This is the author's original manuscript of the article. The version of record was published in the International Journal of Social Research Methodology, and is available online at https://doi.org/10.1080/13645579.2019.1576317. Please cite as Jacobs, T. & Tschötschel, R. (2019). "Topic Models meet Discourse Analysis: a quantitative tool for a qualitative approach." *International Journal of Social Research Methodology*

Topic Models meet Discourse Analysis: a quantitative tool for a qualitative approach

RESEARCH ARTICLE

Keywords: discourse analysis, topic modelling, text analysis, hegemony, digital humanities, computational social science

Thomas Jacobs (corresponding author)

Centre for EU Studies, Ghent University, Ghent, Belgium

Centre for EU Studies (CEUS), Korte Meer 1, 9000 Gent, België

Thomas Jacobs is a PhD candidate at the Centre for EU Studies, Ghent University. His main research interests include EU trade policy, international political economy, and political communication. Most of Thomas' work focuses on discourse, meaning-making, signification, and strategy in political praxis, departing from a poststructuralist perspective, more specifically the discourse-theoretical tradition of Laclau and Mouffe. Prior to entering academia, he worked as a consultant at a Brussels-based communications and public affairs firm.

Robin Tschötschel

Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, The Netherlands

Amsterdam School of Communication Research (ASCoR), Postbus 15793, 1001 NG Amsterdam, Nederland

Robin Tschötschel is a PhD candidate at the Amsterdam School for Communication Research, University of Amsterdam, and holds a teaching position at the Politics, Psychology, Law, and Economics College at the same university. His research is focused on the structure and effects of public debates about climate change and economic growth, and utilises a range of mixed methods approaches, many of which draw from the Computational Social Science toolkit.

Topic Models meet Discourse Analysis: a quantitative tool for a qualitative approach

Quantitative text analysis tools have become increasingly popular methods for the operationalization of various types of discourse analysis. However, their application usually remains fairly simple and superficial, and fails to exploit the resources which the digital era holds for discourse analysis to their full extent. This paper discusses the discourse-analytic potential of a more complex and advanced text analysis tool, which is already frequently employed in other approaches to textual analysis, notably topic modelling. We argue that topic modelling promises advances in areas where discourse analysis has traditionally struggled, such as scaling, repetition, and systematization, which go beyond the contributions of simpler frequency and collocation counts. At the same time, it does not violate the epistemological premises and methodological ethos of even the more radical theories of discourse, we will demonstrate. Finally, we present two small case studies to show how topic modelling — when used with appropriate parameters — can straightforwardly enhance our ability to systematically investigate and interpret discourses in large collections of text.

Keywords: discourse analysis, topic modelling, text analysis, corpus linguistics, methodology, hegemony

Introduction

This paper contends that topic modelling, a method for text-mining in large corpora, can resolve part of the methodological troubles haunting discourse analysis, one of the main theoretical frameworks for studying meaning-making in text and speech.[1] Discourse analysis aims to understand how ideas and realities are socially and

[1] The authors would very much like to thank Julie Birkholz, Anke Wonneberger, Christophe Verbruggen, and Jan Orbie and the anonymous review for their comments on earlier drafts of this article. We are also very much indebted to the participants of the workshop ‘Quantitative Tools for Qualitative Analysis: Computational Social Science meets Discourse Analysis’ organised by us, Julie Birkholz, Anton Törnberg, and Petter Törnberg during the

discursively constructed, yet the insights it can achieve within the rich theoretical frameworks that fall under its banner are often limited by practical barriers to the empirical study of discourse. Whereas many other popular forms of qualitative text analysis, such as content analysis, achieve an impressive methodological meticulousness but undertheorize the process of meaning-making, discourse analysis suffers the reverse problem: some types of discourse analysis have been alleged to suffer from a fully-fledged methodological deficit (Howarth & Torfing, 2005, 25, 316-22), and the field in general has been claimed to direly need more systematic and rigorous operationalization (Antaki *et al.*, 2003). We argue that topic modelling can help discourse analysis conquer some of the practical barriers standing in its way, and contend that it can contribute to the achievement of more methodological rigour and systematicity in the study of meaning-making.

The most important methodological perks offered by the use of corpora and large-scale text analysis tools are well known: they reduce alleged researcher prejudice, allow for the precise study of more fine-grained and subtle aspects of language use, facilitate methodological triangulation, and make possible systematization, large-scale analysis, and the study of repetition and incremental change (Baker, 2006, 10-14). The potential of automated text processing tools for discourse analysis follows directly from these advantages, as they are all situated in areas where 'artisanal' discourse analysis based on close reading has certain hard limits (Antaki *et al.*, 2003). That is not to disparage careful manual study of the text, which will always be the core business of

---

'European Symposium on Societal Challenges in Computational Social Science' at the Alan Türing Institute in the British Library (London) in November 2017. Without their stimulating and enthusiastic comments, ideas, suggestions, and insights, this article would never have materialized.

discourse analysis. But the human brain can only absorb so much text, detect a certain level of subtlety and nuance, and notice evolutions in language use up to a certain scale. In every one of these areas, automated tools help us transcend these limits, making new insights and novel forms of discourse analysis possible. Computer-assisted corpus analysis, in other words, rather than altering the nature of discourse analysis, breaks down and pushes forward the boundaries of what it can do.

Yet despite their considerable added value, the actual usage of corpora in discourse analysis is not very advanced in terms of sophistication. Most discourse analyses that study large text corpora employ fairly simple tools that count words, collocations, and concordances. More complex models and algorithms such as topic modelling have only entered into the consideration of discourse analysts very recently, and to a limited degree (Levy & Franklin, 2014; Tornberg & Tornberg, 2016a; 2016b; Munksgaard & Demant, 2016; Jaworska & Nanda, 2016 are some of the few examples of the explicit use of topic modelling for discourse analysis). This is noteworthy, since topic modelling has been around since 2003. In defence of the discourse analysis community, though, the ignorance between topic modelling and discourse studies is mutual. Scholars and computer scientists specialised in topic modelling have, with few exceptions, shown little interest in developing a deeper understanding of how their algorithms model language use and what theory of meaning-making topic modelling implicitly postulates.[2] This paper therefore constitutes an attempt to put a halt to the

[2] A few exceptions notwithstanding (e.g. DiMaggio, Nag, and Blei, 2013), reflection about the model of language implied by topic modelling is relatively underdeveloped in the methodological literature. The most prevalent ideas include metaphors like the bag-of-words model and the notion that documents are composed of combinations of topical discussions (e.g. Mohr and Bogdanov, 2013).

reciprocal disregard between the topic modelling and the discourse analysis communities.

