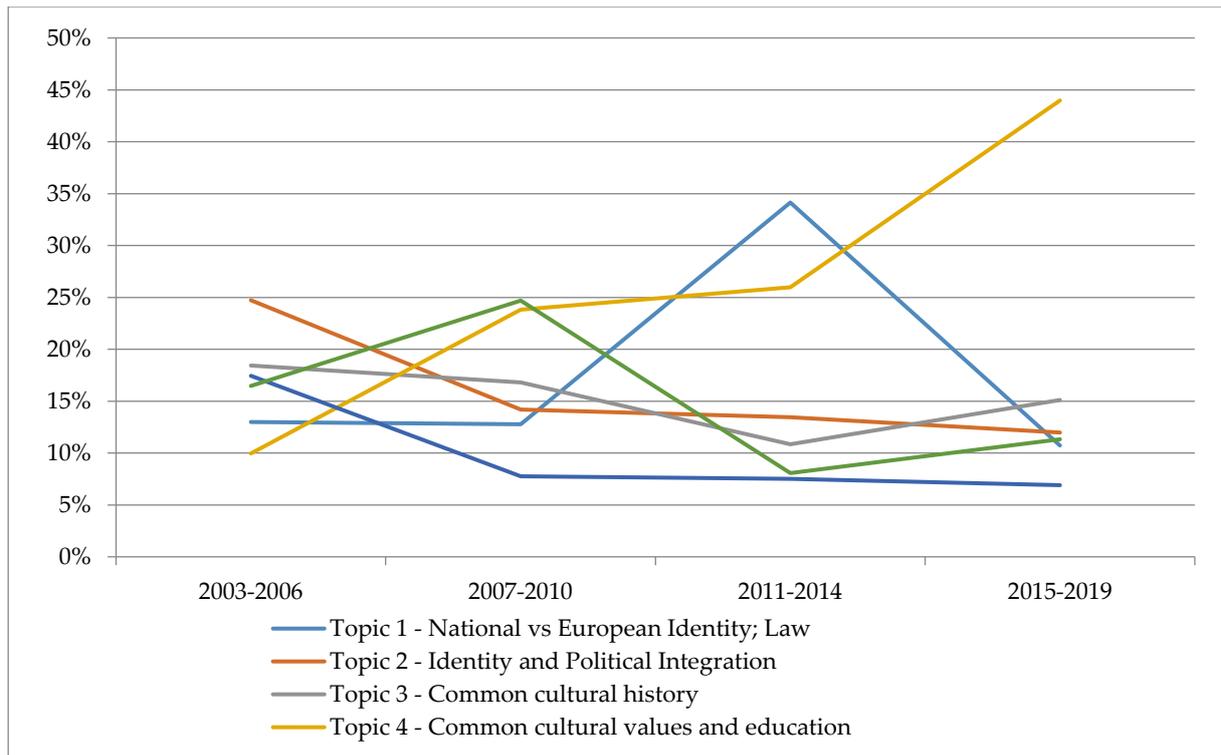

Figure 4.2. Longitudinal average use of topics.

After having analysed the meanings of each topic and having described their use, we can now move to a longitudinal analysis, focusing on how much the use of each topic changed over time. As publications, and paragraphs, are not distributed uniformly in our sample, we divided them in four temporal segments to calculate the average prevalence of each topic in each segment. Figure 4.2 reports the results of this analysis. In the first period, 2003–2006, the most used topic was Topic 2, about political integration. The average prevalence of this topic declined in the following periods. The second temporal segment, 2007–2010, was mostly characterised by Topic 6, borders and minorities. This topic was relevant also in the first period, but reached a peak in the second period, while quickly declining afterwards. In the third period, 2011–2014, the most used topic was number 1, regarding the rule of law and the national and European identities. This peak was probably the effect of the studies and reports analysing the Union after the approval of the Lisbon Treaty by all member states. Indeed, the use of Topic 1 quickly declined afterwards. The final period, 2015–2019, is characterised by the fact that Topic 4, about common cultural values and education, is not only the most important one, but reaches alone an average prevalence of about 45%. All the other topics account for

less than 15%. This data confirms that this topic characterises a forward-looking attitude of the EU, aiming at fostering shared values among its citizens trough education. The average prevalence of this topic, indeed, increased dramatically throughout the four periods that we analysed. As for the two topics that we did not mention, Topic 5 was relevant in the first period, but then but then its average usage lowered significantly. On the contrary Topic 3, about common cultural history, did not experience significant peaks in its usage, but was twice the second most used topic, and in the other two periods was the third and the fourth topic, according to its average prevalence.

