

Sermons as data: Introducing a corpus of 11,955 Danish sermons

Anne Agersnap, Ross Deans Kristensen-McLachlan, Kirstine Helboe Johansen, Uffe Schjødt, Kristoffer Laigaard Nielbo

Anne Agersnap, Ross Deans Kristensen-McLachlan, Kirstine Helboe Johansen, Aarhus University

Uffe Schjødt, Kristoffer Laigaard Nielbo, Aarhus University

Peer-Reviewer: Elizabeth Maddock Dillon

Dataverse DOI: 10.7910/DVN/WIQ6CB

ABSTRACT

In this article, we present a newly established corpus of 11,955 sermon manuscripts written by pastors in the Evangelical-Lutheran Church in Denmark (ELCD) in 2011-2016. We argue that this corpus provides a resource for studying how pastors within the same religious institution attend to general themes in church and society, respond to contemporary events, and represent social worlds. The aim of the article is twofold. 1) To present and discuss our approach to acquire and assemble the sermons corpus. This approach entailed sampling sermons directly from Danish pastors, and cleaning the corpus and annotating it with metadata manually. 2) To demonstrate the research potential of the corpus through a case study on gender representations in the sermons. We find that male and female pastors differ in their use of fundamental linguistic components, namely gendered pronouns and associated verbs. This affects how they assign agency to male and female characters in the corpus, and indicate that male and female pastors shape the social worlds in sermons in quite different ways. This case study therefore illustrates just one of the ways in which corpus-based research of Danish sermons may provide novel insights in the field of religion and society.

The role of religion in contemporary democratic societies is a complex topic. A tendency within the study of religion has been to assert that religion today emerges primarily in individualized spiritual forms that have displaced the relevance of the old religious traditions in society – in the European context, of the old Christian majority churches in particular.¹ In Denmark, the Evangelical Lutheran Church in Denmark (ELCD) still represents the majority religion, inscribed in the constitution and subsidized as such by the state through church taxes. The ELCD is present in the public sphere in Denmark, but as an institution, the church has no distinct agency that speaks in public on behalf of the ELCD: it has no unified voice to respond to societal change or crises, to address ethical matters, or to prioritize particular societal or religious topics over others. It is individual pastors who speak for the ELCD, and

these voices are at their most distinct when they write their sermons for the weekly service – a synchronous event in the ELCD, with pastors interpreting the same biblical passages within a context they deem pertinent to assembled congregations all over the country. In this endeavor, pastors are not mere agents of an organization, sending a consistent and uniform signal to their congregations. The content of preaching of course depends upon tradition and doctrine but preaching is also shaped by contemporary contingencies (local and global news, natural disasters, elections, and cultural discourses) and by the profile of the individual preachers and recipients (gender, age, personality, and sociodemographics). This practice results in the production of approximately 1,500 written sermons a week, a collective production of texts which remain the property of the individual pastors after the service. From these unique and otherwise detached documents, we can uncover the public voice of a religious majority institution.

In this article, we present a newly established corpus of 11,955 ELCD sermons. The corpus will provide a resource for studying how pastors within the same religious institution attend to general themes in church and society, respond to contemporary events, and represent social worlds. After an introduction to the religious and societal context of which this corpus is a product, we will briefly review digital archival resources for studying religion and sermons more generally. Against this background, we will discuss our own careful considerations in regard to sampling, cleaning, annotating, and archiving our corpus, a process that was predominantly conducted manually. We further provide basic corpus statistics in order to describe our data set. Finally, we present a case study demonstrating how pastors perform the ELCD's public voice when they represent gender in this newly established corpus of sermons.

The public voice of a religious community: the Danish context

In the second half of the twentieth century, secularization theories became important perspectives for predicting and interpreting the role of religion in democratic societies. The secularization thesis – in its early form associated with sociologists such as Max Weber, Émile Durkheim, and Karl Marx – stated that over time, as rational modes of thinking in societies increased, religion would lose its ability to

set societal agendas. By the end of the century, secularization discussions were influenced by new concepts, such as “deprivatization,” “desecularity,” and “postsecularity,” and it was argued that societies were becoming increasingly religious once again.² In recent years, past secularization perspectives have been criticized for viewing religion and the secular as two clearly distinguishable and separate domains.³ Rather, religion in public spheres is a complex phenomenon that unfolds in relation to other aspects of society, and not just singularly. It is therefore pertinent to attend to the concrete cultural and societal context in which religious phenomena unfold.

In the Danish context, the Evangelical Lutheran Church in Denmark is the majority religion. Most Danish citizens are members of the ELCD (74.7% of the Danish population at January 1 2019). However, membership rates are slowly but steadily declining (from 81.5% at January 1 2009). Church members tend to participate primarily in church activities that are connected to life events – in particular funerals, but also baptisms, confirmations, and weddings – while participation in Sunday services is low, with an estimated 11% of members attending church services regularly in 2017.⁴ From a secularization perspective, these numbers could indicate that the church is losing its ability to remain relevant in society. Nevertheless, in view of the societal structure of the ELCD, the church is intricately involved in the public sphere of Danish society.

On a macro level, the church is inscribed in the constitution and as such subsidized by government through church taxes, with church legislation handled by parliament. The ELCD is composed of ten dioceses and bishops (four female, six male). Dioceses are divided into deaneries, and deaneries further divided into geographically determined parishes, each with one or more pastors employed. Approximately two thousand pastors are employed by the ELCD. At parish level, small congregational councils are formed in civil society to share in the running of the local churches, a structure that was devised to rule out the possibility that a congregation might be led by a pastor whose theology did not accommodate that of the local population.⁵ The confession of the church is Evangelical–Lutheran, its framework liberal theological. In 1948, the ELCD was the first Lutheran church in the world to ordain women as pastors; in 2014, 55% of pastors employed by the ELCD were women. Same-sex marriages within the ELCD were authorized in 2012.

This authorization is a good example of how the branching structure of the church works in practice: the bill was passed in parliament; bishops were required to develop a ritual for performing the wedding ceremony; but in the end, individual pastors could decide whether their own theological attitude was compatible with their actually performing the ritual.⁶ However, accounting for the societal structures of the ELCD gives little insight into the responses and attitudes of the church to contemporary time and society. The structure in which the ELCD is embedded entails that many agencies are involved in regulating church affairs and that the church has no single, unified public voice with which to respond to societal change or standardize theological, ethical, political, or other attitudes. Instead, this voice is governed collectively by the pastors employed in the ELCD.

Every Sunday, individual pastors form a collective voice as they preach simultaneously to congregations all over the country. Preaching in the ELCD is a practice embedded within a church liturgy that includes performing the Creed, prayers, hymn singing, the Lord's Supper, and readings from the Bible. In the ELCD, the biblical readings are ordered in a prescribed lectionary which determines readings for each church holiday. There are two lectionaries: one for even years, one for odd. The biblical readings include a text from the Old Testament; a text from the epistles in the New Testament; and a text from one of the gospels in the New Testament. Texts for the same holiday share thematic similarities. The pastor reads the gospel passage aloud immediately before preaching and is obliged to preach in consideration of this gospel passage. Compared with more strictly regulated parts of the liturgy, though, preaching is a practice in which pastors address their congregations from a contemporary point of view. Pastors do not merely paraphrase the biblical passages but may instead use them to raise or reflect on existential, societal, or philosophical issues, and they are inclined to implicitly produce, reproduce, or divert from different discourses. As they preach, pastors are thus engaging in a dialogue with both Christian tradition and contemporary time. Even though attendance rates at regular ELCD services are low, this collective voice uttered by pastors is documented in the thousands of written sermon documents that pastors prepare in the week before the service. We argue that these documents are immensely valuable for uncovering the attitudes and discourses that exist within a majority Christian institution.