As topic modelling is inherently a method, and as discourse analysis is principally conceived of as a theoretical framework in this paper, our argument will predominantly take the form of explaining how the former can help the latter achieve its research objectives. Our core aim is to demonstrate how topic modelling can extend what discourse analysis can empirically achieve, and dispel some theoretical, methodological, and practical objections against cross-pollination between both traditions. In addition, we equally maintain that users of topic models can benefit from engaging with theories of discourse, as they help them interpret their results and explicate their often-implicit understanding of meaning-making in language. This way, we seek to broaden the prevalent understanding in digital text analysis of text as a unit of analysis, instead of as a unit of meaning. In this double effort, the emphasis will be on compatibility, mutual added-value, and theoretical fit.

As for the structure of this paper, we first outline the two approaches in detail, providing an overview of their ontological and epistemological premises and characteristics. These characteristics serve as a basis for the refutation of a number of theoretical objections against the use of topic modelling for discourse analysis in the second section. We contend that the premises of topic modelling in fact fit remarkably well with the ontological and epistemological stances taken by most discourse theories. The third section offers several arguments as to how topic modelling extends what discourse analysis can see and argue: we will explain why topic modelling is particularly suited to study questions of hegemony; that it assists verification; and that the level of systematicity it achieves helps us track change and continuity in language use. The fourth and final part of this paper contains two practical examples of the operationalization of topic modelling for discourse-analytic purposes. The case studies are corollary to the argument that topic modelling can make tangible, effective contributions to discourse analysis and show that the operationalization of topic modelling can be fairly straightforward on a practical level.

What are Discourse Analysis and Topic Modelling?

Drawing on large synoptic overviews of the tradition by Jorgensen & Phillips (2002), Blommaert (2005), Rogers (2013), and Gee (2014), we can say that discourse analysis is essentially concerned with studying communication and meaning-making in context. A discourse analysis is an attempt to describe and understand the processes through which meaning is formed, conveyed, and interpreted in a concrete situation. Often, this analysis is accompanied by a critical and normative assessment of how these communicative processes affect the social world in which we live our daily lives — Critical Discourse Analysis (CDA) is a prime example of this.

More specifically, many forms of discourse analysis, such as the Essex School of Discourse Theory or Derridaean deconstruction, are indebted to a poststructuralist understanding of the generation of meaning-making, seeing it as *relational open practice*. The relational component of this definition entails that concepts only become meaningful in relation to other concepts, rather than by corresponding to some external reality. The openness component implies that these relations are not necessary or pre-determined, but contingent, non-necessary and fundamentally incomplete. They only exist in the form they acquire in the articulations of speakers. Finally, the "practice" component implies that meaning is generated and achieved in a specific context, that it is something that is formed, represented, and made by actors, rather than something that exists independent of them.

Despite believing that meanings are ultimately open and shaped by the actors articulating them, all forms of discourse analysis recognize that some meanings do seem to be so common and conventional that they appear as normal and natural. This is explained through a final crucial concept that is key to many forms of discourse analysis, *hegemony*. One could say that a hegemony entails the privileging of one mode of interpretation over all other possible modes of interpretation within a particular field (e.g. “responsible fiscal policy” is usually understood as debt reduction, even though it could conceivably also mean taxing the rich more and the middle and lower classes less). More simply, hegemony refers to a dominant, normalized way of understanding the world which in turn renders some ways of talking and acting more conventional, acceptable, and seemingly logical.

Which type of data and class of questions discourse analysis tackles, depends on the variety and flavour of discourse analysis one uses. CDA, for instance, mostly looks at very concrete and tangible interaction or statements that involve an (implicit) political dimension, whereas Discourse Theory reflects on large-scale systems of thought such as racism or neoliberalism. Yet, broadly speaking, most forms of discourse analysis involve the empirical study of text, inspired by a set of assumptions about how meaning-making works, aimed at deconstructing and understanding how the ideas formulated in a text are constituted.[3]

---

[3] Whereas most discourse-analytic approaches involve the study of text, some stress the need to go beyond what is captured in text and speech, and to look at practices and actions. For obvious reasons, we leave this type of multimodal discourse analysis aside in this paper.

Topic modelling

Topic modelling, meanwhile is a method that aims to reduce the complexity of a large corpus by representing each text as a combination of 'topics'. The name is slightly misleading though: topics are clusters of words that reappear across texts, but the interpretation of these clusters as themes, frames, issues, or other latent concepts (such as discourses) depends on the methodological and theoretical choices made by the analyst – as we will discuss below. While topic modelling does not have an in-built model of how humans use language, the following intuitive idea helps understand how the method works.[4]

Humans have diverse patterns of language use at their disposal to cover different subjects. The number of ways in which we communicate is non-deterministic and nearly infinite, and not all of the words associated with a subject, nor all the different ways of talking about it, are used in every situation. Furthermore, there are many words that can obtain different meanings, depending on their context and usage. Using this idea, a piece of text (a written document, or a transcript of speech) can be represented as the outcome of first selecting subjects, then selecting ways of speaking about them, and finally selecting some words associated with that manner of speaking. Topic modelling can be understood as a reversal of this process in which the algorithms use the observed distributions of words across texts in the corpus to infer non-exclusive clusters typically used in common — each representing a mode of speech about a specific subject.[5]

---

[4] The existing literature usually explains the method using a simpler, yet similar heuristic (eg. Mohr and Bogdanov, 2013)

[5] That is, if the right parameters are set. If too few topics are chosen, topics might cover a subject in total, a more abstract meta-subject (e.g. politics, rather than foreign policy), or a genre of text. We will develop this central point in greater detail below.

Practically, a topic modelling analysis returns three main results to the user (for examples, see the illustrative cases presented below). The first result assigns all words in the corpus a probability for each topic, by ranking them (using the heuristic discussed above) according to the probability that they represent the topic in the corpus (the topic-term matrix). Depending on the parameters used, the first five to twenty words are seen as roughly representative of a topic, and the topic is essentially equated to this list of 'top words'. This output is the main resource to interpret topics and study the relations between them. The second output, the so-called document-topic matrix, specifies how much of each text is made up of each topic. This information can be combined with contextual data about the texts (author, date) to facilitate comparisons across actors or diachronic analysis. Finally, the algorithms produce a precise overview of which topic each individual word in each text has been assigned to. This helps the analyst grasp the topic-specific meaning of each word and the contextual meaning of each topic.

As it departs solely from the texts, the method is fully theory-agnostic and inductive. Hence, a topic model is completely open to interpretation in function of the model's parameters and the larger theoretical framework it operationalizes. This feature is shared across the various statistical models and algorithmic procedures available to scholars that want to use topic modelling. The most common models build on Latent Dirichlet Allocation (LDA), a method developed by Blei, Ng, and Jordan (2003), but in the past years, this model has been extended and elaborated. One of the cases presented in this paper uses the original model, sometimes called *vanilla LDA,* while the other one draws on *structural topic modelling*, which integrates more recent advancements in computer-assisted text processing (Roberts, Stewart & Tingley, 2013).

