Making new from old?
Assessing regulatory novelty in EU renewable energy policies

Colin Walder*

October 10, 2021

1 Introduction

Effective design of renewable energy policies is essential to adequately address the challenges posed by climate change. Renewable energy policies, just like other governance instruments, are subject to change over time. They evolve to address new issues as they emerge and dismiss those that have become obsolete. However, while scholars acknowledge the importance of renewable energy policies' temporal dynamics in terms of their content, there is little systematic evidence in this regard (Schmidt & Sewerin 2019).

This paper addresses this research gap with a text-as-data approach. More specifically, it deals with the concept of *regulatory novelty* of policies. Considering policies as permanently evolving entities implies that they build upon each other (Lindblom 1959). Yet, the way policies are layered does not always follow an incremental path. Pagliari & Wilf (2020), for example, show how (global) financial crises entailed new types of regulation for the banking industry. Thus, it is pertinent to ask whether such events associated with significant regulatory novelties have also occurred for energy policies in recent years.

Widely regarded as a resounding success in the fight against climate change, the Paris Agreement (COP-21) offers the opportunity to apply the measure of cosine similarity to reconstruct the textual evolution of the European Union's (EU) renewable energy policies. I expect that the proclaimed paradigm shift has brought about a change in the nature of policy-making, reflected in regulatory novelty (Kern & Rogge 2016). Thus, the research question of the paper reads: *Did regulatory novelty manifest itself in the EU's renewable energy policies after COP-21?*

This paper has two objectives. First, it explores how EU renewable energy policies have evolved since 2000 using a text-as-data approach. Second, focusing on *regulatory novelty*, it asks if in the aftermath of the landmark COP-21 a paradigmatic textual change in the EU's renewable energy policy-making occurred. A timeline informs the reader about the evolution of renewable energy policies in the EU. This is followed by the rationale behind my approach before I introduce the concept of regulatory novelty. The metric (cosine similarity) used to operationalise regulatory novelty and the data are presented in section 5. Section 6 presents the results. Finally, I discuss the findings and elaborate on the usefulness of cosine similarity for further work in the field of policy studies.

*The scripts to reproduce the results are available in the following GitHub repository: get_termproject. The text corpus must be requested directly from the Energy Politics group at ETH Zurich. Special thanks go to Dr. Sebastian Sewerin and Dr. Lynn Kaack, whose continuous inputs ensured that the methodology did not neglect the theoretical rationale and that the paper is presented as a coherent ensemble. In addition, the comments of Buket Buse Demirci and Felix Zaussinger significantly improved the final draft.

2 Outlining the evolution of renewable energy polices in the EU

EU policy-making concerning renewable energies is no exception in that it has been characterised by the tension between (regaining) national sovereignty and deepened EU-wide coordination from its inception. Just as the oil crisis in the 1970s opened a window of opportunity for a paradigm shift in energy generation, global crises such as the financial crisis in 2008 hampered the implementation of far-reaching measures needed for more stringent renewable energy policies. Besides, particularly in recent years, nationalistic tendencies within the EU have undermined the intergovernmental cooperation needed to overcome the collective action problem posed by climate change. The following section provides a brief overview of how renewable energy policy-making has evolved in the EU over the past decades.[1]

The late 1970s marked the beginning of limited and poorly coordinated renewable energy policy-making in the EU. Energy dependency within the EU had to be rethought due to the two oil crises, resulting in the first funding programmes for R&D to promote new energy sources (Nilsson 2011). For the next decade, however, these would remain the EU's only efforts towards sustainable energy. It was not until the ALTENER programme in the 1990s that a further important step was taken about renewable energies. This time around, the focus shifted to reducing CO_2 emissions. Owing to opposition from various member states, however, the programme had largely been stripped of its substance (Boasson & Wettestad 2016).

Around the turn of the millennium, more consistent policy-making gradually took shape. The document *Energy for the future: renewable sources of energy* constituted an essential step towards the promotion of renewable energy and, for the first time, renewable energy credits were mentioned (Solorio & Bocquillon 2017). Shortly afterwards, the first directives explicitly dealing with renewable energies were issued: Directive 2001/77/EC on the promotion of electricity produced from renewable energy sources and Directive 2003/96/EC promoting renewable forms of energy. However, it soon became apparent that these directives lacked substantial impact, partly owing to the fact that the directives were formulated too sectorally and did not follow a more comprehensive approach.

Next, the 20-20-20 targets entered into force. That is, a 20% reduction in greenhouse gas (GHG) emissions, a 20% share of renewable energy sources in final consumption and 20% less energy consumption by 2020. These targets were intended to give renewable energy sources further importance in light of the global financial crisis and the recently adopted Bali Roadmap of the Kyoto Protocol Parties. Within the EU, these efforts culminated in a new directive in 2009 to demonstrate its pioneering role in renewable energy sources promotion on the eve of the climate conference in Copenhagen (COP-15).

[1] This section is heavily based on the contributions of Solorio & Bocquillon (2017) as well as various chapters in the book edited by Birchfield & Duffield (2011). Both sources guide the interested reader in a much more comprehensive manner through the evolution of (renewable) energy policy-making in the EU.