5.   Conclusion

In this paper we analysed the publications by the Publication Office of the European Union tagged with the label 'European Identity' to understand which dimensions of the European Identity are leveraged by the European Union in its communication. We found six topics, which deals with laws in relation to national and European identities, the process of political integration, common history, education and shared values, nationalism and the conflicts of the previous century, and borders and minorities. We can discuss those dimensions against the different dimensions of European identity recognised by literature. As said, literature dealing with European integration recognises three main symbolic dimensions of the process: first, a dimension dealing with laws and the governance of the European Union; second, a discourse on governance, and finally citizens' identification with the European Union (Christiansen *et al.*, 1999; Risse, 2009; Waever, 2009). The general issue of integration is directly tackled by our Topic 2, Identity and Political Integration, while Topic 1 specifically deals with the first dimension, as it tackles national and European identities under the legal environment of the rule of law. Taken together, Topic 1 and Topic 2 account for about one third of our textual sources, signaling the importance of these issues in the communication by the EU. Yet, both topics were not uniformly used, as Topic 1 is relevant especially in the third period, while Topic 2 in the first one. Therefore, we can speculate on the fact that these topics are mobilised by the EU in correspondence with specific events, and are not part of the daily communication by this Institution. The first period of our analysis, 2003–2006, is indeed a period of enlargement of the Union, directly following the introduction of the euro, and we can imagine that in those years the issue of political integration was particularly relevant. As for Topic 1, as said, it reached a peak after the approval of the Lisbon Treaty by all member states. It is then interesting that the second symbolic dimension of the process of integration, which is the discourse about the forms of governance, is not present at all among our topics: we argue that the reason is that the EU wanted to present itself as a cohesive institution, rather than stressing different possible forms of governance, at least in its official publications.

It is then very interesting to move toward the third symbolic dimension of political integration, which is citizens' identification with the European Union. As said, literature identifies two main drivers of identification: first, a civic component of European identity, encompassing the existence of common laws, institutions, social systems, and the free circulations of goods and people. Second, a cultural component of European identity, encompassing a common history and heritage, values and ideas, shared languages and a shared religion (Bruter, 2003). As for the civic component, the already discussed Topic 1 and 2 focus on national and European laws and institution, and tackles their integration. Topic 6,

borders and minorities, deals with the (free) circulation of people, and it mainly analyses group identities in relation with geographic identities. Interestingly, no topic deal with the circulation of goods, although market criteria are one among the Copenhagen criteria, which are the requirements a country need to comply with to be accepted as a member of the European Union: 'Countries wishing to join need to have: i) stable institutions guaranteeing democracy, the rule of law, human rights and respect for and protection of minorities; ii) a functioning market economy and the capacity to cope with competition and market forces in the EU; iii) the ability to take on and implement effectively the obligations of membership, including adherence to the aims of political, economic and monetary union'.[5] While the rule of law (Topic 1), institutions (Topic 1 and Topic 2), and minorities (Topic 6) are communicated by the EU, as reflected in the elicited topic, the common currency, and the circulation of goods is absent from the debate. One possibility for explaining the lack of this issue is that maybe the EU is avoiding sensitive topics in its communications. A strong criticism towards the EU portraits the Union as a union of banks and economics, more than a union of people and citizens. A possibility is that, lacking evidence to prove this criticism wrong, the EU is avoiding bringing the issue in the debate. Communication on social media by European institutions tend to be neutrally connoted, avoiding emotionally charged tones (Barberio *et al.*, 2020); we can imagine that the translation of this stance in the written communication by the Publication office is avoiding delicate issues.