In the following, we briefly review already existing digital archival resources for the study of religion, in order to substantiate our own approach to acquiring data.

Digital archival resources in the study of religion

Gathering and storing data on religion in databases for research purposes has been in progress for some time. The “Index Thomasticum,” the project to index the works of Thomas Aquinas, led by the priest Roberto Busa, is widely seen as the first large-scale project within computationally informed humanities research. Newer initiatives have been launched to develop impressive database resources for comparative studies of religion – such as the “Database of Religious History” at University of British Columbia⁷, and the “World Religion Database”⁸ and its companion the “World Christian Database”⁹ at Boston University. Whereas these databases provide thorough statistical information on various different religious traditions, other attempts to develop more topic-specific databases have also emerged, for example collections of Christian hymns in an American context (<https://hymnary.org/>)¹⁰ or a collection of images of supernatural agents drawn by children in different parts of the world.¹¹

In comparison with these digital resources for the study of specific aspects of religion, relatively little effort has been devoted to developing digital archives of sermons for research purposes. Recently, though, a Pew project¹² has begun to document the online platforms where churches in the United States share sermons publicly after church services, by creating a database of 49,719 sermons given between April and June 2019. The sermons in this collection have been posted to websites either as texts or in audio or video format, with the material in the latter formats transcribed for the Pew collection. The project defines a sermon in the following terms: “an ‘online sermon’ refers to a portion of a religious service posted to a church website that contains a commentary from the pulpit but sometimes may include other parts of the service as well.”¹³ This definition allows scope for the texts to include excerpts from additional sequences of the church services. In their analyses, the research group have made comparative studies of four Christian traditions (mainline Protestant, Catholic, Evangelical and historically black

Protestant denominations in the US), focusing on variations in word frequencies, in mentions of books of the Bible, and in sermon lengths.

Rather than comparative samples of religious traditions, in our project we were interested only in one case, namely preaching in the ELCD. We found it imperative to invest considerable effort to delimiting the manuscripts in our corpus so as to include only sermons, excluding any other content uttered during the church service, as we wanted to be able to attend not just to word frequencies in the sermons but also discursive content and semantic contexts. We therefore chose to sample sermons as personal documents and we found that a manual approach to assembling the corpus was the best solution to ensure adequate data quality.

Methods: acquiring sermon data and building a text corpus

To build our corpus, we collected contemporary sermon documents directly from ELCD pastors. This procedure resulted in a collection of 11,955 sermons, written by 95 different pastors primarily between 2011 and 2016. There are 2169 ELCD parishes in Denmark. If we conservatively assume that each of these parishes delivers one sermon per week for each of the six years – which may not actually be the case – this means that our corpus comprises roughly 2% of all possible sermons delivered in Denmark during this period. The sampled material forms a case containing comprehensive knowledge about contemporary preaching within a national context. The corpus collection is based on opportunity sampling and therefore is not strictly speaking a representative sample of the population of Danish pastors. Instead, it represents preaching as collective but varied practices within the ELCD.

Sampling, cleaning, and annotating data

Sermons challenge the traditional distinctions of social science research, as it is not straightforward whether they should be considered as personal or official documents. As orally delivered speeches, they are conveyed to the public on behalf of an official institution. Nevertheless, the written manuscripts themselves remain the property of individual pastors, meaning that sermon manuscripts are therefore ultimately personal documents. In some parishes, where the local church publishes

the sermon online after church services, the pastors can be regarded as having authorized their sermons as official documents. However, in contrast to American practice, publishing sermons online after services is not customary in Denmark – still less, recordings of church services. If pastors disseminate sermon manuscripts online, they might be inclined to edit their original drafts, whether because of new ideas or considerations since the service or because they are in principle addressing a much broader audience online than in the local church. We chose to sample sermons solely as personal documents in order to secure three objectives: to secure access to the original, rather than the post-edited sermons; to ensure that a broad group of pastors could provide documents; and to ensure the consent of the pastor to assign us ownership of the material. A further criterion was that the sermons sampled should be from an uninterrupted time span that was as recent as possible. As we began to collect data in 2017, we delimited the sermon period as 2011–2016. This meant that the collection would represent six recent chronological years, as well as representing both lectionaries three times.

We were not able to obtain contact information for all Danish pastors and we therefore employed snowball sampling to gather data. Bishops and rural deans acted as gatekeepers for getting in contact with pastors. Nine out of ten bishops supported our project by signing a short approval statement. We forwarded this statement to the rural deans under the nine bishops, along with a request for them to send out an invitation to the pastors under their supervision, inviting them to submit over email their sermons from 2011 to 2016 in digital versions for a research project¹⁴. We chose to let our sampling strategy be open-ended, insofar as we extended the invitation to all ELCD pastors, while basing our response rate on self-selected participants. We thus allowed the corpus to take on its own shape. In the first part of the process, we did not approach the pastors ourselves, which reduced researcher biases in sampling subjects, as the rural deans mediated our invitation. After a couple of months, we evaluated our response rate. We then ourselves approached individual pastors in dioceses that were under-represented in our sample, which enabled us to control somewhat for possible ‘gatekeeper’ biases. In this second wave of sampling, we also targeted female pastors in particular, as male pastors had been more inclined to respond in the first wave. In the end, though, our corpus is still sampled opportunistically. Even with a more controlled sampling method, there are a number of confounding factors: some pastors preach more frequently than others; some

pastors may not have stored or written sermons throughout the entire period; and we were depended on the motivation of pastors to invest their own time in finding, compiling and sending hundreds of sermons to us. We therefore found that by allowing the sample to take its own shape, we could best achieve adequate volume of data and most variation within the sample – even though it would not be an entirely representative sample of sermons in the ELCD.

The confounding factors mentioned above can be seen in the material we received, with pastors sending in differing amounts of data. Further, some pastors sent in sermons outside the requested time period; others sent in non-sermon material (wedding speeches, talks for various occasions) stored along with their sermons in a heterogeneous file directory. The sermons also differed in length; in the amount of self-composed metadata the pastors had included; and in their structuring of content. Overall, then, the raw data was messy and unstructured. We factored these aspects into archiving and annotating the sermons. Since we could not assume structural consistency in the received material, the best method to accommodate all variances was to archive and annotate manually. Five student assistants helped with the cleaning and archiving process. The cleaning process consisted of a series of steps: inspecting each document, to determine if it was indeed a sermon manuscript; discarding superfluous text from the document such as hymns, biblical readings, prayers; and finally filing each document under a unique document ID. We then assigned metadata to each document using stand-off annotations for each cleaned sermon document.

Metadata and data access

When the pastors sent us their original sermon documents, they provided only content and no personal background information regarding their documents. However, they unanimously gave consent to gather metadata and store both sermons and metadata in a research database. The pastors gave this consent under the terms that we would not disseminate material that could identify individual pastors publicly. For this reason, we are required to grant access to the data only to approved research purposes and we had to construct metadata without disclosing direct links between sermons and individually identifiable pastors.

