IN PRESS FOR CORPORA (17.2, 2022)

Operation Heron – Latent topic changes in an abusive letter series

Lucia Busso, Marton Petyko, Sarah Atkins, Tim Grant

Aston Institute for Forensic Linguistics, Aston University

Abstract: The paper presents a two-part forensic linguistic analysis of an historic collection of abuse letters, sent to individuals in the public eye and individuals' private homes between 2007-2009. We employ the technique of structural topic modelling (STM) to identify distinctions in the core topics of the letters, gauging the value of this relatively underused methodology in forensic linguistics. Four key topics were identified in the letters, *Politics A* and *B*, *Healthcare*, and *Immigration*, and their coherence, correlation and shifts in topic evaluated. Following the STM, a qualitative corpus linguistic analysis was undertaken, coding concordance lines according to topic, with the reliability between coders tested. This coding demonstrated that various connected statements within the same topic tend to gain or lose prevalence over time, and ultimately confirmed the consistency of content within the four topics identified through STM throughout the letter series. The discussion and conclusions to the paper reflect on the findings as well as considering the utility of these methodologies for linguistics and forensic linguistics in particular. The study demonstrates real value in revisiting a forensic linguistic dataset such as this to test and develop methodologies for the field.

Content: Readers are advised that the letters analysed contain offensive language and that the concordance lines given in Section 5 of the paper provide short excerpts of this content, including racist and hateful language directed at particular groups.

1. Introduction

In September 2009, the *BBC Crimewatch* programme, working with Hampshire Constabulary, launched a viewer appeal to try to identify the writer of around 60 racially and sexually abusive letters[1]. The first identified letter in the series had been sent in January 2007 and early letters in the series were sent both to individuals in the public eye and to individuals' private homes. Geographic mapping showed that initial letters and most letters in the series clustered in the Portsmouth and Southampton area, in the South of England but with later letters being sent across the UK. In January 2009, Hampshire police as the force with responsibility for these areas set up a dedicated investigation team, called *Operation Heron*, to coordinate the investigations of local forces across the country and

[1] PM targeted in hate mail campaign. Accessed 27 May 2020. Last updated 29 Sept 2009
http://news.bbc.co.uk/1/hi/england/8279478.stm
PM hate mail plea gets 86 calls. Accessed 27 May 2020. Last updated 30 Sept 2009
http://news.bbc.co.uk/1/hi/england/8282151.stm
Language experts study hate mail. Accessed 27 May 2020. Last updated 30 Sept 2009
http://news.bbc.co.uk/1/hi/england/8282817.stm

to intensify investigative efforts into finding the writer or writers. In April 2009, letters were sent to Prime Minister Gordon Brown at his local constituency office in Fife in Scotland, after which *Operation Heron* was given further impetus.

As part of the intensification of the investigation professor Tim Grant was approached to provide investigative assistance. Their analysis drew on research on associations between language variables and demographic variables (e.g. Kemper et al., 2001 on age; Berman, 2008 on social background) and early profiling work on predicting gender (e.g. Argamon et al., 2003). They profiled the writer as likely to be female, aged 60 or older, from the South of England and although a competent writer, not of a high level of formal education. During the media appeal on *Crimewatch* it was advised that as many of the letters as possible should be released in their entirety – including those which contained idiosyncratic cartoons, and unusual turns of phrase and forms of racist abuse. The aim of the appeal was also to encourage further recipients of similar letters to come forward. In the event six letters were released to the media (see Figure 1 for an example letter) - although a decision was made to exclude from the broadcast the worst of the abuse.

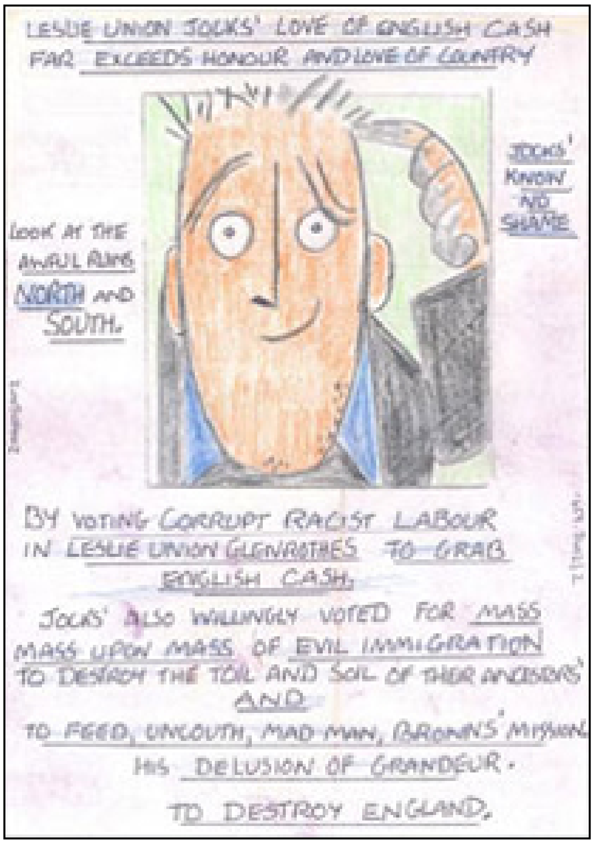

Figure 1. Example letter from the Heron series

After the *Crimewatch* appeal on the 30th September 2009 the programme received 86 calls from viewers who had either received similar letters, or who thought they knew the identity of the writer.

None in the latter category made a correct identification. One of the new victims who came forward provided an envelope from which DNA was extracted and which matched an individual on the police DNA database. That individual turned out to be the son of 68-year-old Margaret Walker and when their family home was searched evidence of draft letters were found in her desk[2]. Walker was convicted under the Malicious Communications Act 1988 and was supported by the British National Party in an unsuccessful appeal under Article 10, the freedom of expression provision of the European Convention on Human Rights. She received a non-custodial sentence which included an Anti-Social Behaviour Order (ASBO) which included a court order to prevent her sending more abusive letters. In 2012, when Walker was aged 73, she was again prosecuted and admitted to sending a total of 500 letters, as a result the ASBO was converted to a lifetime order stating that she should sign every piece of post that she sent and to always provide a return address.[3] There are no further Court appearances in the public record, nor any further press mentions of her activities.

This paper does not focus on the profiling task associated with the original linguistic analysis. There has been considerable research in this area in the decade since *Operation Heron*, notably a significant review and testing of methods by Nini (2015), and a number of PAN[4] computational competitions focussing on different aspects of linguistic profiling (e.g. Rangel et al., 2018). Rather, we here use the corpus of Margaret Walker's letters as a testbed to explore other investigative insights that might be brought to similar investigations by applying current computational linguistic and corpus techniques. In particular, we focus on topic modelling techniques (Blei, 2012), combined with a more qualitative corpus analysis to explore how the interests of the writer change over the time series of the letters. The reason why studying topic change over time – including the statements that make up these topics – can be useful to an investigation is that it can feed into broader police intelligence and analysis efforts both in terms of identifying the author, through understanding their interests, and perhaps how these interests evolve in response to events in the news; and also in terms of analysis of the imminence and severity of any threat they might pose.

Topic Modelling (TM) is a computational text mining methodology used to infer the thematic organization of a collection of texts (Anandarajan et al., 2019). In this paper, we adopt a recent version of TM called Structural Topic Modelling (STM, Roberts et al., 2016), which allows the modeling of data according to document-level covariates. We combine STM with the qualitative corpus analysis

---

[2] The present paper names the offender as she was convicted in open court and all documents are therefore public.
[3] Racist grandmother Margaret Walker gets lifetime Asbo for offensive letters. Accessed 27 May 2020. Last updated 26 Sept 2012.
https://www.dailyecho.co.uk/news/9948166.racist-grandmother-margaret-walker-gets-lifetime-asbo-for-offensive-letters/
[4] PAN is a series of scientific events and shared tasks on digital text forensics and stylometry. https://pan.webis.de/

and annotation of topic-related concordance lines. Through this analysis we address the following research questions:

1. What are the main latent topics in the *Operation Heron* letter series?
2. How do these topics change over time and across addressees?
3. What are the specific statements (complaints and demands) within each topic?
4. How does the prevalence of these statements change over time in the letter series?

In addressing these questions, this paper sets out in Section 2 describing the theoretical background to the present study. Specifically, we outline linguistic approaches to analysing racist and abusive language (2.1, 2.2), and describe STM (2.3) and its use in the social sciences, corpus linguistics and in forensic linguistics (2.4).

Section 3 describes the data for our study and subsequent sections go on to present the findings, distinguishing four key topics over the timespan of the letters (Section 4). Further corpus analyses are used to better understand these findings (Section 5). Finally in Section 6 we draw together these two analytic strands, discussing the significance of our findings and commenting on the potential value of this approach for the fields of corpus and forensic linguistics.

2. Background

2.1 *Linguistic methods for analysing social and ideological representations of identity*

Lexical and thematic patterns in representations of race, ethnicity, religion, and nationality have been extensively analysed in communicative contexts, such as news media, policy, and political communications (see van Dijk, 1991; Poole & Richardson, 2006; Baker et al., 2008; 2013). These have largely taken a Critical Discourse Analysis (CDA) approach, identifying underlying ideologies and societal discourses in texts or genres, rather than overtly motivated acts of hate speech as the study here examines. Nevertheless, many of these studies have relevance to our approach, particularly in corpus identification of topics and lexical patterns.

Baker et al. (2008) trace the intersection of corpus linguistic and more qualitative CDA approaches and then apply this in analysing a 140-million-word collection of British news articles about refugees, asylum seekers, immigrants, and migrants (‘RASIM’). They establish patterns of topics, metaphors and *topoi*[5] through keyword and concordance analysis of the data. Baker et al. (2013) similarly apply corpus methods to representations of Muslims in the British news media. While they do identify some overtly Islamophobic articles ('BBC PUT MUSLIMS BEFORE YOU!' - Daily Star, 18 October 2006), the majority pattern in the data characterised an ambivalent picture indirectly contributing to negative stereotypes. Frequent lexical items were employed to depict Muslim figures such as Abu Hamza (e.g. *radical*, *fanatic*, *scum*, p. 183-6) and topic associations between 'terrorism' as well as the claiming of UK benefits (*scrounger* and *handouts* p. 177-9). These are of relevance when considering the lexical items used in directing racist and offensive insults in the *Operation Heron* letters. More general topic associations around Muslims, immigration and asylum seeking, as well as British government policy (particularly references to issues of finance, such as *taxation*, *benefits*, *spending* (p. 61)), were identified through corpus methods of collocation and concordancing.