Typically, analysts using topic modelling seek to identify a number of topics of interest and use them to quantitatively investigate the corpus of texts, measuring the space devoted to specific topics over time or by different actors. In this procedure, topics are mostly treated as measures of content or issue salience (e.g. Jacobi, van Atteveldt & Welbers, 2015), or as framings of issues (e.g. Boydstun *et al.*, 2013; DiMaggio, Nag & Blei, 2013). Yet, as we have argued, discourse analysis focuses more fundamentally on the discursive constitution of issues and frames, rather than on their prevalence. The first question we have to answer then, is if and under what circumstances topics can contain bits of discourses instead of bits of content?

The compatibility of topic modelling and discourse analysis

While the above description of how topic modelling disassembles and represents text might already sound promising to scholars familiar with discourse-analytic views of meaning-making, we want to render this promise explicit and show that the theoretical underbelly of topic modelling indeed warrants its use as a tool for discourse analysis. We follow two lines of argument in this regard. At a meta-theoretical level, we find that there is good match between the assumptions underlying topic modelling, and the view of discourse as a relational, open practice of meaning-making. At an epistemological level, we argue that the methodological idea behind topic modelling — how it is designed to generate knowledge about the texts and the words in the corpus — fits the analytical process of doing discourse analysis.

The large effort we make to stress the theoretical compatibility of discourse analysis and topic modelling may seem like a rather philosophical exercise, but we strongly believe it is not. Since many forms of discourse analysis adhere to the idea that meaning is exclusively symbolic and generated solely in language and practice, external validation is often epistemologically impossible for discourse-analytic studies, as they deny that discourses necessarily correspond to an external reality. As there is no possibility for external validation, internal validity is crucial if discourse analysis is to avoid the pitfall that 'anything goes' in analytical practice (Antaki *et al.*, 2003). This is achieved by demonstrating that, while the assumptions upon which the analysis rests are inevitably subjective, they are mutually supportive, form a coherent theory, and, crucially, are applied in a methodologically cogent and correct way to the case at hand (Marttila, 2015, 105-114). Our argument over the following pages intends to make this type of demonstration of internal validity for discourse analyses that work with topic modelling methods.

Meta-theoretical fit

As discussed, most forms of discourse analysis consider the meaning of words to be relational and open. This entails that meaning arises from the context a word is employed in and that it is not an inherent feature of the word itself. Topic modelling corresponds well with this view of language and meaning, we argue. As a topic is a probability distribution over all the words used in the original corpus, each word in principle figures in each topic, and its meaning varies between topics. It is the analyst's task then, to interpret the meaning of a topic based on how it ranks terms and how it relates to other topics. Similarly, the meanings of a word are topic-specific and based on the other words that appear in the topics in which it features prominently. These points make that topic modelling as a method aligns well with discourse analysis' assumptions of relationality and openness, as

(1) topic modelling explicitly models 'polysemy' (cf. DiMaggio, Nag, and Blei, 2013), the notion that words can obtain multiple meanings depending on the context they are used in. In fact, what topic modelling does can be summarized as tracing the multiplicity of contexts of every word in the corpus — independent of the meaning a word obtains in other fields or most commonly

assumes. This implies that topic modelling shares the idea of openness of meaning inherent to discourse analysis.

(2) topics themselves obtain their meaning through i) the relations they establish between the words contained in them, ii) the relations words appearing in multiple topics establish between these topics, and iii) through frequent co-occurrence with other topics. Similarly, words obtain meaning by being linked to other words in multiple topics. Thus, topic modelling shares the idea of the relationality of meaning inherent to discourse analysis.

While introducing topic modelling, we mentioned that topics could be interpreted as frames, themes, et cetera, but stressed that the most appropriate interpretation depends on how the method is used – in other words, on the analyst's choices. While this blurriness regarding the status of a topic and what a topic model actually represents may be seen as a disadvantage, this paper argues the opposite, claiming that it gives topic modelling a remarkable methodological polyvalence. We maintain that topic models should be constructed with specific research objectives in mind, rather than with statistical optimization, because we believe that what a topic model tells us and shows us, depends to a large degree on the research questions one tries to answer through the model and on the data one analyses. Hence, the parameters of the model should be chosen so that they facilitate the best possible answer to those research questions, rather than to achieve maximal statistical fit and significance.[6] Instead of adapting research questions so that they can be answered through topic modelling, topic models should be built and interpreted in a way that answers the research question.

[6] In fact, statistical limitations of the technique lead to multiple "local modes", which need to be investigated and compared by the analyst (cf. Roberts, Stewart, & Tingley, 2016).

The variety of interpretations for what a topic represents is strongly interlinked with the fact that the number of topics is usually selected a-priori by the analyst[7]. This last point has created a great deal of controversy over how to select the "right" or "natural" number of topics (Arun *et al.*, 2010; Wallach *et al*., 2009; Zavitsanos *et al.*, 2008). In our view, this controversy cannot be solved by using quantitative measures of statistical topic quality alone; the choice ultimately depends on how the analyst wants to interpret the topics. While some of these statistical measures are still useful (for making a pre-selection of candidate models), we stress the role of qualitative interpretation and of the demands of the research design when selecting the number of topics. No matter how fine-tuned the parameters are, some choices always remain subjective calls to be made by the researcher. A reflexive, conscious handling of subjective choices is the best the analyst can achieve, and this paper aims to provide a blueprint for doing so when using topic modelling as a method for discourse analysis.

Having established that topic modelling as a method fits the way discourse analysis wants to study meaning-making as an open and relational practice, the crux is now to design topic models so that they can trace discourses. Our hypothesis is that this becomes possible if a corpus is coherent enough thematically and stylistically, and if the overall number of topics is made large enough. In these circumstances, most topics will no longer list the various themes or subjects covered in the corpus, but will instead contain more fine-grained and nuanced aspects of language use. No matter which higher-level entities the analyst favours at a more aggregate level (subjects, frames, narratives, etc.), by increasing the number of topics or the thematic and stylistic

[7] In some statistical approaches, the analyst chooses other parameters that influence the outcome of the modelling process in similar ways.

coherence of the corpus, these can be decomposed into topics containing combinations of words that can be interpreted as the various discursive units through which those higher-level entities are constructed and composed.

This process of decomposition will start at a lower number of topics, the more coherent a corpus is. If a corpus only contains texts from a single genre and discussing a specific set of subjects, there will be fewer higher-level entities and thus the process of decomposition will start at a lower number of topics. The number of themes present in a corpus containing only trade policy speeches (as in the first case study) is different from the number of themes in a random collection of journalistic articles, opinion pieces, and advertisement about a variety of issues. Hence, the decomposition of thematic and issue topics into topics containing fragments of language use will start earlier in the former than in the later corpus, if we gradually increase the number of topics.

Simply put, we maintain that by using a high number of topics, by focusing on one well-delineated meta-subject (such as trade policy or the national economy), and by using a corpus that features only a single genre of texts (speeches, newspaper articles), topic modelling becomes a useful tool for discourse analysts. This hypothesis is demonstrated by the case studies at the end of this paper, and it has already implicitly applied in the literature (Tornberg & Tornberg, 2016a, 6-7; 2016b; Munksgaard & Demant, 2016), but our most important arguments to back up this claim, are theoretical.