Metadata consists of pastor pseudonym, sermon date, holiday, size of parish, name of diocese, birth year of pastor, gender of pastor, and where the pastor was educated. We applied a simple pseudonymization technique to mask every pastor by numbers 1 to 95. As timestamps, we supplied the corpus with two annotations: first, the date of the sermon, represented by year, month, day; and second, the holiday and church year of the sermon, followed by either the letter “U” (odd years) or the letter “L” (even years) to determine the relevant lectionary.¹⁵ We used two online resources to provide the remaining metadata for the sermons corpus. For personal background information, we used an online yearbook, *Teologisk Stat*.¹⁶ This resource is updated by and available to Danish ELCD pastors and also to researchers at Aarhus University through the Royal Danish Library. Through *Teologisk Stat*, we obtained information about the birth year of our informants; their place of education; and their employment history in the ELCD. This last piece of information was essential for pastors who had changed job during the period 2011–2016. In these cases, personal information on age, gender, and place of education remains the same for all sermons given by the same pastor, while information relating to location varies dependent on the dates of the sermons. We extracted information relating to location for all parishes in which our informants had been employed throughout the 2011–2016 period from www.sogn.dk – a publicly accessible web page that contains demographic information and statistics about ELCD parishes.

| ID-DOC | ID-PASTOR | DATE | HOLIDAY | PARISH SIZE | DIOCESE | BIRTH YEAR | GENDER | PLACE OF EDUCATION |
|-----------|-----------|--------|----------|-------------|-----------|------------|--------|--------------------|
| pr108_11 | 11 | 161030 | 23Trin_L | 5731/10011 | København | 1977 | 2 | Aarhus |
| pr2096_35 | 35 | 120409 | 2PåsD_L | 3951/4639 | Haderslev | 1961 | 1 | Aarhus |
| pr4385_41 | 41 | 130106 | Hel_U | 11043/15093 | Fyn | 1962 | 2 | København |
| pr5734_88 | 88 | 160925 | 18Trin_L | 6535/8330 | Aarhus | 1975 | 1 | København |

Table 1. Four random rows from the sermons metadata. From the left the columns contain: document ID, pastor ID, date, holiday, parish size (number of ELCD-member/total number of inhabitants), diocese, birth year, gender (1 is male, 2 is female), place of education.

The label “diocese” represents the geographical area of the sermons on a macro scale, while “size of parish” is a micro-scale indicator of demographic variation between congregations in the parishes. We chose these labels as locative indicators in order to meet the condition that links between sermons and metadata capable of disclosing the identities of individual pastors should not be provided. Had we applied

finer-grained information about geographical areas through labels such as the name of the deanery or parish, individual pastors could have been identifiable by combining locative variables with personal ones. Table 1 above shows an excerpt of the sermons metadata¹⁷.

As pastors by default disseminate religious attitudes while preaching, the individual sermon manuscripts in themselves contain personal and sensitive information and are by no means anonymized by our efforts. Further, pastors are quite likely to mention the name of the parish or names of congregants – for instance at baptisms and confirmation services – and their personal experiences as they prepare their sermons for a local and delimited audience. These factors would make it possible to identify concrete pastors by scrutinizing individual sermons. We were therefore obliged to provide a secure solution for storage of the sermons corpus, including metadata, and access to the data has to remain restricted to protect our informants. Nevertheless, we are able to provide access to the dataset for valid purposes of research and teaching.

Cleaning, structuring, and archiving the sermons corpus was naturally time- and resource-consuming, insofar as manual assessment of the content and metadata required considerable domain expertise. However, the process ensured thorough insight into the quality of the material and enabled us to supplement the corpus with finer-grained information. For example, we had defined most of the annotation tags beforehand. However, during the process we had to construct further tags, such as a tag to identify sermons not intended for the ELCD’s prescribed holiday services, e.g. services in care homes or hospitals, or local events such as “hunting services” or “Halloween services.”¹⁸ The cleaning process thus enabled us to construct a structurally uniform text corpus, while remaining responsive to the received data in the process.

Corpus statistics

Following the discussion above of the metadata associated with the corpus, this section provides a brief breakdown of what we perceive to be some of the most salient features of the corpus. In particular, we wish to draw attention to some of the

imbalances inherent in the corpus in its current format. This metadata can be broken down into metadata related to the individual pastors and textual metadata generated using corpus linguistic methods.

At the general level, the most pertinent corpus feature is the distribution of sermons by year. A snapshot overview of this data can be found in Figure 1 below. There are a number of things to be noted here. Firstly, this figure shows only those sermons which come from the period 2011–2016. While this includes the vast majority of the sermons in the corpus (96%), a small number of sermons fall outside of this time period. Moreover, there is a very small number of sermons for which we have no available year (roughly 0.2% of the full corpus).

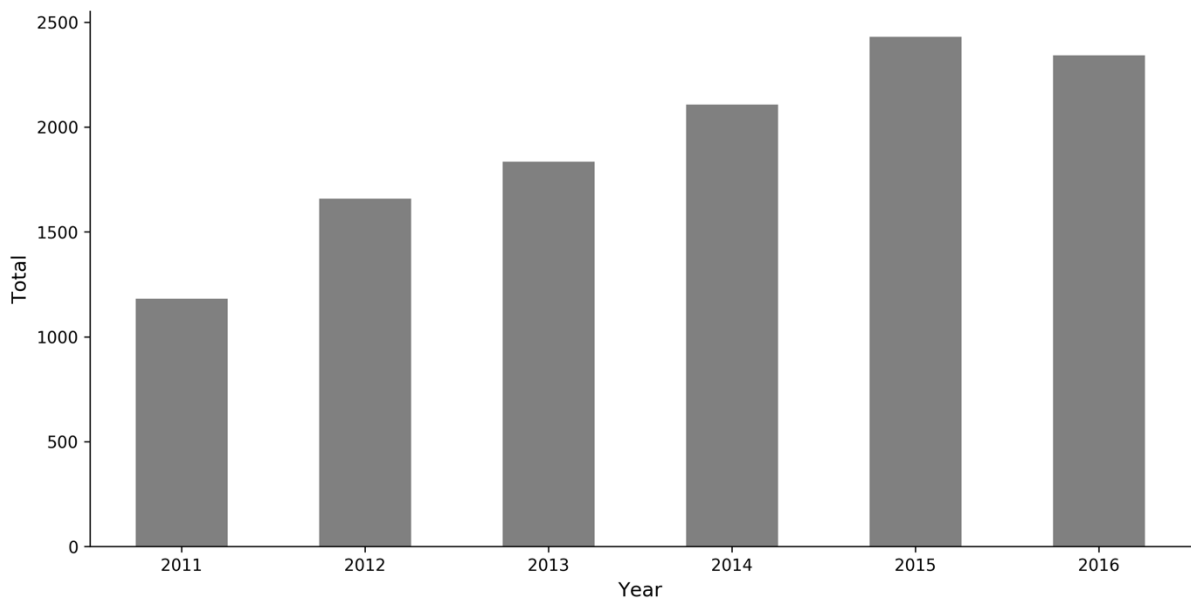


Figure 1. Distribution of sermons by year

Within the distribution, we can see that there is a small imbalance in the data toward later years, with 2015 being the year most represented. For the purposes of the case study presented below, we have chosen to work with the full corpus of texts, including those which lie outside the main 2011–2016 window and those for which we do not have year data. This was in order to make the most of all available data for the purposes of illustrating the analytical value of the corpus. For future studies,

however, this discrepancy can be taken into account by focusing solely on sermons from 2011–2016.

With regard to metadata related to individual pastors, one of the most relevant features for further study is the distribution of sermons in the corpus relative to the registered gender of the pastor. There is a noticeable imbalance in the corpus, with the total number of sermons by female pastors being around 75% of the total number of sermons by men (M=6,823; F=5,132). Grouping pastors by decade of birth creates a more nuanced picture of the individuals behind the corpus. In Figure 2 below, we see that men outnumber women across all birth decades with the exception of the 1950s. That is to say, there are more sermons by women born in the 1950s than men born in the 1950s. A second point that is immediately clear is that the 1960s is by far the most prominent birth decade for both genders. This is especially true for men, however, with nearly as many sermons by men born in the 1960s (3,202) as men from all other decades combined (3,621).