3 COP-21 as a paradigm shift

It is well known that COP-15 did not achieve any significant progress, even being widely regarded as a failure in dampening the pace of climate change (Christoff 2016, Falkner 2016, Savaresi 2016). Thus, instead of discussing its shortcomings, it is more interesting to look at the renewable energy policy framework preceding the pioneering COP-21. One year before Paris, at the end of 2014, the EU member states adopted the 2030 Framework: by then, a 40% reduction in GHG should be achieved, at least 27% of energy consumption must come from renewable sources and energy efficiency must be increased by 27%, with the last target being merely indicative.[2] These EU efforts were further strengthened by the resounding success of COP-21 and subsequently operationalised in the Katowice package.[3]

COP-21 started on 30 November 2015 in Paris and ended with an agreement involving 196 parties. The parties agreed not only on ambitious climate targets but, more importantly, on joint action to combat climate change for the first time. "All Parties are to undertake and communicate ambitious efforts", according to their nationally determined contributions, with the aim of "holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C above pre-industrial levels, [...]" (p.3).[4]

Various literature advances reasons why COP-21, unlike Copenhagen, was such a groundbreaking success in terms of its specific goals and the large number of parties involved. Kinley (2017), former Deputy Executive Secretary of the UNFCCC, points to China's critical role, whose cooperative stance contributed to the conference's success. Others highlight the diplomatic effort of a wide range of parties (Viñuales et al. 2017). The experience of the failed Copenhagen conference also played a role (Savaresi 2016). Lastly, Paris was a success as the global North-South dichotomy was no longer part of the agreement. Both historical emitters and emerging economies were now held accountable for mitigating climate change. At the same time, inequities were taken into account with the nationally determined contributions, resulting in a paradigm shift in the way climate policy should be undertaken in the following decades.

Against this background, the term paradigm shift needs to be conceptualised more clearly. According to Hall (1993), this requires a simultaneous change of the instrument settings, the instruments themselves, and the policy's hierarchy of goals. Regarding the instrument settings, for the first time, we see with the Paris Agreement new and explicitly formulated targets of 2°C and 1.5°C, respectively, above pre-industrial levels. Second, the instrument settings changed in that the annex/non-annex dichotomy was abolished. From now on, all parties had to curb their emissions. In addition, the Kyoto Protocol's top-down approach was replaced by individual national targets that countries could set themselves. While the preparation and assessment of these nationally determined contributions are legally binding, their implementation is not. Thus, the Paris Agreement also differs in this respect from its forerunner, which contained legally binding emission reduction targets for its annex countries.

[2] Link to the 2030 framework for climate and energy policies (HTML)
[3] Link to the Katowice package (HTML)
[4] Link to the Paris Agreement (PDF)

Finally, the rationale underlying the instruments changed fundamentally with the focus shifting from "cost-minimising to opportunity-seizing, and thus from a focus on emissions to a focus on technologies" (Schmidt & Sewerin 2017, p.1). With the Paris Agreement, the fight against climate change is no longer characterised as a burden whose negative consequences have to be shared, but as an opportunity for opening up new economic sectors and even gaining advantages with a greener economy. Consequently, the Paris Agreement can be seen as the cornerstone for a transition period in which the shift away from fossil fuels entails new economic opportunities.

4 Regulatory novelty in policy-making

However, how does such a paradigm shift manifest itself in regulatory novelty? Defined as the "extent to which a new international regulatory initiative builds upon or departs from previous regulations" (p. 4), Pagliari & Wilf's (2020) operational conceptualisation of regulatory novelty shows two ways in which vocabulary can reflect a paradigm shift. To illustrate these textual changes, I will refer to the EU's renewable energy directives (RED) from 2009 and 2018, respectively.

First, *regulatory scope* can result in a different choice of words. For broader policy areas, broader wording is expected. For example, the term "biofuel" was used much more frequently in the 2009 RED than in its successor from 2018, which, in contrast, employed the term "fuel" more extensively. This may indicate either a broadening of the scope or a lower weighting of the policy area of biofuels in the new directive.

Second, new *policy instruments* can be reflected in vocabulary. Different instruments lead to different terms, which indicate relevance, use and other characteristics of the instrument. For instance, in the 2009 RED, *National Renewable Energy Action Plans*, which outline the individually defined targets for the share of energy from renewable sources, are featured extensively. In the 2018 RED, these action plans are only mentioned in the preamble as a basis for increased transparency and effective monitoring. To what extent these action plans are already considered a prerequisite in the newer directive or replaced by another instrument can only be speculated. On the other hand, texts containing similar or even the same instruments hardly differ in their wording. The formulations regarding *joint support schemes*, a policy instrument found in both renewable energy directives, is almost identical in both RED. Thus, these example show that, based on the words used, text similarity serves as a reliable proxy for regulatory novelty.

Consequently, I assume that a text with very similar wording to a previous text shows hardly any traits of regulatory novelty while policies with a very different vocabulary suggest that they cover new policy instruments. Therefore, I expect that after COP-21, with its apparent paradigm shift towards global climate policy, a textual shift will also manifest itself.

