Collective Testimonies to Christianity and Time

A collection and large-scale text study of 11,955 Danish sermons from 2011-2016

Anne Agersnap
PhD dissertation

Graduate School, Arts,
PhD programme for Anthropology,
International Area Studies
and the Study of Religion

Aarhus University
March 2021

Main supervisor: Kirstine Helboe Johansen, Associate Professor, Aarhus University
Co-supervisor: Kristoffer Laigaard Nielbo, Associate Professor, Aarhus University
Co-supervisor: Uffe Schjødt, Associate Professor, Aarhus University
Acknowledgments

A wise woman told me some time ago that conducting a PhD project is like doing a big puzzle, while you are sitting in the box with all the pieces. You manage to find some of them and put together smaller or larger parts of the whole picture – maybe some of the sky, the contours of a forest or a little flowerbed. And even though you are not able to put down the whole puzzle this way, at least you know next time to step out of the box, when you begin. This dissertation is my own testimony to the pieces I have found and placed together. I am fortunate to have had a range of good people around me to help me do the puzzle. I owe my gratitude to all of them.

The one who deserves it the most is my supervisor Kirstine Helboe Johansen. We got to know each other about eight years ago, when we first began to talk about making a sermon database. Somehow, I still feel like the project is brand new and that we still have so many more paths to take together. I hope some of them will be realised in the future. I have enjoyed every session of supervision, whether at your office, over lunch, in company of Laura and Anna or most recently in the form of walking tours in Riis skov. You have an extraordinary ability to infer from my stream-of-consciousness-kind-of-talks exceptionally coherent perspectives, supplemented by your own sharp observations. I value your sincerity – both when you provide criticism, share in my enthusiasms and calm down my frustrations. Thank you for following me on this path into unknown methods, digital technology and corpus building, which was also new to you. You have followed it not only from the side, but also by taking part in it with our partnership at the Digital Literacy course and our course on programming at Oxford University. These experiences have been some of the highlights of my time as a PhD student. I do not have enough words to tell you, how much I have learned from you, because it concerns every aspect of my learning curve these past couple of years. I truly consider you to have been a mentor to me.

I owe a special thanks to the department of the Study of Religion at Aarhus University, who has been my workplace these last four years. It has brought many smiles to my face, when I from my open office-door have heard whistling in the hall, conversations on the proper amount of coffee in the coffee machine and in general pleasant conversations among peers. In particular, I am thankful for having been a part of the Centre for
Contemporary Religion. I have enjoyed the regular lunch meetings and the interest in junior researchers that the group exhibits. Thank you to Uffe Schjødt for being my co-supervisor at the department, and especially for the many regular meetings about the sermon database, when I was still a bachelor student. I have enjoyed being part of the department along with a group of talented and kind PhD students. I have also enjoyed the small PhD community I have had with Laura Bjørg Serup Petersen and Anna Døssing Gunnertoft with lunch meetings and especially our road-trip with Kirstine to Kassel in 2018. I also have very fond memories from the conference we attended in Helsinki, Laura. I would like to thank the PhD students in particular, with whom I have been fortunate enough to share the most awesome office. To Anne Buch Møller and Sara Jul Jacobsen, who were my office mates the first years of my project, and to Sanne Andersen Hansen and Anne Lundahl Mauritsen, who have been there for me at the office these last couple of years. I have enjoyed sharing the magnificent view with you all. The last couple of years it has truly been a sisterhood, where we have identified ourselves with each other’s ups and downs. Thank you for creating a room for sincerity, support and friendship. Thanks for making the office festive and for over-decorating it with ornaments at Christmas. And a special thanks to you Anne, who has been in my life both scholarly and personally since our days at the student editorial team TOTEM. Thanks for stepping up so often – for taking part in cleaning and archiving sermons and for proofreading my dissertation. You are a star and an awesome scholar.

I am very grateful to the Humanities Centre for Computing Aarhus, in particular to Kristoffer Laigaard Nielbo who has been my co-supervisor and initiator of the collaborate sermon project several years ago. I must admit I was a bit uneasy, when you changed job to the University of Southern Denmark in 2017, because I relied on your technical expertise as supervisor. However, when you came back to Aarhus University, it was with a completely new research environment in the form of CHCAA that opened up my own project in new ways. I enjoyed the short research stay I had at SDU in 2018, where we kick-started the analytical phase of my project. Very much, I owe thanks to Ross Deans Kristensen-McLachlan, who has collaborated with me in most of my analytical endeavours. I have enjoyed our meetings about the corpus, which has included a lot of creative thinking as well as pragmatic solutions. Thank you Max Roald Eckardt for taking on the job of designing the sermon database. I am looking forward to seeing the final product.
I also owe Mads Rosendahl Thomsen my thanks for including me in the Digital Literacy course in 2018. When I commenced my PhD project, I did not know of many other digitally informed humanities projects at Aarhus University, but with the DigLit course, all of a sudden we were a whole community across departments, who were asking similar question and requiring similar needs. This course brought renewed energy to my own project. Thank you also to Ulf Dalvad Berthelsen for the afternoons of informal python-programming over coffee at your office with Kirstine. None of us reached the master-level, but it was a very pleasant way to end a day at the office.

I am extremely thankful to Frank Fischer, who was a wonderful host, when I was in Moscow in 2019 for my research stay abroad. My months in Moscow was the absolute highlight of my PhD project. Although you, Frank, were busy with work and travelling, you still found time to discuss my project with me, and to spending time with me in your spare time. Thank you for showing me around, for joining Nastya and me at midnight mass and eating Easter cake afterwards at 2 a.m. Thank you for going to the Bolshoi ballet, for travelling to Saint Petersburg, and for watching and hating the last season of Game of Thrones. Thank you for being a true friend, while I was abroad.

Thank you to the 95 pastors who invested time and goodwill in my PhD project, by searching out their hard-drives and drop-boxes to send me a collection of their sermons. I also owe a great thanks to Line Viborg, Anne Lundahl Mauritsen, Marlene Olesen, Metteline Knudsen and Jakob Kokkenborg for your comprehensive work of cleaning and archiving my 11,955 sermons by hand, and for following the rather tedious and meticulous procedure, I outlined.

I am personally thankful to Jytte, who has provided me encouragement and life-lessons these past years. Thank you to my friend and buddy Maj, who has always showed interest and enthusiasm towards my project, and who knows how to help me find my balance, when I need it. Thank you to my dear friend Line who has taught me that there is always a reason to drink champagne and cheer. Last, but not least, thank you to my partner Esben, who has provided support and cocktails in the order and amount needed these last months. Thank you for turning up in my life in the middle of my project period and offering me a whole other box of puzzle pieces that I did not know I was looking for. Thank you for stepping out of that box with me to find the pieces together.
Content
Chapter 1: Introduction ........................................................................................................................................... 9
  - Project thesis and research questions ........................................................................................................... 9
  - The Evangelical Lutheran Church in Denmark .......................................................................................... 11
  - Aims and contributions: Entering a new research arena ........................................................................... 14
  - Outline of the dissertation ........................................................................................................................ 17
Chapter 2: The collective and intertextual nature of sermons ................................................................................. 20
  - Sermons as a living practice ........................................................................................................................ 20
  - From a performative to a textual understanding of sermons ..................................................................... 24
  - Digital humanities: Theoretical and analytical developments .................................................................... 30
  - A collective text system in complex interactions ..................................................................................... 36
Chapter 3: Methodological Framework .................................................................................................................. 38
  - Data collection ........................................................................................................................................... 39
  - Methods for analyses ................................................................................................................................. 45
Chapter 4: The sermon corpus (Product) .................................................................................................................. 49
  - The research product ................................................................................................................................. 50
  - Background: Corpus linguistics ............................................................................................................... 50
  - Metadata .................................................................................................................................................. 52
  - Cleaning and annotating sermons ............................................................................................................. 55
  - The relational structure of the sermon corpus ......................................................................................... 57
  - Ethical considerations ............................................................................................................................... 59
  - Conclusion ............................................................................................................................................. 60
Chapter 5: Background to articles .......................................................................................................................... 62
  - Background to chapter 6 ........................................................................................................................... 63
  - Background to chapter 7 ........................................................................................................................... 65
  - Background to chapter 8 ........................................................................................................................... 66
  - Concluding remarks ................................................................................................................................. 69
Chapter 6: Sermons as data: Introducing a corpus of 11,955 Danish sermons (article 1) .......................................... 70
  - Abstract ................................................................................................................................................ 70
  - Introduction ........................................................................................................................................... 70
  - The public voice of a religious community: The Danish context ............................................................... 71
  - Digital archival resources in the study of religion .................................................................................... 74
  - Methods: Acquiring sermon data and building a text corpus ................................................................. 75
Collective Testimonies to Christianity and Time

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus statistics</td>
<td>80</td>
</tr>
<tr>
<td>Case study: Gender representations in sermons</td>
<td>85</td>
</tr>
<tr>
<td>Conclusion</td>
<td>95</td>
</tr>
<tr>
<td><strong>Chapter 7</strong></td>
<td>97</td>
</tr>
<tr>
<td>- Unveiling the character gallery of sermons – A social network analysis of 11,955 Danish sermons (Article 2)</td>
<td>97</td>
</tr>
<tr>
<td>Abstract</td>
<td>97</td>
</tr>
<tr>
<td>Introduction</td>
<td>97</td>
</tr>
<tr>
<td>The Evangelical-Lutheran Church in Denmark</td>
<td>98</td>
</tr>
<tr>
<td>Sermons: A polyphony of voices</td>
<td>99</td>
</tr>
<tr>
<td>Methods</td>
<td>102</td>
</tr>
<tr>
<td>The social worlds of sermons</td>
<td>109</td>
</tr>
<tr>
<td>Narrative and thematic dynamics in sermons</td>
<td>113</td>
</tr>
<tr>
<td>Discussion: The rhizome revisited</td>
<td>118</td>
</tr>
<tr>
<td>Conclusion: Characters in a chain of memory</td>
<td>120</td>
</tr>
<tr>
<td>References</td>
<td>121</td>
</tr>
<tr>
<td><strong>Chapter 8</strong></td>
<td>125</td>
</tr>
<tr>
<td>- “What is love…?” – A study of thematic constructions in 11,955 Danish sermons (Article 3)</td>
<td>125</td>
</tr>
<tr>
<td>Abstract</td>
<td>125</td>
</tr>
<tr>
<td>Introduction</td>
<td>125</td>
</tr>
<tr>
<td>The Evangelical-Lutheran Church in Denmark</td>
<td>126</td>
</tr>
<tr>
<td>Sermons: Collective testimonies to Christianity and time</td>
<td>127</td>
</tr>
<tr>
<td>Transformational patterns in contemporary Christianity</td>
<td>129</td>
</tr>
<tr>
<td>Methods and observations</td>
<td>131</td>
</tr>
<tr>
<td>Thematic contexts in sermons</td>
<td>141</td>
</tr>
<tr>
<td>Discussion: Theological dynamics in sermons</td>
<td>147</td>
</tr>
<tr>
<td>Conclusion</td>
<td>149</td>
</tr>
<tr>
<td><strong>Chapter 9</strong></td>
<td>150</td>
</tr>
<tr>
<td>- Conclusion</td>
<td>150</td>
</tr>
<tr>
<td>Access to Danish sermons</td>
<td>150</td>
</tr>
<tr>
<td>Interactions with intertextual systems</td>
<td>151</td>
</tr>
<tr>
<td>In between biblical and cultural contexts</td>
<td>155</td>
</tr>
<tr>
<td>Collective testimonies to Christianity and time</td>
<td>160</td>
</tr>
<tr>
<td>Bibliography</td>
<td>162</td>
</tr>
<tr>
<td>English summary</td>
<td>165</td>
</tr>
<tr>
<td>Dansk resumé</td>
<td>169</td>
</tr>
<tr>
<td>Appendix I</td>
<td>173</td>
</tr>
<tr>
<td>Appendix II</td>
<td>176</td>
</tr>
</tbody>
</table>
Chapter 1

- Introduction

“In sermons, pastors show they respond to the reality surrounding us.”

“As a preacher, you are vulnerable. You are in it yourself – it is your lifeblood. The sermon is something very personal.”

“With the sermon, there has to be something to share in for those attending church.”

“I look for the contestation in the gospel passage – where I have a problem with what is being said.”

“You need to begin with the things that is troubling for your congregation. That is why you need to know your congregation well.”

“The sermon expands between two poles you cannot lose yourself to: Either, you may lose yourself in the text and lose sight of reality, or else, you may lose yourself in reality and lose the anchoring of the gospel.”

“I am so tired of the “natural” themes in the liturgical texts – I seek to find new themes in the texts.”

“It is important that the text speaks to me. If it does not, it can be hard to incorporate it into my sermon. But there might still be a corner of it I can grab onto.”

“Among pastors, it is interesting whether they articulate what the congregation wants to hear: Do they come to hear about television programmes or about Jesus? Or can you combine both?”

“The sermon is very personal – it exhibits some personal attitudes and feelings towards what I talk about, but I never draw in personal experiences.”

“It is about finding common reference points that we can share. Once, it was easier with fewer media and a television monopoly, but today, these reference points are filtered through all sorts of channels.”

“You can enter the sermon as any other human being – the pastor does not invoke a special authority.”

“The pulpit is the freest of lecterns in the country. You can speak freely. Some may be scared by it – words can be dangerous and you can end up being too cautious. Others may feel no limitations and might claim a truth that is nothing more than their own truth.”
Around the turn of the year between 2016 and 2017, I visited several pastors in Denmark to talk to them about sermons. The cause of these visits was my ambition to assemble a large corpus of contemporary Danish sermons. At that time, one of my greater ponderings was how feasible it would be to acquire such material, and so, I sought out a couple of pastors to talk to them about my idea. Besides getting a sense of pastors’ interest in contributing with sermons to the corpus, these meetings gave me a chance to hear about pastors’ own thoughts on writing sermons to their congregations. The quotes above, originating from these meetings, clearly illustrate that pastors have a range of different perspectives on sermons, and on what it entails to prepare and deliver sermons. Although they may not necessarily disagree with one and other, they accentuate different aspects of sermons in different ways: sermons as anchored in the biblical texts, sermons as relevant in contemporary culture, sermons as personal speeches, or sermons as an attentive address to congregations. These different assertions in themselves demonstrate a multitude of different voices behind the sermons pastors write in the Evangelical Lutheran Church in Denmark (ELCD). It is obvious that behind every sermon there is an individual voice.

Every week a sermon is delivered in each of the approximately 2000 ELCD parishes in the country, making preaching in church a synchronous event. This event presupposes a process in which pastors consider, choose, assemble or discard material in order to compose the weekly address. The vast majority of pastors in the ELCD compose their sermons in the form of manuscripts, which means that in parallel to the synchronous event, a synchronous material emerges from pastors’ written sermons. This knowledge makes us aware that sermons are not merely an unfolding of individual pastors’ singular interactions with gospel, cultural fields, self and congregation, but of a multiplicity of these interactions. Therefore, even though there is a multitude of individual voices behind sermons, an important question arises: Where do these voices intersect? Where do they differ? What collective statements about Christianity and contemporary time assemble from these voices?

Etymologically, the terms “sermon” and “preaching” originate from the Latin terms “sermo” and “praedicare” – the former meaning “discourse” and the latter “proclaim”. Writing and delivering sermons is a distinct religious practice that can be traced back to the first Christian congregations, and that may be found in similar genres in other religions. As a a regular practice, it falls upon a specialised group, ordained pastors, to
compose sermons. Sermons exhibit a continuous engagement with scripture and tradition in a contemporary situation. Whether the contemporary setting was yesterday, fifty years ago or five hundred years ago, that engagement is at the core of preaching. However, the contemporary aspect of sermons in itself makes preaching an interesting cultural practice, because sermons are also a recurring practice in society, where a large but distinct group of people poses statements about the world in the world. Therefore, sermons are unique sources to historical periods in society, and in this way, documentations of Christianity and time. Yet, as enhanced, they are not merely individual sources; they are the outcome of a shared practice and therefore parts of a collective text production.