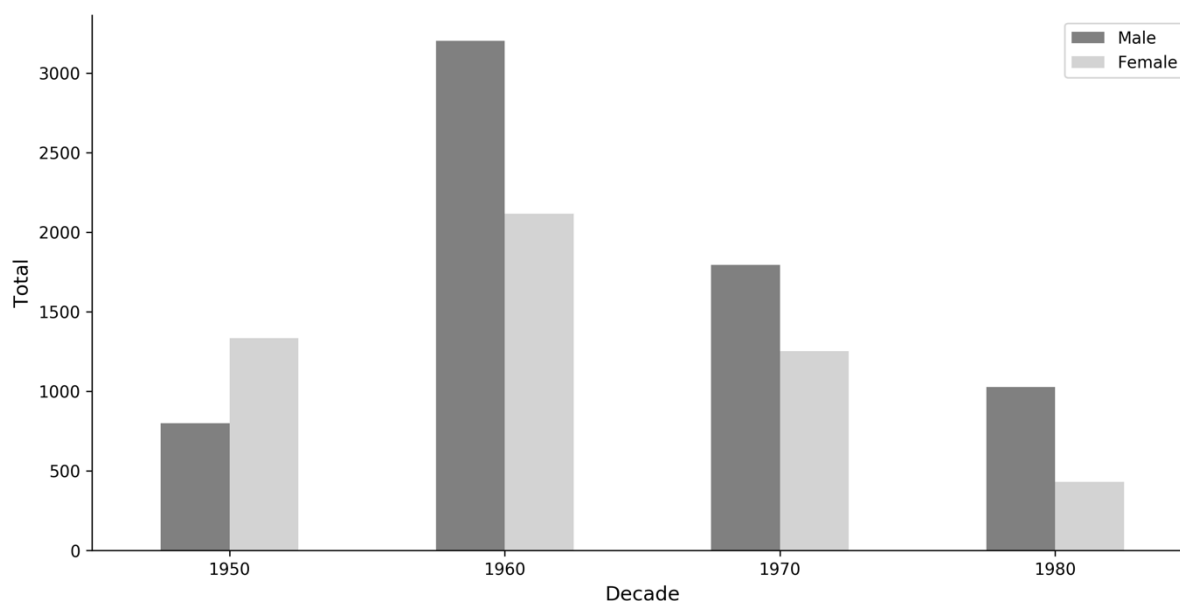


Figure 2. Distribution of sermons by pastors' birth decade and gender

Alongside age and gender, the metadata also contains geographical information relating to the diocese where the pastor and their church are located. A summary of this geographical information can be seen in Figure 3 below. This summary reveals that there are comparatively few sermons from those dioceses containing three of

the four largest cities in Denmark (Copenhagen, Aarhus, Aalborg). In this respect, larger urban areas are not overrepresented in our corpus, relative to rural parishes. It should be noted, though, that the label “diocese” is not necessarily an adequate measure for distinguishing between urban and rural parishes. The diocese of Copenhagen is the only diocese whose boundaries mostly correspond to the city boundaries. In the dioceses of Aarhus and Aalborg, large surrounding areas outside of the city boundaries also form part of the dioceses. Hence, as the label follows the ELCD’s own geographical subdivision of the church, this metadata is important in the study of whether content of sermons diverges according to pastors’ regional affiliations.

As previously mentioned, the corpus content was collected on the basis of opportunity sampling. The imbalances presented here were therefore expected and do not diminish the relevance of the collection. For future research projects, subsampling and balancing will still be possible owing to our well-structured metadata.

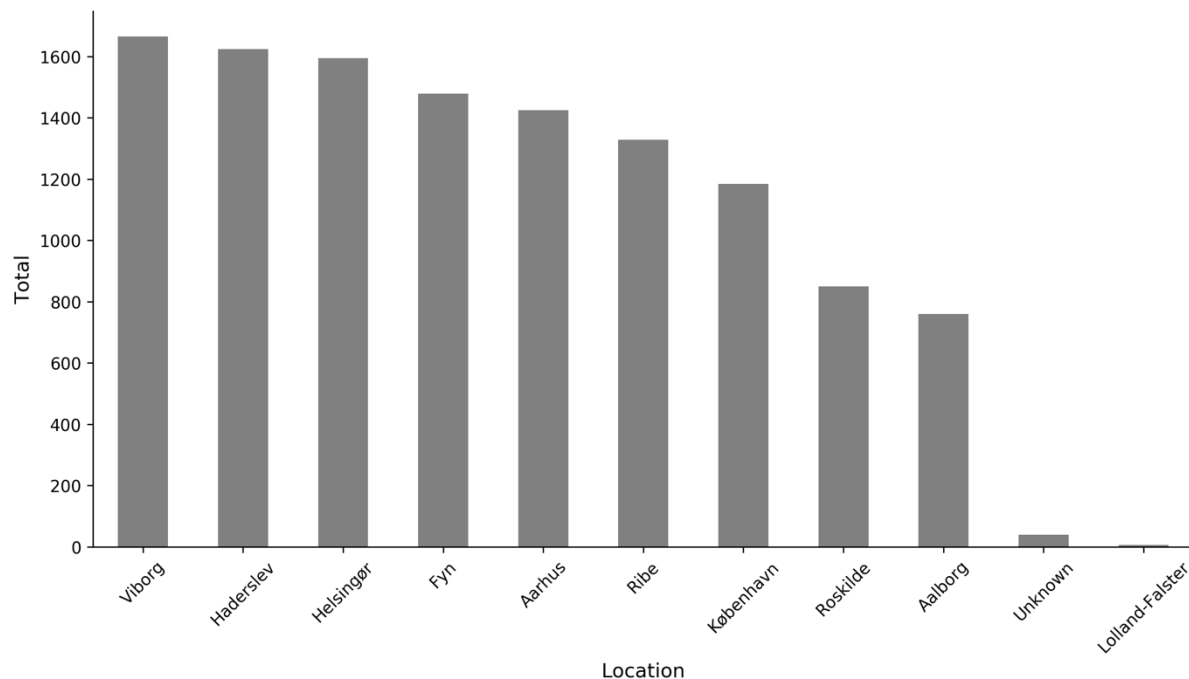


Figure 3. Distribution of sermons by diocese location

| | SENTENCES | WORDS | TTR | MTLD |
|-------|-----------|-------|------|-------|
| ALL | 74 | 1453 | 0.36 | 72.24 |
| WOMEN | 69 | 1406 | 0.37 | 73.50 |
| MEN | 78 | 1489 | 0.36 | 71.28 |

Table 2. Linguistic measures in sermons for all pastors, female pastors and male pastors

The textual metadata about the sermons are general linguistic measures extracted from the corpus using standard Natural Language Processing (NLP) measures. We illustrate a number of these in Table 2 above.

The table presents four linguistic measures against the gender of the pastors, with a third group “All” representing the full corpus. The measures themselves are fairly uncomplicated, and include the average number of sentences per sermon, the average number of words, the average Type-Token Ratio (TTR), and the average Measure of Textual Lexical Diversity (MTLD).¹⁹ We find, generally speaking, surprisingly little variation in these numbers relative to the gender of the pastor. The largest difference would be that between numbers of sentences in sermons written by women. However, this difference is somewhat offset by the fact that the average number of words differs by only around 5% from the average number of words written by men. Similarly, both TTR and MLTD seem to suggest that there is only a small degree of variation between the genders in terms of lexical diversity.

Of course, the above methods of extracting information are somewhat crude NLP measures of underlying textual phenomena. In the remaining sections of this article, we present a case study of the kinds of research that can be conducted on this corpus of sermons if more nuanced NLP methods are used alongside close-reading approaches.

