**Technical Appendix for How Do Speaker Characteristics Influence Use of Rhetorical History?**

While conventional Topic Models only reflect the underlying text, the main advantage of STM is the latter’s ability to incorporate the effects of metadata into its estimation of topics. As mentioned in the main text (section 3B), this allows us to directly test the hypothesis that speaker characteristics influence his or her rhetoric, including the usage of rhetorical history. In our model, frequency with which a topic is discussed is, therefore, defined as a function of the speaker’s party affiliation, gender, age, educational attainment and Brexit stance, as well as the date variable, which records the day of the debate.[[1]](#footnote-1) Next, a 15-topic STM is estimated based on the Brexit corpus presented in section 2.[[2]](#footnote-2) Yet, how was this number of 15 topics defined? Recently, statisticians have introduced different metrics that can be used to compute the number of topics, but the resulting range of numbers only provides an indication of the optimal number of topics in a statistical sense. Social scientists, and particularly historians, agree that the specific research question should determine the number of topics and, therefore, simply test different models and compare them in terms of interpretability (cf. Wehrheim 2019; Ferri et al. 2018). We followed a similar trial-and-error process in this analysis, finally opting for a 15-topic model.

The final STM includes 15 topics, 6,720 documents, and a 16,075 word dictionary, with the estimated topics being displayed in the table below. The standard metric for characterising an estimated topic is plotting its highest probability words, which are based on the probabilities that each word is generated from each topic. Put simply: when discussing topic A in his or her speech, there is a high likelihood that politician B is using words listed in the word profile for topic A. Finding a group of words from topic A’s word profile in one of the speeches, we can, therefore, use them as a proxy and conclude that this speech addresses, at least partly, topic A. In our case, we calculate for each topic not only those highest probability words, but also three additional types of word metrics (FREX, Lift, Score) that characterise the respective topic in a certain statistical way.[[3]](#footnote-3) Combining these different word metrics leads to a more coherent and more easily interpretable word profile, which is why we recommend that other researchers use Topic Modeling to consider more metrics than the highest probability words when determining topic labels. After reading the word profiles, each consisting of four different word metrics, we attached a label to each topic.