Crucially, we can illuminate the process through which ‘subject’ and ‘theme’ topics decompose into ‘discourse’ topics by drawing attention to the fact in topic modelling, documents are not assigned to one topic, but are seen as a combination of a number of topics (Grimmer & Stewart, 2013). How to interpret topics hinges then on the number of topics selected, as this number affects the “granularity” of the decomposition (cf. Maier et al. 2018). The logic behind this is simple: if the number of topics ascribed to a single document increases as a result of an increase in the overall number of topics in the topic model, this obviously does not increase the number of subjects or issues discussed in a document. Rather, the number of topics covering each subject mentioned in the document increases, with the different topics in which one subject features each containing different aspects of this subject, different ways of representing it, and different ways of talking about it.

Hence, increasing the number of topics present in a document by increasing the overall number of topics in the corpus turns that document from a collection of themes into a collection of patterns of language use representing those themes, each pattern featuring in a topic.[8] In other words, the higher-level entities topic modelling recognizes in a corpus, such as subjects, frames, or narratives (which appear when the number of topics is small), can be decomposed into constitutive smaller-level entities (which appear when the number of topics is large) by increasing the overall number of topics. Our claim is evidently, that in some cases, it is possible to interpret these smaller-level entities as discursive elements with the help of discourse analysis.

If that is indeed the case, we can trace how the various discourses in a corpus are constructed, and where and when they feature. As will be apparent from the first case study, a single discourse often exists out of discursive elements that appear in several topics. This means that by studying the relations between topics (both in terms of quantitative co-occurrence and in qualitative connection), we can lay bare how

[8] While partially dependent on the source material used and the corpus' pre-processing, it is very common that some language patterns contained in a topic have little interpretative value. This number will evidently increase as the overall number of topics increases, but as these 'meaningless' topics are commonly ignored in small-k models, there is no reason not to skim over them in large-k models.

discourses are assembled and configured out of smaller discursive elements. Similarly, if we study when discourses-qua-topics are used by whom, we can reveal patterns of speech used at particular points in time by particular groups.

It is important to note that topic models do not automatically conduct a discourse analysis when the number of topics are increased; the topics of larger topic models do not by definition contain discursive elements. We merely contend that what they contain can be interpreted as discursive elements, if we understand the relations between words they reveal through a discourse-analytic lens. Increasing the number of topics thus does not necessarily decompose thematic topics into discourse topics. In some cases it decomposes them into something discourse analysis can work with, but discourse-analytic interpretation is needed to make sense of them and to tease out the discursive elements they contain. As such, topic modelling does not do the discourse analyst's work for her or him, it is merely a tool facilitating his efforts.

Epistemological fit

In addition to fitting the idea of language use and meaning-making that discourse analysis abides to, and containing the practical possibilities to operationalize this idea, topic modelling as a method also allows room for subjective interpretation by the analyst, which is equally a core element of discourse analysis.

As an unsupervised method, topic modelling is an inherently inductive approach to corpus analysis. This is opposed to supervised techniques, where the analyst pre-defines categories or scales, trains an algorithm to accurately reproduce them, and then extends the scoring/classification to the full corpus in a deductive fashion (Grimmer &

Stewart, 2013).[9] Topic modelling merely represents patterns of language use within the corpus, ignorant of anything outside of the texts it is fed for analysis.

It is therefore the analyst's task to interpret and make sense of what the topic model shows him or her about the semantic relations and meaning-making processes at work in the corpus. When interpreting the results of the model, analysts can and should draw on their reading of (some of) the texts, and their knowledge of the context from which the corpus stems. The subjective input of the analyst thus continues to play a crucial role, as is warranted in discourse analysis. One could say that instead of doing analytical work on its own, the algorithm provides the analyst with a condensation or transformation of a large corpus upon which the analyst then releases the analysis itself. The algorithm suggests that certain words have multiple meanings by situating them in different topics, and that certain words are linked to each other to form a larger unit of meaning. But it is the analyst's job to interpret how the different meanings of a word are shaped and how discourses are constructed through combinations of words.

We can render this idea more concrete by illustrating how the method outlined above lends itself to the study of the type of questions typically investigated in discourse analysis. For instance, if a term is solely attributed to one specific topic (ie. its probability in other topics is negligibly low), that topic arguably contains the hegemonic interpretation for this term within the corpus: the other words contained in that topic form the exclusive context in which this term is given meaning, a meaning which within the corpus is dominant and normalized as no alternative interpretations are present. In a topic model where the word "profit" appears in only a single topic, surrounded by

---

[9] Some of the more advanced varieties of topic modelling allow the algorithm to be used as supervised tool as well (McAuliffe & Blei, 2008). These versions are obviously excluded from the argument we make here.

words like “greed”, “exploitation”, “boss”, “capitalist”, and “profiteering”, it is clear that the hegemonic interpretation of profit-making in the topic is a negative, anti-capitalist one.

A similar logic can be used if a concept reappears in many topics pertaining to a certain issue: the concept is in this situation presumably co-constitutive of a hegemonic discourse, provided its meaning remains stable throughout the different contexts contained in the different topics. Were the term “growth” to re-occur in five different topics, respectively about fiscal prudence, societal well-being, government objectives, sound economic policy, and classical economics, each time with a similar and positive connotation, it would probably be an important part of the hegemonic economic view articulated in the corpus.

If its presence in different topics would lead to different meanings being ascribed to a concept, however, we are probably witnessing a struggle over its interpretation. If “growth” is negatively connoted in topics about climate change and inequality, but positively connoted in topics about consumer welfare and business health, the corpus most likely contains a debate over how to signify the term.

The argument we developed here concerning the epistemological and the meta-theoretical fit between topic modelling and discourse analysis also implicitly contains the reason why we think topic modelling can benefit from engaging explicitly with theories of discourse. Automated text analysis tools always contain an implicit and necessarily imperfect model of how language and the generation of meaning through language work (Grimmer & Stewart, 2013, 3-4). Discourse-analytical theories of meaning-making help us explicate how we think about what this necessarily imperfect model looks like for topic modelling, and allow us to reflect on how to reconcile it with theoretically rigorous empirical text analysis.

The added value of combining Topic Modelling and Discourse Analysis

We have already foreshadowed some reasons why using topic modelling for discourse analysis may be desirable when we established the theoretical basis for doing so. In the following, we make these suggestions more explicit and provide a more forceful argument of how using topic modelling pushes the boundaries of what discourse analysis can achieve empirically. The broad benefits of using large corpora and simple software tools have been discussed in some depth already (Baker, 2006; Kennedy, 2014). This section revisits some of these themes, but awards special attention to why topic modelling in particular stands to benefit discourse-analysts willing to engage with it. It raises at least three dimensions where this is the case: the study of hegemony, the study of language in context, and verification and systematization.