Revisiting these methods, Baker and McEnery (2019) look again at representations of Muslims and Islam in the UK press, collecting a parallel corpus that enabled an examination of changing discourses 1998-2009 and 2010-2014. Though some features remained stable over time, their methods identify important shifts in the discourse, particularly increasing references to 'radicalisation' and explanations for its cause. The forensic linguistic dataset of letters examined by this study are not as easily available as electronic corpora collecting news media, but there may be similar value in revisiting these historic datasets, originally collected synchronously by the police during an investigation, to then examine potential diachronic changes. Methods for identifying these subtle shifts in discourse may be of assistance to future police investigations into ongoing hate mail incidents.

News media articles are not of the same genre as the hate mail analysed in this paper. Nevertheless, although this individual author's style may be shown to be idiosyncratic and idiolectal (section 4 and 5) and the particular crime unusual in the level and scale of abuse, the *Operation Heron* letters do not exist in a cultural vacuum – some topics and lexical patterns identified in these earlier CDA studies on racism and Islamophobia provide evidence of broader societal discourses that can be drawn on when composing offensive tropes about particular social groups. Some research has argued the media hold considerable power to influence public opinion (van Dijk, 1991; Lido, 2006). Baker et al. (2013) take a more complex view of this, with one influencing the other and newspapers adapting stories to their readership. In tracking the change in media representations of Muslims over

---

[5] In classical rhetoric, 'topoi' (τόποι) are argumentation strategies used for a particular subject. It has been an important concept in discourse historical approaches to CDA, representing 'the common-sense reasoning typical for specific issues' (Van Dijk, 1991).

time, the authors indicate a potential link in the increased experiences of overt prejudice and violence towards British Muslims (CPS 2008) alongside some changes in media representations. Some of the similarities and thematic links between media representations are perhaps important to consider in the context of the *Operation Heron* data, since Margaret Walker used article ‘cut outs’ within some letters and was later found to have drawn on news media in identifying victims.

Corpus methods such as those outlined above are now well established in CDA but in 2008 Baker et al. made the important case for adopting and interrogating quantitative methodologies to assist with discourse analytic approaches, as our study now similarly proposes for STM. STM is a recent innovation in TM techniques, which allows the modelling of themes of a collection of texts according to document-level covariates. Interest for STM techniques is growing in the linguistic literature (see section 2.3 below), although not without critics. For example, Brookes and McEnery (2019) highlight issues with the lack of definition or theory underpinning the definition of a ‘topic’ in STM and go on to test this in a relatively small corpus of online patient feedback for the UK’s National Health Service. However, of the 20 topics retrieved, only 6 exhibited coherence in the theme or other commonality throughout all the texts and of these only 3 of these provided useful thematic starting points (p.16-17). Furthermore, the authors highlight that these inferences and ‘topics’ did not hold up when the texts themselves were examined. Given the difficulties of the approach, we seek to address whether STM and qualitative corpus methods can be successfully combined in providing a more nuanced analysis of topic within a small dataset, outlined further in 2.3. Here, in working with a much smaller dataset than the RASIM corpus, we make the case for STM and subsequent concordance analysis as a means of systematically identifying somewhat similar themes. Though corpus methods have become widely used in forensic linguistics, the possibilities for STM in combination with these approaches has not been fully explored. STM in particular allows for methods of comparative analysis over time, which this study seeks to explore further.

Beside media representations and discourse, another important element to consider in analysing the *Operation Heron* corpus is the specific characteristics of abusive language. In fact, the letter series is dominated by abusive and racist remarks based on national identity and ethnicity. Hate speech has received considerable attention in law, policy and social psychology, including the structural contexts and ‘routine character’ of its enactment (Iganski, 2008a; 2008b). However, hate crime has received less attention in linguistics, as Nieto (forthcoming 2022) notes, highlighting how such acts are not as unusual or deviant as they have perhaps been characterised. Indeed, linguistic studies highlight the widespread social practices of online ‘trolling’ (Hardaker 2010; Hardaker and McGlashan 2016) and public abuse (Mullany and Trickett, 2018) and the value of approaching our understating of these from pragmatic and sociolinguistic perspectives. These and other findings regarding abusive language – especially in online contexts – are extensively discussed in Section 2.2 below.

2.2 Abusive language online

While abusive letters, emails, and private messages have received very limited academic attention outside forensic linguistics (Grant, 2008), the detection and analysis of abusive language in user-generated online discussions have become an issue of increasing importance in corpus linguistics, computational linguistics, and natural language processing (NLP) (Nobata et al., 2016). This growing academic interest in abusive online content is of course a direct result of the significant role that social media plays in our everyday lives (Park & Fung, 2017).

Abusive language is generally defined as language produced with the intent to hurt individuals or social groups (Wiegand et al., 2018). However, as Vidgen et al. (2019) point out, this definition is rather vague, which might be one of the reasons why abusive language is often used as a catch-all term that encompasses communicated aggression, hateful, offensive or derogatory language, profanity, and hate speech (Nobata et al., 2016). Another problem is that the above definition heavily relies on the speaker's intention to hurt others, which makes the concept of abusive language practically impossible to operationalise (Vidgen et al., 2019). To avoid this, Lee et al. (2018) focus on perceived abusive language by looking at content that causes aggravation in others. Unfortunately, a clear weakness of this approach is that perceived abusiveness is inherently subjective, which is why human annotators often vary considerably in what they consider to be hateful or abusive (Vidgen et al., 2019). Additionally, discourse participants may accuse others of being abusive in bad faith in order to reach various discursive goals, such as to discredit others or draw sympathy to themselves.

Most prior work has focused on the detection of abusive language in publicly available online discussions. The typical blueprint for these studies is that researchers build a dataset from social media platforms, utilise human annotators to manually label the posts within these datasets as either abusive or non-abusive, and develop a classifier to detect abusive language in the analysed data sets (Castelle, 2018). The classifiers for abusive language detection normally rely on a blacklist of abusive terms, character n-grams, word n-grams, part of speech n-grams, and various syntactic features (Clarke & Grieve, 2017; Wiegand et al., 2018). There are three main issues with the above approach. Firstly, due to the practice of using a list of abusive terms, many classifiers are only able to detect explicit abusive content (Nobata et al., 2016). Secondly, as most datasets used in abusive language detection focus on a single social media platform with strong contextual constraints, attempts at detecting abusive language across various online domains has so far remained largely unsuccessful (Vidgen et al., 2019). Finally, as most classifiers work as 'black boxes', they give minimal insight into the linguistic features of online abusive content, which is a clear gap in the current literature (Vidgen et al., 2019).

A recent innovation in computational linguistics, STM, may balance the need for quantitative reliable analyses with a more thorough insight into the linguistic and thematic organization of a text or a collection of texts.

2.3. Methodological background: Structural topic models

TM techniques comprise various unsupervised learning algorithms that retrieve a set of 'topics', i.e. clusters of words that co-occur according to certain probabilistic patterns across a collection of documents (Blei, 2012). These methods use both modelling assumptions (defined by the researcher) and text properties (bootstrapped from the corpus) to estimate general semantic topics within unstructured corpora, and to organise texts on the basis of words co-occurrences (Combei & Giannetti, 2020). TM is a useful data-driven methodology for text analysis, as it inductively discovers patterns in the data, rather than assuming them *ex-ante* (Laver et al., 2003). However, TM still requires a crucial interpretative role of the researcher, and as such it is best referred to as a *computer-assisted* methodology.

The role of the researcher emerges in two aspects: determining the meaning of the topics' consistency, and finding the number of topics to be retrieved. Consistency refers to the fact that automatically retrieved topics comprise words that occur together in text, and/or do not appear much outside that topic. However, it is for the researcher to understand and decide how the topics should be interpreted (Chang et al., 2009; Roberts et al., 2014). Another crucial user-specified parameter is the number of topics (K) in which to cluster the text. Although TM is a statistical procedure, many scholars argue that the selection criteria of K should focus on achieving interpretability of the output rather than mathematical fit (Chang et al., 2009). In fact, a viable value of K heavily depends on the nature of the corpus, and on ease of topic interpretation. Some automatic processes can help define a feasible interval of K values (see 2.4).

The classical TM algorithm is Latent Dietrich Allocation (LDA, Blei 2012), which assumes that all texts in the modeled corpus are generated by the same underlying process (Murakami et al., 2017 provide a useful discussion LDA and its usefulness in linguistics). Thus, LDA cannot be used to estimate differential usage, as it does not allow to model topic variation according to external variables, such as time, author identity, or genre. In linguistics, and especially forensic linguistics, the effect of external factors is however often very important to estimate. To embed the modelling of topic distribution according to external factors, Structural Topic Modelling (STM) is needed.

STM addresses the limitations in the LDA algorithm by introducing the possibility of modelling the distribution of topics as a function of document-level covariates (Roberts et al., 2014). The model can be investigated in function of covariates for both topical content (i.e. the lexical content used within-topic) and topical prevalence (i.e., the frequency with which a topic is discussed). The relationship between covariates can be further explored statistically in regression-like schemes.

2.4 TM algorithms in the literature, between social science and forensic linguistics

Despite being first developed in computational linguistics, TM is still surprisingly under-used in linguistic research, as its usefulness is still debated among researchers. As mentioned above (2.1), Brookes & McEnery (2019) find TM approaches to be generally unreliable, and argue that a traditional qualitative approach is overall more successful. However, their analysis is – in the authors' opinion – problematic. In fact, they define the number of topics not based on data-based criteria, but rather capitalising on previous similar research done on different corpora. As mentioned (2.3), a viable K value needs to be determined using corpus-dependent methodologies, combining computational analyses with researcher intuition (Roberts et al. 2014). Defining an arbitrary number of topics solely based on the literature and without a rigorous selection method is tricky, as TM algorithm are extremely data-dependant. Different corpora – even though thematically similar – might need different settings. The need for flexible settings leads us to the second problematic node of their analytical approach: the use of an "off-the-shelf topic modelling program" (Brookes & McEnery, 2019:9) rather than a more flexible programming language. While user-friendly programs are useful to run complex analyses without coding, they can also be quite limiting in their settings.

If researchers rigorously retrieve K and use a software which allows for the needed flexibility, TM approaches can yield – in the authors' opinion – very interesting results. This holds true especially for STM, which is more powerful, adaptable, and interpretable than LDA (Roberts et al. 2016).