Case study: gender representations in sermons

Uncovering gender in text sources

In comparison with other text genres, sermons represent social worlds. Sermons provide access to a distinct cultural context – preaching in the ELCD – in which

pastors represent social structures and engage in public discourses: an engagement that unfolds both consciously and unconsciously. Susan Brown and Laura Mandell argue similarly in an article on identity issues that literature is a window onto social structures, social positions, and social change in culture, and that the representation and construction of identities are underlying processes of literary writing. They demonstrate, further, that computational analyses provide unique tools for discovering the signifiers that mediate identities and their discursive constructs in large text collections.²⁰ These approaches have already shown immense potential for the study of gender identities in regard to the social worlds of literary characters portrayed by authors in the genre of literary fiction.²¹ We find that uncovering similar structures in text productions by religious communities such as the ELCD can provide valuable insights for the study of contemporary religion.

The ELCD is a liberal theological church and known in Denmark as “the inclusive people’s church”²². However, there are no prominent threads in the ELCD of gender oriented theologies, such as queer or feminist theologies, which is in contrast to other similar church traditions, such as is found in Norway and Sweden. In fact, among pastors in the ELCD, convictions can be found that an increasing feminization could dilute the church²³, or that gender is not interesting in a theological context²⁴. A qualitative study of twenty sermons investigated whether female and male pastors differ in how they represent God and found that there were no significant difference²⁵. These perspectives indicate a tendency for pastors to frame the ELCD as a rather gender-neutral space. This framing could resemble a more general attitude in the Danish population. A survey from 2019 disclosed that 43% of Danes agreed and 36% disagreed with the statement that the struggle for equality between genders has gone too far²⁶, suggesting that gender differences do not appear very topical in the public discourse. Meanwhile, the Global Gender Gap report, which ranks countries in terms of gender equality, shows that Denmark has moved from a fifth place in 2014 down to a fourteenth place in 2020²⁷.

From a theological perspective, Else Marie Wiberg Pedersen has argued that the reason why we do not see distinct feminist theological wings in the ELCD is not in contrast to but *because of* the liberal theological framework of the church. She claims that feminist theology is an inherent but unarticulated core component of the church tradition, which has enabled, for example, an even distribution of male and female

pastors²⁸. In contrast, theologian Lone Fatum has argued that a consequence of claiming gender equality and balance in the ELCD is overlooking gender differences. She believes that such strategies cause gender blindness²⁹.

In this section, we investigate whether a genre such as the Danish sermon is a gender-neutral document. As there is no profound tradition for engaging with gender theologies in the ELCD, we design our study as an investigation of how pastors implicitly represent and discursively construct gender identities through their use of third person pronouns. Gender theorist Judith Butler accentuated in her performativity theory that pronouns can be considered to signify content insofar as the structure of signification (the discursive context) in which they are embedded is taken into account.³⁰ If a consistent structure of signification appears around a gendered pronoun, then a gendered identity can be inferred. In the case study below, we study the signification structure surrounding male and female pronouns in the sermons, while accounting for variations between male and female pastors. By focusing on male and female pronouns in nominative compared with oblique position and their associated verbs, we can learn general agency patterns for male and female agents in the corpus. We acknowledge fully that this approach does not allow us to venture into non-binary constructions of gender. However, as there is no established Danish language practice for challenging these limits of language, this approach is the best way of uncovering both conscious and unconscious constructions of gender, inasmuch as we are able to work simply with the pastors' use of everyday categories in language.

Methods

For this study, we combine distant and close-reading strategies. First, we extracted raw text from each of the documents in the corpus. We then tagged the corpus for parts of speech, and extracted collocations of gendered pronouns in either nominative (Danish *han* and *hun*; English *he* and *she*) or oblique case (Danish *ham* and *hende*; English *him* and *her*) as well as the verbs collocating with these pronouns in their immediate sermon context. We did not include reflexive pronouns in this study because, in contrast to English usage (*himself*, *herself*), the Danish reflexive (*sig*) does not explicitly mark gender. Similarly, we chose not to lemmatize the extracted verbs. This allowed us to distinguish between important grammatical

differences, such as tense and aspect, along with the use of active and passive voice.³¹

This extraction process was repeated on two sub-corpora, one comprising sermons by female pastors and another by male pastors. From these extracted collocations, we calculated pointwise mutual information scores (PMI scores) for every pronoun–verb collocation across each of the three sets of results. In order to filter the results further, we kept only those verbs that appear more than ten times in sermons by male and female pastors respectively. We then sorted the list by strength of PMI score, which identifies the verbs most strongly associated with male and female nominative singular pronouns (*han* and *hun* respectively). We refer to these results as our *single cases*, insofar as they suggest the most important kinds of verbs assigned to men and women individually. In addition, we identified verbs that were highly associated with both male pronouns in nominative case and female pronouns in oblique case – and vice versa – in order to represent the kind of actions that were most likely performed by a subject toward an object of opposite gender. We call these types of representations our *relational cases*.

Detecting gender representation through pronouns and verbs

We find that male and female pastors represent pronouns very similarly in terms of raw frequency, but with a slight tendency for female pastors to represent female agents more than male pastors do (see Table 3 below). This representation mirrors the appearance of male and female characters in the two lectionaries that each contains texts from both Old and New Testament. A recent numeration has shown that in the first lectionary, 90.04% of the characters mentioned by name are male characters, and the remaining 9.96% female. In the second lectionary, the difference is smaller but still noticeable, with 20.2% of all characters mentioned by name being female, compared to 79.8% male³². Though numbers in the sermons refer to pronouns and numbers from the lectionaries refer to named entities, this finding suggests that the character gallery of the biblical passages has a clear influence on the characters represented in their sermons by male and female pastors alike. Whether or not this means that pastors predominantly represent narratives with a male protagonist (e.g. *Jesus*) and replicate the ancient social worlds of the biblical sources, these numbers accentuate that the representation of gender in contemporary

sermons is skewed. From this perspective, we see no break with biblical biases and a direction toward a more equal representation of gender.

| Pronouns | Male pastors | | Female pastors | |
|--------------|----------------|--------------------|----------------|--------------------|
| | Raw frequency | Relative frequency | Raw frequency | Relative frequency |
| He | 128,400 | 71.4% | 90,014 | 70.2% |
| She | 13,200 | 7.3% | 12,390 | 9.7% |
| Him | 31,802 | 17.7% | 20,223 | 15.8% |
| Her | 6,348 | 3.5% | 5,625 | 4.4% |
| Total | 179,750 | 100 % | 128,252 | 100 % |

Table 3. Distribution of male and female pronouns in sermons by male and female pastors

However, by focusing on the discursive constructions of male and female characters, we can obtain additional perspectives for understanding gender representations in the sermons. We therefore performed an inductive coding of the two types of data (single cases and relational cases) in the two datasets (male and female sermon corpora). Here we find four meaningful analytical categories: *active-competent agency* (verbs that express acts of vigor); *cogitative-perceptive agency* (verbs that express acts of sensing and interpreting); *communicative-expressive agency* (verbs that express extroverted or contact-oriented acts); and *religious agency* (verbs that express acts connected to a religious context in particular). The full distribution of analytical categories is shown in Tables 4a, 4b, 5a and 5b below³³.