Table 1. Topics Discovered Through Our Analysis

|  |  |  |
| --- | --- | --- |
| **Topic** | **STM Output** | **Label** |
| 1 | Highest Prob: people, bill, house, us, European, referendum, many  FREX: democracy, voted, elected, voters, sovereignty, election, democratic  Lift: —none, bring, democracy, desirable, enemy, God, ordinary  Score: democracy, referendum, voted, people, sovereignty, election, conservative | *Democracy and Sovereignty* |
| 2 | Highest Prob: European, citizens, minister, union, EU, security, prime  FREX: Munich, citizenship, foreign, security, warrant, arrest, citizens  Lift: citizens, strong, adjective, agencies, ambivalent, Atlanta, attaches  Score: citizenship, Munich, arrest, security, foreign, data, citizens | *Citizenship and Security* |
| 3 | Highest Prob: devolved, government, Scottish, Scotland, UK, powers, Wales  FREX: JMC, Scotland’s, devolution, frameworks, Scotland, administrations, devolved  Lift: normally, SNP’s, distant, tanks, deep, agricultural, devolve  Score: Scottish, devolved, Scotland, administrations, devolution, Welsh, Wales | *Devolution* |
| 4 | Highest Prob: EU, UK, agreement, bill, exit, period, withdrawal  FREX: implementation, period, transition, transitional, certainty, businesses, continue  Lift: exit, ongoing, allow, cart, co-operation, contours, demonstrable  Score: agreement, exit, implementation, UK, transitional, transition, negotiations | *Post-Brexit Transition* |
| 5 | Highest Prob: rights, children, EU, family, UK, equality, government  FREX: children, children’s, UNCRC, family, child, refugees, Euratom  Lift: anti-trafficking, babies, Bowes, breastfeeding, dependants, experimental, FGM  Score: children, child, children’s, Euratom, rights, family, equality | *Family, Children and Refugees* |
| 6 | Highest Prob: environmental, EU, environment, government, standards, bill, public  FREX: environmental, sentience, precautionary, watchdog, animal, environment, animals, pollution  Lift: high, deserve, environmental, high, offer, pay, picks  Score: environmental, animal, environment, watchdog, welfare, animals, sentience | *Environment* |
| 7 | Highest Prob: right, friend, clause, members, government, gentleman, new  FREX: Rushcliffe, Beaconsfield, gentleman, east, constituents, Nottingham, Broxtowe  Lift: -tank, ah, Beaconsfield, Beckett, Blomfield, bogged, contumacious  Score: gentleman, Beaconsfield, clause, Rushcliffe, learned, constituents | *Parliamentary Rhetoric I* |
| 8 | Highest Prob: amendment, minister, government, learned, hope, friend, may  FREX: helpful, moved, Mackay, points, learned, something, grateful  Lift: shortly, sub-delegation, chilling, exhausted, fuzzy, helpful, non-devolved  Score: learned, amendment, Mackay, Pannick, Clashfern, Wigley, friend | *Legal Rhetoric I* |
| 9 | Highest Prob: union, trade, market, customs, single, EU, European  FREX: India, tariffs, customs, trade, single, free, tariff  Lift: Allander, Australians, Barry, Brent, business, Canada-plus-plus-plus, catastrophe  Score: customs, market, trade, EEA, tariffs, billion, India | *Trade* |
| 10 | Highest Prob: parliament, government, deal, vote, house, amendment, bill  FREX: meaningful, motion, mandate, vote, extension, date, notification  Lift: neutral, dandelion, overt, rigidity, timed, duke, next  Score: motion, vote, meaningful, grieve, article, parliament, agreement | *Parliamentary Rhetoric II* |
| 11 | Highest Prob: clause, act, EU, amendment, law, exit, day  FREX: section, schedule, page, insert, provision, subsection, modify  Lift: provisions, saving, EU-derived, including, primary, section a  Score: retained, clause, schedule, insert, section, exit, paragraph | *Legal Rhetoric II* |
| 12 | Highest Prob: law, rights, EU, charter, court, courts, European  FREX: charter, Francovich, damages, judges, courts, supreme, court  Lift: disapply, ECHR, sex, see, general, allowed, another  Score: charter, rights, law, retained, courts, court, human | *EU Law* |
| 13 | Highest Prob: Ireland, northern, border, agreement, good, Irish, Friday  FREX: republic, border, Gibraltar, Irish, Friday, Ireland, northern  Lift: normality, Browne, Ireland, Taoiseach, Trimble, island, agreement  Score: Ireland, border, northern, Gibraltar, republic, Irish, Friday | *Irish Border* |
| 14 | Highest Prob: UK, EU, European, government, industry, research, Brexit  FREX: trials, clinical, Erasmus, patients, medicines, funds, funding  Lift: additionality, ageing, airlines, artists, attractions, audio-visual, Bangor  Score: trials, clinical, Erasmus, funding, medicines, industry, patients | *Lack of Funding* |
| 15 | Highest Prob: powers, legislation, bill, committee, house, scrutiny, power  FREX: affirmative, sifting, scrutiny, secondary, instruments, procedure, delegated  Lift: affirmative, including, tax-like, among, appropriate, enemies, made  Score: affirmative, powers, sifting, secondary, scrutiny, sis, delegated | *Parliamentary Rhetoric III* |