In the field of political and social science STM and LDA algorithms are extensively used (Bauer et al., 2017; Combei & Giannetti, 2020). TM algorithms have also been successfully adopted in related disciplines, such as literary and journalistic studies (DiMaggio et al., 2013; Jacobi et al., 2016). A growing interest in TM has arisen in linguistics as well (Murakami et al., 2017; Busso et al., 2020; Skalicky et al., 2020). TM algorithms are widely employed in these areas to analyse vast corpora of social network data (see Mishler, 2015; Hong and Davison 2010).

The great potential of such models has not gone unnoticed in forensic linguistics. Already in 2008, de Waal and colleagues recognised the usefulness of TM algorithms for "discovering the semantic context of text documents in a forensic corpus and for summarising document content" (de Waal et al., 2008: 125). In forensic linguistics the use of corpora is well-attested (among others, Coulthard, 1994; Blackwell, 2009; Kredens & Coulthard, 2012; Wright, 2017; Nini, 2018), as well as the use of computational techniques (Sousa-Silva et al., 2010; Grieve et al 2019). However, the use of TM in the field is still virtually non-existent for forensic linguistic studies. Only very recently a limited number of studies in the field of forensic computer science – although not yet in forensic linguistics – have successfully applied TM to forensic texts. For example, Guarino and Santoro (2018)'s iterative n-grams STM model enabled the discovery of key information hidden in an apparently "mono-thematic" dataset of TOR drug marketplaces. Kuang et al. (2017) similarly used a TM algorithm to thematically classify crimes using police crime reports. The model helped to discover meaningful latent crime classes, providing new insights about "the structural relationships between different formally recognised crime types" (Kuang et al., 2017: 18).

The present paper aims at partially filling this gap by employing STM in combination with a qualitative analysis of topic-related concordance lines to explore the corpus of 50 abuse letters from *Operation Heron.*

3. Data - The *Operation Heron* Corpus

The *Operation Heron* corpus includes letters sent by Margaret Walker up until the *Crimewatch* appeal. The corpus consists of 47 letters with their accompanying envelopes, as well as 3 letters without envelopes and 2 envelopes without letters (total: 50 letters and 49 envelopes)[6]. The letters are directed to private individuals (50%), healthcare professionals – especially doctors (28%) and to other categories such as Imams, city county officials, hairdressers etc. (22%).

These data represent an incomplete subset of the entire collection of abusive letters sent by Margaret Walker and it is rather unbalanced in its composition across a number of variables. It spans from January 2007 to April 2009, with the majority of texts condensed between August 2008 and January 2009 (see Figure 2). Given that the corpus comprises a selection of the whole writings by Margaret Walker, several time gaps between letters are also to be expected.

[6] Following the involvement in the investigations, the police has given permission to one of the authors and their team to use the letters in teaching and research. The documents that compose the corpus are available on *Textcrimes* (www.textcrimes.com) a free online repository held by the Aston Institute for Forensic Linguisics at Aston University. All following references to specific letters follow the numbering originally gave by the police.

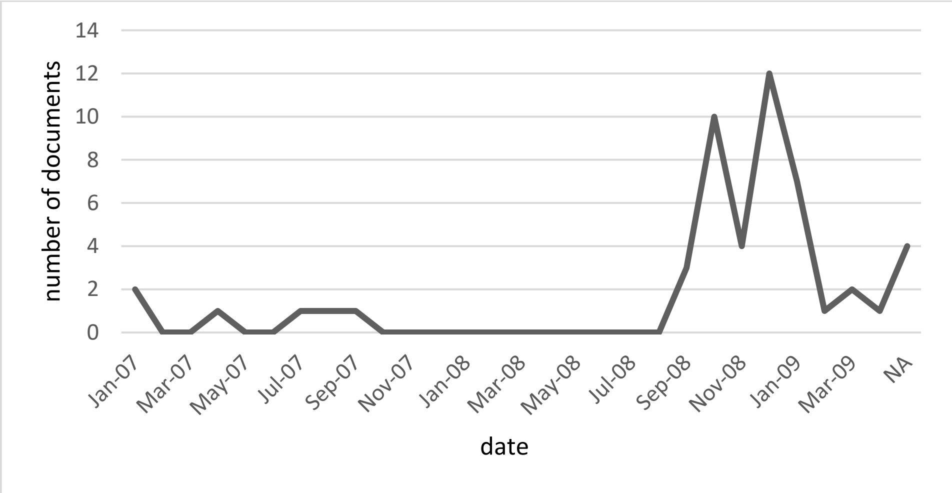

Figure 1. Distribution of letters and envelopes throughout the time span of the corpus

The letters are also extremely variable in size, ranging from some very short messages (17 words) to longer texts (351 words), with a median length of 184 words (Figure 3). The total corpus size amounts to 10,650 tokens and 9,165 words.

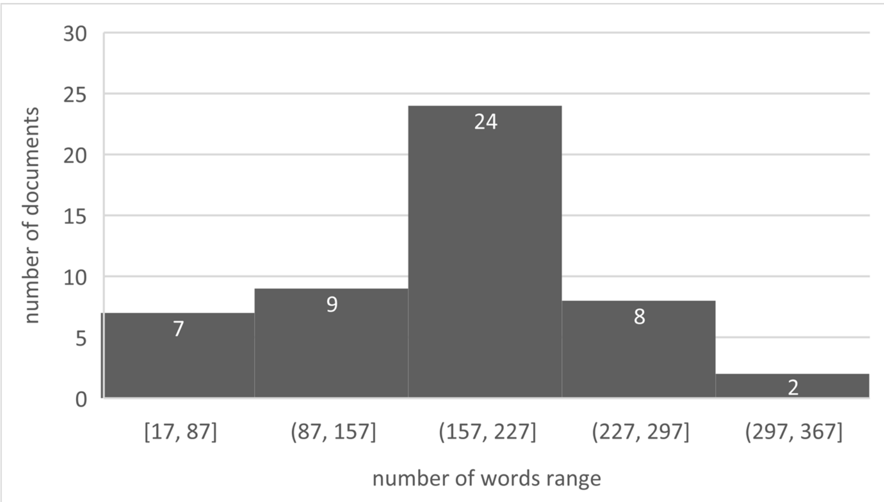

Figure 3. Distribution of letters for number of words

STM algorithms – like all machine learning – function like “black boxes”. Therefore, the limited and historical nature of this corpus makes it the ideal testbed for the potential usefulness of STM in forensic linguistics (see 1). In fact, the corpus is not only small, but comes from a resolved criminal case in which one of the authors was directly involved. Therefore, the accuracy of the STM is easily verifiable by the researchers: not only do they have an in-depth knowledge of the whole corpus, but also have had direct experience with the original investigation.

To get a more thorough picture of the thematic organization of the corpus, we further analyze it using a qualitative in-depth concordance analysis. The next sections will present the topic modelling analysis (Section 4) and concordance analysis (Section 5). In Section 6 we outline concluding remarks by discussing the significance of our findings.

4. An STM analysis of the *Operation Heron* corpus

The analysis was conducted on the statistical software *R* (R core teams, 2019), using the package *stm* (Roberts et al., 2016). All texts were first manually checked. The standard information contained in the envelopes (i.e. addressee and posting date) was used as metadata for the corresponding letter. Envelopes that contained additional text were also kept in the corpus. For example, envelope 5 (August 20th, 2007) reports, beside the addressee, the text "FLY OUR ENGLISH FLAG HIGH IN OUR SKY".

We also included as metadata the original number of the document (as given by the police) and the type of document (letter or envelope). For example, the abovementioned envelope – addressed to a doctor – is coded as "05_env_doctor_082007". The metadata are outlined in Table 1.

Table 1. metadata

Metadata	Description
Document number	The original document number, ordered cronologically from January 2007 to April 2009.
Document type	Envelope or letter.
Addressee	3 groups: "doctor", "private citizens" (pvc) and "other". [7]
Date	Month and year of the posting. When date was missing or unintelligible, we extrapolated a probable date from the letters' chronological order. For example, letter 13 (no date) was given the date October 2008 as letters 12 and 14 were also dated October 2008.

Before conducting the analysis, the necessary text preprocessing was performed using the *textProcessor()* function from *stm*. This function automatically removes stop words, punctuation,

[7] This categorization was chosen to disambiguate between letters sent to health officials in their official role, with letters often sent to their workplace, and private individuals (addressed as "mr(s)")

numbers, additional white spaces, and *hapaxes* (i.e. words that occur once in the entire corpus).[8] Stemming (i.e. reducing words to their root form) was also applied to the corpus, as this procedure has been shown to improve STM results for English (Singh and Gupta, 2016). After a preliminary data exploration, we removed the word *English* as well, whose high frequency would skew the model (out of 9,165 words, *English* has a frequency of 397). Since the word was only excluded for statistical reasons (like function words or hapaxes), it will be obviously present in the qualitative corpus analysis (Section 5). We obtained a final text corpus of 463 types and 3823 tokens.

To retrieve K (i.e. the number of topics, see 2.3) we relied on a combination of data-specific mathematical criteria and qualitative assessments. There is no *a priori* "correct" number of topics (Grimmer and Stewart 2013), but the optimal (i.e. interpretable, useful and statistically sound) value of K should be inductively found through a rigorous selection process that involves both quantitative and qualitative assessments. (Chang et al., 2009).

We perform a number of automated tests on a range of possible values using the *searchK* function, which measures quantitative criteria of 'good fit' for the values specified, such as held-out likelihood[9] and residual analysis[10]. After examining quantitative criteria for an interval of values between 2 and 10, we selected 4 as the best K value.[11]

After careful examination of the top 50 words per each thematic cluster (top 10 words are reported in Figure 4), all authors collectively derived labels based on topic content and interpretability. This methodological approach is consistent with previous research (Di Maggio et al. 2013; Roberts et al. 2014). The labels chosen (see figures 4 and 5) are: (1) Immigration[12], (2) Politics A, (3) Healthcare[13], (4) Politics B.

Although K=4 was chosen as the optimal value, some overlaps across topics are expected. This is predominantly due to the corpus's limited dimensions and the interrelatedness of all thematic nodes in Margaret Walker's writings: the discourse around immigration is tied both to Healthcare (as she condemns England for hiring immigrant doctors), and to Politics (because "English tax is for the

[8] As STM is based on word *co-occurrences*, hapaxes are not useful in identifying the topics.
[9] i.e. the likelihood of the same model on documents initially excluded by the model. The smaller held-out likelihood is, the better the model fits the actual data.
[10] The analysis of the differences between predicted output and measured data. It is used to validate the model by inspecting if model assumptions are justified.
[11] We kept the interval of K low given the small dimension of the corpus.
[12] One anonymous reviewer correctly points out that ethnic minorities' terms ("black", "Asian", "wog", "nigger", "paki" and "muslim") do not necessarily relate to Immigration. However, the writer considers all "foreigners" – including Scottish nationals – and non-Caucasian as "immigrants". The consistency of the label "Immigration" for Topic 1 is reinforced by the presence of the word "tax", as one of the recurring themes within this topic is that English tax should only serve for English nationals (Table 3).
[13] One anonymous reviewer points out that beside "hospit" and "doctor" no other terms in figure 4 seems to refer to Health. However, all the adjectives present in the top ten words ("dirti", "filthi", "disgust", "ancestr") always co-occur with terms referred to hospital structures, which are considered to be in appalling state. Moreover "foreign", "immigr" and "wog" are commonly used derogatory terms to refer to non-Caucasian health professional (Table 5).