There are quite likely pastors who still store their sermons as print versions in archive folders, but the form of this material has changed considerably over time due to basic digital resources. Today, this ever-growing material exists to a large extend in pastors’ individual digital archives throughout the country. In the form of manuscripts, sermons therefore remain the property of individual pastors, disengaged from the historical and collective production to which they contribute.

Project thesis and research questions
The thesis of this PhD project is, in line with the above perspectives and considerations, that sermon manuscripts are outcomes of a collective practice, and that they can be studied this way; as collective phenomena. It is further the argument of the thesis that pastors through this practice facilitate an encounter between religion and contemporaneity.

This thesis has led to two sets of research questions. The first addresses the issue of Danish sermons and accessibility:

*How can empirical access be achieved to a material of sermon manuscripts, owned by individuals and in this form disengaged from the collective production it takes part in?*

Thus, the first question concerns the empirical ambition of the project and illuminates that it has been a distinct project objective to assemble a text corpus of ELCD sermons from across the country in order to make the material come to exist as a collective production. As an outcome of this objective, I have assembled a text corpus of 11,955
Danish sermons from 2011-2016 written by 95 ELCD pastors, which I include in the dissertation as a research product.

The second question concerns the analytical ambition of the project:

*How do pastors’ collective encounters with Christianity and cultural fields unfold in sermons?*

This question is naturally far too comprehensive to exhaust within the boundaries of one PhD project and therefore, I have operationalised it into three analytical cases that I have enclosed in three individual articles. In the first case, I explore how cultural codes in society unfold in sermons, by attending to the presence of gender discourses in the corpus. In the second study, I attend to the character gallery that appears from the sermon production, in order to investigate which cultural and religious contexts pastors seem to interact with in particular. The last article is a study of theological dynamics in the corpus, where I examine emergent themes as related to liturgical texts and contemporary cultural codes. In combination, the articles scrutinise the dynamics in three different relations between Christianity and cultural fields: 1) How everyday use of language embed cultural discourses in sermons; 2) How cultural and religious contexts interact; and 3) How theological themes relate to contemporary cultural codes.

**The Evangelical Lutheran Church in Denmark**

Preaching is a core practice in Christian traditions, but the concrete prescriptions and role of sermons differ to a great extent among Christian communities. The sermons at the centre of this PhD project are all written by pastors in the Evangelical Lutheran Church in Denmark, and thus composed within the same institutional framework.

The ELCD is the majority religion in Denmark with high membership (73.8% of the entire Danish population as of 1 January 2021)\(^1\). However, the population’s affiliation to the church is in stable decline (down from 80.4% of the population as of 1 January 2011)\(^2\).

The ELCD has never had a constitution of its own, and while it is not a state-church, the ELCD is strongly affiliated with the Danish state. Section 4 of the Danish constitution

---

\(^1\) Statistics regarding membership in 2021 can be found at: [https://www.folkekirken.dk/om-folkekirken/folkekirken-i-tal/medlemmer](https://www.folkekirken.dk/om-folkekirken/folkekirken-i-tal/medlemmer) (visited 4 March 2021).

\(^2\) Statistics regarding membership in 2011 can be found at: [https://www.km.dk/folkekirken/kirkestatistik/](https://www.km.dk/folkekirken/kirkestatistik/) (visited 4 March 2021).
The Evangelical Lutheran church is the Danish people’s church and as such subsidised by the state. In this way, the ELCD has no legal independence or a national synod, and formal regulations regarding church affairs are handled by parliament. However, it has been a standing principle that neither government nor parliament interfere with internal church affairs.

The ELCD is organised in ten dioceses corresponding to delimited regional areas in Denmark. The dioceses are supervised by ten bishops and each diocese subdivided into deaneries. In each of the 103 deaneries, a rural dean supervises the parishes adhering to the deanery. As of 2016, the ELCD had 2169 parishes led by one or more local pastors.

ELCD pastors are as a rule university graduates with a theological degree from either Aarhus University or the University of Copenhagen, and after academic education, they have undertaken the ELCD’s practical pastoral training. In the parishes, parishioners form parochial church councils that take part in running the local churches. The parishes enjoy in general a considerable amount of autonomy, as pastors and counsels are free to host activities or perform tasks in the parish as supplement to the official ELCD practices. Supplementary to the local parishes, the ELCD also employs a group of pastors with distinctive tasks, who for example perform their pastoral duties for patients in hospitals, inmates in prisons or students at educational institutions.

Although the ELCD has high membership, regular church attendance is considerably low. The most recent survey disclosed that 11% of the population attends church once a month at the least. However, the ELCD is considerably more visited when it comes to life ceremonies. In 2020, 55.6% of all newborns were baptised, 66% of all 13-15 year old teenagers had a confirmation performed, and 81% of all deceased had a church funeral.

The church adheres to the Evangelical-Lutheran confession. While the ELCD encompasses different theological positions, the church in general is known for a liberal

---

13 Section 4 of the constitution in Danish: “Den evangelisk-lutherske kirke er den danske folkekirke og understøttes som sådan af staten.”

4 Numbers regarding the ELCD’s structure can be found at: https://www.folkekirken.dk/om-folkekirken/organisation (Visited 19 February 2021).

5 In rare cases, the minister of ecclesiastical affairs and a bishop may permit applicants without a theological degree to apply for pastoral offices. However, this dispensation has only been granted 38 times from 1999-2019 (https://www.folkekirken.dk/om-folkekirken/hvem-arbejder-i-kirken/praesten/paragraf-2-ordningen - visited on 18 February 2021).

6 Statistics regarding participation in life ceremonies can be found at: https://www.folkekirken.dk/om-folkekirken/folkekirken-i-tal (Visited on 19 February 2021).
theological stance. The ELCD was the first church community to ordain women pastors in 1948. In 2014, numbers showed that 55% of the employed ELCD pastors were women (Henriksen 2014). Seven of the bishops are men and three of them women as of January 2021. In 2012, parliament passed a bill that permitted same-sex marriages to be performed in church, but pastors are free to choose whether they want to perform the ritual themselves. Within the ELCD, there is a general openness towards critical research of Christianity, the church and its practices. The church has researchers employed at their centre for education and knowledge⁷ to undertake sociologically informed research projects, for example to remain informed about and responsive to members’ use of, interests in and requirements to the ELCD.

The ELCD’s liturgical calendar includes 69 official church holidays, and the official church services adhere to a prescribed liturgical order. The current liturgical order was approved by bishops and authorised by the Danish Queen in 1992. Pastors are supposed to follow the order, but the liturgy includes a number of variations they can choose between. Preaching is an integral part of the liturgical order that consists of additional ritual acts such as prayers, hymn singing, the Creed, the Lord’s supper and of two or three readings from the Old testament and/or Epistles and from the Gospels. The readings for each holiday is prescribed in two lectionaries that apply respectively for even years and uneven years. Preaching falls in the middle of the service, and commences immediately after the pastor reads aloud the Gospel pericope of the day. Pastors are supposed to base their sermon on the prescribed periscope, but they are free to choose themselves how to include the text, and in general, they are free to choose the content they deem relevant to present for their congregations. In this way, the sermons for the church services are always in principle formulated anew.

**Aims and contributions: Entering a new research arena**

My project thesis is based on the observation that sermons have not been studied extensively as a collective historical text material at large scale before. I presume one of the main reasons is a practical one; namely, that such a project requires an empirical material of thousands of documents, and that it would be infeasible to study this amount of texts from traditional text perspectives. I further presume it is due to this seemingly

---

⁷ In Danish: Folkekirkens Uddannelses- og Videnscenter.
inability to process such material that we have so far not really noticed or considered the knowledge potential regarding religion and contemporary culture that sermons at large scale hold. I do therefore not suppose it seems particular radical to claim that sermons comprise a vast collective material – rather, it is the epistemological and methodological implications following the claim that may appear so.

In line with these considerations, the aim of this PhD project is threefold. The first aim is to establish a coherent and well-structured collection of contemporary ELCD sermons in order to solve the lack of empirical access to the collective nature of sermons. The second aim is an epistemological and ontological endeavour of presenting a horizon of inquiry to this type of material. This endeavour entails commencing a process of comprehending sermons as a collective production that is intertextually related to each other and to religious and cultural fields. The third aim is a methodological venture of outlining an analytical framework to study this material based in the presented horizon of inquiry and with the intertextual aspect of sermons as point of departure.

These aims presupposes each other, and they ought to demonstrate that the text material has been at the centre throughout every aspect of my research: the epistemological, ontological and methodological considerations are rooted in the existence of this material. This matter is also why, as already stated, I will not be able to or pretend to exhaust the analytical potential of this text collection, as my project has required entering a field of research that has not really been established epistemologically or methodologically so far.

I have found my way into this research arena through currents from different disciplines, which has guided me in deriving a horizon of inquiry. One of these comes from the sociology of religion and more precisely from the trend known as lived religion or everyday religion. This trend represents a cultural turn in the study of religion, where attention has been increasingly towards how cultural contexts influence the unfolding of religious phenomena. In this way, religion is approached as phenomena interacting with and responding to cultural and societal surroundings. Another current runs from the empirical turn in homiletics, which has provided descriptive approaches to study preaching practices. Homiletics is a theological discipline that has traditionally been normative with a focus firstly on what preaching ought to be. With the empirical turn, a focus has arisen on how preaching and sermons unfold in practice, which means that
Collective Testimonies to Christianity and Time

descriptive analyses are becoming more noticeable in homiletics. The last current runs from the broad discipline known as digital humanities. This discipline has prompted computationally informed methodologies for humanities research, which have enabled detections and studies of patterns in large text collections. However, digital humanities is not merely a methodological toolbox, but a discipline that increasingly contributes with epistemological considerations of cultural phenomena when we approach this at large scale.

My analytical aim is obtained through the application of computational tools. In line with the considerations above, the applied methodologies entails formulating research questions by taking the shape of the text material and the appertaining epistemological considerations into account. The methodological ambition of my project is therefore to study the material on its own terms as far as possible. In this endeavour, I investigate how sermons’ dialogue with each other in the text collection reveals a collective dialogue with Christianity and cultural codes in society. Central analytical concepts in this effort are intertextuality, distant reading and close reading. I have on the basis of these concepts sought out components in the text collection that are carriers of cultural information, and as such establish cultural and religious representations in sermons. I have with this method strived to approach cultural and religious codes as anchored in concrete textual components.

With the encounter between disciplines, this PhD project contributes to outlining a research arena centering on the collective and textual dimensions of Christian sermons. My particular project takes its point of departure in sermon practices in the ELCD, but I believe that the theoretical and methodological framework of the project can be adjusted to target sermon practices in other national contexts and in other Christian traditions. However, the individual outcomes of my research contribute to ongoing studies of contemporary Christianity in Denmark. The corpus I have assembled of 11,955 sermons will in itself be a lasting research resource for continuous studies of religiosity in the ELCD from the point of view of religious specialists. My approach to studying the corpus follows current research objectives, focusing on the ELCD as a cultural actor in society and the practices that unfold within or revolves around the church. These objectives comprise a wide-ranging field of interest, concerning aspects such as the funerary culture in Denmark (Kjærsgaard 2017), local churches’ response to migrants (Petersen 2019), the roles of pastors with distinctive tasks (Kühle & Christensen 2015), ideas of
“everyday” Christianity among cancer patients (Johanessen-Henry 2015) and the social and empirical realities of Nordic churches (Schmidt 2020). This tendency follows sociologist of religion Nancy Ammerman’s approach to everyday religion, where she argues it is important to investigate the expression of everyday religion within institutional religious frameworks (Ammerman 2016).

As a project where the theoretical object is sermons, I further see my project contributing to the empirical turn in homiletics, by outlining new strategies to study sermons as texts at large scale. Other studies have enhanced the intertextuality aspect of sermons, however, often focusing on the use and role of the concrete biblical texts expounded in sermons (Malmström 2015). I expand in my project the intertextuality aspect to include religious and cultural codes more broadly. Moreover, I see my project as a contribution to the so-called second wave of digital text scholarship, since the digital technology in this project serves to outline the epistemological framework, as the technology allows us to reconsider the nature of sermons. I do not contribute with analytical models or tools novel to the broad field of digital humanities, but I invite already existing digital resources into the study of sermons. I have not myself been the technological expert in the studies included in this dissertation, but all of my analyses are conducted in close cooperation with skilled programmers at the Centre for Humanities Computing Aarhus at Aarhus University. This cooperation has facilitated a creative and innovative environment that has been essential for the outcome of my PhD project.

Outline of the dissertation

This dissertation is not a traditional monograph, since I present my research outcomes in the form of a research product (the sermon corpus) and three individual research articles. The first part of the dissertation is dedicated to presenting the theoretical and methodological framework of the PhD project as a whole, while the latter part includes a presentation of the sermon corpus that holds the empirical outcome of the project and the articles that hold the analytical outcomes of the project. While the articles are composed under consideration of the overall framework of the project, they appear as independent studies, as they target different journals and audiences. This means that the dissertation
as a whole will repeat some content more than once. The independent nature of the articles also means that the chapters in which they occur (chapters 6, 7, and 8) hold their own bibliographies according to the journals’ styleguides, while the references in the remaining chapters are all included by the end of the dissertation. In the following, I will briefly present the remaining chapters of the dissertation.

Chapter 2 is a state-of-the-art chapter in which I present the currents from lived religion research, the empirical turn in homiletics and the digital humanities that have informed the epistemological considerations behind the PhD project. I enhance aspects of these disciplines individually, while I combine them through cultural and textual perspectives from Courtney Bender, Marjorie Garber, Paul Ricoeur, Julia Kristeva, Franco Moretti and Andrew Piper.

Chapter 3 is a presentation of the methodological framework of the project. First, I present my methods for data collection, which are informed by an opportunistic sampling approach in order to obtain access to a comprehensive collection of ELCD sermons. Secondly, I present the analytical outline of the PhD project. This outline is deeply rooted in the theoretical background, as I operationalise key analytical concepts: *intertextuality*, *topological reading*, *distant reading* and *close reading*.

Chapter 4 is a product report in which I present the sermon corpus of 11,955 sermons and the process of transforming the sampled sermons into a coherent and well-structured text corpus. This process has been conducted manually in order to ensure consistency in the material. In this endeavour, perspectives and techniques from the discipline corpus linguistics have informed the process. The corpus itself is included through a link to an online data-repository.

Chapter 5 is an introduction to the three articles of the dissertation, where I relate each article to the theoretical and methodological framework of the project. The purpose of the chapter is to provide an intermediate stage behind the PhD project’s horizon of inquiry and analytical cases.

Chapter 6 holds the article entitled “Sermons as data: Introducing a corpus of 11,955 Danish sermons”. The article introduces the ELCD as a public actor and sermons as

---

8 For example, the ELCD will be presented as research context in each article corresponding to the presentation already given in this chapter, and different aspects of the overall theoretical and methodological framework of the dissertation will recur in the articles.
access to the public voice of the church. The first part of the article is devoted to presenting the corpus in terms of sampling, metadata and corpus statistics. The second part is a case study of gender representations in the corpus, where my co-authors and I investigate how female and male pastors respectively construct social worlds in sermons. Bibliographic information occur in footnotes.

Chapter 7 consists of an article called “Unveiling the character gallery of sermons – A social network analysis of 11,955 Danish sermons”. Here, my co-authors and I study the representation of named entities – characters – in the corpus, as we regard these entities as access to uncovering a broad scope of the biblical and cultural domains pastors navigate between when they compose their sermons. Through character networks, we explore how various domains meet and reconfigure in the sermon corpus. Citations are included in parentheses (Author-Date), and references occur by the end of the chapter9.