| Category | Single case: Male subject | | | Single case: Female subject | | |
|---------------------------------|---------------------------|-------|-----|-----------------------------|-------|-----|
| | Verb | Freq. | PMI | Verb | Freq. | PMI |
| Active-Competent agency | Undertook | 54 | 3.5 | Gave birth | 110 | 5.3 |
| | Accomplished | 73 | 3.2 | % | % | % |
| | Expounds | 41 | 3.0 | % | % | % |
| | Commanded | 50 | 2.9 | % | % | % |
| Cogitative-Perceptive agency | % | % | % | Sense | 313 | 3.3 |
| | % | % | % | Looked | 149 | 3.2 |
| | % | % | % | Recognize | 272 | 3.2 |
| | % | % | % | Understood | 583 | 3.5 |
| Communicative-Expressive agency | % | % | % | Address | 149 | 3.6 |
| | % | % | % | Smiles | 86 | 3.9 |
| | % | % | % | Laughs | 66 | 3.8 |
| | % | % | % | Interrupts | 40 | 4.0 |

| | | | | | | |
|------------------|-------------|-----|-----|----------|----|-----|
| Religious agency | Resurrected | 60 | 3.1 | Anointed | 80 | 4.4 |
| | Heals | 326 | 2.7 | % | % | % |
| | % | % | % | % | % | % |
| | % | % | % | % | % | % |

Table 4a. Distribution of analytical categories in sermons by men. The table presents verbs representative of one of the four categories in single cases.

| Category | Relational case: Male subject and Female object | | | | Relational case: Female subject and Male object | | | |
|---------------------------------|---|-------|------------------|-------------------|---|-------|--------------------|-----------------|
| | Verb | Freq. | PMI Male Subject | PMI Female Object | Verb | Freq. | PMI Female Subject | PMI Male Object |
| Active–Competent agency | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| Cogitative–Perceptive agency | % | % | % | % | Recognized | 78 | 3.4 | 3.5 |
| | % | % | % | % | Appreciates | 43 | 3.6 | 2.6 |
| | % | % | % | % | Ascribes (to) | 40 | 2.9 | 2.7 |
| | % | % | % | % | % | % | % | % |
| Communicative–Expressive agency | % | % | % | % | Laughs | 66 | 3.8 | 3.1 |
| | % | % | % | % | Offer | 95 | 2.1 | 2.9 |
| | % | % | % | % | Praise | 105 | 2.0 | 4.4 |
| | % | % | % | % | Approach | 162 | 1.5 | 3.3 |
| Religious agency | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |

Table 4b. Distribution of analytical categories in sermons by men. The table presents verbs representative of one of the four categories in relational cases.

| Category | Single case: Male subject | | | Single case: Female subject | | |
|------------------------------|---------------------------|-------|-----|-----------------------------|-------|-----|
| | Verb | Freq. | PMI | Verb | Freq. | PMI |
| Active-Competent agency | Undertook | 21 | 3.2 | Gave birth | 72 | 5.2 |
| | Promise | 388 | 2.9 | Involves | 20 | 4.0 |
| | Overcame | 57 | 2.9 | Managed | 37 | 3.4 |
| | Direct | 82 | 2.8 | Achieve | 39 | 3.3 |
| Cogitative-Perceptive agency | Sensed | 29 | 3.0 | Misunderstand | 28 | 4.1 |
| | Regrets | 63 | 2.8 | Dreamt | 82 | 3.7 |

| | | | | | | |
|--|------------|-----|-----|------------|-----|-----|
| | Decides | 87 | 2.9 | Imagined | 42 | 3.6 |
| | % | % | % | Recognizes | 201 | 3.3 |
| Communicative- Expressive agency | Discussed | 32 | 3.1 | Greeted | 22 | 4.3 |
| | Propagate | 32 | 2.9 | Kisses | 72 | 4.3 |
| | Declare | 32 | 3.1 | Cries | 366 | 3.4 |
| | Asserted | 27 | 3.2 | Expressed | 83 | 3.3 |
| Religious agency | Sermonized | 129 | 2.8 | Anointed | 110 | 3.2 |
| | Heals | 213 | 2.8 | % | % | % |
| | Preached | 101 | 2.8 | % | % | % |
| | Revealed | 101 | 2.8 | % | % | % |

Table 5a. Distribution of analytical categories in sermons by women. The table presents verbs representative of one of the four categories in single cases.

| Category | Relational case: Male subject and Female object | | | | Relational case: Female subject and Male object | | | |
|--|---|-------|------------------------|-------------------------|---|-------|--------------------------|-----------------------|
| | Verb | Freq. | PMI Male Subject | PMI Female Object | Verb | Freq. | PMI Female Subject | PMI Male Object |
| Active- Competent agency | Protects | 84 | 1.6 | 4.5 | Gave birth | 72 | 5.2 | 2.6 |
| | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| | % | % | % | % | % | % | % | % |
| Cognitive- Perceptive agency | Treasures | 23 | 2.2 | 5.1 | Recognizes | 201 | 3.3 | 4.0 |
| | Observes | 26 | 1.3 | 4.9 | Misunderstands | 28 | 4.1 | 3.5 |
| | Overlooked | 25 | 2.4 | 4.9 | Notifies | 29 | 2.9 | 3.0 |
| | Accepted | 32 | 2.5 | 5.6 | (Is/was) surprised | 73 | 2.7 | 2.1 |
| Communicative- Expressive agency | Declare | 32 | 3.1 | 3.1 | Praised | 27 | 3.7 | 5.0 |
| | Shoved | 21 | 2.2 | 2.2 | Kisses | 72 | 3.8 | 4.4 |
| | Enhances | 108 | 2.7 | 2.7 | Greet | 27 | 3.5 | 2.3 |
| | Reveals | 47 | 1.8 | 1.8 | Laughed | 72 | 2.7 | 2.2 |
| Religious agency | Preached | 101 | 2.8 | 2.8 | Anointed | 110 | 3.2 | 2.6 |
| | Revealed | 67 | 2.8 | 2.8 | Is redeemed | 104 | 2.3 | 3.8 |
| | Washed | 69 | 2.7 | 2.7 | Worships | 54 | 3.0 | 4.1 |
| | % | % | % | % | % | % | % | % |

Table 5b. Distribution of analytical categories in sermons by women. The table presents verbs representative of one of the four categories in relational cases.

Male and female agency in sermons

Tables 4a, 4b, 5a and 5b clearly illustrate how the prevalence of the four categories differs between male and female pastors. The tables indicate that male pastors most

significantly represent male subjects as active–competent agents and female subjects as cogitative–perceptive and communicative–expressive agents. Female pastors seem more inclined to associate both genders with all agency types. Furthermore, male pastors are more restricted in their use of verbs, whereas female pastors show higher variation within the presented categories.

In the single cases, male and female pastors appear to represent male subjects quite similarly in the category active–competent agency: that is, as capable and independent agents. As mentioned, a similar full category does not appear for female subjects as represented by male pastors. For male and female pastors alike, the most significant verb according to PMI score for female subjects in the corpora is “gave birth,” suggesting that an important role for female agents in the corpus is that of mothers. In the relational cases, female pastors establish a relation between female subjects and male objects with this verb, suggesting that they might reserve the mother role for the Virgin Mary, mother of Jesus Christ. The same relationship is not significant in sermons by male pastors, as they tend to associate the verb with male as well as female objects (him and her).³⁴ This observation may indicate that male pastors focus more generically on female agents as mothers.

In the categories cogitative–perceptive agency and communicative–expressive agency, female and male pastors associate rather similar verbs with female subjects; male pastors, however, use verbs that are slightly more emotionally neutral. When female pastors ascribe these types of agency to male subjects, they tend to use verbs such as “decides,” “accepted,” “declare,” “asserted,” which may support the independence and capability also found in the category of active–competent agency.