English", and not for immigrants). At the same time, Politics and Healthcare are also strictly related, as she blames the government for the poor state of healthcare facilities, and she encourages Scotland to vote for independence and pay for their own welfare structure. This interrelatedness makes the use of STM even more interesting, as a human coder would not have picked up probabilistic regularities across documents that the algorithm is instead able to discern. This does not mean that the distinctions across topics are meaningless, as the criterion of exclusivity is still met, i.e. high-probability words for each topic tend to either not appear in other topics or to appear as low-probability words (Roberts et al. 2014).[14] For example, the word "tax" appears in three topics. However, in Topic 1 it has the highest-ranking probability (0.07), while in Topics 3 and 4 its probability is only slightly higher than 0.01 (0.013 and 0.014 respectively).

Politics A and Politics B have a more fine-grained distinction than the other topics, and they share several words. They were kept as distinct topics since further analyses reveal that the distinction between these two latent thematic dimensions – which human coders would probably not have picked up – between the two carries important considerations (5).

As Figure 5 depicts, political themes (Politics A and B) are the most discussed in the corpus. Thus, distinguishing fine-grained latent features in the political 'discourse' of the author can help identifying potentially important argumentative nuances. Particularly, the two topics exhibit slight differences in lexical choices and sub-themes, and the model captured what - in our opinion – can be considered a stylistic progression, as will be thoroughly explained in Section 5. Concordance analysis reveals that Politics A mainly deals with complaints about the past and present situation of the UK (mainly immigrants and Scots taking advantage of English taxpayers), while Politics B mostly revolves around a demand for (future) Scottish independence.

After having thus defined the topics, we model the effect of the covariates. To explore all possible dimensions, we estimated effects for all our covariates (Table 1), for both topical content and prevalence.

[14] For a corpus this small, "high-probability" – as is apparent in figure 4 – refers to the top 1 or 2 words.

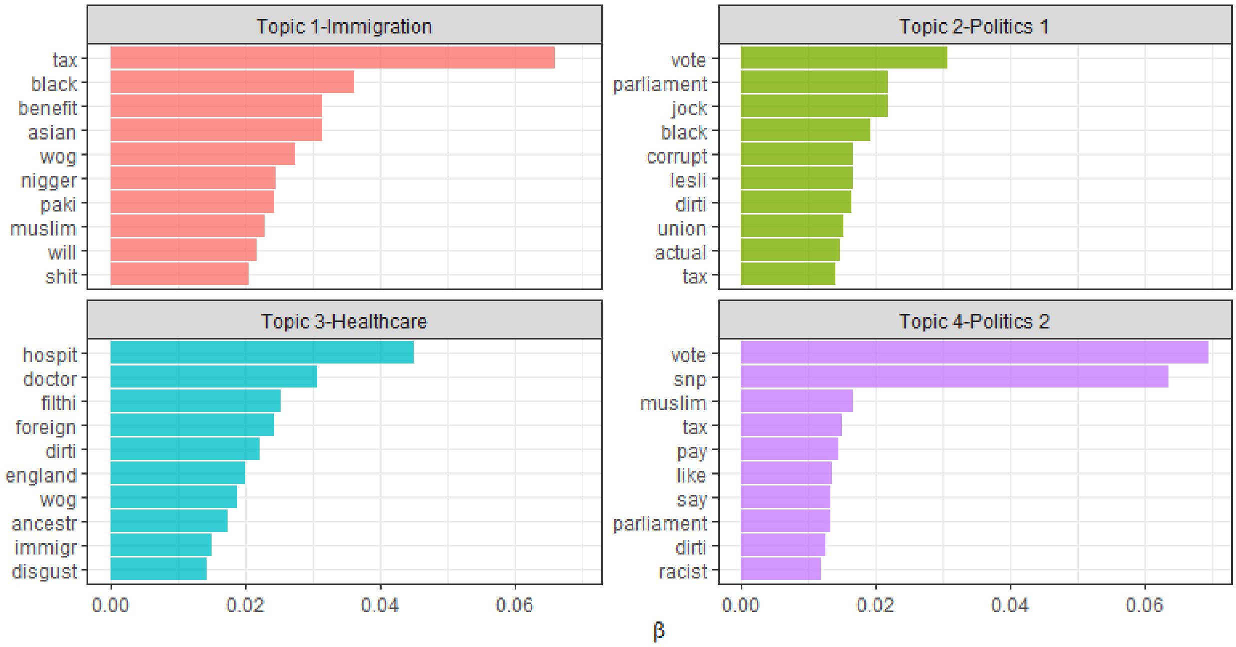

Figure 4. The top 10 word probabilities for each topic

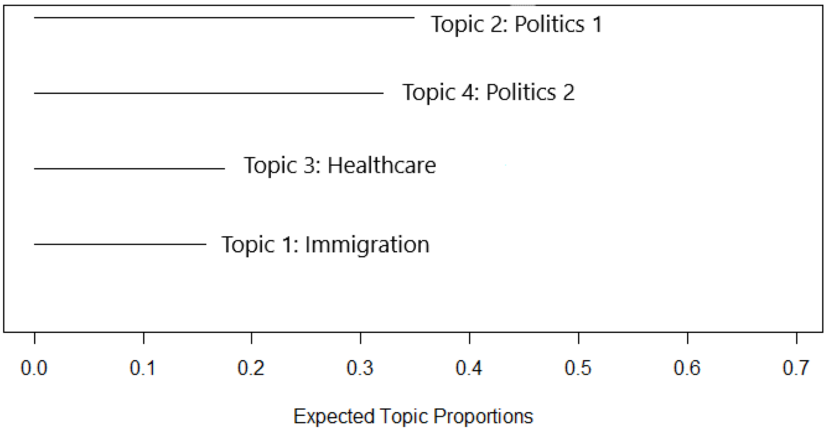

Figure 5. Proportion of topics in descending order

4.1. Topical content and prevalence estimation

Preliminary qualitative exploration suggested high consistency of word choices within topics. To explore possible latent dimensions, we conditioned topical content on the covariates of "addressee" (pvc – doctor – other), "document type" (envelope - letter) and date. That is, we investigated whether content change across topics is affected by different addressee, by time, or by the difference in document type (envelope or letter). No significant effects for topical content were found for any of the covariates. In other words, within-topic lexical choices do not differ significantly as a function of either addressee, date, or document type. This indicates a consistent vocabulary in each topic.

While topic content is consistent across covariates, some thought-provoking differences arise when estimating effects for topical prevalence – i.e. frequency with which a specific topic is used – on the covariates of "addressee" and "date". That is, the frequency with which topics are discussed varies significantly depending on who the letter is sent to, and topic distribution changes over time.

Addressee: while topics 1 (Immigration) and 4 (Politics B) are not addressee-specific, we find that topic 2 (Politics A) is significantly more likely to be addressed to private citizens ($p < 0.05$), while topic 3 (Healthcare) is mainly addressed to doctors, and is hence less likely to appear in letters directed to private citizens ($p < 0.05$) and (marginally) to other recipients ($p < 0.06$)[15]. Figure 6 plots the probability of each topic per addressee.

[15] The regression assessing covariates' significance codes the first level alphabetically as the intercept. In this case, 'doctor' is the intercept, and hence all other levels are assessed compared to it.

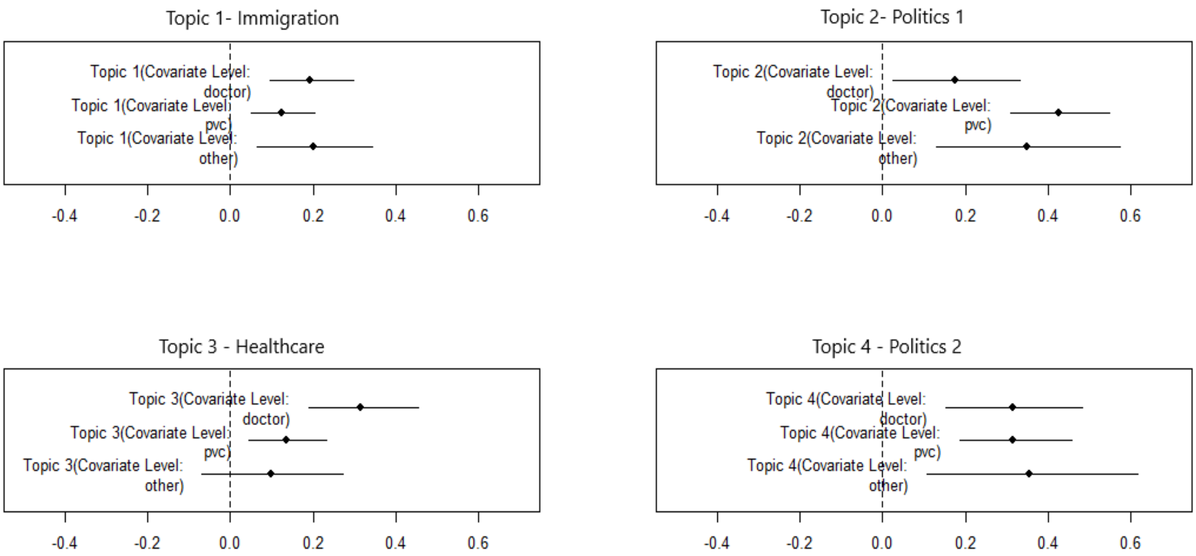

Figure 6. Probability of each topic per addressee

Date: While within-topic vocabulary is consistent, the distribution of topics changes over time, as can be seen in Figure 7.[16] We can see that Immigration (pink) is distributed evenly across time, albeit with a low probability (Figure 3). Politics A (green) is positively affected ($p < 0.0001$) by date, i.e. tends to appear in later letters. In fact, the plot reveals that Politics A increases significantly from December 2008 to February 2009 ($p < 0.005$). Healthcare (light blue) shows an inverse trend: it is significantly more discussed in early letters ($p < 0.005$). That is, earlier letters were predominantly sent to health officials and dealt with health matters. Politics B (violet) increases in September and October 2008 (respectively, $p = 0.005$, $p < 0.05$), which suggests a shift of priorities for the writer during 2008: from predominantly health-related to more political concerns.