5 Research design and methods

5.1 Cosine Similarity as a measure for regulatory novelty

Cosine similarity is an established metric in text analysis and is becoming increasingly popular in political science for document classification based on textual resemblance (Blumenau 2019, Wilkerson & Casas 2017, Garrett & Jansa 2015). As its name implies, cosine similarity compares two given texts in a semantic vector space based on the cosine value of the angle between the vectors. The vector's length corresponds to the number of unique terms in the matrix, while each value within the vector represents the number of occurrences of a term. Below is a simplified example:

$$\begin{array}{lc} \text{Terms} & \begin{array}{ccc} \textit{Text}_1 & \textit{Text}_2 & \textit{Text}_3 \end{array} \\ \begin{array}{l} \textit{term}_1 \\ \textit{term}_2 \\ \textit{term}_3 \\ \textit{term}_5 \end{array} & \begin{pmatrix} 3 & 1 & 4 \\ 0 & 5 & 2 \\ 4 & 0 & 0 \\ 0 & 1 & 2 \end{pmatrix} \end{array}$$

To calculate the cosine similarity of two vectors $(\vec{p}, \vec{v})$, their dot product (or inner product) is divided by their multiplied norms:

$$cos(\vec{p}, \vec{v}) = \frac{\vec{p} \cdot \vec{v}}{\|p\| \|v\|}$$

Due to this normalisation, different text lengths do not matter for the calculation of the metric. Also, with occurrences assigned as vector values, cosine similarity can only range between 0 and 1, as negative word frequencies are impossible. In this case, 0 means that the vectors are orthogonal (maximally dissimilar) to each other, while 1 implies that they are identical. Its simple calculation, the straightforward interpretation and other useful properties make cosine similarity a popular metric. Most importantly, it is also semantically robust. The more words two given texts share, the more similar they are in meaning – a fundamental principle from distributional semantics, which is all the more important for the research question at hand (Firth 1957). Despite these merits, the approach is not without weaknesses. For instance, Li & Han (2013) showed that the metric is biased towards very frequently occurring terms. Furthermore, cosine similarity is based on a bag-of-words approach. That is, the order of words is not taken into account, which results in a loss of information. However, given the question at hand, the metric is ideal for operationalising the concept of regulatory novelty.

Term-weighting is of central importance, especially in the case of texts with a highly similar structure, such as EU policies. The high textual similarity stems from the fact that the structure and the choice of words follow a clear pattern in these legal provisions. The preamble is followed by the policy's actual articles, while standardised formulations, depending on the discretionary power, are used. Thus, these legal terms or even entire passages hardly convey any relevant information. Various weighting methods can be applied in this case, whereby the choice for the present analysis falls on the *TF-IDF* scheme. By multiplying the term frequency within a document by the inverse document frequency, relevant terms are identified. Terms that occur very frequently only in certain documents are weighted the most, while those terms that occur in (almost) all documents in the corpus are

weighted the least. Virtually all law-specific and general formulations are thus assigned very little to no weight. In contrast, policy-specific terms are weighted much more strongly. This not only enables a more substantial comparison in terms of content but also alleviates the problem of bias towards high cosine similarity values as described by Li & Han (2013).

After weighting, the cosine similarity with all previous policies is calculated to see how the given policy resembles the other regulatory instruments. Higher values represent higher similarity and thus less regulatory novelty and vice versa. However, this general comparison with all other policies would be flawed as this approach does not take policies' incremental nature sufficiently into account. Policies build on each other, requiring consideration of the temporal component. They can appear very dissimilar if they are separated by an extended period, even if they deal with the same issue. Over the years, new issues arise, and old provisions become obsolete. Conversely, the similarity between two policies that follow one another closely in time is assumed to be greater as it is unlikely that entirely new issues will arise or new policy instruments will be needed in a relatively short period of time.

For this reason, following the example of Pagliari & Wilf (2020), I introduce the metric of maximum cosine similarity (*MaxSim*), defined as the highest cosine similarity value shared by document d with a given document from the corpus, except for itself. Again, an example helps to illustrate the point, with the vector representing the cosine similarities between document d and five other documents:

$$d_i = \begin{bmatrix} 0.21 \\ \mathbf{0.89} \\ 0.33 \\ 0.80 \\ 0.54 \end{bmatrix} \tag{1}$$

In this case, only the maximum value of 0.89 is assigned the cosine similarity of document d compared to all other texts and not the averaged value across the whole vector. Thus, with *MaxSim*, very (temporally) distant texts with very low cosine similarity and seemingly high regulatory novelty do not cause the similarity metric to appear artificially high. Instead, novelty is assessed solely based on the textually closest text, measured by the weighted frequency of its terms. This procedure ensures that the cosine value provides a more accurate picture of how novel the given policy is.

5.2 Data

The analysis is based on data that was prepared as part of an SNF Spark project.[5] The raw data comprises a text corpus of 495 renewable energy policies of the EU, retrieved from EUR-Lex, the EU's official portal for legal texts.

In the first step of pre-processing, I selected those texts from the corpus that contained the words "renewable" or "energy" at least once. This rather simplistic procedure is justified by MaxSim selective characteristics as it discards all but one policy after calculating their textual similarities.

[5] See: https://epg.ethz.ch/news-and-events/EPGNews/2019/11/epg-researchers-acquire-snf-spark-grant.html

However, this keyword approach does not entirely prevent thematically unrelated policies from ending up in the resulting smaller corpus with 175 policies. Therefore, I manually inspected the headings of the new corpus' respective policies and removed those texts that had no topical relation to renewable energy policies. For example, Directive 2014/89/EU to establish a framework for maritime spatial planning or Regulation 2019/2152 on European business statistics were removed. The period covered by the relevant policies spans from 2000 to 2019 with a moderately fluctuating number of policies per year, as illustrated in Figure 1. Next, stopwords, punctuation, symbols and numbers were removed, and the words stemmed. Lastly, the the document-term matrix was weighted according to the *TF-IDF* weighting scheme.