The category of religious agency is a small category of verbs which, as with communicative–expressive agency, mainly indicate contact-oriented actions, but by means of a vocabulary associated with biblical narratives. In sermons by male pastors, the few verbs indicating religious agency do not imply a distinct male–female relation, as there are no religious verbs represented in the relational cases, only in the single cases. In contrast, this category is more demonstrable in sermons by women, where the actions seem to mediate a relation between male and female agents. The religiously related verbs seem to be closely linked to biblical characters. “Preaching” and “healing” are actions strongly associated with Jesus in the gospels;

whereas the story of the female sinner who washes and anoints Jesus' feet (Luke 7: 36-38) seems to be lingering behind the close connection between female religious agency and the word "anoint." Thus, it seems that female pastors not only represent religious actions with a bit more variety, but also tend to represent the agency of biblical characters as more relational across genders.

In general, in sermons by men we find extremely few verbs that are significant for both male agents in nominative case and female agents in oblique case – and none that are representative of the four categories in the relational case "Female subject and Male object". The relational cases generally indicate that female pastors tend to use verbs that establish links between genders. This finding seems to correspond with the observation that female pastors are in all cases more inclined to include verbs signaling communicative or interactional behavior.

Male and female language

The gender patterns found in the sermons corpus seem to mirror observations from socio-linguistic research on male and female language users. To characterize tendencies among female language users, Julia Wood and Deborah Tannen has respectively proposed the terms "feminine speech community" and Deborah Tannen the term "rapport talk". Here, language users seek to establish contact and connections between conversational partners as a strategy for maintaining relationships by means of communicative acts expressing support, equality, responsiveness, and tentativeness. In addition, Wood and Tannen find that male language users often seek to establish distinctions between conversational partners and to use language instrumentally to provide solutions for concrete objectives and to signal status and independence. Here, language tends to be more direct and assertive. Wood uses the term "masculine speech community" to conceptualize this strategy, and Tannen characterizes it as "report talk". Both Wood and Tannen explain these strategies by different socialization processes typical for girls and boys, while Wood emphasizes that any child, regardless of biological sex, can be socialized into either of the speech communities³⁵.

In our findings, male pastors seem to distinguish female and male agents in terms of traits from feminine and masculine speech strategies cf. Wood and Tannen, with

communicative agency characteristic of female agents, and active agency associated with male subjects. In contrast, female pastors appear to attribute active and communicative agency to both genders in all cases; that is, female agents and male agents share these traits, whereby the traits do not appear to be gender dependent in sermons by women. We find a similar tendency when we look at the category cogitative-perceptive agency: male pastors reserves this trait for female subjects; whereas female pastors are likely to attribute it to male and female agents alike. In a study on inward- versus outward-oriented behavior in classic nineteenth-century novels, Andrew Piper demonstrates cogitative and perceptive behavior as prominent traits of female main characters in novels by female authors. He argues that these traits established a new type of subjectivity for female literary characters in this period.³⁶ In contrast, our analyses indicate that these traits are more likely to occur as feminine in contemporary sermons by men, than in sermons by female pastors. In general, female pastors seem to represent social worlds in which agents acquire both female and male traits regardless of the gender of the pronoun, thus facilitating a symmetrical social structure. In comparison, male pastors tend to provide an asymmetrical structure when representing gender, by maintaining different traits for female and male agents – a tendency that resembles aspects of report talk and masculine speech communities. Therefore, this analysis illustrates that sermon content is not gender neutral: male and female pastors differ in their use of fundamental linguistic components, gendered pronouns and verbs, and thereby in the actions they let characters perform. As such, they shape the social worlds of sermons in quite different ways.

We acknowledge that pastors could use language to represent more diverse gender identities than just male and female and that our approach does not capture such constructions. Other methods would be more appropriate for exploring nuances to such non-binary structures. Whereas the personal pronouns we have used are already “pregendered” binary in language, extracting personal entities through Named Entity Recognition would provide corpus characters that can potentially be constructed more freely in language. This approach would enable us to study agency variations of individual characters across the corpus. For example, whether the agency of God would fit within a binary gender pattern or divert completely from such a structure. However, before embarking on investigating possible deconstructions of binary structures and more pluralized gender identities in a new and otherwise unknown

corpus such as ours, we found it imperative first to investigate whether and how binary gender structures exist in the corpus. Furthermore, our study indicates clearly gendered differences in the context of the ELCD, where such topics tend to gain rather little attention.

Conclusion

We hope that this article serves to accentuate the potential that large-scale sermon studies can offer. Our contribution to this endeavor has been to sample, clean, and archive a corpus of Danish church sermons that have never before been read. These sermons constitute a collective text production containing in-depth cultural information about the voice of a religious community. Our corpus is especially noteworthy for the sampling and archiving approach adopted. We sampled the sermon manuscripts directly from their authors as private documents that were not necessarily intended for dissemination. One of the consequences of sampling unedited sermons was that the pastors did not attain comparative structural consistency in their documents, which convinced us to clean and annotate the documents manually. While time-consuming, this approach allowed us to remain responsive to the material and also to familiarize ourselves with it. This provided us with an opportunity to gain insights into the quality of the material before commencing data analyses.

One particular research challenge concerns the actual contents of the corpus. Given that these are contemporary documents, written and spoken by living pastors and often containing references to individual parishioners, they constitute potentially sensitive data, with the consequence that the corpus cannot be fully open. We find it commendable that communities within the computationally informed humanities encourage initiatives to share data and provide open-access data. Constructing data sets is both theoretically and methodologically fraught, not to mention time-consuming, and barriers to access can also be a barrier to progress. However, it is also necessary to promote the creation of data sets that do require restricted access. We believe data sets such as ours can provide unique insights into living cultural fields. Access to these fields, though, requires that we attend to the agents behind the documents and to their incentives to provide data. In return, collecting data

directly from contemporary contributors and obtaining their consent allows researchers to acquire full ownership of data.

We have included a case study in this article in order to illustrate the research potential of the data set. We find that mentions of male and female characters by female and male pastors are proportionally similar both to each other and to mentions in the lectionaries. Nevertheless, we also find that representations of gender binaries are more complicated when comparing tendencies in male and female sermon corpora on a discursive level. The public voice of the ELCD is thus collective but not “homophone”. The case study thereby demonstrates the value of combining quantitative and more qualitative approaches to this new corpus, and it illustrates that sermons are the outcomes of intricate dynamics between scripture, individual, and society – dynamics that need to be investigated more thoroughly in the future.