According to the STM, the main topics discussed by British politicians surrounding Brexit were democracy and sovereignty; citizenship and security; devolution; post-Brexit transition; family, children and refugees; environment; trade; EU law; the Irish border; and post-Brexit lack of funding (e.g. medicine, education). In the following part, we will discuss the individual topics and explain the reasoning behind their respective labels. Usually, topics evoke specific associations, so that coherent labels can be inferred relatively quickly. Luckily, this was the case for most material in our sample, since our corpus was large enough to produce robust word profiles without obvious outliers. This effect might also be credited to the fact that, in contrast to most historical studies using Topic Modeling, we did not have to deal with Optical Character Recognition (OCR) errors (for this common problem, see: Romein et al. 2020), since the parliamentary debates on Brexit could be downloaded in a machine-readable format directly from the UK Government’s website, rather than having to be transformed from scans or pictures.[[4]](#footnote-4) However, it is worth re-emphasising that the labels were manually attached by the researchers and are, to a certain degree, subjective. Nevertheless, since the word profiles are automatically generated by the algorithm underlying the STM, it is hoped that the resulting topical structure of the debate is more objective than a mere cherry-picking of quotes related to a certain theme. With a large corpus consisting of thousands of speeches, employing machine learning tools like STM helps in structuring an otherwise difficult to comprehend mass of semantic data.

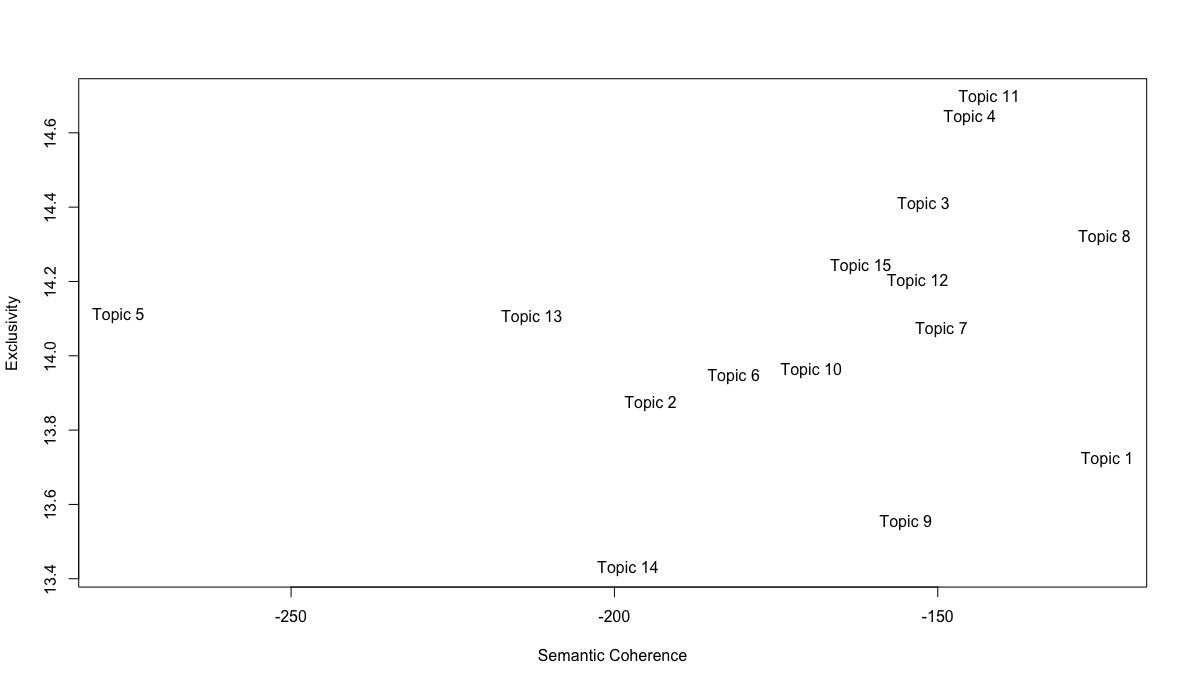
The first topic encompasses words such as ‘voters,’ ‘referendum,’ ‘democracy,’ and refers to concepts such as ‘sovereignty’ and ‘election.’ It has thus been labelled *Democracy and Sovereignty* (Topic 1). The existence of this topic can be seen as a reflection of the Brexit referendum campaign, during which much emphasis was placed by ‘Leave’ supporters on the lost parliamentary sovereignty, which could not be restored unless the UK exited the EU (Ringeisen-Biardeaud 2017). Next, words such as ‘citizenship,’ ‘citizens,’ ‘foreign,’ and ‘security’ refer to the discourse on *Citizenship and Security*, which is why Topic 2 has received this label. Noteworthy is the apparent inclusion of the ‘Chatham’ house rules and the annual ‘Munich’ security conference in the word profile. Coming to Topic 3, we encounter terms that deal with *Devolution* (‘devolved,’ ‘powers,’ ‘scottland,’ ‘wales,’ ‘administrations’). In England, devolution is the transfer of power and funding from national to local government. The impact of UK withdrawal from the EU on the UK’s devolution settlements is one of the most technically complex and politically contentious elements of the Brexit debate,[[5]](#footnote-5) so it makes sense that STM estimates a corresponding topic. The next topic deals with the *Post-Brexit Transition* (Topic 4) period and the settlement that would need to be negotiated with the EU during the one-year transition period. Topic 5 deals with the rights of *Family, Children and Refugees* in the light of Brexit. This topic refers to the legal situation of unaccompanied children stranded in Europe, who were reliant on EU rules to be legally and safely reunited with a family member in the UK.