Chapter 8 encompasses the last article entitled “What is love? – A study of thematic constructions in 11,955 Danish sermons”. In this article, my co-author and I investigate the role of theological discourses in emergent thematic fields in the corpus. We explore the overall thematic framework of the corpus, the construction of specific theological concepts (love and sin) and the relationship between thematic constructions in the corpus and thematic content in the liturgical texts that pastors expound in sermons. Bibliographic information are included in footnotes.

Chapter 9 is the concluding chapter, where I summarise the outcomes of the PhD project and discuss the analytical findings in the three articles.

---

9 Due to the blind peer-review process for this article, it has not been possible to upload code and data in an online data-repository. Therefore, code and data can be accessed upon request to supervisor Kirstine Helboe Johansen (kp@cas.au.dk).
Chapter 2

- The collective and intertextual nature of sermons

At its core, this project is empirically driven. My research therefore takes its point of departure in sermons as a theoretical object that can be accessed and analysed empirically. In defining sermons as theoretical object, it has been vital to consider a range of sermon dimensions: religious dimensions, cultural dimensions, text dimensions and proportionality dimensions. I have, however, not been able to take all dimensions under consideration from the point of view of one discipline, and therefore, this project relies on an interdisciplinary approach informed by three turns in recent scholarship: the cultural turn in the study of religion, the empirical turn in homiletics and the computational turn in the humanities. Through perspectives from lived religion research, empirical homiletics and digital humanities combined with concrete culture and text theories, I have arrived at an understanding of sermons as a collective and intertextual text production.

Sermons as a living practice

Around the turn of the millennium, studies of lived religion set forth new perspectives for understanding and defining religious phenomena in the study of religion. Lived religion has largely been defined by researchers in the sociological study of religion, who problematised some of the scholarly implications brought about by the secularisation paradigm in Western societies. With the secularised view on society as segmented into different domains, religion became a domain separate from other parts of culture and public life. Secularisation seemed apparent on a macro level, as Christian majority churches lost cultural and societal influence. This development affected approaches to study contemporary religion, as religion seemed to be a phenomenon to be studied in terms of a macro context and within religious institutional frameworks. Under the concept of lived religion, sociologists of religion argued that religiosity does not unfold in terms only of official teachings and practices within institutional frameworks. Instead, lived religion implied attention to the religious practices of individuals that unfold outside the
context of traditional religion, and to the ways in which they construct beliefs and practices by themselves. From this perspective, Robert Orsi argued it was important to investigate religion as a phenomenon interacting with immediate surroundings. He described lived religion as: “religious practice and imagination in ongoing, dynamic relation with the realities and structures of everyday life in particular times and places” (Orsi 2002, xxxi). This meant revising our understanding of religion and (re-)positioning religion in culture and not separate from it. With this shift of perspective, religion was emphasised as a context sensitive phenomenon, influenced by and responding to surroundings. Sociologist of religion Meredith McGuire emphasised the individual level of practitioners. Through interviews, she became aware of how differently worshippers of the same religious community described their religiosity even though they professed to the same faith (McGuire 2008, 3). She thus proposed a turn to study religion “as an ever-changing, multi-faceted, often messy – even contradictory – amalgam of beliefs and practices that are not necessarily those religious institutions consider important” (ibid., 4). In line with this view, McGuire distinguishes between official religion and nonofficial religion. She characterises official religion as institutional specialisation with clearly defined doctrines, world view and practices, where organisational programs are promoted (ibid. 2002, 99). In contrast, she explains nonofficial religion as.

(…) a set of religious and quasi-religious beliefs and practices that is not accepted, recognized, or controlled by official religious groups. Whereas official religion is relatively organized and coherent, nonofficial religion includes an assortment of unorganized, inconsistent, heterogeneous, and changeable sets of beliefs and customs (ibid., 113).

The characteristics of nonofficial religion corresponds to the objectives of lived religion, whereby McGuire clearly discerns lived religion from institutional contexts. McGuire’s focus on religion at the individual level and Orsi’s emphasis on religion as part of surrounding culture imply an understanding of religion as a constructive and emergent phenomenon. In this way, religious symbols and practices are not stable and unequivocal but consist of the content that individuals provide through their engagement with these idioms. Religion is thus viewed as a product of individuals’ negotiations with beliefs, practices and culture that happen situationally. In order to capture the discursive dynamics and practices through which practitioners construct their lifeworlds and place religious
content in it, lived religion research has entailed a turn to interviews and ethnographic work as central methodological instruments.

Overall, a turn to practice has been central for lived religion research. Sociologist Courtney Bender has distinguished between two ways of understanding religion as practice – as religious practices or as practicing religion. The former category resembles an attention to religion as a separate domain, where religious practices are characterised as highly scripted, repetitive and associated with specific situations, times or places – acts that are encapsulated within a domain clearly discernible as ‘religious’. They are a type of self-referential acts, pointing to the religious context they are part of and reproductive of. These practices would belong to McGuire’s category of official religion. Against this description, Bender’s category of practicing religion resonates with the lived religion paradigm, as she speaks in favour of an approach to study religious phenomena outside contexts – or in crosscutting contexts – of where we already expect religion to unfold. Bender argues that research needs approaches to study contexts where religion emerges sporadically and in connection with other aspects of culture. From this perspective, religious phenomena are not necessarily separate or alien to other cultural domains and settings. Instead, religion is perceived as a hybrid, dynamic and interacting phenomenon embedded in culture (Bender 2011).

Sociologist of religion Nancy Ammerman shares the overall understanding of lived religion and the emphasis on religion as emergent and context sensitive. However, she is sceptical of a too narrow focus on the individual level. She has introduced the term everyday religion and argued that attention should be given to unconventional and traditional religious forms alike, since majority churches still hold a significant role in many Western societies (Ammerman 2007, 5-8). More concretely, she has problematised definitions of lived religion that depend on a binary opposition to a traditional form of religiosity such as McGuire’s categorisation. With binary categorisations, Ammerman claims that lived religion is defined first and foremost in terms of what it is not. However, if lived religion is truly about what people do, then the organisational context of religion should not be discarded by default. In many European countries, such as Denmark, churches are still affiliated with the state and in some way interwoven in the everyday life of citizens – for example in relation to life ceremonies (ibid. 2016). Ammermann argues:
We have learned a great deal by intentionally looking beyond institutions, beyond elites, and beyond beliefs, but we may learn more by putting those “outside the box” insights into dialogue with a similar set of questions addressed to the settings in which societies have institutionalized religion (ibid., 93).

It is Ammerman’s conviction that it is important to follow developments within religious institutions and to explore practices within collective settings from a lived or everyday religion perspective. Sermons in the Evangelical Lutheran Church in Denmark present a pertinent case for this endeavour. While inscribed in a liturgical order and in this way performed on a regular basis and following specific prescriptions, ELCD sermons hold distinct practicing features in line with Bender’s category. The content of sermons are always formulated anew and addressed to a congregation at a specific time in history. Biblical narratives, Christian tradition and doctrinal teachings are consequently expounded by pastors from a contemporary point of view and as such conceived in culture. Whether or not pastors address contemporary issues explicitly, sermons are discursive products emergent in time. Therefore, sermons represent a meeting between religious tradition and historical cultural contexts.

Marjorie Garber, scholar in English and visual and environmental studies, has considered the ways in which cultural productions give access to prevailing cultural discourses and narratives. She describes such discourses as symptoms of culture and defines a symptom as:

(…) a kind of code, a way in which a body – or culture – signals something that lies beneath or within. For the human psyche, a symptom is precisely that: a coded or ciphered message, a withheld narrative performed by the body (Garber 1998, 3).

In her theory, Garber emphasises that symptoms of culture are historically contingent. As signs, they have no stable meaning, but are given meaning and function based on the context they occur in – this in contrast to the symbol, which she describes as a sign with a timeless content. She demonstrates this difference under reference to literary works. If understood as symbol, the work is thought to hold some universal truths, whereas

---

10 I will prospectively use the term “lived religion” to refer to an overall theoretical background in which I also include Ammerman’s concept “everyday religion”.

literature as symptom is emphasised as a product of its own time reflecting a contingent cultural climate\textsuperscript{11} (ibid., 7). Similarly, sermons can be grasped in terms of their symbolic as well as symptomatic content. With the former, one would emphasise their ability to claim and connect to a universal truth of the religion; with the latter, attention is brought to the sermon as a window to the cultural codes of its time. We can combine this perspective with Ammerman’s view that institutional forms of religion do not only hold stabilised doctrines and practices – they can also be studied in terms of change and dynamics caused by a variety of surrounding factors. The meeting between liturgical prescriptions and a contemporary point of view means that sermons hold the potential of stabilisations as well as destabilisations of religious traditions. Pastors may represent some aspects of the symbolic system over others, and their engagement with religious idioms may disclose some cultural discourses over others – discourses that may renew or uphold religious traditions. With the symptomatic and lived religion understanding of sermons, we can perceive sermons as interactive configurations of Christianity and time. With this perspective, we are able to take the religious and cultural dimension of sermons into account from a general practicing point of view. However, in order to apply this understanding to sermons, it is pertinent to consider sermons in terms of its own genre and immediate role in the church service. Therefore, I will in the following introduce sermon perspectives from the homiletical discipline.

\textbf{From a performative to a textual understanding of sermons}

Sermon studies have traditionally belonged within the theological discipline homiletics. It is a field that is focused both on the preparation of sermons and on the performance in itself, namely sermons as preaching. While I am not a homiletician myself, recent homiletical developments and endeavours have in many ways paved the way for this project. Therefore, the field of homiletics and its recent directions are pivotal in consolidating my approach to sermons.

\footnote{Garber’s definition of symptoms of culture implies an approach to uncover hidden power structures in societies. I do not emphasise this aspect in my project, but instead, I use the symptom as a descriptive tool to emphasise compounded discourses represented through cultural codes in sermons.}
The empirical turn in homiletics

In the last decades of the preceding century, an empirical turn occurred in homiletics. This development was driven by newfound considerations of sermons as communication in German and North American contexts. These considerations differed from the *kerygmatic model* that entailed a clear emphasis on preaching as an event, where the salvific narrative of scripture is proclaimed to the congregation – an idea originating from the theology of Karl Barth and Rudolph Bultmann. The kerygmatic model is still influential in terms of sermon theology. However, in the 1970’s in Germany this approach was problematised, as it did not seem to work well in terms of communication, since the kerygmatic approach tended to privilege a focus on the content in sermons as expressed by the preacher, and not as experienced by the listener. This meant that the focus on gospel exposition came to emphasise representation of scripture over personal experience of the word of God. New observations from social sciences demonstrated that such approaches were based on naïve assumptions of communication, namely as a monological conveyance of content from an active speaker to a passive listener (Immink 2004, 93-97). Theologian Marlene Ringgaard Lorensen has described how this communicative strategy presupposes the transfer model, which was originally developed by mathematicians Claude Shannon and Warren Weaver in 1948, where the purpose is to transmit a message as uninterrupted as possible (Lorensen 2013, 69-70). Based on a large-scale survey among congregants in Hannover, theologian Karl-Fritz Daiber demonstrated that listeners do not passively absorb the message of preaching, but actively interpret what they hear from the preacher (Daiber 1980/1983). These insights from empirical studies of preaching brought about an increased focus on the subjective experience of sermons from the perspective of the listener. In the North American context, this perspective was essential to the movement called New Homiletics. Proponents of this movement held that the sermon is a medium that is not just kerygmatic in its message, but is also endowed with a transformative effect upon congregants. Therefore, the purpose of preaching should not only be for the preacher to transmit insights from the gospels to the listeners, but to facilitate a newfound experience for the congregants through the act of preaching. For this purpose, theologian David Buttrick promoted the use of symbols and metaphors as elements to create an affective experience for listeners, and in this way enable a human understanding of the kerygmatic event (Buttrick 1987, 115). Correspondingly, homiletician Fred Craddock presented an inductive preaching
strategy, where preachers were endorsed to incorporate anecdotes and narratives from contemporary situations so that congregants could relate personally to the biblical texts. This approach to preaching meant that sermons relied on an identification between congregants and preacher (Craddock 1971, 54-59). The New Homiletic’s approach to preaching has since then been targeted by a postmodern criticism, arguing that a pastor’s dialogue with congregants should be based on the awareness of the differences between the communicating parties, because the assumption of identification ends up undermining the subjectivity of the listener. In this criticism, the presumably dialogical ideal of New Homiletics ends up disguising an underlying monological principle, where the preacher remains in control of the sermonic plot, whereby the listeners de facto are not invited to draw their own conclusions (Lorensen 2013, 87-88). This scepticism towards the consequences of the New Homiletic’s preaching has been a motivation behind the newer branch of the empirical turn in homiletics called Other-wise Preaching. In this movement, the asymmetry between the communicating parties in preaching is acknowledged, and so the preacher ought to consider the congregants as co-authors, who possess other perspectives than the preacher, which are crucial for what they hear and interpret from the pastors’ sermon. Therefore, the purpose of preaching should not be to transform the listener, but as a preacher to allow oneself to be moved in new directions, whereby a mutual enrichment of all parties can occur in the preaching situation (ibid., 89-90). American theologian John McClure has concretised this idea in the roundtable pulpit, by encouraging pastors to invite congregants in the week before the service to share their perspectives on the pericope for the forthcoming sermon, in order to allow other voices than the preacher’s to be heard in the sermon (McClure 1995).

The kerygmatic model, New Homiletics and Other-wise Preaching demonstrate different accentuations of the act of preaching. With the kerygmatic model, we find a focus on preaching as proclamation of the word of God, and an emphasis on what is being said rather than heard. New Homiletics provides a listener-oriented focus by focusing not on the kerygma as extraordinary but as experiential, and thus as universally human. With Other-wise preaching that understanding of the human being becomes too abstract, since human beings are viewed as multifarious that must be addressed in consideration of their otherness.

For a long time, homiletics has not been prominent in Denmark as an academic discipline. Rather, pastoral training in preaching has been a part of the short practical course that
theological candidates follow when they have obtained their degree at the university. However, in recent decades, homiletical dissertations and works have emerged with an emphasis on preaching as communication, and with a listener-oriented focus (Thøissen 2005; Nielsen 2004; Lorensen 2012). Most recently, the PhD dissertation of homiletician and now bishop Marianne Gaarden has combined the perspectives from Other-wise Preaching with empirical research on preaching in the ELCD. Through interviews with congregants and pastors, Gaarden studies preaching as a performative act, where the sermon creates a third room in-between pastor and congregants. In this understanding, the sermon does not just correspond to the words as uttered by the preacher, but just as much to the reflections that arise in the listener. Congregants become co-authors of the sermon, and the sermon is presented as an interactive phenomenon in-between speaker and listener (Gaarden 2014).

Overall, the empirical turn in homiletics has implied a break with a normative and narrow focus on doctrinal standards for preaching and introduced the discipline of homiletics for descriptive approaches to explore how preaching practices actually unfold. However, the empirical approach has primarily been adopted in the attempt to shift focus from sermons as conveyed by the preacher (the kerygmatic model) to sermons as acquired by congregants (New Homiletics and Other-wise Homiletics). This shift has provided an increasing awareness of the dynamic and subjective dimensions of preaching. Therefore, the homiletical interest in sermons tends to be in terms of their communicative context in church.

**The sermon manuscript: From singular text to intertext**

Instead of capturing the dialogue between preacher and congregants, I emphasise in my perspective on sermons - in line with my previous argument - that sermons are religiocultural documents which encapsulate pastors’ dialogue with time and Christianity. Therefore, I focus on the documentary potential of the sermon – not the sermon as intended by the preacher, nor as heard by the listener, but as the written document produced prior to the service and which remains after it. This attention to sermon documents draws on aspects of Paul Ricoeur’s text theory from his work *Interpretation theory: Discourse and the surplus of meaning* from 1976. Ricoeur distinguishes between oral and written forms of communication, and he emphasises that the document as artefact
provides a fixation of discourse through inscription (Ricoeur 1976, 26-29). Meanwhile, this inscription of discourse is not equal to the intentional message by the sender. The medium, which inserts a layer between sender and receiver, as compared to the oral speech situation, obtains a semantic authority:

Inscription becomes synonymous with the semantic autonomy of the text, which results from the disconnection of the mental intention of the author from the verbal meaning of the text, of what the author meant and what the text means. The text’s career escapes the finite horizon lived by its author. What the text means now matters more than what the author meant when he wrote it (ibid., 29-30).