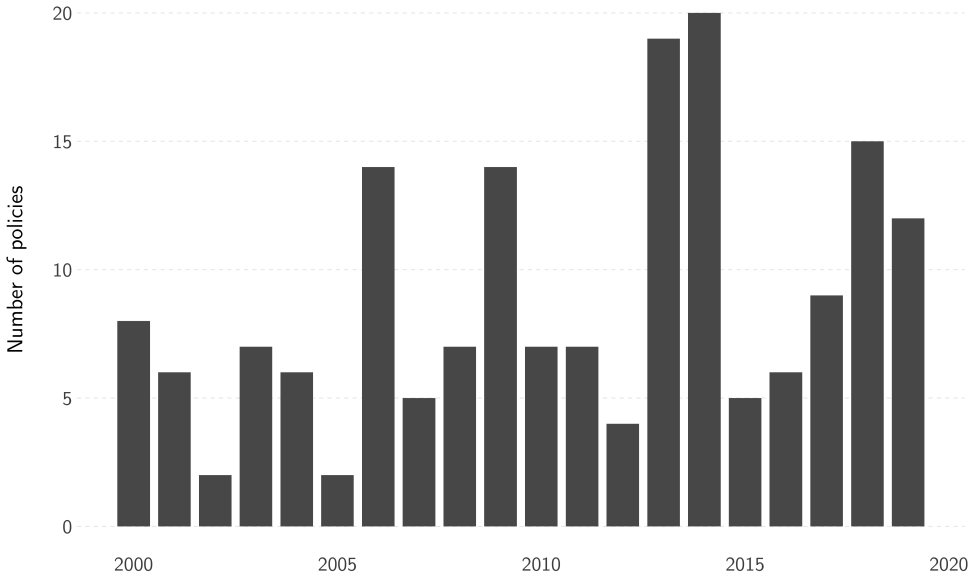

Figure 1: Number of renewable energy policies (2000-2019): In total, the final corpus comprises 175 policies, which to a lesser or greater extent contain legal provisions on renewable energies.

6 Results

Figure 2 shows the evolution of MaxSim from 2000 to 2019 with the red dots representing policies written after COP-21. The smoothing function (LOESS) shows a slight wave movement of MaxSim values with its peak around 2006 and a relatively linear trend since 2010. There seems to be no difference between the MaxSim values before and after COP-21. Both very similar and almost maximally dissimilar policies are present in the selected corpus, resulting in a high standard deviation of 0.25. As regards almost identical policies, these are often revisions of an older version. For example, the recast regulation on a community energy efficiency labelling was expectedly textually closest to its older version. These pairs confirm the ability of the cosine similarity and the MaxSim procedure, respectively, to identify thematically related texts merely based on weighted terms. It should be added that these pairs are not always symmetrical. The pool of paired policies is composed of 63% of all policies. For example, Directive 2009/72/EC concerning common rules for the internal market in electricity paired with four other policies.

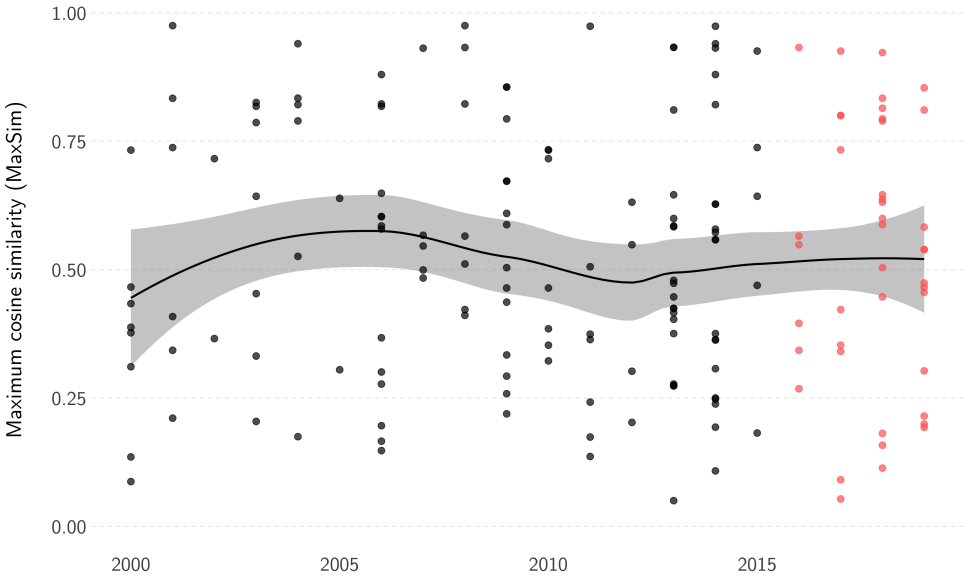

Figure 2: Evolution of *MaxSim* (2000-2019): The variance of MaxSim, the maximum cosine similarity of a policy compared to all other legal provisions, remains high over the entire period. There are both almost identical policies (MaxSim ≈ 1) and completely new policies (MaxSim ≈ 0) in the corpus. The smoothed trend line shows that MaxSim peaks around 2006 and has been following a relatively linear trend since 2010. Policies that were drafted after 2015, i.e. after COP-21, are coloured red.