Topic 6 deals with ‘pollution’ and ‘environmental’ ‘standards’ as well as the ‘welfare’ of ‘animals’ after Brexit and has thus been identified broadly as a discourse on the *Environment*. Topic 9 incorporates deliberations about *Trade* issues. It includes words such as ‘trade’, ‘single’ ‘European’ and ‘market’, but also ‘tariffs’ and ‘customs,’ thus referring to potential trade distortions that might occur after the transition period had ended. It also encompasses the countries with which British politicians aim to intensify trade relations in a post-Brexit world, such as ‘India,’ ‘Australia‘ and “Canada“, thus reflecting the ‘Global Britain’ trading approach referred to by Theresa May. Likewise illuminating is Topic 12, which includes words like ‘law,’ ‘EU,’ ‘charter,’ ‘supreme,’ ‘court,’ and ‘courts,’ thus referring to the field of *EU Law*. Removing the supremacy of EU law was a core element of the ‘Leave’ agenda, and the presence of this estimated topic implies that this concern was reflected in the parliamentary debates. Besides the terms listed above, the word profile for Topic 12 refers to the well-known *Francovich* decision of the European Court of Justice (ECJ) in 1991, which mandated that if a Government broke EU law then a claim in damages would lie if the breach had been sufficiently serious and had caused loss. The entry of ‘Francovich’ into the topic’s word profile, reflected the fact that the Withdrawal Bill would remove, with some very limited exceptions and a two-year grace period up to March 2021, any action for Francovich damages (Craig, 2019). According to the STM, the difficult question of the Irish Border (Topic 13) was also raised in the Brexit debates (‘ireland,’ ‘northern,’ ‘border’), alongside the question of Brexit’s impact on the Good Friday Agreement (‘agreement,’ ‘good,’ ‘Friday’). Finally, Topic 14 includes examples from the field of medicine (‘clinical,’ ‘trial,’ ‘patients,’ ‘medicines’) and R&D (‘research,’ ‘industry,’ ‘Erasmus’), reflecting concern that Brexit would preciptate a decline in funding, arising from inability to access certain EU funding schemes. This topice has, therefore, been labelled post-Brexit *Lack of Funding*.