Even though Ricoeur still characterises the text in terms of a communicative context, he indicates that the text becomes a kind of objectivised reality not equal to that of the sender, nor that of the intended receiver. The text is free to exist outside of its intended communicative situation. Ricoeur argues that when this occurs, the absolute of a here and now is dissolved, which enables the text to meet readers in an indeterminate time. Ricoeur further argues that the implications of writing is the productive potentials of the act itself. When a matter is given form – here through inscription – it is also being shaped and produced in other ways than the oral utterance. Therefore, the text itself presents a coherent world for the reader, which is constituted of the references and the relations in the text. When the text has left the communicative situation, the discourse of the text must be understood as a projection of a world and of being in the world, which can only be accessed through the process of reading and interpretation (ibid., 35-37).

Ricoeur’s perspective emphasises that the text – the medium – gives access to “something else” than the immediate communicative situation it is intended for. In the production of a sermon document, the pastor also projects a matter that comes to exist objectively and apart from the preacher through the document. Through the medium, the sermon is not a subjective reality of the preacher, nor of the congregant, and the sermon document does not exist only in and through the communicative situation.

Like Ricoeur, literary critic and philosopher Julia Kristeva has opposed the view that the subjective voice of the author can be inferred from literary texts. However, her opposition against this subjective dimension in texts is different than Ricoeur’s, which appears from her theory of intertextuality. In her work Desire in Language: A semiotic approach to
literature and art, she argues that a text can never stand alone as a singular unified object – it is always permeated with utterances and discourses pre-existent of the individual text and adopted from what she calls the cultural or social text. The cultural text may be understood as other concrete texts or as a larger and abstract cultural context. As such, the individual text is always also an assemblage of cultural discourse. Kristeva therefore argues that a text communicates on two levels simultaneously. On a horizontal axis, the inscribed words belong to writer and reader, corresponding to the dimension within which Ricoeur attends to texts. Meanwhile, in Kristeva’s expounding, the text also has a vertical axis through which the text is communicating with past texts or with the abstract cultural text. As such, she juxtaposes the intersubjective dimension (the horizontal axis) with an intertextual dimension (the vertical axis) (Kristeva 1980, 36-37; 65-66). In her later work Revolution in poetic language, Kristeva uses instead the term phenotext to define the text that appears as a coherent unit represented through a grammatical logical language system, and she applies the term genotext to describe a dimension of the text which disrupts the sense of internal coherence (ibid. 1984, 86-87). However, the genotext does not necessarily have its own language system, and thus, it permeates the text through the language of the phenotext. Kristeva describes this intertextual dynamic as a transposition of one sign system into another where it is stabilized anew. Thus, through the new text, the cultural text gains a new place and context, and in this, an “otherness” inhabits the textual object through the process of transposition (ibid., 59-60).

For the context of sermons, Ricoeur’s text theory posits a perspective for grasping the sermon as a text object that is not equivalent to its performative context in church. However, when he describes the text in terms of its semantic autonomy in order to make the transition from oral to written object, he comes to define the text solely as a phenotext. With Kristeva, we can perceive the sermon document in terms of its vertical axis – as a text in an intertextual relationship to the cultural text. Kristeva thus provides a perspective to attend to the interaction between sermon, Christianity and culture – as I have explicated under reference to Bender and Garber – in sermon documents. Further, Kristeva’s concept of intertextuality concept the same idea of dialogue, which runs through Other-wise Homiletics. In both contexts, a dialogue implies a co-existence of difference without a dissolving of differing horizons into a synthesis as in Hegelian dialectics. However, where the dialogue in Other-wise Homiletics unfolds in an inter-subjective relationship,
it unfolds in Kristeva’s text theory as an intertextual relationship. Therefore, with Kristeva, we can attend to otherness as latent in the textual dimensions of sermons.

The recent developments within homiletics have opened a field for empirical studies of sermons, which brings attention to plurality and relational aspects of sermons. With Kristeva’s idea of intertextuality, we can focus on these aspects in sermons as text phenomena. However, relationality as a text feature should not only imply a relationship between singular documents and cultural context. Sermons are the outcome of a shared practice, where pastors simultaneously and continuously compose their manuscripts for church services, which means that as texts, sermons are also a collective phenomenon. Therefore, the intertextual dimension of sermons ought to concern their relationship to each other as well as their relationship to cultural contexts. In order to grasp the collective aspect of sermons further, I will turn to the field of digital humanities, which has provided new perspectives and tools for attending to cultural phenomena in terms of scale and complexity.

**Digital humanities: Theoretical and analytical developments**

Digital humanities is a term that has come to cover a wide scope of research. The common thread in digital humanities projects is the integration of computational resources in humanities research. However, the role of technology and computer power in research projects has developed over time. Jeffrey Schnapp and Todd Presner proposed in their *Digital Humanities Manifesto 2.0* in 2009 to understand these changes in terms of two waves. In their description, the first wave of digital humanities emerged around the millennium and was strongly influenced by a quantitative approach to searching and extracting data from large digitised collections. In this endeavour, computers were predominantly considered as toolboxes that could solve this task faster and more efficiently than individual researchers could (Presner & Schnapp 2009). This meant that it was possible to cope with enormous amounts of texts, which allowed a greater breadth of analysis, and it was possible to gain insights to patterns in the material that would not be discernible with the naked eye (Evans & Rees 2012, 23). With the second wave, ‘the digital’ was not just emphasised as providing new technological solutions to old problems in the humanities. Rather, the digital was regarded as a generative component in contemporary culture. While the focus of the first wave in particular was on digitising
and assembling text collections from print, the second wave broadened the scope with an attention to cultural phenomena that unfold digitally. Presner and Schnapp demonstrated that this development has raised new paradigms for humanities researchers, as the digital directly influences the general object in focus for the humanities: culture and the human condition. They describe the second wave as:

(...) qualitative, interpretive, experiential, emotive, generative in character. It harnesses digital toolkits in the service of the Humanities’ core strengths: attention to complexity, medium specificity, historical context, analytical depth, critique and interpretation (Schnapp & Presner 2009, 2).

Leighton Evans and Sian Rees elaborate that with the second wave, a distinction is made between the digital humanities as boiled down to a range of quantitative tools and methods, and the digital humanities as a paradigm where computation provides new approaches to “accessing, interpreting, and reporting the world itself” (Evans & Rees 2012, 29). Computational resources are not neutral toolboxes, and their role in the research process therefore prompts a higher level of reflection than in the first wave, which means that the use of computational resources in humanities research facilitates new epistemological considerations: as new types and amounts of cultural data become possible to research, new types of perspectives and research questions emerge (ibid., 22-23). In Evans and Rees’ argument, technological developments have always affected the human condition. They refer to Nietzsche’s use of the typewriter, which marked a change in style compared to his earlier works. However, his attitudes and conclusions did not seem to change due to typewriting, but Nietzsche himself found that the typewriter took an important part in shaping his thoughts. They argue that similarly to the existence of typewriters at Nietzsche’s time, digital technology is a concrete and existing thing in the world today. As such, digital technology is not a neutral access to the world, but if we are aware of this, technology can be used constructively to represent our research objects in new ways, which allows for different perspectives on cultural phenomena. With their description of the second wave, Evans and Rees emphasise that digital humanities is not in opposition to the humanities, but should be considered an integral part of it, as the new wave provides new perspectives through which to continuously interpret the world and seek insights to cultural phenomena (ibid., 34-38).
Sociologist of religion Tim Hutchings has argued that the study of religion has not been very notable in digital humanities research so far. However, he identifies three modes of engagement with digital technology in religious studies: digital technology as tool, as object of study and as an expressive medium (Hutchings 2015, 284). The first mode, the “tool mode”, entails building large digital archives of primary sources to religion and using digital technology as a tool to analyse documents. The second mode has entailed a focus on religious practitioners’ use of digital media, in particular social media, to track trends in contemporary religion and to study online platforms’ influence on religious content. The third mode involves the use of digital technology as a strategy for researchers in the study of religion to communicate their research to other researchers and the broader public. Hutchings finds that mode two and three clearly follow the trajectories of Schnapp and Presner’s second wave. Meanwhile, he argues that the “tool mode” began as a first wave approach to religious sources, corresponding to Presner and Schnapp’s description, but has since then come to follow the intents of second wave research. The impressive archives made during the first wave has contributed to the development of new, innovative and dynamic reading strategies for digital text studies (ibid., 284-286).

It is this development, I ascribe my research to. My ambition has been to assemble contemporary sermon manuscripts for a digitised corpus – manuscripts that are digitally born, but not intended for a life on digital platforms – and using innovative reading strategies compliant with second wave objectives.

_Distant reading and collective text productions_

Literary scholar Franco Moretti has been renowned in the digital humanities for the concept _distant reading_ – an attempt to explicate what it means to study texts through the aid of computational resources, when working with large literary corpora. This approach entails applying computational and quantitative visualisation techniques to represent large collections of texts in different ways – for example in the forms of graphs, maps or trees (Moretti 2007). With distant reading, Moretti problematised that literary studies had taken close readings of literature for granted, with attention to language, detail and depth in individual works. In cases such as the New Criticism, close readings were based on the

12 An example would be The Religious Studies Project that conveys various material concerning contemporary religion, such as podcasts, essays and conference reports (https://www.religiousstudiesproject.com/).
assumption that a work can be self-explanatory, holding an organic unity between form and content. This assumption implied a focus on the intrinsic features of a text and how they combined present a coherent world. In Moretti’s view, such text assumptions tend to overlook how individual works take part in a wide historical context. He finds the limitations of close reading to be that the method cannot cover the comprehensive extension of the literary field in historical periods. He argues that close reading has prompted a canonised approach to literary history, where few works are being read over and over again, while the extensive historical productions remain “the great unread” (ibid. 2000). He therefore favours an approach of reading many works distantly, instead of reading a few works closely. This entails applying quantitative measures to the study of literature. With distant reading, he proposes to remove focus from individual texts and instead attend to texts as products of a given historical context. Individual texts thus form parts of a collective system consisting of a broad production of works from the corresponding historical period. Therefore, the individual text is of less interest compared to the entire collective production it is part of. He argues:

(…) a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn’t the sum of individual cases: it’s a collective system, that should be grasped as such, as a whole (ibid. 2007, 3-4).

Similarly to the lived religion perspective, Moretti enhances an attention to the cultural and temporal context a study object, here literary texts, are part of. Meanwhile, his premise is an attention to the collective qualities over the singular and unique qualities of the objects in focus. Therefore, Moretti’s text perspective provides another layer to Ricoeur’s and Kristeva’s text theories. Ricoeur focuses on a text’s semantic authority – a perspective that in Moretti’s explanation would privilege a close reading approach. With Moretti we may instead emphasise semantic coherences in a text system. As such, attention to text structures is still immensely important, however, not within individual texts, but within text collections. Yet, we do not need to think of such systems as closed self-explanatory systems, but instead as systems that are intertextually linked to cultural discourses in historical periods, in line with Kristeva’s text theory.

These combined perspectives are immensely relevant for sermon studies. As emphasised previously, the content of individual sermons are provided by individual pastors.
However, in the ELCD, the task itself – producing sermons – is not just an individual, but a shared practice, since pastors in parishes all over the country engage in this production on a weekly basis. This synchronicity situates the production of sermon manuscripts in a collective context. By combining Kristeva’s idea of intertextuality with Moretti’s text perspective, we can explore how cultural discourse is transposed not into singular sermon manuscripts, but into a collective text system. A text system, that has a distinct intertextual basis in the prescribed lectionaries in the liturgical year, a presumed intertextual basis in the bible and Christian tradition, but also an implicit intertextual basis in the cultural historical context in which the sermons are conceived. With this combined perspective, I argue that sermons are documents in-between different contexts, where pastors may incorporate various references and make significant encounters between theological and contemporary discourses.

Topological reading of cultural codes

Since Moretti introduced the concept, distant reading is often used as a generic term that has become almost synonymous to digital text studies. Andrew Piper has used the idea of topology to articulate what the act of reading entails when working computationally and quantitatively with digital corpora. Topological reading challenges traditional humanist ideals of reading, which Piper calls bibliographic reading and traces back to the church father Augustine. Bibliographic reading illustrates a process of individuation, where the reader through the act of reading becomes a transformed human being. Therefore, the bibliographic mode is oriented towards what is new and unique, and it is rooted in the materiality of the book through which the direction of reading is predefined: we read word by word, sentence by sentence, page by page. In contrast, topological reading is oriented towards multiplicities over singularities and redundancy over exclusivity (Piper 2013, 377). A literary topology is constituted of relations between text units, where the sentence is no longer the structuring principle. The topology’s basic units can be smaller than the sentence, such as letters or words, or larger structuring principles, such as genre (ibid., 378). This decomposition of the text transgresses the materiality of the text page or the book, which in turn facilitates attention to the relations of text units throughout a text collection. The relational understanding invites a numeric logic into text reading, whereby topological reading is not a distinct domain for literal sign systems. Instead, this
way of reading seeks out “the contiguities that run between literal and numerical reasoning” (ibid., 380). In the bibliographic mode, focus is on the literal system through which the reader will search for meaning. Piper illustrates how the content of a concept such as love in this logic can be determined through an infinite vocabulary – the logic is therefore that “x means y”, where love is something else than the sign itself. In topological reading, a term is understood through its relations to other terms. Piper uses as example that love in *The Sorrows of Young Werther* is 0.00065% of the full vocabulary, meaning that it is compared and contrasted to every other word in the work. The word is thus not described in terms of meaning, but in terms of likeness and difference to other words. Where bibliographic reading is oriented towards substitution (x means y), topological reading is preoccupied with succession and scale (x is more/less than y). This means that the topological text field is non-teleological – it dissolves the two dimensions of the text page and the three dimensions of the book. As such, Piper asks us to think instead of reading as multidimensional, where a text can have as many dimensions as it has words – each outlining different directions. In his argument, he does not see close reading and distant reading necessarily as oppositions, but as a reading spectrum that facilitates different focalisations (ibid., 380-383).

Another pivotal point of topological reading is attention to redundancy and repetition of word-sets. Here, Piper asserts that topological reading is still concerned with inferring meaning from texts: the basic principle is that recurrences of word relations condense meaning. Whatever the theoretical object in focus is (a book, a character, a concept), it is based in lexical regularities revealed as patterns across texts – patterns of how language repeats itself (ibid., 383; 394). Bibliographic reading is based on an immediacy, where one does not have to ponder every word to make sense of a sentence. Topological reading challenges this premise of reading and forces us to focus particularly on repetitive – and otherwise banal – text units and the complex relations, they partake in. By focusing on patterns across text collections, we can discover underlying structures and discourses that would not be discernible from a teleological oriented reading. In Piper’s words:

(…) topology draws our attention to a very different idea of meaning and the meaningful. It combines a sense of the latency of meaning, what a text says without saying it (that which is signified beyond the text) with a sense of the manifest nature of textual meaning, that a text says nothing more than what it says (…). In so doing, topology privileges the latency of the manifest (ibid., 394).
As such, text meaning is based on observations of complex relationships between linguistic phenomena, which uncover latent semantic structures that could not be discerned through a bibliographic reading. This approach complies to some extend with Marjorie Garber’s call for new approaches to reading culture – approaches to observe symptoms of culture. Inspired by Freud’s dream analysis, she seeks to look behind a culture’s conscious surface of rational thoughts and coherent narratives. She finds the cultural unconscious to be structured like a network of representations that may not be apparent on the surface. Codes can be found in the network, which consists of connections between seemingly disparate elements. Concretely, she favours strategies that can uncover a multiplicity of associations and linkages instead of identities and essences (Garber 1998, 9-10). This follows Piper’s argument that a sign should not necessarily be defined through its reference to an essentialist content (x means y), but instead in terms of its relationship to other signs. However, Garber strongly problematises the use of quantitative measurement of cultural phenomena. She finds that quantification reduces our ability to understand cultural complexities, because it is based on logical inferences and therefore fails to capture phenomena that is linked based on unconscious associations (ibid., 5-6). Meanwhile, topological reading demonstrates an approach where numeric as well as literal reasoning serves the same purpose, namely unveiling textual structures, by mapping complex relationships between linguistic phenomena. This approach discloses topological reading as a second wave approach to textual scholarship in the digital humanities. With the idea of topological reading, I find an approach that enables a study of sermons as a collective text production in which we can seek out cultural and religious codes represented as intertextual links throughout the text collection.