6.1 Regression

Next, I examine whether the difference in MaxSim similarity before and after the Paris Agreement is significant and thus indicative of regulatory novelty. Since MaxSim assumes continuous values in the standard unit interval [0, 1], I use a beta regression to estimate the determinants of regulatory novelty. Not only considering that the outcome variable follows a beta distribution, these regression models also naturally account for features such as heteroskedasticity and skewness in the data (Cribari-Neto & Zeileis 2010). To simplify interpretations, I define regulatory novelty as 1-MaxSim in the regression specifications. This means that the higher the outcome variable, the higher its regulatory novelty. I define the dummy variable *afterparis* as the primary regressor, assigning it the value 1 if the policy in question was drafted after 2015. Next, I introduce three complexity measures as control variables. They account for text-based characteristics that could influence the regulatory process and were adopted from Hurka & Haag (2020) and Katz & Bommarito (2014), respectively. First, structural complexity indicates how many provisions the corresponding legal instrument contains. For this purpose, I compute the structural complexity of a policy based on the number its articles and their items. The title page of the document and the preamble, which lists objectives and timelines, are ignored. While some of these passages might be pertinent, in most cases their inclusion would artificially inflate the complexity without returning substantial relevance. Second, *relational complexity* describes the number of references within a document that refer both to articles within the same document and articles in other policies. The third control variable is that of *linguistic complexity*, inspired by information theory (Shannon 1948). Here the concept of entropy, a quantity describing the signal (or uncertainty) of a given text is applied. The formula is as follows:

$$-\sum_{w \in W} p_w log_2(p_w),$$

where p_w is the probability p of a token's occurrence in the given text W (Katz & Bommarito 2014, p.357). Thus, the more diverse a text is in its word choice, the higher its entropy and the corresponding uncertainty. As a final control variable, I add the cumulative logged number of policies issued up to that point. This considers that the higher the number of policies, the greater the probability that there is already a similar policy among them. The variable is additionally logged because, despite the higher overall probability of an already existing policy, its marginal effect is limited (Pagliari & Wilf 2020). This results in the following regression specification:

$$y_i = \alpha + \beta' afterparis + \gamma' C_i + \epsilon_i,$$

where y_i represents the transformed *MaxSim* value of document i, α is the constant, $\beta' afterparis$ is a dummy indicating if the policy was drafted after COP-21, $\gamma' C_i$ is a vector with the aforementioned control variables, and ϵ_i is the error term.

Table 1 shows the results of the model specifications. The first model specification considers the entire period from 2000 to 2019. It shows that the years following Paris are not associated with a higher regulatory novelty. The corresponding coefficient fails to meet the threshold of the 5% significance level. The same applies to the remaining regressors. None of the complexity measures correlates significantly with regulatory novelty. To better understand the evolution in more recent years, I narrowed the time window to the years after 2010 in the second model. With the relatively flat trend in the last decade, it is pertinent to examine how policies have changed since Copenhagen (COP-15). However, policies in the years after Paris are not associated with a statistically significant difference in regulatory novelty in this shortened period either. Again, the same is true for the other regressors. These results do not change when one considers that the drafting of new policies might have a certain time lag. If *afterparis* is coded so that only the years from 2016 take on the value one, the results do not change.

Table 1: Regression models

	Dependent variable:	
	Regulatory novelty	
	(2000-2019)	(2011-2019)
After COP-21	0.053	0.241
	(0.197)	(0.334)
Structural complexity (n articles)	−0.007*	−0.004
	(0.003)	(0.004)
Relational complexity (n references)	0.00004	0.00000
	(0.0003)	(0.0003)
Entropy	−0.250	−0.415*
	(0.176)	(0.252)
Number of previous policies (log)	0.044	−0.938
	(0.105)	(0.774)
Constant	1.858	7.883*
	(1.375)	(4.372)
Observations	175	97
Log Likelihood	18.229	9.065
Note:	*p<0.1; **p<0.05; ***p<0.01	

7 Discussion

Strictly based on the regression models, one cannot infer regulatory novelty after COP-21. Both models showed no significant differences in the maximum textual similarity of policies before and after Paris. Therefore, some aspects need to be examined more closely to evaluate the validity of cosine similarity for empirical operationalisation of textual similarity. For this purpose, the first part of the following discussion focuses on two policies in particular, namely the EU's two renewable energy directives (RED). The focus lies on these policies due to their pre-eminence concerning the EU's energy transition, but they are emblematic of the corpus' other policies. In the second part, I elaborate on broader conceptual difficulties and shortcomings. These include the thematic breadth of the corpus, the structure of the legal instruments at hand or the quantitative nature of cosine similarity. Finally, I show how this metric could be useful for future research.

Figure 3 shows the most common bigrams (terms composed of two words) of the 2009 and 2018 directives. Terms such as "energi", or "sourc" result from word stemming, i.e. the reduction to the part that gives words their lexical meaning. Furthermore, only the actual articles were taken into account for the analysis; the preambles and annexes were excluded. Based on the most frequent bigrams, the directives hardly differ. Renewable sources ("renew sources") and renewable energy ("renew energi") are mentioned most frequently in both directives, followed by energy sources ("energi sourc") in the 2009 directive and biomass fuel(s) in the 2018 directive. Furthermore, with this directive ("thi direct"), the many internal and external references (see relational complexity) become apparent.