Topic modelling and the study of hegemony

The most innovative way in which discourse analysis can benefit from topic modelling, is in the latter facilitating a new way of studying hegemony in text. Discourse analysis often looks at a fairly small body of data, due to the limitations of the manual, close reading methods it employs. In combination with its inductive approach, this means that the study of hegemony in discourse is often forced to focus on moments where a hegemony breaks down or is established to learn the most about its nature (Wood & Kroger, 2000, 34; Gee, 2014, 37-38; Jorgensen & Phillips, 2002, 138-174). An inductive logic attaches greater demonstrative value to an observation that breaks or creates a pattern, than to one that confirms it. If you don't know anything about swans, spotting a group of ten white swans gives you a lot of information. The next twenty white swans you see don't add that much to your knowledge of swans, but spotting a spotting a single black one does.

In the same vein, observing a few instances where an apparently hegemonic interpretation is reproduced unproblematically is not very telling of how a hegemonic logic works; whereas the one instance where it is instituted, rejected, or contested is far more informative. For example, in the economic sphere, scholars have studied the characteristics of the current hegemony of liberal ideas about finance and capital by looking at the historic process through which the liberal interpretation triumphed over alternative conceptions in the 18th and 19th century. They have also paid great attention to the scarce moments in contemporary history when the contingent status of this interpretation briefly reappeared as the smooth reproduction of its hegemony briefly glitched, either due to external dislocation or active resistance (De Goede, 2005; Gibson-Graham, 2006).

While this approach makes sense epistemologically, it does not sit together all that well with how discourse analysis fundamentally understands hegemony on an ontological level. Studying the nineteenth-century triumph of capitalist globalization over its alternatives and capitalism's recovery after moments of weakness like the crisis of 2008 indeed tells us a lot about its characteristics, much like the first white swan and the rare black swan do. But hegemony carries in it the notion of normalization and standardization. It is about the unquestioned acceptance as common sense of an idea that is not by nature given or unchangeable. Hence, moments where a consensus is uprooted or founded are secondary to what hegemony actually is supposed to be about, notably unproblematic and unquestioned repetition. Only looking at exceptional but informative instances of breakdown or institution means we study hegemony in a rather indirect and derivative way: we assume its existence, and then look for its roots or its momentary breakdown.

Searching for the patterns, routines, logics that form the regular and normalised grammar of our daily life is an approach more true to how hegemony is understood in discourse analysis (Glynos & Howarth, 2007). Yet since these regular and normal cases contain less unique information (they are white swans ten through thirty), they are less instructive. This is a problem for close reading discourse analysis, which for practical reasons only looks at a small number of cases and therefore risks generalizing from an overly limited amount of information.

This is where topic modelling comes in, as it provides us with a way of solving this catch-22. It allows us to complement those few highly insightful cases with numerous normal, unexceptional, and individually uninstructive ones where hegemony is reproduced without a hitch. The latter type of data might be less educational, but they are far more abundant, and with topic modelling we can overview a large quantity of them comprehensively (to continue the swan metaphor, we can look at thousands and thousands of swans). As such, since topic models can help us to detect what is continuously repeated (or continuously absent but assumed) in a corpus of texts, they render it possible to study hegemony directly by analysing its reproduction, its normalization, and its subtle transformations and adaptations over time.

While other, more simple quantitative tools enable similar procedure, they require some assumption to be made by the researcher about the nature and content of the hegemonic discourse. Keyword frequency analysis, for instance, only works if one knows the keywords that drive a hegemonic interpretation. Topic modelling, on the other hand, allows us to explore the corpus in its entirety without prior manual analysis or a priori assumptions on what might be considered as normal. This helps us find routines and normalized logics which we might not have spotted otherwise, precisely because of the degree to which we see them as given and take them for granted.

Topic modelling and the study of language use in context

Topic modelling additionally facilitates the study of words in their textual context. First of all, most topic modelling tools do not just provide the analyst with an overview of which topics are present in which documents (the document-topic matrix), but also with a detailed annotation of which topic was allocated to every word in every text in the corpus. This creates a fast and practical procedure to switch between the topic model as the aggregation of language use in the corpus and the documents themselves as actual instances of language use in the corpus, thereby helping the analyst avoid the common pitfall of under-analysis through summary of the context (Antaki *et al.*, 2003, 13-16).

Secondly, topic modelling equally helps us avoid the reverse problem, over-analysis by awarding too much attention to idiosyncratic contextual detail (Antaki *et al.*, 2003). Crucial in this regard is that topic modelling allocates each and every word to a topic. As such, we cannot only easily jump back to the textual context, the textual context itself is also quantified. This facilitates a systematic approach to the study of the textual context, as it becomes possible to integrally track which topics dominate the texts featuring a keyword, a topic, or a discourse. As such, through topic modelling, the study of textual context can be quantified and systematized as well. This helps the analyst to avoid drawing hasty conclusions from one specific statement, and lets him or her overview with ease the variety of contexts in which a term, topic, or discourse is used.

Note that this possibility constitutes an important advantage over simpler text analysis tools which quantify the (co-)appearance of selected keywords, but do not quantify words appearing around them — which means that the context of the term(s) under analysis still needs to be read, interpreted, counted, and analysed manually, creating the risk of summative under-analysis or localized over-interpretation. As it forecloses these pitfalls by offering the possibility to get a complete image of the textual context in which words and discourses appear, topic modelling is a valuable methodological asset to an approach like discourse analysis, which emphasizes the importance of context in meaning-making.

Topic modelling and validation and systematization in discourse analysis

Third of all, topic modelling addresses to the need for replicability and systematization in discourse analysis. The first of these two notions might be reminiscent of a positivist demand for verification at odds with the interpretivist roots of discourse analysis. But even within an interpretivist framework, there is a need to demonstrate that one's context-bound interpretation is indeed representative of the context in question and not just the product of subjective selection or 'cherry picking', whether intentional or unintentional (Johnston, 2002; Louw, Todd & Pattamawan, 2014; Baker & Levon, 2015; Mautner, 2015). Indeed, it has been suggested that discourse analysis is in fact quite vulnerable to making the mistake of using its data to make a pre-existing point (Rogers, 2013, 74; Antaki *et al.*, 2003, 19-21; 27-30).

Topic modelling helps out in this regard in two ways. First of all, it evidently creates the option of quantification. The data's representativeness and the interpretation's significance and reliability can be demonstrated statistically. Secondly, and perhaps more importantly given the ethos of discourse analysis, topic modelling facilitates qualitative validation of whether our interpretation makes sense, even when working on a very large corpus. As we argued, the interpretation of a topic as a meaningful unit remains the task of the analyst. This interpretation draws on close reading of the texts, knowledge of the subject, and personal perspective and experience. But the previously discussed possibility to jump back and forth quickly between the topic-term matrix and the concrete incidence of words belonging to this topic throughout texts in the corpus also allows the analyst to verify whether his or her interpretation of the topic at face value strokes with his or her interpretation of this topic when he or she encounters the words allocated to it in the texts. This way, the analyst can easily check whether the conclusions he or she draws from studying the co-appearance of words and topics in the topic model hold up when confronted with concrete formulations in the texts under analysis.