The cultural component of European identity is way more present in our sample of communications by the European Union: Topic 3, Common cultural history, and Topic 5, European, German, Austrian identities, and the war tackle in different ways the existence of a common history: the former positively emphasises the common roots of European citizens and states in the legacy of the Greek democracy. The latter builds an idea of Europe in contrast which the conflicts of the previous century and the idea of Nationalism. These topics, also, recall researches analysing the nature of European Identity as historically rooted (Delanty, 1995). Yet, while Topic 3 remained relevant, in terms of average prevalence, in all four periods analysed, Topic 5 tends to fade away. The case of Topic 4, Common cultural values and education, is completely different. This topic is the most used one, the one that is most often the more important, and its usage is strongly increasing over time. We argue that this trend reflects a strategic positioning by the EU, which wants to stress positive values (i.e., common cultural values), as the basis for European identity. Moreover, paragraphs coded to this topic explicitly describe education, both for adults, and young, as a mean to obtain active European citizenship. In other words, especially by looking at the longitudinal trend, we can imagine en effort by the EU to describe itself mostly with positive cultural values. It is interesting to note that the cultural dimension of a shared religion, although recognised by literature, is completely not present in our sample. Once again, we think that the EU wants to avoid a delicate issue. Finally, Topic 6, borders and minorities, is not recognisable in Bruter's (2003) framework, as it is not clearly linked to a civic or cultural component. Yet, we can better understand this topic with reference to the literature that addresses European identity in relation to 'others'; these others can be reference groups, which permits to define an identity by difference, or minorities with specific identity claims (Wodak and Boukala, 2015).

---

[5] Conditions for membership described online at https://ec.europa.eu/neighbourhood-enlargement/policy/conditions-membership_en (last accessed: April 11, 2020).

So, in a nutshell, how is the European Union communicating its idea of European Identity to European citizens? Topics that we elicited refer to several of the components of European Identity described by literature, but most delicate themes are avoided. Also, while at the beginning of the period under analysis, different themes are present, over the time this form of written communication tends to focus on a definition of identity which is defined by shared values, which can be taught to citizens. As the power of words permits defining the reality, we can say that the reality portrayed by these communications has a positive image, and is forward looking, to the extent that it can be learnt. Of course, we cannot say whether elicited topics are strategically used or not, and we are not making such a claim. What we can say is that there is a clear trend toward the use of Topic 4, about shared values and education.

Naturally this study presents few limitations: the most important one is that, although we focused on a vast sample of publications by the Publication Office of the European Union, we were able to capture only part of the debate, which is the part described as related to European Identity by the Publication Office itself. But the theme of European identity can characterise also other documents, even if those are not categorised as dealing with 'European identity' by the Publication Office itself. Also, we focused on a very specific class of textual sources, as the publications that we analysed are mostly long studies and reports. While this kind of document permits in-depth analysis, and the invocation of a wider amount of ideas and arguments, it has rhythms and rules that differ from the daily communication that takes place on newspapers and social media. Each communication channel has its peculiar rules and habits, and we can imagine that major events, polemics and reactions would reflect more on communication on newspapers through interviews and editorials. Also, the specific nature of social media permits a dialogue with citizens that is not always fruitful, but that is characterised by a bi-directional communication. The documents that we analysed have a different aim, and this must be taken into consideration while reflecting on our results.

Finally, as we said, the European Union can be conceived as an organisation, but it is not the 'typical' organisation that we usually find analysed in organisational literature. This difference poses a question about how this study would be useful or replicable for industrial organisation, individual companies, or public organisations. We think that, from a methodological point of view, Topic Modelling can be applied to other forms of organisations, exactly as we can apply qualitative content analysis regardless of the empirical object of our analysis. Of course, we need to find reliable documents to be analysed, and this task would be easier dealing with big organisations (i.e., analysing investors' communications, or websites), or with the public administration, which drafts a wide amount of documents. As for the results, our case can be compared to the organisations that have the resources to draft studies and reports that are not meant for a daily consumption and communication. Once again, it is hard to imagine individual organisations dealing with this issue, but there is a plethora of organisations that produces positioning papers and reports that can be analysed with this methodology, such as public administration, consultancy agency, governmental bodies, and also organisations in the stock market.

Reference list

Ahonen, P. (2015), "Institutionalizing big data methods in social and political research", *Big Data & Society*, 2 (2): 1–12.