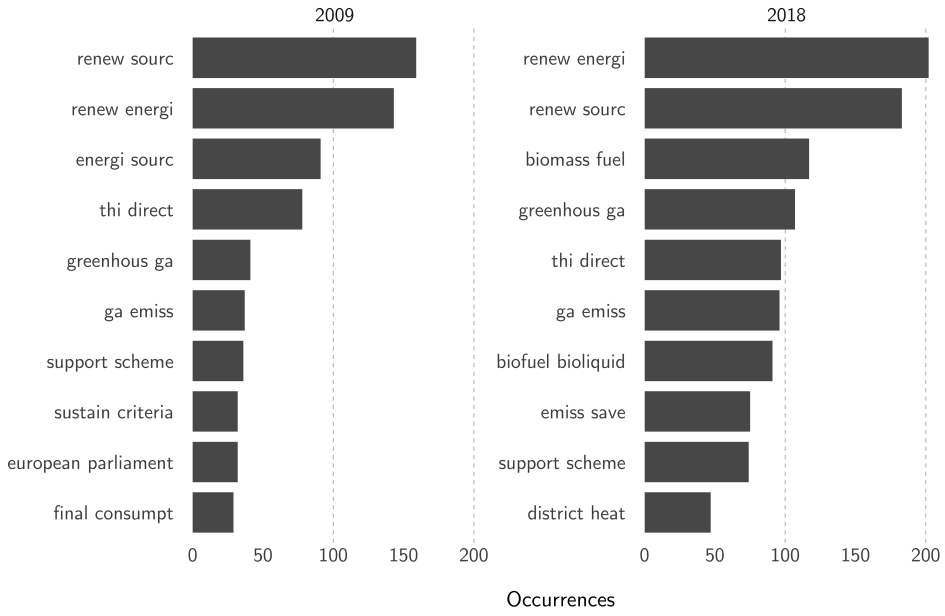

Figure 3: Top Bigrams: Based on the terms composed of two words, the 2009 and 2018 REDs are very similar. Renewable energy/sources are the most common terms, followed by internal references ("thi direct") and some directive-specific terms.

However, bigrams often do not adequately represent more complex composite words, which is why Figure 4 shows the most common trigrams. Differences in wording or terms now become apparent. In the 2009 directive, the trigram renewable energy sources is the most frequently mentioned term, while in the 2018 directive, this term was shortened to renewable sources (see previous figure). Another important term in the 2009 directive, spread across three trigrams, is *national renewable energy action plans*. At the time, these served to define the national targets for the share of energy from renewable sources by 2020. In the newer directive, greenhouse gas emissions/savings are mentioned more than twice as often as in its predecessor. The prominent role of Regulation (EU) 2018/1999 is also evident. That is, the Regulation on the Governance of the Energy Union and Climate Action, implementing the strategies and measures designed to meet the objectives and targets of the Energy Union consistent with the Paris Agreement. I will return to this at a later stage. Lastly, the trigram "renew energi commun" trigram, which stands for the *renewable energy communities* (REC) as a legal entity, deserves particular attention. RECs were introduced in the 2018 directive to provide environmental, economic or social community benefits for the (local) stakeholders of renewable energy projects. Specifically, this new type of provision explicitly enables local communities "to consume, store or sell RE generated on their premises" (p. 1), making it an essential cornerstone for the successful energy transition, based on a decentralised system of energy production and use (Lowitzsch et al. 2020).

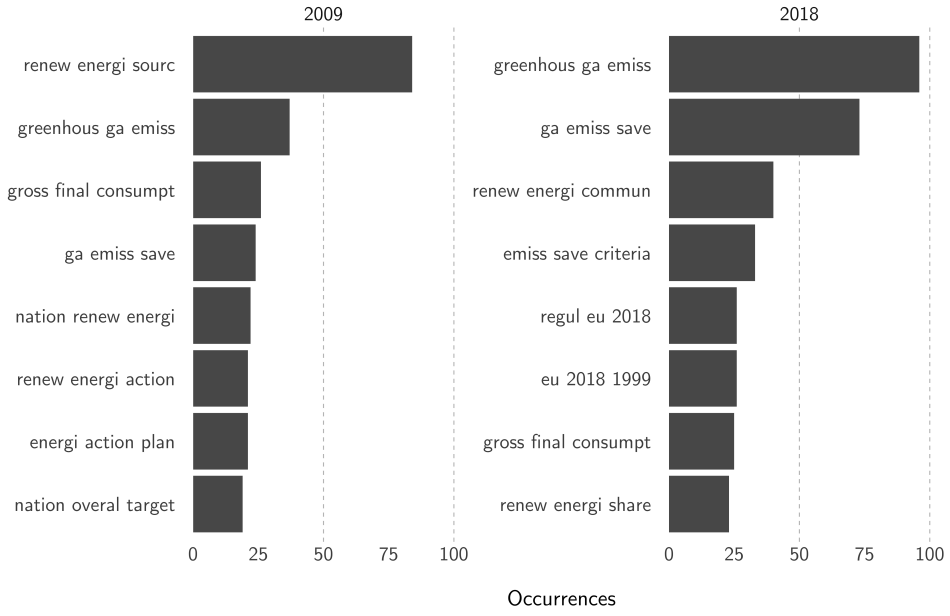

Figure 4: Top Trigrams: Looking at the most frequent trigrams, differences between the two directives become apparent. While the most frequent terms are still similar, the directive-specific focuses become clear. The RED 2009 introduced the national renewable energy action plans, while the RED 2018 puts a focus on renewable energy communities.