References

- ¹ Linda Woodhead and Paul Heelas, *Religion in Modern Times: An Interpretive Anthology* (Oxford: Blackwell, 2000).
- ² José Casanova, *Public Religions in the Modern World* (University of Chicago Press, 1994): 19; Peter L. Berger, ed., *The Desecularization of the World: Resurgent Religion and World Politics* (Wm. B. Eerdmans Publishing, 1999): 2-6; Jürgen Habermas, “Notes on Post-Secular Society,” *New Perspectives Quarterly* 25, no. 4 (2008): 17–29
- ³ Inger Furseth, *Religious Complexity in the Public Sphere: Comparing Nordic Countries*, Palgrave Studies in Religion, Policy. (Cham: Palgrave Macmillan, Springer International Publishing AG, 2018); Henrik Reintoft Christensen, Brian Arly Jacobsen, and Signe Engelbreth Larsen, "Introduktion: Religion i det offentlige rum - mellem myter og blinde vinkler", in *Religion i Det Offentlige Rum et Dansk Perspektiv*. (Aarhus: Aarhus Universitetsforlag, 2019).
- ⁴ Peter B. Andersen, Jakob Erkmén, and Peter Gundelach, “Udviklingen i (ikke)religiositet,” in *Usikker modernitet: danskernes værdier fra 1981 til 2017.*, ed. Morten Frederiksen (Kbh: Hans Reitzel, 2019), 231–64.
- ⁵ Marie Vejrup Nielsen, “Folkekirken Og Det Offentlige Rum,” in *Religion I Det Offentlige Rum – Et Dansk Perspektiv*, ed. Henrik Reintoft Christensen, Brian Arly Jacobsen, and Larsen, Signe Engelbreth (Aarhus: Aarhus Universitetsforlag, 2019), 48–51.
- ⁶ Kirstine Helboe Johansen and Helene Helboe Pedersen, “The Power over the Marriage Ritual: Positions in the Administrative Consultation Prior to the Implementation of Same-Sex Marriages in the Evangelical–Lutheran Church in Denmark” *Journal of Church and State* 58, no. 4 (September 1, 2016): 731–52.
- ⁷ University of British Columbia, “The Database of Religious History”, 2020, <https://religiondatabase.org/landing/>.
- ⁸ Todd M. Johnson and Brian J. Grim, eds., *World Religion Database* (Leiden/Boston: Brill, 2020), <https://www.worldreligiondatabase.org/>.
- ⁹ Todd M. Johnson and Gina A. Zurlo, eds., *World Christian Database* (Leiden/Boston: Brill, 2020), <https://worldchristiandatabase.org/>.
- ¹⁰ Tina M Schneider and W Harry Plantinga, “Hymnary.Org: The Use of a New Database in Hymnology”, *Advances in the Study of Information and Religion* 2 (2012): 155–77.
- ¹¹ Zhargalma Dandarova Robert et al., “A Web-Based Database for Drawings of Gods: When the Digitals Go Multicultural” *Archive for the Psychology of Religion* 38, no. 3 (2016): 345–52, <https://doi.org/10.1163/15736121-12341326>.

- ¹² Pew Research Center, "The Digital Pulpit: A Nationwide Analysis of Online Sermons," December 16 (2019), <https://www.pewforum.org/2019/12/16/the-digital-pulpit-a-nationwide-analysis-of-online-sermons/>.
- ¹³ Pew Research Center, "The Digital Pulpit: A Nationwide Analysis of Online Sermons," December 16 (2019): 5, <https://www.pewforum.org/2019/12/16/the-digital-pulpit-a-nationwide-analysis-of-online-sermons/>.
- ¹⁴ We contacted the rural deans under the last of the ten bishops without a statement of approval.
- ¹⁵ For instance, All Hallows' Day in odd years is represented: AHelg_U.
- ¹⁶ "Teologisk Stat", 2020, <http://teo.apolloweb.dk/public/login.asp>.
- ¹⁷ Due to GDPR concerns, we are unable to share the full metadata, owing to the fact that individual pastors can be identified by a unique ID.
- ¹⁸ During the process, we became aware that many pastors had also supplied extra material that we wished to preserve separately for other projects. This consisted predominantly of prayers given just after the sermon and pastoral speeches intended for other events (funerals, visits from schoolchildren). Both types of material contain valuable information for the research questions we pursue. Our choice of a manual approach to cleaning and archiving the sermons allowed us to realize that we had received much more cultural information than requested from our informants.
- ¹⁹ P. M. McCarthy and S. Jarvis, "From Intrinsic to Extrinsic Issues of Lexical Diversity Assessment: An Ecological Validation Study," in *Vocabulary Knowledge: Human Ratings and Automated Measures*, ed. S. Jarvis and M. Daller, Studies in Bilingualism; 47 (Amsterdam: John Benjamins Publishing Company, 2013), 45–79.
- ²⁰ Susan Brown and Laura Mandell, "The Identity Issue", *Journal of Cultural Analytics* (February 2018), <https://doi.org/10.22148/16.020>.
- ²¹ Matthew Jockers and Gabi Kirilloff, "Understanding Gender and Character Agency in the 19th Century Novel", *Journal of Cultural Analytics* (December 2016), <https://doi.org/10.22148/16.010>; William E Underwood, David Bamman, and Sabrina Lee, "The Transformation of Gender in English-Language Fiction", *Journal of Cultural Analytics* (February 2018), <https://doi.org/10.22148/16.019>; Eve Kraicer and Andrew Piper, "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction", *Journal of Cultural Analytics* (January 2019), <https://doi.org/10.22148/16.032>.
- ²² Marie Vejrup Nielsen, "Transformationer i Folkekirkekristendommen i Danmark i Dag", *Religionsvidenskabeligt Tidsskrift*, no. 53 (2009): 68.
- ²³ Sørine Gotfredsen, "Debat: Kvindelige præster diskrimineres ikke - de tager magten", *Berlingske Tidende*, January 27 2019, <https://www.berlingske.dk/kommentatorer/kvindelige-praester-diskrimineres-ikke-de-tager-magten>.
- ²⁴ Ingrid Ank, "Jeg mener ikke noget", *Grundtvigsk tidende*, 2018, <https://grundtvig.dk/jeg-mener-ikke-noget/>.
- ²⁵ Sanne Thøisen, "At lede efter kønsspecifikke gudsbilleder", in *Kvinde, mand, kirke: Folkekirken og den lille forskel*, ed. Eva Holmegaard Larsen (Frederiksberg: Anis, 2012), 169–82.
- ²⁶ Tommy Byrne, "Danskere: Ligestilling? - det har vi da", *Århus Stiftstidende*, November 5 2019, <https://stiften.dk/artikel/danskere-ligestilling-det-har-vi-da>.
- ²⁷ The World Economic Forum, "The Global Gender Gap Report 2014", 2014, <https://reports.weforum.org/global-gender-gap-report-2014/>; The World Economic Forum, "Global Gender Gap Report 2020", 2020, <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>.
- ²⁸ Else Marie Wiberg Pedersen, "Feminisme eller feminisering?", in *Kvinde, mand, kirke: folkekirken og den lille forskel*, ed. Eva Holmegaard Larsen (Frederiksberg: Anis, 2012), 59–75.
- ²⁹ Lone Fatum, "Kønsblind - En Diagnose", *Kritisk Forum for Praktisk Teologi*, no. 142 (2015): 61–70.
- ³⁰ Judith (Judith P.). Butler, "Conclusion: From Parody to Politics," in *Gender Trouble: Feminism and the Subversion of Identity* (New York: Routledge, 2002), 181–90.
- ³¹ A Github repository for this project can be found here: <https://github.com/centre-for-humanities-computing/sermon-research.dk>
- ³² Dina Holm Larsen, "Kvindeskikkelser i den danske gudstjeneste - Et religionsvidenskabeligt speciale om bibelske kvinders repræsentation i gudstjenestens tekstrækker" (Unpublished master thesis, Aarhus University, 2019): 29-36.
- ³³ Tables 4a, 4b, 5a, 5b presents verbs representative of one of the four analytical categories in single and relational cases for male and female pastors respectively. The verbs are translated from Danish to English for this publication. Frequency scores (Freq.) represent how often the verbs appear in total in sermons by men (Tables 4a, 4b) and women (Tables 5a, 5b). PMI scores (PMI) measure the associative relationship between verbs and a pronoun. % indicates no representation of verbs or scores in a category.
- ³⁴ PMI scores for the verb "gave birth" associated with female subjects and respectively male and female objects in sermons by male pastors are "She": 5.3, "Him": 3.0, "Her": 3.0. In comparison, PMI scores for the same variables in sermons by female pastors are "She": 5.2, "Him": 2.7, "Her": 0.0.

³⁵ J.T. Wood, *Gendered Lives - Communication, Gender and Culture* (Cengage Learning, 2014): 102-122; Deborah Tannen, *You Just Don't Understand: Women and Men in Conversation.*, Repr. (London: Virago, 1992): 23-95.

³⁶ Andrew Piper, "Characterization (Constraint)," in *Enumerations: Data and Literary Study* (Chicago/London: The University of Chicago Press, 2018), 118-46.