A collective text system in complex interactions

My theoretical point of departure for this project is in the convergence between lived religion, the empirical turn in homiletics and digital humanities that allows for a relational perspective on sermons. From a lived religion perspective, we can perceive sermons as phenomena emergent in time and in the intersection of Christianity and cultural contexts within the framework of a religious organisation. This means that sermons are bound to interact with and adopt a variety of different discourses. In the newer trend in homiletics
research, Other-wise Preaching, sermons are already targeted as interactive phenomena. Even though this interaction concerns the relationship between preacher and congregant, this development in the empirical turn in homiletics provides an understanding of sermons as context sensitive phenomena that can be observed and described. On this background, I argue that we can also target sermons as relational text phenomena, by combining the text perspectives of Paul Ricoeur and Julia Kristeva, and that we can target these text dimensions, while attending to the collective nature of sermons. The interactions between Christianity, cultural fields and time in sermons unfold collectively, since pastors on a regular, continuous and synchronous basis construct sermon documents that add up to a comprehensive text production. This allows us to think of sermons as collective documentations of Christianity and time – as textual imprints of the encounters between contexts that pastors produce, when writing their sermons. The digital humanities has enabled approaches, whereby we can attend to collective aspects of cultural phenomena through empirical studies. Therefore, with an approach similar to topological reading, we can map and interpret pastors’ interaction with different contexts, which can unveil traces of a collective voice – a voice that cannot be unfolded from a linear reading one sermon at a time, but must be uncovered from a multiplicity of sermons and the intertextual links they share. From this overall relational approach to sermons, my project is not firstly a study of what sermons are or what they symbolise – rather, it is a study of how sermons connect. To each other, to Christianity and to cultural contexts through textual codes.
Chapter 3

- Methodological Framework

The methodological choices I have made throughout the course of my project have been highly influential on its outcome. Of course, the results of any project depend upon the methods applied – that is almost a tautological statement. However, the theoretical background and my approach to sermons as a collective phenomenon make distinctive demands to the nature of the data-set in this project: it has to be a comprehensive and varied collection of contemporary sermons from the Evangelical Lutheran Church in Denmark and accessible for digital analyses. Striving to fulfil these demands has been time-consuming and taken a lot of effort, but the result is a well-structured text corpus of 11,955 Danish sermons from 95 ELCD pastors. The sampling process took place predominantly from February to June 2017, but was completely finalised by the end of the year, whereas cleaning, structuring and archiving the material as a text corpus was commenced in August 2017, but took place primarily from January to June 2018. Acquiring viable sermon data thus proceeded throughout a year and a half.

It has been vital and necessary to develop a methodological framework aimed specifically towards large-scale collection and analyses of sermon manuscripts. The methodological framework are thus deeply rooted in the theoretical perspectives on sermons presented in chapter 2. In this way, the focus on sermons as relational phenomena are essential on several levels: in regards to collecting sermons, which entails gathering otherwise disparate text objects stored privately by pastors; in regards to assembling a text corpus of the sermons, which entails cleaning and assigning metadata to the material, thus facilitating discernible intertextual links between the texts; and in regards to analysing the material, which entails extracting interrelating text components in the entire text collection.

Since methodological decisions have been a core aspect of the project, this chapter will repeat some of the content inherent in the articles in chapter 6, 7 and 8. In particular, the outline of the data collection below is to some extend a repetition of the outline in chapter 6. However, in this chapter, I will unfold the broader spectrum of complexities and considerations associated with sampling sermons. Meanwhile, I will only briefly touch
 Collective Testimonies to Christianity and Time

upon issues in regards to cleaning and annotating the received material, as the strategy
applied for creating a text corpus from the sampled material will be introduced in depth
in the product report in chapter 4.

Data collection

Background to sermon sampling

Data in the form of full text documents that are freely constructed by individuals has
traditionally been considered a qualitative data material in the social sciences. Qualitative
text studies normally imply analytical approaches where the data is read meticulously.
Therefore, acquiring the adequate amount of data tends not to be a considerable
methodological problem, since you normally need a rather small sample. In digital
humanities text studies, the situation is different. Here, access to large text collections is
a huge premise of scholarship, which is why different resources for accessing digital texts
have been developed. Web-scraping is a valuable tool for retrieving digitally-born texts
from for example social media platforms. Open access to digitized text collections online
are vital resources in particular for literary or historical scholarship13. In efforts of
providing open access, it has been more feasible to ensure access to historical text
collections as opposed to material from contemporary authors, where matters of copyright
make it difficult to enable free access.

So far, there has not been many attempts to collect contemporary sermons for large-scale
text studies. However, in the US, a recent Pew study analysed a sample of 49,719 sermons
(Pew Research Center 2019), and at THEOLAB at Heidelberg University, Germany, a
sample of 20,000 sermons was collected for a new homiletics project, focusing on
religious communication (Karcher 2020). These projects demonstrate some of the
difficulties in acquiring viable data for sermon analyses. The Pew project used a web-
scraping method in order to collect sermons from 6,431 church websites posted
throughout April and May 2019. In their sample, they juxtapose written sermons with
transcriptions of audio- or video-recorded sermons, and they allow other excerpts than
the actual sermon part of a text/recording to be included in the data, because the tools
applied cannot take these variations into account, when retrieving the material. Therefore,

13 For example the DraCor-platform, where collections of drama manuscripts from different historical
periods and countries are hosted (Fischer et al. 2019)
the sampled sermons end up with a considerable amount of uneven variance. The THEOLAB project also collected sermons online, not from church websites, but from eight online sermon databases in Germany. They have thus been able to rely on already existing archival resources, but the sampling criteria are not explicated well, which is problematic since the databases seem to have different features. In one database, only two distinctive pastors provide sermons, whereas in another database, pastors in general are free to upload sermons, as they like. Some of the databases host sermons from up to twenty years ago, and others only sermons written within the last couple of years. One of the databases targets a broad “online congregation”, whereas most of the databases target pastors seeking preaching inspiration. Further, there is a big discrepancy between the sermons retrieved from the databases – from one database 8850 sermons was retrieved, while from another only 200 sermons. It is therefore difficult to discern what aspects of sermon practices the material in full will be able to represent.

Similar to my project, the attempt in the Pew study and the THEOLAB project is to obtain a broad sample of sermons from a delimited national context. The clear advantages of the sampling strategies in the American and German projects are the researchers’ ability to acquire very large sermon collections rather easily by sampling online. This way, they have also been able to retrieve metadata automatically, since each sermon was already annotated for the web-formats.

It has so far not been a common practice in the ELCD for local parishes to publish church sermons online, or for other stakeholders to host sermon databases online. However, even if such online resources had been available, I find – as outlined above – some considerable limitations to the strategies applied, as you risk allowing for some variables that cannot be controlled adequately. The only sustainable way of obtaining a broad collection of ELCD sermons was in my opinion to receive them directly from the authors themselves. I chose consequently to sample data as personal documents, which meant that I had to ensure informed consent from every pastor – in return, the consents provided ownership of the material. Compared to web-scraping, this approach of receiving and managing the material as delivered by individual pastors is substantially slower, which is one of the reasons why my final data-set is considerably smaller than in the Pew study or the THEOLAB project. Another reason is that since the received material are personal

---

14 However, the project is still very new, and it is therefore possible that the project holders will make targeted sub-samples for specific analyses.
documents and not intended for reading, the texts shared no consistent structure and had no standardised metadata annotations. Therefore, I had to manually clean up and annotate the material. This process was time-consuming, but it ensured a consistency in the document structures so that no other text excerpts than the actual sermons were included in the final dataset. Further, the process gave me a chance to get an insight into the material prior to analyses. These aspects of sampling (retrieving personal consent) and cleaning the material (thorough inspection of the data) resembles qualitative data preparation strategies, while the datasets final corpus structure resembles a quantitative data format. Yet, by large, there has been no standard design to follow – such as interview design or participant observation – when it comes to collecting text documents that are outcomes of a living practice and stored in the private homes or offices of thousands of individuals. I have therefore designed the strategies used to acquire data specifically for my project.

*Sampling ELCD sermons: Access and phases*

A important aim and premise of this project has been to gain access to an otherwise inaccessible material. Even though sermons as delivered in church are public events, sermon documents are the property of individual pastors. When commencing the process of data collection in 2017, I therefore knew very little about the manners in which pastors store their sermons and about their willingness to share their manuscripts. In some parishes, local pastors publish their sermons online at the parish website after the service, but this is not a general or standard practice. Therefore, my first criteria was to sample sermons directly from pastors in a standard digital text format such as word-files. I chose to target sermons from 2011-2016, as I wanted a large text collection from a coherent and as recent as possible time span in order to ensure that the sermons belong to the same temporal context in line with Franco Moretti’s argument. Based on these criteria, I had to rely on pastors to invest time in sending me sermons. However, as I did not know of their concrete storing systems and their personal attachment to their sermons, I could not assume whether it would be feasible for them to provide material and whether they would

---

15 I will explain the quantitative structure in chapter 4.
16 The corona-crisis as of 2020 may have altered ELCD pastors’ use of digital technology and platforms today. The lock-down of churches throughout the spring of 2020 made many pastors record and stream services online. This conduct implies that sermons came to exist as digital phenomena and openly accessible (Folkekirkens Uddannelses- og Videnscenter 2020).
be motivated do so. Therefore, I sought to make the sampling strategy as accommodating as possible towards the pastors. This meant that I asked pastors to send via email sermons from 2011-2016 in the amount they individually found feasible.

I chose to address the supervising authorities in the ELCD, bishops and rural deans, and ask them to help circulate my request for sermons to the ELCD pastors under their supervision. This approach was chosen, since I was not able to obtain contact information for pastors without manually extracting it from all parish websites, and since I estimated that my request would be more noticeable in pastors’ inboxes if it was sent by ELCD authorities and as such signalled their approval. All bishops except for one contributed with a brief statement of support to my project as a whole. The statement signed by the bishops read:

Through the construction of a sermon database, Anne Agersnap’s PhD project will contribute with a nuanced understanding of sermons as interpretations of society and Christianity today. I would therefore highly recommend everyone to submit sermons for the project17.

I forwarded to all rural deans in the ELCD this statement and my invitation addressed to pastors, asking the deans to circulate the invitation in their deanery18. In my mail to the rural deans, I clarified that I did not expect them to impose pastors to submit sermons, but I hoped they would help bring pastors’ attention to the project and encourage them to provide sermons. In my address to the deans (co-signed by my supervisor Kirstine Helboe Johansen), I clarified this issue:

We hope you also find the project interesting and pertinent, and that you will help us introduce the project in your deanery. In this e-mail, we have attached an invitation for project participation addressed to ELCD-pastors in Denmark. In this, we provide a presentation of the PhD-project, a description of data management and an instruction for submitting mails for the project. It would be of great importance, if you would forward the attached invitation to the pastors in your deanery. Naturally, it is entirely up to pastors individually to decide whether they wish to participate in

17 The original wording in Danish: "Ved at opbygge en prædikendatabase vil Anne Agersnaps ph.d.-projekt bidrage med en nuanceret forståelse af prædikenen som samfunds- og kristendomsfortolkning i dag. Jeg vil derfor varmt anbefale alle at indsende prædikener til projektet."

18 The invitation addressed to ELCD-pastors are enclosed in Appendix I
the project. You are, however, very welcome to motivate the project yourself in your
mail to pastors and at gatherings in your deanery.

We would love to hear from you, and you are very welcome to contact us if you have
any questions or comments.19

A couple of rural deans never responded to my request and two rural deans replied they
would not forward my invitation. In these cases, I retrieved contact information for the
pastors in these deaneries online and contacted them directly along with the approval
statement of their bishop.

In this first phase of sampling, my aim was to extend my invitation to all pastors. Here, I
did not control for any demographic variables, as I allowed the pastors to be self-selected
participants for the project. However, I monitored how basic demographic variables were
dispersed (gender, diocese, deanery) in the material I received. The response rates were
comparatively higher for male pastors than female pastors. Therefore, in the second phase
of sampling, I approached individual pastors – female pastors in particular – in dioceses
and/or deaneries that were underrepresented comparative to the material received in the
first phase. This approach helped balance the received material on its own premises, but
not in terms of external representative criteria. Therefore, the sampling process has
overall been open-ended and based on an opportunistic sampling strategy. This was the
most appropriate strategy, considering that I had to rely on pastors to invest rather a lot
of time to find and submit material. Further, I could not assume how much material each
pastor would be able to submit, bearing in mind that some pastors wrote long sermons,
some wrote short sermons; that some pastors preached regularly, some did not; and that
some pastors were employed throughout the entire period, some were not. These factors
illustrate that sermon documents are an outcome of a lived practice that unfolds in terms
of a range of variances. These variances would be very difficult to take into account and

19 The original wording in Danish: ”Vi håber, at du ligeledes finder projektet interessant og relevant, og at
du vil hjælpe med at præsentere det i dit provsti. Vi har i denne mail vedhæftet en invitation til at deltage i
projektet, som er henvendt til landets folkekirkepræster. Heri gives en præsentation af ph.d.-projektet, en
beskrivelse af datahåndtering og en vejledning i indsendelse af mails til projektet. Det vil være af stor
betydning, hvis du vil sende den vedhæftede invitation ud til præsterne i dit provsti. Det er naturligvis helt
op til den enkelte præst at tage stilling til, om han eller hun vil deltage i projektet. Ved at udsende
invitationen til dit provsti, er du således blot med til at synliggøre det. Du er imidlertid meget velkommen
til selv at motivere projektet i din mail og ved sammenkomster i dit provsti. Vi vil meget gerne høre fra
dig, og du må endelig tage kontakt, hvis du har spørgsmål eller kommentarer til os.”
balance even with a probabilistic sampling approach. Compared to sampling in other contexts – for example surveys or experimental studies – the data I wish to study exists on its own terms independent of my research design; it is not constructed in and through the process of data collection. Therefore, I find that the open-ended sampling strategy served the project best. In order to transform the received text material in to a text corpus and prepare it for analysis, I chose a thorough manual process. After the process of cleaning and archiving the material, the text corpus consisted of 11,955 sermons from 95 pastors.

It is evident from my sampling strategy that I - similarly to the Pew and THEOLAB projects - allowed for a considerable variance in the final material. Meanwhile, I have strived to reduce the unwanted and distorting noise by inspecting all documents and removing text parts that were not parts of the sermons. Therefore, I would argue that the variance allowed in the final collection by and large are consequences of the practice in focus due to the inherent variances associated with writing sermons as outlined above. More criteria for streamlining the texts either prior to or after sampling would have affected the ecological validity of the material too much.

**Ethical considerations**

The data collection was reported on 1 February 2017 and approved by the Danish Data Protection Agency through Aarhus University’s legal department (Journal no.: AU-2016-051-000001, ID 378). Every pastor who supplied data for the project submitted an informed consent. In this, they stated that their sermons are permitted to be annotated with metadata and used for analysis, provided that no link between sermons and the identity of individual pastors are disclosed publicly.