This brief, slightly more elaborate comparison shows that such novel provisions as the RECs only become apparent through in-depth examination. While the cosine similarity facilitates identifying thematically similar texts and reliably does so, it does not convey information in what respects these texts differ. Only through case studies can it be shown how legal instruments such as the first and second renewable energy directives differ in content. Consequently, the present results once again plead for combining quantitative with qualitative approaches. To conclude that there is a lack of regulatory novelty merely based on the null findings of the regression models would fall short. Although cosine similarity can reveal whether regulatory novelty manifests itself, it can hardly reveal its extent and thematic changes and requires complementary, qualitative approaches to better understand the evolution of their content. Also, with its bag-of-words (BOW) approach, i.e. not taking the sentence structure into account, it neglects critical structural components of policies. While this measure reliably links thematically similar texts, it does not provide information about the relationship between the regulating entity and its subjects. Especially in policy-making, however, this relationship is primordial. How much discretionary power do the agents of a policy have? How is it delegated? Using part-of-speech tagging, Vannoni et al. (2021) could show how bureaucratic discretion and executive delegation have changed over the years in the US. Adapted to other policies, such approaches that consider sentence structure may provide further compelling insights into how legal provision evolves.

Furthermore, thematic delimitation is essential to exploit the potential of cosine similarity. Although more or less related to renewable energy, the policy areas included in the present corpus were so diverse that the thematic differences between policy areas were weighted more heavily at the expense of the smaller textual nuances within a policy area. The MaxSim approach indeed ensures that only not completely unrelated texts are compared. Simultaneously, the accurate recognition of the content based on its word embeddings is very coarse since few texts in the corpus are clustered in numerous and different topics. In this respect, Pagliari & Wilf's (2020) recommendation to expand their approach to different regulatory bodies only makes sense if the analysis is restricted to a narrow thematic framework. Otherwise, the researcher runs the risk of representing the textual evolution within a policy area too inaccurately.

In addition, the metric of cosine similarity, and thus the operationalisation of regulatory novelty, relies on the most specific legal instrument possible. By their very nature, regulations and especially directives of the EU, despite their binding character, are formulated as guidelines rather than specific rules. Even more so as the member states are responsible for their implementation. Their broad and very standardised form leads to low textual variance, making it difficult to reconstruct the individual legal provisions, let alone their textual evolution. While weighting schemes such as the TFIDF scheme used in the analysis help isolate keywords, they remain too crude to inform satisfactorily about regulatory novelty.

Another reason for the lack of novelty may be the delay in policy-making. As already indicated, Regulation (EU) 2018/1999, which outlines the strategies for implementing the COP-21 goals, was only included in the 2018 renewable energy directive. This does not mean that the alleged paradigm shift after Paris is not reflected in policies earlier, but it is well-known that a certain amount of time passes before legal provisions start to implement changes. In this respect, additional years would have to be added to determine whether, with a slight delay, regulatory novelty eventually manifests itself.

Despite these major and minor shortcomings, it is essential to emphasise the usefulness of cosine similarity for text analysis. As has been shown, the metric is strongly dependent on a narrow thematic framework concerning the identification of regulatory novelty. Pagliari & Wilf (2020) succeed in credibly showing regulatory differences after crises due to a narrow thematic frame as well as a narrow scope concerning the regulatory bodies. In the present case, this is much more difficult due to the topical variety of the corpus. Simultaneously, this was indispensable to ensure that the findings carried some external validity. However, coupled with other approaches, cosine similarity can answer important questions, especially in policy studies. For example, Hager & Hilbig (2020) use this metric to show how agenda-setting by the political elite is shaped by public opinion. They operationalise this mechanism by comparing the textual similarity of German politicians' speeches before and after the publication of so-called public opinion reports. Others link the similarity of speeches or legislative texts with network analysis to understand policy diffusion better. Garrett & Jansa (2015) show that interest groups in the USA primarily use model legislation, i.e. easily adaptable legal instruments, to promote their policy agendas in different states. Here, cosine similarity formed the basis for examining the similarity of the adopted legislation in the different states. Lastly, it is important to consider that this data-driven approach enables a standardised examination of a large number of documents, which would hardly be possible with a qualitative approach. It allows to explore longer periods of time or larger volumes of text in an efficient way without relying on any coding schemes or other manual input. Thus, this theory-agnostic metric is ideal, applied as a preliminary to a narrower qualitative analysis, to cope with the ever-increasing volume of texts and their contents for the study of policies' temporal dynamics.

8 Conclusion

Using a text-as-data approach, I wanted to explore the temporal dynamics of policies in this paper. For this purpose, I analysed the renewable energy policies of the EU by tracing their textual similarity over time. More specifically, I was interested in the concept of regulatory novelty. That is, to what extent a new policy builds on or deviates from an existing policy. For this, the Paris Agreement (COP-21) formed the focal point, since it has been widely regarded as a paradigm shift in the joint endeavour against climate change. Thus, I I formed my research question as whether the EU's renewable energy policies' textual similarity had changed significantly in its aftermath. The underlying assumption was that a conceptual shift in the form of new policy instruments and provisions also resulted in a textual shift. New provisions entail new or at least different vocabulary, making textual similarity a reliable proxy for policy change.