By combining the four trends over time, a coherent narrative of the different themes in the letters emerges, with Healthcare (Topic 3) becoming less prominent from 2008 onwards and Political themes (Topics 2 and 4) acquiring more and more importance with time (Figure 8).

If we relate the topics chronological progression to the already discussed differences in addressee, we can further refine our findings. In January 2007, the author begins sending abusive letters especially pertaining to healthcare (Topic 3) to medical doctors. As time progresses, she moves away from healthcare to begin writing what she has referred to as “propaganda leaflets”: the themes become deliberately political (Topic 4), and throughout 2008 she specializes in sending these

[16] The date covariate was converted into factor to estimate the effect on each month separately.

“political leaflets” to private citizens (Topic 2).[17] Beside these topics, the theme of Immigration (Topic 1) is constantly present in her writings.

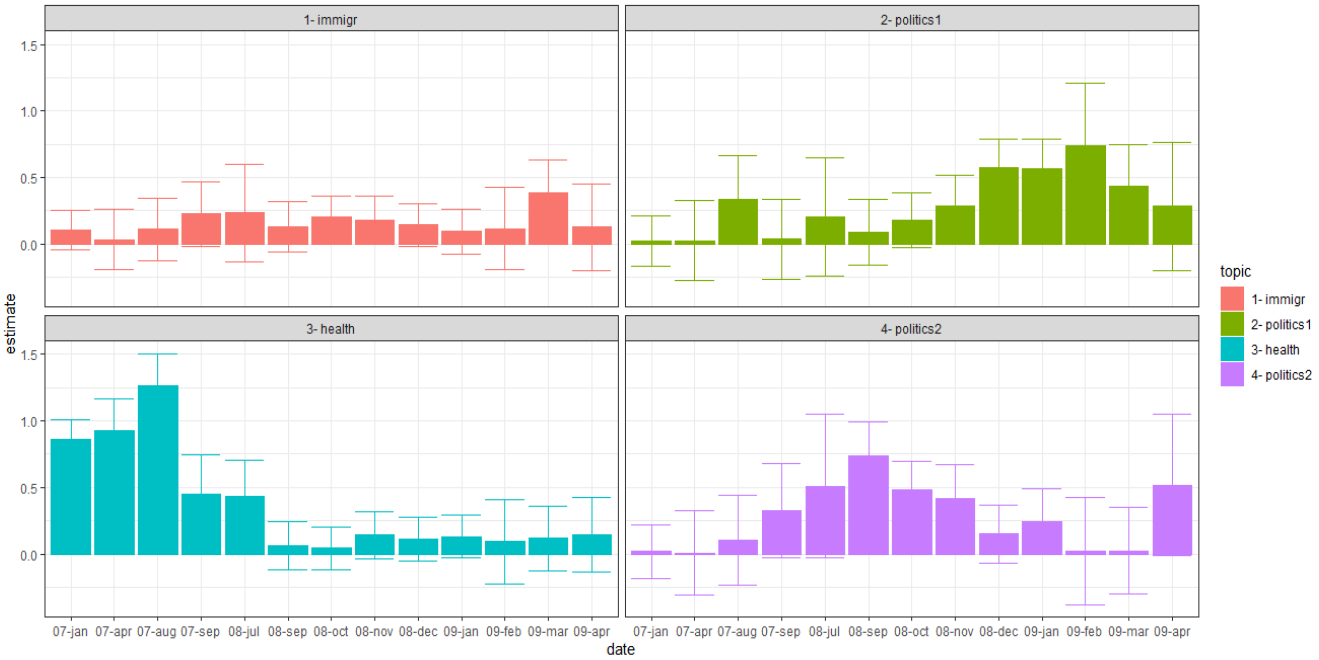

Figure 7. Topic estimates per month. The date is labeled with the YEAR-MONTH convention

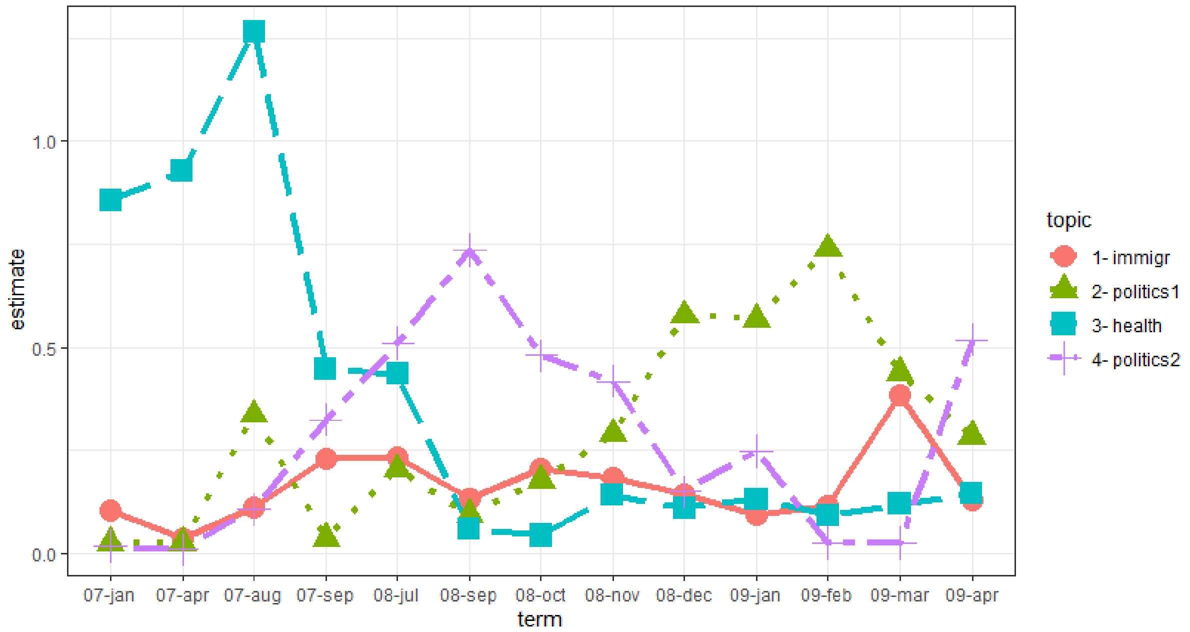

Figure 8. Line plot of the estimates for all four topics per each month.

[17] One anonymous reviewer points out that topic and addressee appear to highly correlate, and that thus topic changes could relate to a shift in addressee (from doctors to private individuals). While it is indeed true that addressee and letter topic are related, this does not suffice in explaining topic changes over time. There are also letters addressed to doctors which contain political propaganda (e.g. letter 29), and letters sent to private citizens that contain health-related issues (e.g. letter 6). Moreover, a letter will often contain a mixture of different topics (e.g. letter 21).

5. Qualitative corpus analysis

As explained in Section 2.3, STM is a data-driven technique for identifying latent topics in texts. The topics that STM algorithms produce are simple lists of words with a high probability to co-occur in a corpus. STM, however, does not give any insight into the main statements these topics entail, which is why we decided to carry out a qualitative corpus analysis to understand the content of the topics in detail. In line with this, the aim of the qualitative corpus analysis was to identify the main statements within each topic and establish how the prevalence of these statements changes over time in the letter series.

We here understand *topic-related statements* as the author's inherently subjective descriptions of her perceived socio-political environment and how this environment should change in her view. In line with this definition, the author's statements can be grouped into two categories: complaints about the current state of affairs and demands for future change. Some of the complaints and demands are related and thus form arguments where complaints work as premises whereas demands form the conclusion of arguments. For example, one of the author's immigration-related arguments is that because immigrants/ethnic minority people live off the state and they are also criminals/terrorists, only the English should benefit from the tax collected in England. This argument consists of two complaints (immigrants/ethnic minority people live off the state and immigrants/ethnic minority people are criminals/terrorists), which serve as the premises, and one demand (only the English should benefit from the tax collected in England), which forms the conclusion of the argument. This example also illustrates that we conceptualise arguments as sets of related statements where the authors' complaints as premises support her demands as conclusions. Each complete argument consists of one or multiple premises and a single conclusion.

Note that some of the author's arguments are incomplete: not all complaints are followed by a corresponding demand and not all demands are supported by complaints. The primary units of our analysis are therefore statements (complaints and demands) whereas arguments serve as secondary analytical units. Also note that neither the topic-related statements nor the arguments are topic-specific as the same statements and arguments can occur across multiple topics (see 4 above). This is because the STM topics involve some inevitable overlaps, as multiple topics can share the same words. As we aim to provide a comprehensive description of each topic, these overlapping words were not excluded from our analysis. In Section 5.1, we discuss the methodology behind the corpus-based analysis of the topic-related statements. Section 5.2 gives an overview of the statements that we identified while Section 5.3 focuses on how the prevalence of these statements changes over time in the letter series.

5.1. Methodology

As a starting point for the qualitative corpus analysis, we extracted the top 20 topic-related words for each topic from the output of the STM algorithm. These words are listed in Table 2.

Table 2. The top 20 topic-specific words

Immigration	Politics A	Healthcare	Politics B
tax, black, benefit, asian, wog, nigger, paki, muslim, will, shit, scot, jew, indian, pay, scum, grab, drug, scrounge, chink, huge	vote, parliament, jock, black, corrupt, leslie, dirty, union, actual, tax, two, racist, year, pocket, vile, scot, asian, long, priority	hospital, doctor, filthy, foreign, dirty, england, wog, ancestral, immigration, disgusting, country, elderly, muslim, bed, make, racist, given, masquerade, destroy	vote, snp, muslim, tax, pay, like, say, parliament, dirty, racist, fly, country, brown, kick, scotland, scum, pension, fund, black

Table 2 and figure 4 suggest that there are some overlaps among the word lists, as explained in the related section. For example, the word *tax* is featured in three different topics, including *Immigration*, *Politics A*, and *Politics B* – although the criterion of exclusivity is still met (see 4). This might be because *tax* is the third frequent content word in the corpus after *English* and *vote*, which are also represented in more than one topic, suggesting that the topic overlaps can be explained by the fact that high-frequency words have a high probability to co-occur with any other word in a corpus.