Furthermore, the systematicity topic modelling furnishes allows for the detection of the recurrence of nuances and subtleties in text at a very large scale. This way, the concern with the details of language use that characterizes discourse analysis can be exercised with an order of magnitude several times that of close reading. This makes topic modelling an appropriate tool for diachronic analysis of how discourses evolve and change incrementally, for instance (Jaworska & Nanda, 2016). If the timespan becomes too long or the change too subtle, such a transformation might be missed if one relies solely on close reading. Similarly, topic modelling makes room for comparative discursive research. It facilitates for example the study of the differences, similarities, and changes in the rhetoric of politicians from different parties.

Case studies

So far, we have discussed on a relatively abstract level how topic modelling can facilitate discourse analysis, discussing their ontological, epistemological, and methodological fit. In the following two case studies, we aim to demonstrate how such a combination works in practice, rendering some of the insights from the previous sections more concrete and tangible.

First of all, the meta-theoretical and epistemological fit of topic modelling and discourse analysis is on display, as both cases clearly show how the topics of large topic models contain collections of words which the analyst can, subjectively and reflexively, interpret as discursive elements. The first one does so mainly qualitatively, the second one mixes qualitative and quantitative interpretation. Additionally, both case studies also allude to how topic modelling facilitates the study of discursive hegemony, and the second case furthermore makes an effort to show its utility for the study of language use in context and for internal validation. Nevertheless, it must be emphasized that these case studies are by no means fully-fledged, stand-alone analyses. They do not present self-sufficient empirical research or results, but merely try to illustrate some of the abstract methodological arguments made above.

Our first case tackles the discourse of political speeches on international trade policy in the European Parliament, by building a topic model of interventions during the Parliament's plenary sessions between 1999 and 2016.[10] The case particularly focuses on how we can interpret the discursive elements contained in a topic as hegemonic and normalized. The second case study, analysing Austrian newspaper articles, does the reverse, and critically interrogates the idea that pro-growth stances are hegemonic in public economic discourse. Here, using a decomposition of the discursive patterns captured in a variety of topics, we use a topic model to identify public discourses about

[10] We used MALLET to build a standard LDA-topic model with 120 topics, built over 3.500 iterations with hyperparameter optimization every 20 iterations and a burn-in of 40, which we validated against other models with higher and lower topic counts. It is based on a corpus of 11.744 pre-processed speeches discussing international trade policy delivered in the European Parliament's plenary session between 1999 and 2016, which we lemmatized and from which we removed the stopwords. Speeches were drawn from the Talk of Europe database using a SPARQL keyword query for 121 terms and phrases specific to international trade policy.

economic growth and to challenge the notion that pro-growth discourses are truly hegemonic.

Case 1: trade politics in the European Parliament

In the European Parliament case study, we are looking for traces of the extant hegemony within trade policy-making. The hegemonic practices of a policy field can be considered as forming the normal, appropriate rules any politician has to follow when acting within this policy field (see Glynos and Howarth, 2007 on social logics for a more elaborate discussion on this). It is on the basis of this normalized and socialized nature, that we can set about developing a heuristic to study hegemony, as it is fair to expect that such a normalization will leave traces in language use. The prime empirical characteristic we anticipate any form of hegemonic language use to display, is, by definition, that it features continuously, regardless of the speaker, his or her ideology, the specific (sub)issue, or the timing. As such, we looked for topics representing a fairly stable share of every speech in the corpus.[11] The topics whose incidence we found be relatively stable across all speech, regardless of when or by whom they were delivered, were then qualitatively interpreted as (fragments of) discourses. This left us with several relevant discourses, of which we analyse two here, consisting of respectively two and

---

[11] The coefficient of variation (CV) of each topic's share in each speech was used as a measure of this stability. Topics with a low CV have a low standard deviation over all texts in comparison to their average share per text (and their share in the corpus). More simply put, topics with a high CV generate their share in the corpus by featuring to a relatively high degree in a relatively low number of speeches, while comprising a relatively low degree of all other speeches. Topics with a low CV get their share by representing a relatively stable share of each speech, without a high number of significant outliers in any direction. The latter are evidently the topics of interest here.

three different topic.

[insert table 1 here]

Topics 15 and 20 were interpreted as establishing trade as a practice revolving around *cooperation* and *partnership*. Systematically linking international trade to terms like “relation”, “partner”, “cooperation”, “relationship” and partnership”, these topics represent the practice of trade as involving a teaming-up, a connection. The terms “agreement”, “benefit”, “support”, “important”, “importance”, “essential” and “promote” furthermore instil this partnership with a positive sentiment. Trade as a relationship between partners is considered to *benefit* those involved, and hence, it is necessarily something to be pursued. The fact that all europarliamentarians draw on the discourse contained in these topics suggest that they all find it evident that trade relations should be promoted, supported and developed further. Trade relations are considered *important*, even *essential*. Of course, parliamentarians do disagree on what those trade relations should look like, or about how trade’s positive potential ought to be realized. In other words, it remains possible to discuss the unwanted negative effects and consequences of a particular trade policy, or debate what commercial policy is necessary to bring about the innate blessings of trade relations, but on a more fundamental level, trade is apparently always presented as something inherently positive.

A second set of topics (55, 90, 99) together contain a discourse of *organisation*, articulating the idea that trade and trade relations always feed into a wider, global system. Terms like “order”, “system”, “world”, “global”, “organisation”, “multilateral” and “framework” are suggestive of this tendency, as is the relatively stable way in which the WTO is referenced. Trade relations are not just isolated connections between partners, they are discursively constructed as constituting a larger whole, a global trading system. Other terms in these three topics, such as “opportunity”, “benefit”, and “prosperity”, suggest some carry-over from the previous discourse of cooperation, which established trade as a mutually beneficial partnership. Similarly, in the topics containing aspects of that discourse of cooperation, we can also find some elements of organisation, through terms such as “order” and “framework”. Partners maintain and develop the benefits of their cooperation in a large whole.

This discourse of organized gives the relationships which trade consists of a logical, ordered character. There is a structure to the network of trade relationships, an organizational coherence, but this structuring does not come automatically. It needs to be “ensured” through “measure[s]”, “rule[s]”, “regulation[s]”, “legislation”, “authority”, “implementation” and “reform”. The structured nature of trade relations is not a fact of nature, the presence of these terms in the discourse suggests, *political intervention* is required to achieve it. Trade thus necessarily involves policy-making, as trade relations and systems need to be built. Again, the type of intervention and political action that politicians want to see presumably differs greatly throughout the Parliament, but the idea that having a trade policy is necessary to reap the benefit of structured, organized trade relations appears to be a given regardless of political ideology or nationality.