However, the various regressions models employed in this study show no significant difference in the policies following the Paris Agreement, compared to the ones before. Possible reasons for this are manifold: due the thematic breadth of the corpus, the topical differences were highlighted considerably more than textual subtleties. Furthermore, the legal provisions' highly standardised form entailed little variance in the terms embedded, which could only be partially controlled for. Subsequently, I compared the EU's two renewable energy directives to illuminate the shortcoming of cosine similarity. Nevertheless, this metric has great potential for policy studies, especially as a preliminary for qualitative approaches, due to its versatility, its data-driven nature and ease of interpretation.

References

Birchfield, V. L. & Duffield, J. S., eds (2011), *Toward a Common European Union Energy Policy*, Palgrave Macmillan US, New York.

Blumenau, J. (2019), 'The Effects of Female Leadership on Women's Voice in Political Debate', *British Journal of Political Science* pp. 1–22.

Boasson, E. L. & Wettestad, J. (2016), *EU climate policy: industry, policy interaction and external environment*, first issued in paperpack edn, Routledge, London New York.

Christoff, P. (2016), 'The promissory note: COP 21 and the Paris Climate Agreement', *Environmental Politics* **25**(5), 765–787.

Cribari-Neto, F. & Zeileis, A. (2010), 'Beta Regression in *R*', *Journal of Statistical Software* **34**(2).

Falkner, R. (2016), 'The Paris Agreement and the new logic of international climate politics', *International Affairs* **92**(5), 1107–1125.

Firth, J. R. (1957), 'A synopsis of linguistic theory, 1930-1955', *Studies in Linguistic Analysis* . Publisher: Basil Blackwell.

Garrett, K. N. & Jansa, J. M. (2015), 'Interest Group Influence in Policy Diffusion Networks', *State Politics & Policy Quarterly* **15**(3), 387–417.

Hager, A. & Hilbig, H. (2020), 'Does Public Opinion Affect Political Speech?', *American Journal of Political Science* **64**(4), 921–937.

Hall, P. A. (1993), 'Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain', *Comparative Politics* **25**(3), 275.

Hurka, S. & Haag, M. (2020), 'Policy complexity and legislative duration in the European Union', *European Union Politics* **21**(1), 87–108.

Katz, D. M. & Bommarito, M. J. (2014), 'Measuring the complexity of the law: the United States Code', *Artificial Intelligence and Law* **22**(4), 337–374.

Kern, F. & Rogge, K. S. (2016), 'The pace of governed energy transitions: Agency, international dynamics and the global Paris agreement accelerating decarbonisation processes?', *Energy Research & Social Science* **22**, 13–17.

Kinley, R. (2017), 'Climate change after Paris: from turning point to transformation', *Climate Policy* **17**(1), 9–15.

Li, B. & Han, L. (2013), Distance Weighted Cosine Similarity Measure for Text Classification, *in* H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li & X. Yao, eds, 'Intelligent Data Engineering and Automated Learning – IDEAL 2013', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 611–618.

Lindblom, C. E. (1959), 'The Science of "Muddling Through"', *Public Administration Review* **19**(2), 79.

Lowitzsch, J., Hoicka, C. & van Tulder, F. (2020), 'Renewable energy communities under the 2019 European Clean Energy Package – Governance model for the energy clusters of the future?', *Renewable and Sustainable Energy Reviews* **122**, 109489.

Nilsson, M. (2011), EU Renewable Electricity Policy: Mixed Emotions toward Harmonization, *in* V. L. Birchfield & J. S. Duffield, eds, 'Toward a Common European Union Energy Policy: Problems, Progress, and Prospects', Palgrave Macmillan US, New York, pp. 113–130.

Pagliari, S. & Wilf, M. (2020), 'Regulatory novelty after financial crises: Evidence from international banking and securities standards, 1975–2016', *Regulation & Governance* .

Savaresi, A. (2016), 'The Paris Agreement: a new beginning?', *Journal of Energy & Natural Resources Law* **34**(1), 16–26.

Schmidt, T. S. & Sewerin, S. (2017), 'Technology as a driver of climate and energy politics', *Nature Energy* **2**(6), 17084.

Schmidt, T. S. & Sewerin, S. (2019), 'Measuring the temporal dynamics of policy mixes – An empirical analysis of renewable energy policy mixes' balance and design features in nine countries', *Research Policy* **48**(10).

Shannon, C. E. (1948), 'A Mathematical Theory of Communication', *Bell System Technical Journal* **27**(3), 379–423.

Solorio, I. & Bocquillon, P. (2017), EU renewable energy policy: A brief overview of its history and evolution, *in* I. Solorio & H. Jörgens, eds, 'A Guide to EU Renewable Energy Policy', Edward Elgar.

Vannoni, M., Ash, E. & Morelli, M. (2021), 'Measuring Discretion and Delegation in Legislative Texts: Methods and Application to US States', *Political Analysis* **29**(1), 43–57.

Viñuales, J. E., Depledge, J., Reiner, D. M. & Lees, E. (2017), 'Climate policy after the Paris 2015 climate conference', *Climate Policy* **17**(1), 1–8.

Wilkerson, J. & Casas, A. (2017), 'Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges', *Annual Review of Political Science* **20**(1), 529–544.