My communication with bishops and rural deans illustrates that I used gatekeepers as a way of gaining access to respondents and of legitimising my project among potential respondents. I believe that their cooperation played a significant part in gaining pastors’ attention and interest in the project. However, I find it unlikely that the approval of bishops and rural deans has made pastors feel obligated to provide sermons. If pastors

---

20 I will devote full attention to this methodological aspect of cleaning, annotating and structuring the material in chapter 4 in the product report.

21 In chapter 6, variable distributions of the sermons will be presented and discussed.
truly felt imposed to do so, I should expect to have had a much higher response rate, considering that the invitation was extended to all pastors in the ELCD. In fact, the approach could just as likely have had the opposite effect in some cases – for example, if there were strained relationships between rural deans, bishops and pastors. Therefore, I believe that I used a sampling approach, where pastors in general experienced participation in the project as a voluntary choice.

**Methods for analyses**

It has been a key principle of my project to base my analyses of the sermon corpus on the application of digital tools, and not on close readings sermon by sermon. The immediate reason is pragmatic, as it would be impossible to read such an amount of sermons and maintain an overview of the material. Further, close readings alone would not be able to detect latent patterns within the text collection any way nearly as systematically and precisely as possible with computational tools. Yet, more substantially, such resources enable us to approach the sermon collection as a collective production and not as individual cases stitched together. With the aid of computational resources, we can capture aspects of the sermons *at once* instead of *one-by-one*.

An important characteristic of the sermon corpus is that it holds a collection of unknown material that has never been read before. In a way, sermons are truly parts of “a great unread” to use Moretti’s words – for one thing, of course, since they were never meant for reading in the first place as compared to the great unread of literature. This project consequently has an exploratory dimension, as it entails a first discovery of the material. Yet, this also means that I do not have an in-depth knowledge of the content of the individual sermons which I can compare and assess my analytical findings by. Since the sermon collection holds unknown texts entirely, it poses demands to the analyses I conduct – I cannot presume specific plotlines, narratives, character galleries or themes in the sermons. Therefore, I need to first gain access to these dimensions, before interpreting their content. To accommodate these initial analytical considerations as well as the theoretical background of my project, my analytical framework is developed under consideration of three methodological questions. First, *how may we observe cultural and religious information in a collective text production of sermons?* This question addresses issues of accessing the knowledge potential in the sermon corpus, which involves
detecting where and how symptoms of culture or the cultural text, to use Garber’s and Kristeva’s concepts, is textually manifested in the sermons. The second question concerns the ontological status of this information in the corpus: how may we understand cultural and religious information as relational phenomena? With this question, I focus on how the observed information links to other textual components in the sermons, and how these links as patterns within documents establish relations of likeness and difference to other documents in the corpus. The third question in turn is epistemological: how may we assess the relations between text components? In order to solve this, I need measures for quantifying the relations in order to understand the strength of the patterns I observe, while qualitatively interpreting the meaning and significance of the observed patterns.

I outline in the following general answers to these questions by operationalising key analytical concepts for this project, namely intertextuality, topological reading, distant reading and close reading. In chapter 5, I will concretise how I have applied the analytical framework and computational tools in the three articles included in chapters 6, 7 and 8, as these chapters demonstrate specific answers to these basic methodological questions.

Analytical concepts and approach

The first two questions outlined above address the dimension of intertextuality, which involves the role of cultural and religious elements in sermons. Thus, the analyses in this project explore how pastors incorporate inter-text in sermons in different ways. Under influence of Kristeva and Moretti, inter-text in my project concerns the vertical axis of texts and in this way, how pastors may transpose and manifest cultural as well as theological discourses in a collective production of sermons.

In my analyses, I have chosen to include all sermons in the corpus and not to segment the samples according to variables. Even though the material is not fully a representative sample of ELCD sermons, the sermons are still thoroughly based within the same religious, cultural and historical context. Therefore, we can presume an intertextual basis through which we may study the sermons as a meaningful collective system. Further, since this project entails a first reading of this newly assembled material, I have found it relevant and pertinent to include all sermons in the analyses. This, in order to be acquainted with the material and provide some general insights to this particular corpus.
As I have included the entire corpus in the analyses, I strive to allow all documents to contribute to the analyses, and I try to study inter-text on the collection’s own terms. I have therefore in general not targeted specific cultural representations, such as God images, societal crises, interpretations of particular pericopes, celebrations of targeted holidays or responses to local and global news. Such aspects of contemporary sermons are all immensely relevant. However, in my opinion, it seems premature to target specific phenomena, before gaining insights about general structures in the material as a whole, given that this is a collection of unknown texts. I have therefore designed my analyses so that cultural representations may unveil themselves in the data. In this way, my approach to discover inter-text in the sermons entails focusing on text components that most likely would be detectable in all corpus texts. The general solution to the first question stated in the paragraph above is consequently to identify and target text components that are naturally occurring in language, while carrying with them cultural information. In chapter 6, for example, the targeted text components are third person pronouns pre-signified as male or female, while in chapter 7, the targeted text components are named personal entities, referencing cultural domains.

The second question stated above involves detecting these targeted text components as parts of patterns in the corpus. Here, the idea of topological reading is relevant, as this way of reading implies detecting how text components connect and how their relations repeat themselves across a collection of texts. In chapter 7, these repetitive structures involve the relationship between the targeted named entities in order to detect how cultural domains are integrated and combined in sermons. In my project, I apply topological reading as a general approach for observing connections between textual phenomena in the sermon corpus, and thus to observe patterns of inter-text. However, to assess the relationships between text components, as required by question three above, I use distant and close reading as conceptual tools.

In a state-of-the-art report on common digital visualisation techniques, Stefan Jänicke et al. introduce close reading as a concept within digital scholarship and not opposed to it. They present distant reading as techniques to summarise large data-sets and they include examples such as heat maps, geospatial maps and a range of graph visualization tools – techniques following the objectives proposed by Franco Moretti. For close reading, they enhance digital tools for analysing one or a few texts, such as annotating word- or sentence-units with colours or labels, while maintaining the original structures of the texts.
Collective Testimonies to Christianity and Time

(Jänicke et al. 2015). My application of distant and close reading does not correspond exactly to Moretti’s or Jänicke’s definitions. In my analyses, distant reading concerns the act of establishing the relations between text components. This act entails retrieving the text components in focus from their immediate text context in the sermons and then assessing the relationship between text components through quantitative measures. For example, in chapter 7, I need to assess how strong the relationships are between the retrieved named entities. This whole process rematerialises the texts and represents them in a tabular format, only including the targeted text components and their quantitative measures. This approach follows the idea of topological reading, as it dissolves the structures, the pages, and the chronology of the sermons. My use of close reading does not imply a specific established reading strategy, nor does it correspond to specific computational visualising tools. In my case, close reading involves reading the text that has materialized in the new format based on the distant reading approach. Therefore, the close readings in the included articles concerns interpreting the structures that have emerged from distant readings in order to understand what these patterns tell us about sermons.

My use of distant and close reading implies a shift between first distancing myself from the individual sermons and observing patterns across the text collection, and then zooming in on these patterns to observe and interpret them closely. These patterns provide discursive content, since discourses depends on repetitive structures in order to be observed. The distant reading analysis enables observations of discursive structures in the sermons, while the close reading analysis allows for interpretation and contextualisation of the discursive content. Thus, my approach is a movement between complexity reductions (removing text abundance to get a clear view of the targeted text features) and complexity increases (interpreting the various dynamics that allow these patterns to emerge in sermons).

Distant reading, close reading and topological reading are not all juxtaposing concepts in my analyses. Instead, I see topological reading as a general approach, which I perform through distant and close readings. I find that the shift between distant and close reading considers the contiguities between literal and numeric reasoning as well as focuses on the latency of the manifest in a text collection – both of which are features of topological reading. As such, the inter-text I seek out is rooted in concrete textual features, while the intertextual structures are latent patterns.
Chapter 4

- The sermon corpus (Product)

Data-access has been a focal point of this project. As described in chapter 3, I managed to obtain a final collection of 11,955 Danish sermon documents from 95 pastors. However, the actual amount of documents I received consisted of 13,032 documents. The reason for the discrepancy between material received and material included in the final data-set is my sampling approach, where I requested pastors to send the amount of sermons possible for them in the manner most feasible. It turned out that the participating pastors in general were not very concerned about submitting private documents – in fact, many preferred to send large folders consisting of a variety of different files (such as other ceremonial speeches, brainstorm documents, unfinished documents and invitations to church activities). Consequently, it was not a uniform material to be archived swiftly, and I had to evaluate the material to ensure that only sermons would be included. Another reason that archiving could not be done quickly was that the sampled sermons were personal documents not created for reading by others than the authors. This meant that I could not presume pastors had a standard way of structuring their manuscripts, and therefore, the received documents had to be cleaned up and annotated manually in order to ensure a consistent structure in the final data-set. Further, given the general inaccessibility of Danish sermons for research, it was a key concern to save the material for research projects beyond my own. In this endeavour, it seemed valuable to supply as much available metadata as possible to the collection and to archive the material in a sustainable format. I have therefore transformed the data received from pastors into a text corpus, which in itself is a product of my research. This chapter is devoted to presenting this product, by unfolding the process undertaken in order to create the sermon corpus. In this way, the chapter addresses the following question: *How can an opportunisticly sampled sermon collection be managed and structured in order to provide a sustainable dataset for current and future projects?*

I will focus in particular on the procedures applied for retrieving metadata for the corpus and for annotating and standardising the sermons in it, and for that reason, I will repeat content presented in chapter 6, where corpus structuring and annotation play a significant
role for the context of that article. However, I will be able to provide more detail and
nuance here than what was suitable and relevant for the article format.

The research product

The sermon corpus is accessible in a data-repository hosted by the Centre for Humanities
Computing Aarhus (CHCAA). The repository is developed specifically for displaying the
corpus as a concrete research product and as an outcome of my PhD project. Due to the
General Data Protection Regulations (GDPR), the repository only includes corpus
metadata and not the sermons in full text\(^\text{22}\). The repository holds a brief presentation of
the metadata, which comprises the point of departure for the remaining of this chapter,
and a short description of the data models that provide a basis for archiving the corpus in
a relational database. Software engineer at CHCAA Max Roald Eckardt will develop a
database for maintaining the corpus, and he has composed the description of the data
models in the repository\(^\text{23}\). Therefore, only the corpus of cleaned sermons and metadata
is a part of my research product, not the database.

The data repository is available here\(^\text{24}\):

https://github.com/centre-for-humanities-computing/praedikener

Background: Corpus linguistics

My approach to building the sermon corpus is influenced by the discipline corpus
linguistics, where corpus assembling is a key premise for linguistic studies. Within this
discipline, John Sinclair has defined a corpus as: “…a collection of pieces of language
text in electronic form, selected according to external criteria to represent, as far as
possible, a language or language variety as a source of data for linguistic research”
(Sinclair 2005, 23). Pieces of language can come in a variety of genres (such as
newspapers, literature, speeches, tweets, conversational recordings), but what is

\(^\text{22}\) Access to the sermons in full text can be provided upon request to Kirstine Helboe Johansen
(khp@cas.au.dk).

\(^\text{23}\) For access to the database generator applied at the Centre for Humanities Computing Aarhus, contact
Kristoffer Nielbo Laigaard (kln@cas.au.dk).

\(^\text{24}\) In order to view the metadata-file in tabular form, download the repository-data as a zip-file by clicking
on the green item called “Code” in the repository.
important is the corpus’ ability to demonstrate linguistic variations within a clearly
defined domain, whether broad or narrow. In general, corpus linguistics tend to
distinguish between three types of corpora. One type is the monitored corpus, which is
designed to survey linguistic developments over time, and data is therefore supplied
continuously over the course of a defined period. Another type is the sampled corpus,
which is a collection of language, representing linguistic variation among a clearly
defined population – for example based on a representative distribution of demographic
variables associated with the corpus texts. The third type is the opportunistic corpus. The
construction of this corpus is based on a pragmatic approach to collecting texts. In
contrast to the sampled corpus, the opportunistic corpus is not a representative collection
in terms of typical quantitative criteria – either because it has not served possible to adhere
to such criteria, or because it has not been the intention to begin with (McEnery & Hardie
2012, 6-13). By juxtaposing the opportunistic corpus to the monitored and the sampled
corpus, it is evident that in order to capture a segment of language as it unfolds in practice,
quantitative criteria does not always suffice. Corpus linguistics provide in this way a
perspective on texts as outcomes of living practices, while embedded in clearly defined
contexts. This perspective gives rise to approaches for maintaining the content and form
of the language pieces as they were collected, while systematically adding additional and
consistent information to every language piece. This perspective is significant for my own
approach to cleaning and structuring the sermon corpus, even though my project does not
entail distinct studies of linguistic variation.

In order to provide additional information to a corpus, corpus linguistics provide
techniques and tools for annotating corpora. Linguistic annotations provide information
about internal text composition. Automated tools such as part-of-speech-tagging (POS-
tagging), named entity recognition (NER-tagging), syntactical tagging and semantic
tagging can annotate the words in every text in a corpus with information based on their
linguistic function. Metadata annotations provide information about external text
contexts, for example basic demographic variables, but may also include more domain
specific information about the environment or circumstances the texts were composed for
or in (ibid. 13; Burnard 2005). Corpus metadata indicates an awareness within corpus
linguistics of how the corpus texts have been detached from their originally intended
situation or purpose and integrated into a research context. With the annotations, one
seeks to preserve as much information as relevant about the original context, but within
a structure that makes the corpus texts internally comparable. Therefore, the metadata annotations also provide a consistent structure for describing and summarising the corpus through quantitative information, applied to a qualitative text material. Linguistic annotations and metadata annotations can be made directly within the assembled texts or in stand-off versions, where the annotated data is linked through the document-id of the corpus texts. The annotation techniques provide approaches for enriching the corpus with information – an enrichment that can be done sequentially as the need for new information about the corpus occurs.

### Metadata

The metadata for the sermon corpus is constructed as a stand-off version, and the data structure is apparent from the metadata-file in the included data repository. The first column shows the corpus texts’ document-id, and the succeeding columns designate the metadata annotations provided for the corpus. The labels can be grouped into three types, namely identification labels (document-id and pastor-id), time labels (date and holiday) and demographic labels (parish size, diocese, birth year, gender and place of education).

#### Identification labels

Both identifiers constitute a random number provided to identify sermons and pastors. The core of the document-id is the digits following the initial ‘pr’-code until the underscore, as they designate the unique number of each sermon in the corpus. The pastor-id is a pseudonym provided to mask the identity of the pastors in order to prevent disclosures of direct links between pastors and sermons, as I promised the pastors. The pseudonym is also included in the document-id following the underscore, which strengthens the link between document and metadata.

#### Time labels

It was important for this corpus to have two timestamps, since sermons adhere to two calendar systems. The timestamp called date consists of six digits and designates the conventional Gregorian calendar system in the following structure: `ymmd`
(year/month/date). Thus, it represents a linear way of measuring time. The timestamp called holiday, on the other hand, is associated with the ELCD’s church calendar. The liturgical year consists of up to 69 holidays, but not all holidays occur in a given year\(^{25}\). Throughout the course of a liturgical year, the Christian narrative of salvation unfolds through the prescribed lectionary texts and illustrates a cyclical way of measuring time. As the ELCD has two lectionaries, there is a text variation between even years and uneven years, since the two lectionaries hold different texts for representing the salvation narrative. I have included this variation in the annotation tags for the holiday-label. The structure of this tag consists of an abbreviation of the Danish holiday-name accustomed for this particular corpus, followed by an underscore and a ‘U’ that represents lectionary number one, or an ‘L’ that represents lectionary number two\(^{26}\). This tag points to the biblical texts associated to each sermon. The conventional calendar year and the liturgical year does not begin and end correspondingly. Instead of commencing at New Year, the liturgical year commences the first Sunday of Advent, and consequently, sermons adhering to this holiday in 2011, for example, are tagged 1Adv_L (and not 1Adv_U), even though the sermons fall in an uneven calendar year.