Case 2: the Austrian public growth debate

In the second case study, we are interested in how the news media make sense of economic growth and a major economic crisis. Economic growth is a prominent concept in politics and academic research alike, with much scholarly work and public attention devoted to its causes and consequences. Somewhat surprisingly though, research on public understandings of and attitudes towards economic growth is quite rare. The sparse literature suggests a "hegemony of growth" (e.g. Schmelzer, 2016), a pro-growth discourse among policymakers and publics that dominates over critical lines of argument, stressing, for example, environmental concerns, or linking economic growth to rising inequality and other social issues. This study puts the hegemony of growth-hypothesis to the test in one particular case: the Austrian media.

We analysed a corpus of newspaper articles concerning economic growth[12], with the aim of dissecting public discourses about the subject and investigating how they evolved over time. First, we identified topics related to economic growth and studied their salience over time. Next, we identified several discourses in these topics, and studied how they represent it as a concept. We then analysed the correlations between topics[13], to investigate how different elements of these discourses are typically combined within articles. This allowed us to analyse the hegemonic and non-hegemonic discourses about growth presented in newspapers in depth. The expectation was that discourses with an explicit or implicit pro-growth stance dominate the corpus, to the detriment of those devoted to a critical view.

---

[12] Gathered by selecting articles from major newspapers *(Die Presse, Der Standard, Kronen Zeitung, Kurier,* and *Kleine Zeitung),* published between September 2006 and end of August 2016, and containing at least one of the following keywords: "economic growth", "inequality", "sustainability", "employment", and "unemployment" — keywords related to the debate about economic growth. The corpus consisted of 52,593 articles in total.

[13] In this case, we used the *structural topic model* (Roberts, Stewart, & Tingley 2013), which allows and models topic correlations, enabling this type of inquiry. After pre-processing (stopword removal, stemming, and dropping words mentioned less than 15 times), we ran multiple models with different parameters. The results presented here are based on a model with 200 topics, and were validated against other models with the same and lower topic counts.

[insert figure 1 here]

We found, unsurprisingly, that the economic crisis was covered in-depth over the period 2008-2011 (top left panel of fig 1). As expected, the topic capturing most of the crisis-related discourse presents the recession in negative terms. For example, the keywords “dramatic”, “severe”, “lost”, and “massive” that characterise the topic give it a negative sentiment and legitimate immediate pro-growth policy intervention. The correlated (corr. coef. = 0.28) *recovery* topic explicitly contrasts the severe crisis with a “recovery” marked by “strong” “growth”. Qualitatively inspecting some of the articles that score highly on these two and other topics correlated with the *crisis* topic (namely *optimism,* and *prognosis*), corroborates this interpretation. This is in line with our conjecture that the public discourse emphasises economic growth promotion as a public good and desirable policy outcome. Thus, in particular during the major uptick of the crisis-related news coverage, the lack of growth was seen as a major problem, revealing a pro-growth stance.

[insert table 2 here]

However, while this overall positive attitude towards economic growth is strongly present during the period 2008-2011, it recedes in later years. The public debate becomes more balanced, giving space to discourses quite critical of economic growth and the global economic “model” in general. The brunt of this discourse is captured in the *growth_critique* topic, which is correlated with others covering *social_justice, inequality,* and *democracy.* The critique can partly be read from the keywords — using terms like “capitalism”, or “neoliberalism” is already indicative of a critical stance — but close reading of a sample of articles shows the critique more profoundly. To illustrate, one article, published in the centre-right newspaper *Die Presse* states: “[t]he decline of growth is thus a necessity for survival. But it demands a different economy, lifestyle, civilisation, and a change of social conditions”. Most articles do not side with the critics as strongly as the example does, but typically present the critique from a well-balanced point of view.

Returning figure 1, presenting topic salience over time, we see that after 2011 pro-growth topics and those more critical and reflective have somewhat equal shares of the corpus over time. This we interpret as a sign that the public discourse about economic growth is currently less hegemonic then it might appear at first sight — at least in Austria, that is. This argument hinges heavily on how we used the topic model in this case: we moved from corpus inspection over qualitative study, interpretation, and validation, to (illustrative) quantification.

Final Remarks

In this article, we contended that topic modelling can be a powerful aid for discourse analysis. We argued the potential benefits of combining discourse analysis and topic modelling, discussed their theoretical compatibility, hypothesised a methodology that would facilitate their combination, and showed the practical feasibility of this combination through two examples illustrating the necessary methodical and analytical steps. What we want to emphasize in this conclusion, however, are the limitations and implications of our proposition. Not all discourse analysis can and should be done using topic models. While topic modelling holds the potential to deconstruct texts into their discursive elements, whether or not this works in a specific case is up to the judgment of the analyst familiar with the theory, the method, and the material at hand. Ultimately, topic modelling does not convert discourse analysis into an exact or a quantitative science; rather than solving all its challenges, the method transforms some of the critical questions that need asking.

A first question in need of reformulation, concerns the issue of representativeness. Artisanal discourse analysis often faces the criticism of working with limited data unsuited to make claims about the discourse of an entire field of practice. Claims about the scope and applicability of an interpretation are often rather vague (i.e. "many of the articles analysed", "a feature rarely found"). As such, readers frequently have to take analysts on their word when they claim that their material is substantial enough to allow for generalization. The systematization topic modelling introduces to discourse analysis helps analysts to win their readers' trust by facilitating bigger corpora, by allowing them to show their entire corpus (rather than a mere sample or an illustration), and by making transparent how much of it is represented by individual topics (thereby revealing the scope of where their argument does and does not apply). Still, improved systematization does not make the question of trust disappear, it merely transforms it. As it is the analyst who picks the model used for further analysis out of a potential limitless number of alternatives, critical readers now have to trust that the model of the corpus is indeed representative, and they can challenge analysts to validate this claim by showing alternative models.

Secondly, topic modelling transforms how we think about interpretation. Traditional discourse analysis typically features illustrative quotes in the text to show the relationship between data and the analyst's work. Whether or not the reader accepts the analyst's interpretation of the data depends on whether he or she trusts that the analyst did not cherry-pick, or was not led astray by confirmation bias. To verify this is not the case and validate an analysis, a critical reader can actively look for counterexamples, for example. Topic modelling simplifies this process by allowing the researcher to locate all text segments that have a high share of a topic, and where a particular interpretation of that topic should thus apply. The reader can now critically evaluate the interpretation of a particular combination of words transparently and exhaustively, rather than having to trust that the snippets offered summarize the analyst's work well. Still, as noted above, the reader has to ask whether similar interpretations and conclusions can be arrived at using alternative model specifications.

Finally, the use of topic models transforms the relationship between the outcomes of a specific study and larger claims concerning hegemony and power. Artisanal discourse analysis typically studies hegemony by looking at critical junctures and intense discursive struggles, claiming that the surviving repertoires are hegemonic. Topic modelling, we argue, allows discourse analysus to turn its focus to the everyday, the normal, and the regular. In this reading, topics consistently appearing across an entire corpus can thus constitute representations of hegemonic repertoires. But a topic is a mere collection of words, and its meaning is contingent on its relation to other repertoires at play. Can we trust, for example, that the discourse captured in one topic doesn't turn self-reflexive? If that were the case, the same words may be used in different places and at different times, but their meaning would not be the same. As such, critical readers have to ask whether the analyst systematically investigated the variety of contexts in which a specific topic was used, and whether the meaning captured in it is indeed stable throughout them.