We used the words in Table 2 to identify the topic-related n-grams within the corpus. We defined these n-grams as sequences of words that included at least two topic-related words from the same topic. We created separate lists of n-grams for each topic. Due to the fact that the STM algorithm we used produced overlapping wordlists for the four topics, some of the topic-related n-grams do occur in multiple topics. To ensure that we provide a comprehensive analysis of each topic, these overlapping n-grams were not excluded. We focused on the n-grams with a length of 2–6 words and a minimum frequency of 2. In total, we found 197 topic-related n-grams, 29 for *Immigration* (e.g., *Muslim wogs*), 73 for *Politics A* (e.g., *English parliament*), 42 for *Healthcare* (e.g., *English doctors*), and 53 for *Politics B* (e.g., *vote SNP*). We decided to focus on n-grams rather than words to keep the number of concordance lines analysed at a manageable level without damaging the comprehensiveness of the study.

Once we had the topic-related n-grams, we produced the topic-related concordance lines for each topic. These concordance lines were 200-character long text chunks from the letters containing a topic-related n-gram. In total, we retrieved 1,047 concordance lines from the corpus, 97 for *Immigration*, 394 for *Politics A*, 196 for *Healthcare*, and 360 for *Politics B*. The different number of concordances retrieved is in line with the prevalence proportion of topics in the corpus (see Figure 7 above). Similarly to topic-related words and n-grams, some of the topic-related concordance lines do occur in the multiple topics.

After this, three members of the research group looked into the topic-related concordance lines independently to find the recurring statements within each topic. As statements are arguably not formal units of language, the analysis heavily relied on the researchers' interpretation of the concordance lines, which is unavoidable in discourse analysis. To improve the reliability of this initial analysis, researchers were given the instruction to read each concordance line and describe each of the author's recurring complaints and demands in one declarative sentence. This was followed by a series of discussions, during which we finalised the list of topic-related statements for each topic.

Finally, two members of the research group, one of whom was not involved in establishing the list of topic-related statements, coded the presence or absence of the topic-related statements in all concordance lines to look at how the prevalence of the topic-related statements changes over time in the letter series.

We employed two annotators to improve the reliability of the annotation process. As the first coding round showed some disagreement between the two coders (percent agreement: 85.3%, Krippendorff's Alpha: 0.648[18]), we decided to run a second coding round after a series of discussions between the two coders on their annotation strategies. By the end of this second coding round, the two annotators reached almost complete agreement (percent agreement: 97.8%, Krippendorff's Alpha: 0.945) when tested on the full set of topic-related concordance lines for each topic.

5.2. Topic-related statements

This section gives an overview of the topic-related statements that were identified during the qualitative analysis of the concordance lines. Table 3 presents the statements within the topic of *Immigration*, illustrates these statements with concordance lines from the corpus, and displays their relative frequency. The percentages show the proportion of those topic-related concordance lines

---

[18] Percent agreement is the proportion of the number of units of analysis on which two coders' categorisations match perfectly to the total number of units coded (Krippendorff 2011a, 2). Krippendorff's Alpha is a chance-corrected reliability coefficient developed to measure the agreement among coders. It varies between 0 (complete disagreement) and 1 (complete agreement) (Krippendorff 2011b, 1).

where a particular statement occurs to the total number of topic-related concordance lines within the topic of Immigration. Table 3 also indicates whether a statement is a complaint or a demand and whether it serves as a premise or conclusion of an argument. Arguments are numbered to clarify which statements belong to which argument. Some of the arguments are incomplete and hence lack either a premise or a conclusion. Tables 4–6, outlining the statements within the other three topics, *Politics A*, *Healthcare*, and *Politics B*, are structured in the same way. Note that the illustrative concordance lines in Tables 3–6 contain highly offensive language, which some readers might find disturbing.

Table 3. The statements within the topic of Immigration

Statement	**Example**	**Frequency (100% = 97)**
Immigrants/ethnic minority people live off the state [complaint] [premise of argument (1)]	*Gun crazy Nigger rapist - grabbing Jews complaining English tax loving dirty Indians benefit demanding Chinkies sly yellows drugged out Niggers & Blacks.*	60.8%
Immigrants/ethnic minority people are criminals/terrorists [complaint] [premise of argument (1)]	*not for dirty blackmailing murderers bombers Asians and Blacks Muslim scum Islam infiltrating our English hospitals silent killers and wogs in city centres breeding like flies*	56.7%
Only the English should benefit from the tax collected in England [demand] [conclusion of argument (1)]	*The English do not pay tax for foreign cock. Nip and tuck. Neither do the English pay tax for foreign cliteros remodel & canal reopening. The English pay tax for their number one priority. English benefit.*	39.8%
Immigrants/ethnic minority people abuse the NHS either as medics or as patients [complaint] [premise of argument (2)]	*Open the ward doors look inside. - scrounging Pakis, grabbing Jews, demanding Chinkies. Complaining Indians. Muslim scum. Gun running Niggers are in our English tax ancestral hospital beds. Kick the filthy foreigners out*	34%
Immigrants/ethnic minority people should not work as healthcare professionals in England [demand] [conclusion of argument (2)]	*Us English abhor untrustworthy dirty Muslim wogs masquerading as doctors in our English ancestral hospitals - they have to get out.*	21.6%

Table 3 shows that the author openly expresses some strong anti-immigration and racist sentiments in her letters, accusing immigrants and ethnic minorities of exploiting the English benefit system, engaging in criminal behaviour and also abusing the National Health Service (NHS). The illustrative concordance lines also demonstrate that the author uses a plethora of derogatory terms when referring to various ethnic and religious communities, such as *Nigger, Chink, Paki, wog*, and *scum*. In general, the author depicts immigration and ethnic/religious diversity as a serious threat to the safety and prosperity of England, demanding that immigrants and ethnic minority people be denied access to social services and excluded from the NHS.

The second topic, *Politics A*, revolves around Scottish and English independence. The author accuses the Scots of taking advantage of the English and urges them to vote for the Scottish National Party (SNP) so that Scotland can become an independent nation, which would end Scotland's supposed reliance on the English taxpayers. She also argues that England should have its own parliament without Scottish MPs and should leave the supposedly corrupt European Union as well. In addition to these demands, the author gradually develops the idea that Scottish politicians, most notably Scottish-born former Labour Prime Minister Gordon Brown, assumed complete control over British politics in an attempt to exploit the English public and eventually destroy England by actively encouraging mass immigration into the UK. The most prominent complaint within this topic, namely immigrants, ethnic minority people and the Scots live off the English taxpayers, is also akin to one of the statements within the topic of *Immigration*. This shows that the topics the STM algorithm identified in the corpus are not entirely distinct from one another as there are some overlaps between them.

Table 4. The statements within the topic of Politics A

Argument	**Example**	**Frequency (100% = 394)**
Immigrants/ethnic minority people/the Scots live off the English taxpayers [complaint] [premise of arguments (1), (2)]	*In Scotland the parasite Scots and their leeching wogs get care homes free paid for by English tax*	64%
Only the English should benefit from the tax collected in England [demand] [conclusion of argument (1)]	*The English are the number one priority. Our tax our benefits.*	27.2%

The Scots should fund their own pensions/lifestyle [demand] [conclusion of argument (2)]	*Vote S.N.P and you work for your gold lined public purse (English tax funded) pension. Act like an English gentleman and fund it by working yourself the English demand*	22.8%
The Scots voted for the Labour Party and hijacked the British Parliament so that they can exploit the English [complaint] [premise of arguments (3), (4)]	*The shame of the free loading Scot hypocrites Leslie Unionist Scots who voted for Labour in Glenrothes did not vote for corrupt anti-English racist patsy Muslim Brown. They did in fact vote for fear*	26.7%
Gordon Brown hates the English and wants to destroy England with mass immigration [complaint] [premise of arguments (3), (4)]	*Not given our permission for immigration into our country England. You voted the scum in, not the English. Gordon Brown anti-English racist scum. With a mission to destroy England. The shit will go. Us English abhor dirty untrustworthy Muslim and Asian Blacks*	14%
England should have its own parliament without Scottish MPs [demand] [conclusion of argument (3)]	*Vote S.N.P and get your dirty, freeloading "Martin" crooks out of our English parliament*	18.8%
Scotland should be independent from England and the Scots should vote SNP to reach independence [demand] [conclusion of argument (4)]	*Vote SNP and get out of our lives, you have destroyed England. Your time is up with us.*	18%
The UK should leave the European Union [demand] [conclusion of argument (5)] [part of incomplete argument]	*The English say, we want out of Europe it stinks of corruption. The English say we want our own English Parliament*	16.5%

Table 5. The statements within the topic of Healthcare

Statement	Example	Frequency (100% = 196)
Foreign/ethnic minority doctors are not real doctors [complaint] [premise of argument (1)]	*Us English have never given a mandate for mass immigration into our country or hospitals, it is therefore illegal for ethnics minorities to masquerade as doctors in our English hospitals.*	32.1%
England should only fund/train/employ English doctors and should not fund/train/employ foreign doctors (demand) [conclusion of argument (1)]	*Us English demand 99.99% English doctors in our English ancestral hospitals*	41.8%
English taxpayers do not receive proper treatment in English hospitals [complaint] [premise of argument (2)]	*Our elderly English after 50 yrs on high-tax and nat insurance payments are denied drugs and care.*	12.2%
Only English people should be treated in English hospitals and foreigners/ethnic minority people should not as they do not pay taxes [demand] [conclusion of argument (2)]	*Kick the filthy foreigners out of our ancestral - English tax payers hospitals*	15.3%
English hospitals are in a despicable state [complaint] [premise of argument (3)]	*Clean up our filthy English hospitals they are disgusting.*	18.9%
English doctors should not be forced overseas to work [demand] [conclusion of argument (3)]	*Train only our own English doctors, and don't force our good English doctors overseas.*	10.7%

When discussing the topic of *Healthcare*, the author's key demand is that England should only fund, train and employ English doctors while foreign doctors should be completely excluded from the NHS as they are not real doctors. The author also expresses a general criticism of the NHS, complaining about the poor standard of care that taxpayers receive in English hospitals, which the author depicts as institutions unfit to serve their purpose due to a lack of hygiene, proper facilities, competent staff and priority given to English patients.