Antons, D., Kleer, R., and Salge, T.O. (2016), "Mapping the topic landscape of JPIM, 1984–2013: In search of hidden structures and development trajectories", *Journal of Product Innovation Management*, 33 (6): 726–749

Barberio, V., Kuric, I., Mollona, E., and Pareschi, L. (2020, forthcoming), "Institutional organizations' use of social media to communicate EU policy and the emergence of a European Public Sphere", *Investigaciones Regionales / Journal of Regional Research*, 46

Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003), "Latent dirichlet allocation", *Journal of Machine Learning Research* 3: 993–1022

Bourdieu, P. (1985), "The social space and the genesis of groups", *Theory and Society*, 14: 723–744.

Bourdieu, P. (1991), *Language and symbolic power,* Cambridge, MA: Harvard University Press.

Bourdieu, P. (2009), *Ragioni Pratiche*, Bologna: Il Mulino.

Bruter, M. (2003), "Winning hearts and minds for Europe: The impact of news and symbols on civic and cultural European identity", *Comparative Political Studies,* 36 (10): 1148-1179.

Bruter, M. (2005), *Citizens of Europe? The emergence of a mass European identity*, New York: Palgrave Macmillan.

Capello R. (2018), "Cohesion Policies and the Creation of a European Identity: The Role of Territorial Identity", *Journal of Common Market Studies*, 56 (3): 489–503

Caporaso, J.A., and Kim, M. (2009), "The Dual Nature of European Identity: Subjective Awareness and Coherence", *Journal of European Public Policy*, 16 (1): 19–42.

Chacha, M. (2012), "Regional Attachment and Support for European Integration", *European Union Politics* 14 (2): 206–227.

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., and Blei, D.M. (2009), "Reading tea leaves: How humans interpret topic models". In *Advances in neutral information processing systems,* Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), *vol. 22.* Retrieved from https://papers.nips.cc/paper/3700-reading-tea-leaves-how-humansinterpret-topic-models.pdf

Checkel, J.T., and Katzenstein, P.J. (2009), "The politicization of European identities". In *European identity,* J. Checkel: Katzenstein (Eds), pp. 1-28. Cambridge: Cambridge University Press.

Christiansen, T., Jorgensen, K., and Wiener, A. (1999), "The social construction of Europe", *Journal of European Public Policy*, 6 (4): 528-544.

Contandriopoulos, D., Denis, J., and Langley, A. (2004), "Defining the 'public' in a public healthcare system", *Human Relations,* 57 (12): 1573-1596.

Delanty, G. (1995), "The limits and possibilities of a European identity: A critique of cultural essentialism", *Philosophy & Social Criticism,* 21 (4): 15-36.

DiMaggio, P., Nag, M., and Blei, D. (2013), "Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding", *Poetics* 41 (6): 570–606.

Eisenhardt, K.M., and Graebner, M.E. (2007), "Theory Building from Cases: Opportunities and Challenges", *Academy of Management Journal*, 50: 25-32.

Fairclough, N. (1995), *Critical discourse analysis: The critical study of language,* London: Longman.

Ferri, P., Lusiani, M., and Pareschi, L. (2018), "Accounting for Accounting History: A topic modeling approach (1996–2015)", *Accounting History*, 23 (1–2): 173–205.

Fligstein, N., Polyakova, A., and Sandholtz, W. (2012), "European Integration, Nationalism and European Identity", *Journal of Common Market Studies,* 50 (S1): 106–122.

Fligstein, N., Stuart Brundage, J., and Schultz, M. (2017), "Seeing like the Fed: Culture, cognition, and framing in the failure to anticipate the financial crisis of 2008", *American Sociological Review*, *82* (5): 879–909.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*, London: Harper & Row.

Hannigan, T., Haans, R.F.J., Vakili, K., Tchalian, H., Glaser, V., Wang, M., Kaplan, S., and Devereaux Jennings, P. (2019), "Topic Modeling in Management Research: rendering New Theory from textual data", *Academy of Management Annals*, in press.

Hardy, C., Palmer, I., and Phillips, N. (2000), "Discourse as a strategic resource", *Human Relations,* 53 (9): 1227-1248.