What these points show, is that topic modelling pushes the methodological boundaries of discourse analysis, without abolishing them altogether. Yet while these transformations create new limitations as well, we believe the potential gains are worth the attempt to use topic modelling, and we encourage researchers working within the discourse-analytic tradition to explore how topic models can enrich their craft. Likewise, we think that researchers regularly working with topic modelling would benefit from critically reflecting upon their praxis of interpretation, and from engaging with the wider theoretical literature on meaning-making. Such mutual engagements can open up avenues for tackling old methodological questions in new ways, and they may even spark a few entirely novel debates that have so far flown under the radar.

References

Antaki, C., Billig, M., Edwards, D., & Potter, J. (2003). Discourse analysis means doing analysis: A critique of six analytic shortcomings. *Discourse Analysis Online 1(1)*, 1-39.

Arun, R., Suresh, V., Madhavan, C. E. V., & Narasimha Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Berlin: Springer.

Baker, P, & Levon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication, 9(2)*, 221-236.

Baker, P. (2006). *Using corpora in discourse analysis*. London: A&C Black.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*, 993-1022.

Blommaert, J. (2005). *Discourse: A critical introduction*. Cambridge: Cambridge University Press.

Boydstun, A. (2013). *Making the news: Politics, the media, and agenda setting*. Chicago: University of Chicago Press.

De Goede, M. (2005) *Virtue, fortune, and faith: A geneaology of finance*. Minneapolis: University of Minnesota Press,.

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

Gee, J. (2014). *An introduction to discourse analysis: Theory and method*. London: Routledge.

Gibson-Graham, J. K. (2006). *A postcapitalist politics*. Minneapolis: University of Minnesota Press.

Glynos, J, & Howarth, D. *Logics of critical explanation in social and political theory*. London: Routledge.

Grimmer, J., & Stewart, B. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21(3)*, 267-297.

Howarth, D, & Torfing, J. (2005). *Discourse theory in European politics: identity, policy and governance*. London: Springer.

Jacobi, C., van Atteveldt, W., & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4(1)*, 89–106.

Jaworska, S., & Nanda, A. (2016). Doing Well by Talking Good: A Topic Modelling-Assisted Discourse Study of Corporate Social Responsibility. *Applied Linguistics*, *14*, 1-28.

Johnston, H. (2002). Verification and proof in frame and discourse analysis. B. Klandermans & S. Staggenborg. *Methods of social movement research* (62-91). Minneapolis: University of Minnesota Press.

Jørgensen, M, & Phillips, L. *Discourse analysis as theory and method*. London: Sage.

Kennedy, G. (2014). *An introduction to corpus linguistics*. London: Routledge.

Levy, K., & Franklin, M. (2014). Driving regulation: using topic models to examine political contention in the US trucking industry. *Social Science Computer Review*, *32(2)*, 182-194.

Louw, S., Watson, T., & Pattamawan, J. (2014). Picking the ripe cherry: Extract selection in qualitative research. *Proceedings of the International Conference: Research in Applied Linguistics* (155-169). Bangkok: King Mongkut's University of Technology Thonburi.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures,* 12(2-3), 93–118.

Marttila, T. (2015). *Post-foundational discourse analysis: From political difference to empirical research*. Berlin: Springer.

Mautner, G. (2015). Checks and balances. How can corpus linguistic contribute to CDA. R. Wodak & M. Meyer *Methods of critical discourse studies* (154-179). Thousand Oaks: Sage.

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

Munksgaard, R., & Demant, J. (2016). Mixing politics and crime–The prevalence and decline of political discourse on the cryptomarket. *International Journal of Drug Policy*, *35*, 77-83.

Roberts, M., Stewart, B., & Tingley, D. (2013). The structural topic model and applied social science. *Paper presented at the NIPS 2013 Workshop on Topic Models: Computation, Application, and Evaluation*. Lake Tahoe.

Rogers, R. (2013). Critical discourse analysis. A. Trainor, & E. Graue, *Reviewing qualitative research in the social sciences* (66-81). London: Routledge.

Schmelzer, M. (2016). *The Hegemony of Growth*. Cambridge: Cambridge University Press.

Törnberg, A, & Törnberg, P. (2016a) Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society*, *27(4)*, 401-422.

Törnberg, A, & Törnberg, P. (2016b) Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media*, *13*, 132-142.

Wallach, H., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th annual international conference on machine learning* (1105-1112). New York: Association for Computing Machinery.

Wood, L., & Kroger, R. (2000). *Doing discourse analysis: Methods for studying action in talk and text*. Thousand Oakes: Sage.

Tables and figures

Table 1. Hegemonic discourses on EU Trade Policy.

ID	Name	CV	Characteristic Words
15	relationality	1,934	trade relation partner agreement economic cooperation trading important union area benefit european investment development political country relationship economy party partnership
20	relationality	1,373	policy development strategy economic european objective support trade union international report social essential promote sustainable cooperation importance order approach framework
55	systematicity	1,341	order system ensure measure important member effective rule commission regulation state information make implementation guarantee time proposal authority provide legislation
90	systematicity	2,229	trade world economy growth market global economic opportunity europe free country job important prosperity open president globalisation create benefit barrier
99	systematicity	2,904	trade world organisation international wto system rule multilateral fair global development country reform framework trading benefit globalisation level developed developing

Table 2. Austrian growth discourses.

ID	Name	Description	Characteristic Words
1	inequality	Inequality and capitalism (critical)	state, wealth, inequality, economy, capital, welfare state, neoliberal, state, market, money, redistribution, schulmeister, private, private, financial market, economic
72	social justice	Social Justice and Citizenship	social, citizen, claims, central, responsibility, contribution, fundamental, strategy, approach, shaping, weak, claim, dependent, independent, access
81	growth critique	Critique of growth and the economy	society, economy, world, welfare, globalisation, economy, capitalism, growth, market economy, progress, more, model, change, resources
87	optimism	Careful optimism about recovery	positive, remarkable, current, that, expectations, development, consequences, despite, negative, still, despite, strong, situation, stable, optimism
94	prognosis	Economic analysis and prognosis	this year, expectations, prognosis, economic growth, meagre, next, expected, after, rise, expects, year, prognosis, economy, sink, experts
115	recovery	Economic recovery across the globe	growth, strong, recovery, global economy, global, economy, boom, transition country, oecd, weak, investment, national economy, recession, globally, slow
126	crisis	The crisis, causes and consequences	crisis, economic crisis, financial crisis, economy, consequences, unemployment, dramatic, severe, lost, time, deep, recession, hard, a lot, massive

Figure 1. Topic salience over time.