Finally, in Topic Modeling studies the estimated topics do not necessarily have to describe a straightforward theme; they can form clusters of methodological words or even days of weeks (Wehrheim 2019). In our example, this happened for five topics, which feature many rhetorical terms that are either standard parliamentary language for starting/finishing one’s contribution to the debate (Topics 7, 10, 15) or denote legal terms that are related to the bill’s passage through the two Houses (Topics 8, 11). These terms were utilised by diverse speakers, irrespective of the particular theme discussed, and are, therefore, not very informative. These topics do not represent ‘errors’ on part of the model, but simply reflect the speeches’ occasion and might be ignored in our analysis. Note that the STM found all 15 topics without knowing that it deals with a set of political speeches and without any pre-coded definitions or lists of key terms that would bias the resulting selection of key topics.

Finally, as a robustness check, the quality of the estimated topics can be measured through a combination of their semantic coherence and exclusivity. Semantic coherence is maximised when the most probable words in a given topic frequently occur together (Mimno et al. 2011), whereas exclusivity of words to topics is included to ensure that high semantic coherence is not the result of a few topics being dominated by very common words. The figure below plots the semantic coherence and exclusivity scores for the 15 estimated topics. The results show that the selected model features desirable properties in both dimensions and that many topics cluster towards the figure’s upper right side.

Figure 2. Illustration of Semantic Coherence



**References**

Airoldi, E.M., & Bischof, J.M. (2016). Improving and evaluating topic models or text. *Journal of*

*American Statistical Association,* 111.516, 1381-1403.

Craig, P. (2019). Constitutional principle, the rule of law and political reality: The European

Union Withdrawal Act (2018). *Modern Law Review,* 82(2), 319-350.

Ferri, P., Lusiani, M., Pareschi, L. (2018). Accounting for *Accounting History*: A topic modeling

approach (1996-2015). *Accounting History,* 23(1-2), 173-205.

Mimno, D., Wallach, H.M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing

semantic coherence in topic models.  *Proceedings of the 2011 Conference on Empiricial*

*Methods in Natural Language Processing*, 262-272.

Ringeisen-Biardeaud, J. (2020). “Let’s take back control“. Brexit and the Debate on

Sovereignty. *Revue Franchaise de Civilisation Britannique.* French Journal of British Studies,

22, XXII-2.

Romein, C.A., Kemman, M., Birkholz, J.M., Baker, J., De Gruijter, M., Merono-Penuela, A.,

Thorsten Ries, R.R., & Scagliola, S. (2020). State of the Field: Digital History. *History: The*

*Journal of the Historical Association, 105(365),* 291-312*.*

Taddy, M. (2013). Multinomal inverse regression for text analysis. *Journal of the American*

*Statistical Association,* 108.503, 755-770.

Wehrheim, L. (2019). Economic history goes digital: topic modelling The Journal of Economic

History. *Cliometrica,* 13: 83–125.

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

1. The variables are entered additively, and the date variable is allowed to have a non-linear relationship in the estimation stage. [↑](#footnote-ref-1)
2. Using ‘spectral initialization,’ which means that, independent of the seed that is set, the same results will be generated. [↑](#footnote-ref-2)
3. FREX indicates words that are frequent and exclusive to each topic (Airoldi and Bischof 2016). Lift’ weights words by dividing by their frequency in other topics, thereby prioritizing words that appear less frequently in other topics (Taddy 2013). Similar to lift, ‘score’ divides the log frequency of the word in the topic by the log frequency of the word in other topics. [↑](#footnote-ref-3)
4. It is worth noting that topics are not always initially recognizable. If a topic lacks a straightforward interpretation, it is helpful to read the speeches that exhibit a large share of this topic in order to get a better sense of the proper interpretation of the word list and, therefore, the appropriate label. While this procedure was, strictly speaking, not necessary in the present case, it was followed for most topics listed in the table above as an additional robustness check. [↑](#footnote-ref-4)
5. For a summary, see: <https://publications.parliament.uk/pa/ld201719/ldselect/ldeucom/9/903.htm> (31 August 2020). [↑](#footnote-ref-5)