Heibi, I., Peroni, S., Ferri, P., and Pareschi, L., (2019, June 27), catarsi/mitao: MITAO first release (Version v1.1-beta). Zenodo. http://doi.org/10.5281/zenodo.3258328

Knights, D., and Morgan, G. (1991), "Corporate strategy, organizations and subjectivity: A critique", *Organization Studies*, 12: 251-273.

Koopmans, R., and Statham, P. (eds) (2010), *The Making of a European public sphere. Media discourse and political contention,* Cambridge: Cambridge University Press.

Laffan, B. (2004), "The European Union and its institutions as 'identity builders'". In *Transnational identities: Becoming European in the EU,* R. Herrmann, T. Risse, M. Brewer (eds), pp. 75-96. Lanham, MD: Rowman & Littlefield.

Levy, K.E.C., and Franklin, M. (2014), "Driving regulation: Using topic models to examine political contention in the U.S. trucking industry", *Social Science Computer Review*, 32 (2): 182–194

Marcussen, M., Risse, T., Engelmann-Martin, D., Knopf, H., and Roscher, K. (1999), "Constructing Europe? The evolution of French, British and German nation state identities", *Journal of European Public Policy*, 6 (4): 614-633.

McCallum, A.K. (2002), "MALLET: A Machine Learning for Language Toolkit", http://mallet.cs.umass.edu

Medrano, J., and Gutiérrez: (2001), "Nested identities: National and European identity in Spain", *Ethnic and Racial Studies,* 24 (5): 753-778.

Meyer, R., and Höllerer, M. (2010), "Meaning structures in a contested field: A topographic map of shareholder value in Austria", *Academy of Management Journal,* 53 (6): 1241-1262.

Miles, M.B., and Huberman, A.M. (1984), *Qualitative data analysis*. London: Sage.

Mimno, D., Wallach, H.M., Talley, E., Leenders, M., and McCallum, A. (2011), "Optimizing semantic coherence in topic models". In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 262–272. Stroudsburg, PA: Association for Computational Linguistics.

Mohr, J.W., and Bogdanov: (2013), "Introduction – Topic models: What they are and why they matter", *Poetics* 41 (6): 545–569.

O'Callaghan, D., Greene, D., Carthy, J., and Cunningham: (2015), "An analysis of the coherence of descriptors in topic modeling", *Expert Systems with Applications*, 42 (13): 5645-5657.

Phillips, N., Lawrence, T., and Hardy, C. (2004), "Discourse and institutions", *The Academy of Management Review*, 29 (4): 635-652.

Reeskens, T., and Hooghe, M. (2010), "Beyond the civic-ethnic dichotomy: Investigating the structure of citizenship concepts across thirty-three countries", *Nations and Nationalism*, 16 (4), 579-597.

Rhodes, C., and Brown, A. (2005), "Narrative, organizations and research". *International Journal of Management Reviews*, 7 (3): 167-188.

Risse, T. (2009), "Social constructivism and European integration". In *European integration theory*, A. Wiener & T. Diez (Eds), pp. 144-159, Oxford: Oxford University Press.

Risse, T. (2010). *A community of Europeans: Transnational identities and public spheres, 29-30*, Ithaca: NY, Cornell University Press.

Risse, T. (2014), "No Demos? Identities in Public Spheres in the Euro Crisis", *Journal of Common Market Studies*, 52 (6): 1207–1215.

Schilde, K. (2014), "Who Are the Europeans? European Identity Outside of European Integration", *Journal of Common Market Studies*, 52 (4): 650–667.

Strauss, A., and Corbin, J. (1990), *Basics of qualitative research: techniques and procedures for developing grounded theory*, Thousand Oaks, CA: Sage.

Waever, O. (2009), Discursive approaches. In *European integration theory,* A. Wiener & T. Diez (Eds), pp. 163-180. Oxford: Oxford University Press

Wodak, R., and Boukala, S. (2015), "European identities and the revival of nationalism in the European Union: a discourse-historical approach", *Journal of Language and Politics,* 14 (1): 87-109.

Zhao, W., Chen, J.J., Perkins, R., Liu, Z., Ge, W., Ding, Y., and Zou, W. (2015), "A heuristic approach to determine an appropriate number of topics in topic modeling". *Proceedings of the 12th Annual MCBIOS Conference,* 16 (13). http://doi.org/10.1186/1471-2105-16-S13-S8.