Since I received sermons directly from pastors, I had to rely on the timestamps they had provided in their documents in order to tag the sermons according to date and holiday. Many pastors had stated both date and holiday in their documents, and in these cases, I could be very confident in time accuracy. If there was a discrepancy between these parameters, it was possible in most cases to infer the correct tags by comparing them to the date of origin in the file’s own metadata. In quite many cases, where pastors had only stated holiday and year, I inferred the exact date from the holiday, while comparing it to the file’s metadata. However, in some sermons, I received no information about the year they were composed, and so, the date-tag is stated this way: “xxxxxx”. The holiday tags were developed prior to the process of annotating the corpus, but the process itself brought attention to some sermon tendencies that my tag-list had not considered, as some sermons were composed for holidays or settings exceptional to the official church calendar. Some were written, for example, for occasions called “children’s services”,

---

\(^{25}\) How the liturgical year unfolds within a Gregorian calendar year depends on Easter. If Easter falls early, the second holiday tide of the liturgical year, the Epiphany tide, is short, while the last tide, the Trinity tide, is long. In parallel, if Easter falls late, the Epiphany tide will be longer, and the Trinity tide shorter.

\(^{26}\) Lectionary one adheres to uneven numbered years, for example 2011, and lectionary two adheres to even numbered years, for example 2012. In Danish, uneven years are called “Ulige år”, and even years are called “Lige år”. The letter ‘U’ in the tag therefore designates the former, while the letter ‘L’ designates the latter.
“music services” or “hunting services”, and other exceptional sermons were written for services in other settings than in church, for example at nursing homes. For these cases, we developed a holiday tag called “\$ÆR”. If the service still fell on the date of an official holiday, we added the tag to the existing holiday tag, for example “AHelg_U\$ÆR”. If the service took place on another date than on an official church holiday, we tagged the holiday solely as “\$ÆR” and tagged the date according to the date stated by the pastor.\textsuperscript{27}

\textit{Demographic labels}

The demographic labels consist of information regarding pastor (birth year, gender and place of education) and information regarding the locative context for which the sermons are written (parish size and diocese). I have retrieved the background variables from an online yearbook called \textit{Teologisk Stat}. The association of pastors in Denmark\textsuperscript{28}, which functions as union and professional organisation for ELCD pastors, hosts this forum. The yearbook holds basic information about all present pastors and their office, but pastors can add supplementing information as they like. This online record on pastors is a supplementary resource to a physical book volume that the association publishes every year. Online access is therefore only provided to those who buy the book volume, but I have acquired access to the online resource through the Royal Danish Library. In \textit{Teologisk Stat}, I could observe the parishes that pastors in my data had been employed in from 2011-2016. I retrieved this information and accessed further locative information about the parishes from the ELCD’s web-resource \url{www.sogn.dk}, which is open-access and holds basic statistical information about ELCD parishes. I chose to include only diocese to label the geographic area of the sermons, since more fine-grained indicators, (name of parish or deanery) would very easily identify individual pastors in the data. I describe, instead, parish in terms of population size, where I distinguish between total population and members of the ELCD. The data I retrieved from \url{www.sogn.dk} in 2017 was applicable for 1 January 2016.

The demographic labels included in the metadata seemed the most suited for describing the immediate circumstances regarding contemporary sermons. In selecting the labels, I

\textsuperscript{27} In appendix II, which consists of a sub-grouping of sermons according to holiday for the analyses in chapter 8, the annotation-tags used to tag sermon holidays appear. The ‘\$ÆR-tag is not included here.

\textsuperscript{28} In Danish: Præsteforeningen.
had to balance availability and detail (how much information can be acquired systematically for all sermons?), clarity (what type of information is most applicable for summarising the sermons?) and personal data security (what and how much information can be included without disclosing pastor identities?).

The metadata in full creates a structure that makes the sermon corpus searchable, and it facilitates that subsamples can be retrieved from the corpus according to different variables. In line with the theoretical background in chapter 2, the metadata further validates that these sermons are interrelated in terms of historical, national and religious contexts, and the metadata thus supplies concrete intertextual links between the corpus texts. In my own project, I have not had the need to consider and use all metadata labels in my analyses. However, since the sermon annotations were supplied manually, it was most reasonable to deliver as much metadata as possible at once in parallel to cleaning and assembling the corpus. Repeating the annotation process at a later point with new metadata would seem tedious, and I therefore included more labels than required for my own project.

Cleaning and annotating sermons
The process of transforming the received sermon material into a text corpus entailed a thorough cleaning and annotation process. I had in total five student assistants to assist me – two of them over a shorter period of two weeks in the fall of 2017 and three of them in a longer employment over five months in the spring of 2018. The student assistants’ tasks were the hands-on work of cleaning and labelling the sermons, whereas my tasks were primarily to structure and supervise their work. The overall purpose of the process was to discard texts that were not labelled as sermons by the pastors, and to discard text bits in the sermon documents that was supplementary text to the sermon itself.

Cleaning
The vast majority of the received documents were written in Microsoft Word, and only a few were in a PDF-format. The text cleaning was done in Microsoft Word, but only in copied versions and never in the original documents. The type of text we discarded from
the sermon documents was content appearing before or after the stated sermon. The amount and type of extra text we had to remove varied a lot, but the elements consisted in particular of metadata as composed by pastors individually and liturgical content other than the sermons. For example, it was very common for pastors to include the holiday’s gospel pericope in their documents, which they, according to the church liturgy, read aloud immediately before preaching. Furthermore, for quite many pastors, their manuscripts were in fact composed as a kind of script for the church service as a whole, as they had included other liturgical elements such as hymn references, prayers and words of praise. As a general rule, we did not discard text elements within the sermon part of the manuscripts, even though there might be notes or comments for the preacher that were not supposed to be uttered. Yet, as my project does not focus on the outspoken sermon, but on the crafted manuscript, we would have influenced the authenticity of the stated sermon, if we had begun eliminating elements within the sermons. However, one pastor tended to begin the sermon prior to the gospel reading, and included a copy of the pericope within the stated sermon. In this one case, I chose to delete the gospel text from within the sermons and marking where the text bit had appeared, as it would have made an imbalance compared to the other documents in the corpus, if I had left the pericopes in the documents.

Labelling and archiving

The cleaned sermon files were filed under the same document-ids as in the metadata. In the top of the document, we labelled the sermons with date and holiday according to the metadata tags. All three labels were provided in a standard-off metadata spreadsheet along with demographic labels corresponding to the structure of the metadata-file in the data-repository. All work was done on a shared and secured drive provided by Aarhus University. Here, the students had individual folders, and in these, I allocated them a portion of the received sermons regularly, ensuring they would not process the same documents. Similarly, I allocated the document-ids they were permitted to assign to the sermons in their folders, and they received a portion of 1000 ids at a time (pr1-pr1000; pr1001-2000; pr2001-pr3000), ensuring they would not be able to assign the same document-ids to different sermons. They filed the cleaned sermons in sub-folders until the sub-folders held exactly 1000 sermons, and they filled out individual metadata
spreadsheets in Microsoft Excel 1000 sermons at a time, corresponding to the sub-folder in which they were currently filing cleaned sermons. It was very important that no document-ids would be skipped, thus making holes in the corpus structure. Archiving sermons 1000 at a time made it easier to survey that such mistakes did not occur – if the students had archived all cleaned sermons in the one same folder to begin with, it would be quite complex to find and correct mistakes of different kinds.

**The relational structure of the sermon corpus**

The result of the cleaning and annotation process was eleven subfolders with the cleaned sermons and eleven appertaining spreadsheets, which combined constituted the final sermon corpus. Figure 1 below demonstrates the filing system applied and the link between documents and metadata. Following the manual cleaning and annotation of the sermon documents, all sermons have been read into a large spreadsheet with 11,955 rows and two columns; the first consisting of document-ids and the other consisting of the cleaned sermons. In this structure, the cleaned sermons are data-objects that are identifiable through data-keys – that is the document-ids. This structure makes it possible to associate data-properties to the data-objects, if the data features are provided in an equivalent structure with document-ids as the data-keys.
Collective Testimonies to Christianity and Time

Figure 1. The filing system for cleaned and annotated documents. Documents and metadata are linked through file-name and document-id.

Following this structure, the metadata we composed forms a tabular layer that points to specific properties of the data-objects through the document-id. This layering is illustrated below in Figure 2. With this data structure, new layers of information can be supplied continuously to enrich the corpus. For example, linguistic annotations of the data can be included as separate stand-off versions as demonstrated in Figure 3 below.

Figure 2. The corpus structure illustrated from the two sets of tabular sermon data linked by document-id.
A future perspective of my project is that the sermon corpus will be stored in a research database. With this data structure, the corpus is fit to be archived in a relational database, which is built in layers of tabular data corresponding to the figures above. One of the advantages of a sermon database will be that it provides a solution for sustaining the corpus, for providing new data enrichments and for allowing a secure access to data for future projects.

**Ethical considerations**

As part of constructing the sermon corpus, it has been vital to adhere to the law on GDPR. Therefore, I have included metadata that would not disclose the identity of pastors in the data-set. My pseudonymisation technique implied assigning each pastor a random number from 1-95. Further, I chose to eliminate geographic indicators, such as names of parish and deanery, which were otherwise publicly available. Parish names could unmask pastors very easily, since most parishes employ only one pastor. Name of deanery would be less revealing, since there are just about twenty pastors in a deanery, but in combination with the personal background variables about pastors (age and gender), individual pastors might be identified from deanery information. I have in this way avoided identity markers in the metadata. However, within the cleaned sermons, personal markers may still appear, as pastors could be likely to mention parishes or parish members directly in the manuscript or to include personal anecdotes. As stated earlier, we were
very careful not to discard text from within the documents in order to maintain the integrity of the original sermons. In this way, the metadata in combination with individual sermons could disclose personal identities. Consequently, the full sermon corpus and the sermon database will not be publicly accessible, and access will have to be granted by Kirstine Helboe Johansen, who is responsible for the full data-set on behalf of Aarhus University.

Conclusion

The approach I have used for transforming the sermons I received from pastors into a text corpus has obviously had its advantages as well as disadvantages. Prior to the cleaning and annotation process, there were many factors I could not account for, since this material is the outcome of a living practice. Therefore, it was necessary to undertake the process manually, as I could not predict what the material would look like beforehand. The consequence of the manual approach was a complex and time-consuming process. There is naturally a margin of error to consider when five different individuals are asked to type in or copy paste in metadata manually. However, I find that I developed a system to prevent and detect errors, as I made random checks of the students’ work regularly to ensure the annotation tags were correct and that there was consistency in the links between sermons and metadata. In connection with the analyses in chapters 6, 7 and 8, we were able to detect and correct some mistakes that had bypassed my attention. Yet, I have so far not discovered any substantial errors as a consequence of the manual procedure.

Though slow, the advantages of the procedure are still significant. The manual approach in itself provided a thorough insight into the material, and by sampling sermons as private documents, I have ensured that the material is the outcome of the same practice conducted on the same premises: Pastors’ writing of sermon manuscripts intended for local congregations in the ELCD. These sermons have not been retrieved from parish websites or from online databases, where the hosts could have specific motivations for selecting sermons to share. Of course, I cannot guarantee that the pastors have not selected some sermons over others or edited the documents they send me. However, my general impression is that the pastors tended not to put effort in editing their sermons or in sending a targeted collection of sermons. The structure of the sermons I received could appear
diffuse and not necessarily readable, and as already mentioned, many pastors tended to send in folders with a variety of material. Furthermore, since we stated in our invitation to the pastors that we did not intend to close-read their sermon documents, but to look for patterns in them, they may not have felt so self-critical about the sermons they sent in. I consider this to be a positive thing and to be an indication of the received material’s authenticity, meaning that the sermons I received, quite likely are the sermons originally intended for the services stated in the documents.
Chapter 5

- Background to articles

The output of my studies of the sermon corpus is three research articles. Each article holds a specific research design, but all of them adhere to the theoretical and analytical framework of my project. My choice of journals demonstrates that this project is in dialogue with different fields. The article in chapter 6 has been published in *Journal of Cultural Analytics*, which is a journal that promotes digital humanities scholarship. Thus, the journal does not target research on religion in particular, and in this context, my contribution on sermons is published as a large-scale project among other computationally informed humanities studies. I have submitted the article in chapter 7 for *Temenos – Nordic Journal of Comparative Religion* that publishes a broad spectrum of research on religion and culture particularly in the Nordic countries. My publication emphasises, in this research setting, Danish sermons as contemporary phenomena in the intersection of biblical and cultural discourses. The final article in chapter 8 is intended for the *International Journal of Homiletics*, which promotes homiletics research across Christian traditions and national contexts. Here, I enhance large-scale studies as a way of unveiling and interpreting theological discourses in sermons.

I am the main author on all articles, but I have had the privilege of cooperating on the studies and the publications with others. I have had the main responsibility for framing and designing the studies and for conducting the close reading analyses. The main responsibility for the computational analyses has been on a co-author, since the technical execution of the analyses lies beyond my own competencies. However, I have been in very close dialogue with the co-authors responsible for the technical aspects in the articles in order to take part in developing the appropriate pipelines. In the following, I will make short introductions to each of the articles and place them in the overall framework of my project. I will not go into detail with the articles’ theoretical backgrounds or methodological pipelines, but I will instead demonstrate their overall research contexts and the analytical principles in my project.
Background to chapter 6

The article for *Journal of Cultural Analytics* in chapter 6 is titled “Sermons as data: Introducing a corpus of 11,955 Danish sermons”, and my co-authors are Ross Deans Kristensen-McLachlan (Centre for Humanities Computing, Aarhus University), Kirstine Helboe Johansen (department of Theology, Aarhus University), Uffe Schjødt (department of the Study of Religion, Aarhus University) and Kristoffer Laigaard Nielbo (Centre for Humanities Computing, Aarhus University). The article is published in the section of the journal called *Data Sets*, where the effort of constructing data is promoted, as researchers get a chance to exhibit their data collections. The article has two primary foci. The first one is data specific, as we outline and discuss the process of acquiring sermons and assembling the corpus, while the other is analytical, as we include a case study on gender constructions.

As research context for data collection and the case study, we enhance the ELCD as a public agency intricately embedded in society. This view follows the lived religion perspective as presented in chapter 2, since we focus on a religious actor as present in culture rather than detached from it. The enhancement of the ELCD as a public actor influences the two focal points regarding sermons. On the one hand, it points to the complicated nature of written sermons, as the documents are oriented towards a public audience, while the manuscripts in themselves are private properties of pastors. This aspect has been important to consider in regards to sampling and cleaning Danish sermons. On the other hand, the lived religion perspective influences the way we consider the discursive content in sermons. Therefore, we focus on the ELCD as an actor that actively takes part in producing social worlds, just like other actors in society. In this way, discourses that may unfold in society as such may just as well take place in the ELCD, and we are interested in how such discourses unfold in the sermon genre. More specifically, we attend to pastors’ representations of gender agency in order to perceive one type of sociality in sermons. Up until now, gender has not been a prevalent theological topic in the ELCD, and therefore, our strategy was to uncover gender discourses implicitly. We based this strategy on the knowledge that gender is something we do never not talk about, since the language we use inherently holds gender categories. Gendered personal pronouns (he/she) are a part of everyday vocabulary, and as such, the basic linguistic components used to designate gender are the same in the ELCD as they
are in other societal spheres. In the case study, we therefore explore what functions these components more specifically have in sermons.