Table 6. The statements within the topic of Politics B

Argument	**Example**	**Frequency (100% = 360)**
Immigrants/ethnic minority people/the Scots live off the English taxpayers [complaint] [premise of arguments (1), (2)]	*Leslie Union Jocks love of English cash far exceeds respect for their English paymaster. Blacks Pakis Asians breeding like flies*	43.1%
Scottish pensions are funded from the English taxpayers' contributions [complaint] [premise of arguments (1), (2)]	*your massive gold lined (English tax funded) public pension*	9.7%
The Scots should fund their own pensions/lifestyle [demand] [conclusion of argument (1)]	*The English say vote S.N.P and ask for your own health care. Vote S.N.P. and stand on your own two feet. Stop fleecing our English taxes.*	34.4%
Only the English and not immigrants/the Scots should benefit from the tax collected in England [demand] [conclusion of argument (2)]	*English taxes are for English benefits*	29.7%
Gordon Brown hates the English and wants to destroy England with mass immigration [complaint] [premise of arguments (3), (4)]	*Us English have never given a mandate for immigration into our country England. It is therefore illegal Brown paranoid schizophrenic racist scot and half caste Blair are hell bent to destroy the English.*	19.2%
The Scots voted for the Labour Party and hijacked the British Parliament so that they can exploit the English [complaint] [premise of arguments (3), (4)]	*By voting corrupt racist Labour in Glenrothes to grab English cash traitor Jocks also willingly and purposely voted for mass immigration*	15%
Scotland should be independent from England and the Scots should vote SNP to reach independence [demand] [conclusion of argument (3)]	*Vote S.N.P for your own parliament and democracy*	47.8%

England should have its own parliament without Scottish MPs [demand] [conclusion of argument (4)]	*The English say we want our own English Parliament*	29.7%

Table 6 shows that the statements within the topic of *Politics B* are virtually identical to those discussed under *Politics A*. This is why we call these topics *Politics A* and *B* in recognition that these are in fact two closely-related versions of the same overarching content topic. The overwhelming similarity of *Politics A* and *Politics B* also highlights a key limitation of structural topic modelling: while STM algorithms provide a principled method for establishing topics in a large set of texts, the differences in word occurrences that they pick up might not be visible, or significant, to the researcher. In this case, the only notable qualitative difference between *Politics A* and *B* is that a call for Brexit is only present in *Politics A* but absent from *Politics B*. There is, however, another key difference between these two topics, which is quantitative in nature. *Politics A* mainly focuses on the past and present as it is dominated by the complaint that immigrants, ethnic minority people and the Scots have taken financial advantage of the English taxpayers (64% in *Politics A* but only 43.1% in *Politics B*). In contrast, *Politics B* is dominated by a demand for future Scottish independence (only 18% in *Politics A* but 47.8% in *Politics B*). This quantitative difference might also be the main reason why the STM algorithm identified *Politics A* and *B* as two separate topics rather than one.

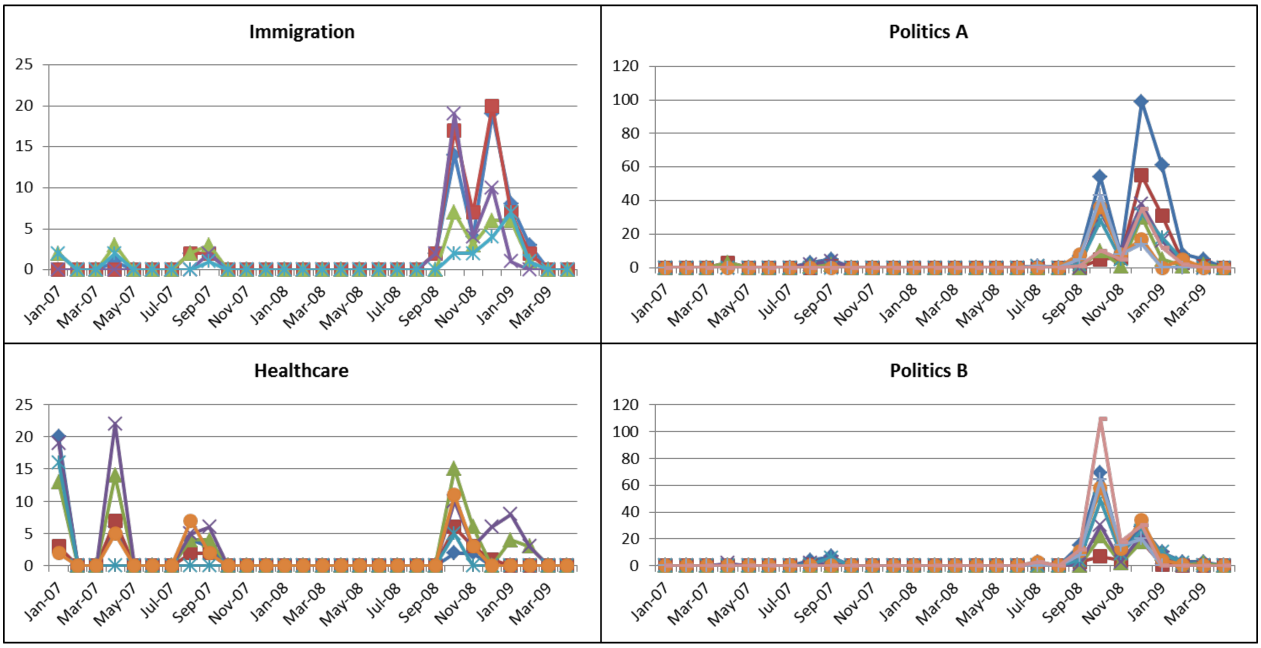

Figure 9. The prevalence of topic-specific arguments over time in the letter series

Figure 9 shows the number of times individual topic-related statements occur in letters sent in a specific month between January 2007 and March 2009. Figure 9 therefore reveals how the prevalence of the topic-related arguments changes over time in the letter series. It shows that the various statements within the same topic tend to move together. For example, within the topic of *Immigration*, most statements are the most frequent in the letters sent in October 2008 and January 2009 while they remain less frequent outside this period. This is a general pattern that is applicable to the other topics as well, suggesting that the author is rather consistent in the way in which she constructs these topics over time in this letter series. Figure 7 also confirms the finding from Section 4 that the early letters mainly focus on the topic of *Healthcare* while the late letters are dominated by political themes. The topic of *Immigration*, in contrast, is present in both the early and late letters albeit with a lower frequency than the dominant topics.

6. Conclusions

The present paper has presented an innovative methodology which combines quantitative and qualitative insights to explore semantic themes in a corpus. Specifically, our main aims were to identify the key latent topics and topic-related statements (complaints and demands) within the *Operation Heron* letter series and to establish how the prevalence of these changes over time and across addressees in the letter series. Using structural topic modelling (STM), we found four recurring topics in the corpus, including *Immigration*, *Politics A*, *Healthcare*, and *Politics B*. We were also able to point out that although the topics of *Immigration* and *Politics B* are not addressee-specific, *Politics A* is the most prominent in the letters sent to private individuals while the topic of *Healthcare* is mainly discussed in the letters addressed to healthcare professionals. We also found that *Healthcare* is the most frequent topic in the early letters from 2007 but it loses prominence over time while *Politics A* and *B* gain prevalence in the late letters from 2008 and 2009. At the same time, the topic of *Immigration* is consistently present throughout the letter series with a low frequency.

The qualitative analysis of the topic-related concordance lines revealed the recurring statements that make up these four topics. The topic of *Immigration* is dominated by the complaint that immigrants and ethnic minority people supposedly exploit the English taxpayers by abusing social services. The key demand within the topic of *Healthcare* is that England should only fund, train and employ English doctors while foreign doctors should be completely excluded from the NHS. Finally, both *Politics A* and *B* revolve around Scottish and English independence. The key difference between these two topics is that *Politics A* is characterised by the complaint that not only immigrants and ethnic minorities but also the Scots exploit the English public whereas a demand for Scottish independence takes prominence in *Politics B*. The above complaint and demand can be conceptualised as the premise and conclusion of the same argument, respectively. Concordance analysis also demonstrated that the various statements within the same topic tend to gain or lose prevalence together over time, confirming that the content of the four topics remains consistent throughout the letter series, as the topical content analysis suggested.

To concludethis paper demonstrated that the qualitative analysis and annotation of topic-related concordance lines can help us make sense of the output of STM algorithms, which would otherwise work as ‘black boxes’. Consequently, one of the key implications of this project for corpus linguistics is that concordance analysis is a crucial addition to STM as it allows for the in-depth exploration of topic-related statements within a wide range of corpora. The combination of STM and concordance analysis can also be used for mapping the evolution of topics over time, suggesting that this paper may serve as a methodological inspiration to other projects focusing on topic changes in other datasets, especially where time is a relevant factor.

Moreover, when it comes to forensic linguistic research, this paper showed that the combination of STM and the qualitative analysis of topic-related concordance lines is extremely useful for uncovering latent topics, identifying topic-related statements and detecting topic changes in texts relevant to forensic linguistics. This is a novel technique in forensic linguistics which would be a useful addition to the methodological toolset currently used in the field. One of the other benefits of the STM + concordance analysis approach is that it provides a principled tool for analysing a macro-level textual phenomenon, namely latent topic change. It can thus supplement the more traditional analysis of micro-level linguistic features (Coulthard et al2017).

Finally, we also argue that the STM + concordance analysis technique has potential use in investigative work as it allows for eliciting non-trivial thematic patterns in a large set of texts. This technique can pinpoint how the prevalence of these patterns changes over time or across text types, which can be especially useful when dealing with a large number of texts from one or multiple authors. We believe that there is potential for further practical use in authorship analysis (Grant, 2007) as stark or subtle topic changes might indicate a change in authorship. However, given that forensic authorship analysis generally relies on a large number of various linguistic features, including character n-grams, word n-grams, spelling, grammatical features, lexis, and higher-level pragmatic features, to capture idiolectal linguistic preferences, further research in this area would be beneficial in establishing the extent to which structural topic modelling can contribute to authorship analysis.

References

Anandarajan, M., Hill, C. and Nolan, T. 2019. 'Probabilistic topic models', *Practical Text Analytics* (pp. 117-130). Springer, Cham.

Argamon, S., M. Koppel, J. Fine, and AR. Shimoni. 2003. 'Gender, genre, and writing style in formal written texts', *Text - Interdisciplinary Journal for the Study of Discourse*, 23(3), pp. 321-346.

Baker, P., C. Gabrielatos, T. McEnery. 2013. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press.* Cambridge: Cambridge University Press.

Baker P., C. Gabrielatos, M. KhosraviNik, M. Krzyzanowski, T. McEnery and R. Wodak. 2008. 'A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press', *Discourse and Society* 19(3) pp. 273–306.

Baker, P. and McEnery, T. 2019. 'The value of revisiting and extending previous studies: The Case of Islam in the UK Press', in Scholz, R. (ed.), *Quantifying Approaches to Spoken Discourse for Social Scientists,* 215–249. London: Palgrave Macmillan.

Bauer, P. C., P. Barberá, K. Ackermann, and A. Venetz. 2017. 'Is the left-right scale a valid measure of ideology?', *Political Behavior*, *39*(3), pp. 553-583.

Berman, R. 2008. 'The psycholinguistics of developing text construction', *Journal of child language*, 35(4), pp. 735–71.

Blackwell, S. 2009. 'Why forensic linguistics needs corpus linguistics', *Comparative Legilinguistics*, 1, pp. 5-19.

Blei, DM. 2012. 'Probabilistic topic models', *Communications of the ACM*, 55(4), pp. 77–84.

Brookes, G. and T. McEnery. 2019. 'The utility of topic modelling for discourse studies: A critical evaluation', *Discourse Studies*, 21(1), pp. 3-21.

Busso, L., C.R. Combei, O. Tordini. 2020. 'The mediatization of femicide: a corpus-based study on the representation of gendered violence in Italian media', *Analisi Linguistica e Letteraria* xxviii, pp. 29-40.

Castelle, M. 2018. 'The Linguistic Ideologies of Deep Abusive Language Classification', in Fišer, D., R. Huang, V. Prabhakaran, R. Voigt, Z. Waseem and J. Wernimont (eds) *Proceedings of the 2nd Workshop on Abusive Language Online*, pp. 160–170. Association for Computational Linguistics.

Clarke, I. and J. Grieve. 2017. 'Dimensions of Abusive Language on Twitter', in Waseem, Z., WH Kyong Chung, D. Hovy and J. Tetreault (eds.), *Proceedings of the First Workshop on Abusive Language Online*, pp. 1–10. Association for Computational Linguistics.

Chang, J., S. Gerrish, C. Wang, JL Boyd-Graber and DM Blei. 2009. 'Reading tea leaves: How humans interpret topic models', *Advances in neural information processing systems*, pp. 288-296.

Combei, CR and D. Giannetti. 2020. 'The Immigration Issue on Twitter Political Communication. Italy 2018-2019'. *Comunicazione politica*, *21*(2), pp.231-263.

Coulthard, M. 1994. 'On the use of corpora in the analysis of forensic texts', *International Journal of Speech, Language and the Law*, *1*(1), pp. 27-43.

Coulthard, M., A. Johnson, D. Wright. 2017. *An Introduction to Forensic Linguistics: Language in Evidence*. London: Routledge.

DiMaggio, P., M. Nag and D. Blei. 2013. 'Exploiting Affinities between Topic Modelling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding', *Poetics* 41, pp. 570–606.

Grant, T. 2007. 'Quantifying evidence in forensic authorship analysis', *International Journal of Speech, Language and the Law*, 14(1), pp. 1–25.

Grant, T. 2008. 'Approaching questions in forensic authorship analysis', in Gibbons, J. and MT. Turell (eds.), *Dimensions of Forensic Linguistics*, pp. 215-230. Amsterdam: John Benjamins.

Grieve, J., Clarke, I., Chiang, E., Gideon, H., Heini, A., Nini, A., & Waibel, E. 2019. 'Attributing the Bixby Letter using n-gram tracing', *Digital Scholarship in the Humanities*, *34*(3), pp. 493-512.

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

Guarino, S., and M. Santoro. 2018. 'Multi-word structural topic modelling of ToR drug marketplaces', in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, Laguna Hills, CA, 2018, pp. 269-273.

Hardaker, C. 2010. 'Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions', *Journal of Politeness Research*, 6(2), pp. 215-242.

Hardaker, C. and McGlashan, M. 2016. '"Real men don't hate women": Twitter rape threats and group identity', *Journal of Pragmatics* 91, pp. 80-93.

Hong, L., BD Davison. 2010. 'Empirical study of topic modeling in Twitter', in Culotta, A. (ed.). *Proceedings of the First Workshop on Social Media Analytics*, pp. 80–88. Association for Computing Machinery.

Iganski, P. 2008a. *Hate crime' and the city.* Bristol: Bristol University Policy Press.

Iganski, P. 2008b. 'Criminal Law and the Routine Activity of 'Hate Crime''. *Liverpool Law Review 29(1)*. Accessed at https://doi.org/10.1007/s10991-008-9033-x.

Jacobi, C., W. van Atteveldt and K. Welbers. 2016. 'Quantitative analysis of large amounts of journalistic texts using topic modelling', *Digital Journalism*, 4(1), pp. 89-106.

Kemper, S., M. Thompson and J. Marquis. 2001. ‘Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content’, *Psychology and Aging*, 16(4), pp. 600–614.

Kredens, K., & Coulthard, M. 2012. ‘Corpus linguistics in authorship identification’, in Tiersma, Peter M., and Lawrence M. Solan (eds.), *The Oxford handbook of language and law*, pp. 504– 516. Oxford: Oxford University Press.

Krippendorff, K. 2011a. ‘Agreement and Information in the Reliability of Coding’, *Communication Methods and Measures*, 5(2), pp- 1–20.

Krippendorff, K. 2011b. ‘Computing Krippendorff’s Alpha-Reliability’, *Departmental Papers of the Anneberg School of Communication at the University of Pennsylvania*, 55, pp. 1–10.

Kuang, D., PJ. Brantingham and AL Bertozzi. 2017. ‘Crime topic modeling’, *Crime Science*, *6*(12).

Laver, M., K. Benoit, and J. Garry. 2003. ‘Extracting policy positions from political texts using words as data’, *American political science review*, *97*(2), pp. 311-331.

Lee, Y., Yoon, S. and Jung, K., 2018. ‘Comparative studies of detecting abusive language on Twitter’ in Fišer, D., R. Huang, V. Prabhakaran, R. Voigt, Z. Waseem and J. Wernimont (eds) *Proceedings of the 2nd Workshop on Abusive Language Online*, pp. 101–106.

Lido, C. 2006. *Effects of the Media Priming Asylum-Seeker Stereotypes on Thoughts and Behaviour*. London: Economic and Social Research Council.

Mishler, AE. Crabb, S., Paletz, B. Hefright, and E. Golonka. 2015. ‘Using structural topic modeling to detect events and cluster Twitter users in the Ukrainian crisis’, in Stephanidis C. (ed) *HCI International 2015 - Posters’ Extended Abstracts. HCI 2015. Communications in Computer and Information Science*, 528. Springer, Cham.

Mullany, L and Trickett, L. 2018. ‘Misogyny hate crime: new research reveals true scale of issue – and how the public are united against it’, available at: https://theconversation.com/misogyny-hate-crime-new-research-reveals-true-scale-of-issue-and-how-the-public-are-united-against-it-100265 (accessed: 27 January 2021).

Murakami, A., P. Thompson, S. Hunston and D. Vajn. 2017. ‘What is this corpus about? Using topic modelling to explore a specialised corpus’, *Corpora*, *12*(2), pp.243-277.

Nieto, V. G. (forthcoming 2022). *Hate Speech: Linguistic Approaches - Foundations in Language and Law*. Berlin: De Gruyter Mouton.

Nini, A. 2015. *Authorship profiling in a forensic context*. Aston University: PhD thesis.

Nini, A. 2018. ‘An authorship analysis of the Jack the Ripper letters’, *Digital Scholarship in the Humanities*, *33*(3), pp. 621-636.

Nobata, C., J. Tetreault, A. Thomas, Y. Mehdad and Y. Chang. 2016. ‘Abusive language detection in online user content’ in *Proceedings of the 25th international conference on world wide web* (pp. 145-153).

Park, J. H. and P. Fung. 2017. ‘One-step and Two-step Classification for Abusive Language Detection on Twitter’, in Waseem, Z., WH Kyong Chung, D. Hovy and J. Tetreault (eds.), *Proceedings of the First Workshop on Abusive Language Online*, pp. 41–45.

Poole, E. and JE Richardson. 2006. *Muslims and the News Media*. London: I. B. Tauris.

R Core Team (2019). *R: A language and environment for statistical computing*. [Computer program]. https://www.r-project.org. Wien: R Foundation for Statistical Computing.

Rangel, F., P. Rosso, M. Montes-y-Gómez, M. Potthast and B. Stein. 2018. ‘Overview of the 6th author profiling task at pan 2018: multimodal gender identification in Twitter’, *Working Notes Papers of the CLEF*.

Roberts, ME, BM Stewart, D. Tingley, C. Lucas, J. Leder-Luis, SK Gadarian and DG Rand. 2014. ‘Structural Topic Models for Open-Ended Survey Responses’, *American Journal of Political Science*, 58 (4), pp. 1064-1082.

Roberts, ME., BM Stewart, D. Tingley. 2016. ‘Navigating the Local Modes of Big Data: The Case of Topic Models’, in Alvarez, RM. (ed.), *Computational Social Science: Discovery and Prediction (Analytical Methods for Social Research)*, (pp. 51–97). Cambridge: Cambridge University Press.

Skalicky, S., N. Duran and SA Crossley. 2020 (preprint). ‘Please, Please, Just Tell Me: The Linguistic Features of Humorous Deception’. Retrieved from osf.io/qdjmn

Singh, J., and V. Gupta. 2016. ‘Text stemming: Approaches, applications, and challenges’, *ACM Computing Surveys (CSUR)*, 49(3), pp. 1-46.

Sousa-Silva, R., Sarmento, L., Grant, T., Oliveira, E., & Maia, B. 2010. ‘Comparing sentence-level features for authorship analysis in Portuguese’, in I*nternational Conference on Computational Processing of the Portuguese Language* (pp. 51-54). Springer, Berlin, Heidelberg.

de Waal, A., J. Venter, and E. Barnard. 2008. ‘Applying topic modeling to forensic data’, in Indrajit R. and S. Shenoi (eds.), *IFIP International Conference on Digital Forensics*, pp. 115-126. Boston: Springer.

van Dilk, T. 1991. *Racism and the Press*. London: Routledge.

Vidgen, B. A. Harris, D. Nguyen, R. Tromble, S. Hale and H. Margetts. 2019. ‘Challenges and frontiers in abusive content detection’, in Roberts, ST, J. Tetreault, V. Prabhakaran and Z. Waseem (eds) *Proceedings of the Third Workshop on Abusive Language Online*, pp. 80–93. Association for Computational Linguistics.

Wiegand, M., J. Ruppenhofer, A. Schmidt and C. Greenberg. 2018. ‘Inducing a Lexicon of Abusive Words – A Feature-Based Approach’, in Walker, M., H. Ji, A. Stent (eds.) *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics* (1), pp. 1046–1056. Association for Computational Linguistics.

Wright, D. 2017. ‘Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem’, *International Journal of Corpus Linguistics* 22(2), pp. 212-241.