The analytical approach we undertake in the study demonstrates the shift between distant and close reading as described in chapter 3. The intertextual components we focus on are gendered pronouns, while the relational structure – the intertextual pattern – we observe, consists of the pronouns’ associations to verbs. By focusing on verbal content as part of the structure, we can interpret the agential spaces of the character types. This co-occurrence of pronouns and verbs are so basic in language that we can expect it to occur in most sermons. Our distant reading thus implies that we retrieve every instance of a gendered pronoun (he/she/him/her) and its associated verb. The only computational tool required in this pipeline is a part-of-speech-tagger that assigns a word class to every word in the corpus. We assess the relationships between the word pairs with a measure called a pointwise mutual score (PMI-score) that measures how significantly each of the retrieved verbs are associated with each of the four pronouns in the corpus. The result of the distant reading is a rematerialised text in a tabular format. The text included in this format are the pronouns represented as column names and the verbs represented as row names. The cells consist of PMI-scores designating the strength of the relationship between the relational components. Our close reading entails that we assess the verbs most significantly associated to the pronouns and that we categorise the verbs in terms of four agency types. With this categorisation, we end up with four matrices – two for verbs related to each of the pronouns in nominative case (he and she respectively) and two for each of the pronouns in nominative case contra the pronouns in oblique case (he-her and she-him respectively). We then use these matrices to interpret the agential dynamics of the various character types. In this, the structures we observe distantly unveil themselves as constructive phenomena in which we can observe and interpret identity constructions.

The approach we apply in the gender study is essentially quite simplistic. In fact, here, we actually consider sentence structure in the analyses as we retrieve the word pairs from individual sentences in the sermon. However, with our approach we are able to explore how these linguistic components are linked and repeated throughout the entire corpus.

---

29 Ross Deans Kristensen-McLachlan has conducted all computational analyses for the article in chapter 6.
Background to chapter 7

The article in chapter 7 is entitled “Unveiling the character gallery of sermons – A social network analysis of 11,955 Danish sermons”, and I have written it in cooperation with Kirstine Helboe Johansen (department of Theology, Aarhus University) and Ross Deans Kristensen-McLachlan (Centre for Humanities Computing, Aarhus University). As the title implies, we here pay close attention to the characters who inhabit the sermon corpus. In the article, we investigate who the characters in sermons are, how they connect to other characters and what the encounters between characters tell about pastors’ engagement with history, culture and bible.

Theoretically, we take our point of departure in the recent developments in homiletics as outlined in chapter 2. From an Other-wise Homiletics perspective, Marlene Ringgaard Lorensen has used Mikhail Bakhtin’s idea of the polyphonic text to enhance how sermons may hold an ‘otherness’. A polyphonic sermon thus implies that the preacher enables recipients to hear other voices than just the preacher’s, so that listeners may find themselves in an internal dialogue with a variety of different voices. We, however, shift the perspective from the dialogue between preacher, recipient and sermon to the dialogue between the characters, the voices, in the sermons. In this, we argue that sermon characters carry an identity with them when pastors represent them in sermons, as pastors do not construct them freely, but incorporate them from various contexts. With this argument, we accentuate the intertextual dimension of sermons, as we find that observing characters allows us to investigate the cultural, Christian and historical domains that pastors collectively transpose to sermons. In the article, we illustrate this intertextuality as a dynamic interaction by applying the Deleuzian theory of rhizomes. The rhizome as conceptualised by Gilles Deleuze and Félix Guattari is an open and ever-expanding system that allows phenomena to connect in unpredictable ways. We use this concept to analyse sermon characters as components in a rhizomatic network who carry links to text external contexts (cultural and biblical domains) as well as links to text internal contexts (other characters in the sermons). As such, the connections between characters in sermons facilitate encounters between domains.

Methodologically, the intertextual components we target are personal named entities in the corpus and the intertextual structures we observe and interpret are the connections

---

30 We have submitted the article for Temenos, but it has not been published yet.
between the entities. Similar to the targeted text components in the previously described article, named entities are a very basic component of language and thus, we can presume that personal entities should be represented very well in the entire corpus.

We switch between distant and close readings more than once, as our analysis falls in two sequences. First, we use a named-entity-tagger to identify personal entities in the corpus and we calculate the frequency scores of unique entities. This procedure provides us with a new text representation (unique characters as rows, frequency score as column). As we closely read this table, we assign information to the characters about the domain they represent. In the second analytical sequence, we perform a social network analysis using the open-source software Gephi. This tool allows us to conduct a distant reading of the relationships between characters in the corpus. Gephi is commonly used as a tool to visualise a network with text components as nodes and the connections between nodes as edges. However, we use instead the tool’s community detection algorithm that assigns the characters in the sermon corpus to sub-communities (modularity classes). With these groupings, we can observe characters that tend to co-occur or to share connections to the same nodes. The measurements performed in Gephi provide extra columns of information to the initial spreadsheet and therefore an extra layer to the text for close reading. With the close reading, we attend to the character sub-groups and interpret the intertextual dynamics in the character groups. Here, we focus on the labels we provided before the network analysis, but we are also able to get a more complex understanding of the roles of the characters. Their relation to each other in the group offers a more thorough insight into the identity or function they seem to have in the sermons, as compared to observing them as singular phenomena. They mutually provide each other with context, which informs our understanding of the groups.

**Background to chapter 8**

The article for the *International Journal of Homiletics* is entitled “What is love…? – A study of thematic constructions in 11,955 Danish sermons”, and I have written it with Kristoffer Laigaard Nielse (Centre for Humanities Computing, Aarhus University). The article is a study of thematic fields in the corpus, where we investigate how pastors

---

31 Ross Deans Kristensen-McLachlan has conducted all computational analyses for the article in chapter 7.
32 We have submitted the article for the International Journal of Homiletics, but it has not been published yet.
collectively represent and construct theological content in contemporary time. In particular, we are interested in how theological discourses in sermons may connect to cultural codes. The movement between religion and culture is thus different from the case study in the first article, where we focus on how a public discourse from “without” unfolds “within” the church – in principle, our approach to discern gender discourses can be replicated in other types of texts to discover other genre specific constructions of gender. In the last article, we focus on sermons as the theological discipline of making the Christian tradition pertinent in the present. We thus study whether pastors link discourses from “within” the church to broader cultural contexts “without”. Comparatively to these articles, we explore in chapter 7 which ramifications towards Christianity and cultural fields the sermon corpus charts from “the middle” – as such, we do not frame a discursive movement from within or without.

The empirical turn in homiletics and its developments in the Nordic countries are a vital part of the research context in the article in chapter 8. As in chapter 2, we seek out the contiguities between this development and the digital humanities’ perspectives and methodologies. We introduce the move from the Other-Wise focus on sermons as performatives to sermons as documenting objects taking part of a collective production in line with Franco Moretti’s perspective on literature. Further, we follow the traits of lived religion, as we again emphasise the ELCD as an actor embedded in culture. We do this under considerations of the concept displacement of theology that Linda Woodhead has used to describe the effect of societal changes upon Western majority churches. With this concept, Woodhead points to tendencies in the 20th century, where the theological expertise of majority churches has lost its influence in society. We use the concept to explore the role of theological discourse and dynamics in sermons, and we place our study in line with a current trend of studying ELCD religiosity from a lived religion perspective, namely as an organisation in dialogue with contemporary culture.

Although we are specifically interested in observing theological structures in the sermon corpus, we still make a design for studying the structures on the text material’s own terms. However, we use three different methodological pipelines to target theological content from different angles, which means that shifts between distant and close reading occur in each of the sub-designs.

Kristoffer Laigaard Nielbo are responsible for the computational analyses in in chapter 8.
First, we use a word count analysis to observe the words most frequently used in the corpus. As such, we perform a distant reading by calculating frequency scores for each word in the corpus and filtering the list from stop-words (such as adverbs and conjunctions). This list provides a very basic insight into the content that this corpus seems to focus on in particular. With our close reading of the list, we interpret what the main themes of the sermon collection are and we consider the prevalence or absence of authoritative theological concepts in these thematic fields. In this approach, we do not target specific intertextual structures, but we allow them to emerge in the data bottom-up. However, in the second sub-design, we use our thematic insight from the word list to target specific theological concepts, namely love and sin that respectively are considerably high and low on the frequency list. With a semantic network analysis, we perform a distant reading by measuring the targeted concepts’ proximity to other terms in the corpus. This process results in three network visualisations: one with the concept love and its most proximate word associations, one with sin and its most proximate word associations and one that holds both of the former networks, including secondary associations to the concepts. These networks constitute the rematerialized texts on which we perform a close reading. Here, we consider the semantic structures the targeted concepts seem to be a part of, as we interpret and contextualise the theological and mainstream cultural dynamics appearing from the networks.

With the last sub-design, we investigate the relationship between thematic content in the sermon corpus and thematic content of the liturgical seasons in the ELCD. In this endeavour, we use a topic model, which is a tool to explore thematic content bottom-up in large text collections. The topic modelling of the corpus accounts for our distant reading in this design. From the output of the model, we can observe 11 topics that each adhere to sermons from one of the holiday seasons in the ELCD. In this way, the model indicates that thematic content in the sermon corpus very much follows the narrative structure of the lectionaries. The topics consist in a list of words that according to the model are very probable to share a topical context in the corpus. Each topic represents a text that we can read closely, and here, we interpret how the vocabulary in each topic relates to the narrative themes of the liturgical year.
Concluding remarks

In each article, we investigate dynamics between Christianity and cultural domains. All analyses include a shift between distant and close reading to access and interpret intertextual structures in the corpus. When I argue that we may understand the outputs of our distant readings as rematerialised text, I acknowledge that the “texts” are represented as spreadsheets and as such in a quantitative format. However, I argue that in contrast to survey data or experimental data, the elements we target in our spreadsheets through close readings are not the quantitative measures but the textual phenomena we have measured. In this endeavour, we rely on our competencies of reading and interpreting text in accordance to text contexts. With this approach, we do of course not gain detailed insights to the individual texts as composed by pastors, but instead, we are able to obtain thorough knowledge about the retrieved intertextual structures of the sermons.
Chapter 6

- Sermons as data: Introducing a corpus of 11,955 Danish sermons (article 1)

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.

Introduction
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.34 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.\textsuperscript{35} In recent years, past secularization perspectives have been criticized for viewing religion and the secular as two clearly distinguishable and separate domains.\textsuperscript{36} 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.\textsuperscript{37} 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 still 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. 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. 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 Columbia\(^{40}\), and the “World Religion Database”\(^{41}\) and its companion the “World Christian Database”\(^{42}\) 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/)\(^{43}\) or a collection of images of supernatural agents drawn by children in different parts of the world.\(^{44}\)

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 project 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.” This definition allows scope for the texts to include excerpts from additional sequences of the church services. In their analyses, the research group has 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 project. 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.

47 We contacted the rural deans under the last of the ten bishops without a statement of approval.
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.48 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*.49 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–

48 For instance, All Hallows’ Day in odd years is represented: AHelg_U.
2016 period from [www.sogn.dk](http://www.sogn.dk) – a publicly accessible web page that contains demographic information and statistics about ELCD parishes.

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 below shows an excerpt of the sermons metadata.

<table>
<thead>
<tr>
<th>ID-doc</th>
<th>ID-pastor</th>
<th>Date</th>
<th>Holiday</th>
<th>Parish size</th>
<th>Diocese</th>
<th>Birth year</th>
<th>Gender</th>
<th>Place of education</th>
</tr>
</thead>
<tbody>
<tr>
<td>pr108_11</td>
<td>11</td>
<td>161030</td>
<td>23Trin_L</td>
<td>5731/10011</td>
<td>København</td>
<td>1977</td>
<td>2</td>
<td>Aarhus</td>
</tr>
<tr>
<td>pr2096_35</td>
<td>35</td>
<td>120409</td>
<td>2PåsD_L</td>
<td>3951/4639</td>
<td>Haderslev</td>
<td>1961</td>
<td>1</td>
<td>Aarhus</td>
</tr>
<tr>
<td>pr4385_41</td>
<td>41</td>
<td>130106</td>
<td>Hel_U</td>
<td>11043/15093</td>
<td>Fyn</td>
<td>1962</td>
<td>2</td>
<td>København</td>
</tr>
<tr>
<td>pr5734_88</td>
<td>88</td>
<td>160925</td>
<td>18Trin_L</td>
<td>6535/8330</td>
<td>Aarhus</td>
<td>1975</td>
<td>1</td>
<td>København</td>
</tr>
</tbody>
</table>

*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.*
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.”

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.

---

50 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.
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).

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.

*Figure 1. Distribution of sermons by year*
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.
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). 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 MTLD 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.

---

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

<table>
<thead>
<tr>
<th></th>
<th>Sentences</th>
<th>Words</th>
<th>TTR</th>
<th>MTLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>74</td>
<td>1453</td>
<td>0.36</td>
<td>72.24</td>
</tr>
<tr>
<td>Women</td>
<td>69</td>
<td>1406</td>
<td>0.37</td>
<td>73.50</td>
</tr>
<tr>
<td>Men</td>
<td>78</td>
<td>1489</td>
<td>0.36</td>
<td>71.28</td>
</tr>
</tbody>
</table>

---

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. 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. 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”. 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 church, or that gender is not interesting in a theological context. A qualitative study of twenty sermons investigated whether female and male pastors differ in how they

---

54 Marie Vejrup Nielsen, "Transformationer i Folkekirkekristendommen i Danmark i Dag", Religionsvidenskabeligt Tidsskrift, no. 53 (2009): 68.

85
represent God and found that there were no significant difference\textsuperscript{57}. 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 far\textsuperscript{58}, 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 2020\textsuperscript{59}.

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 pastors\textsuperscript{60}. 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 blindness\textsuperscript{61}.

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.\textsuperscript{62} 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

\textsuperscript{57} 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.
\textsuperscript{60} Else Marie Wiberg Pedersen, "Feminisme eller feminisering?", in Kvinde, mand, kirke: Folkekirken og den lille forskel, ed. Eva Holmegaard Larsen (Fredricksberg: Anis, 2012), 59–75.
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.

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

63 A Github repository for this project can be found here: https://github.com/centre-for-humanities-computing/sermon-research.dk
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% male64. 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.

---

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

<table>
<thead>
<tr>
<th>Pronouns</th>
<th>Male pastors</th>
<th>Female pastors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw frequency</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>He</td>
<td>128,400</td>
<td>71.4%</td>
</tr>
<tr>
<td>She</td>
<td>13,200</td>
<td>7.3%</td>
</tr>
<tr>
<td>Him</td>
<td>31,802</td>
<td>17.7%</td>
</tr>
<tr>
<td>Her</td>
<td>6,348</td>
<td>3.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>179,750</strong></td>
<td><strong>100 %</strong></td>
</tr>
</tbody>
</table>

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 below.65

---

65 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.
Table 4a. Distribution of analytical categories in sermons by men. The table presents verbs representative of one of the four categories in single cases.

<table>
<thead>
<tr>
<th>Category</th>
<th>Verb</th>
<th>Freq.</th>
<th>PMI</th>
<th>Single case: Male subject</th>
<th>Verb</th>
<th>Freq.</th>
<th>PMI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active–Competent agency</td>
<td>Undertook</td>
<td>54</td>
<td>3.5</td>
<td></td>
<td>Gave birth</td>
<td>110</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>Accomplished</td>
<td>73</td>
<td>3.2</td>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Expounds</td>
<td>41</td>
<td>3.0</td>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Commanded</td>
<td>50</td>
<td>2.9</td>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Cognitive–Perceptive agency</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
<td>Sense</td>
<td>313</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
<td>Looked</td>
<td>149</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
<td>Recognize</td>
<td>272</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
<td>Understood</td>
<td>583</td>
<td>3.5</td>
</tr>
<tr>
<td>Communicative–Expressive agency</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>Address</td>
<td>149</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>Smiles</td>
<td>86</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>Laughs</td>
<td>66</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>Interrupts</td>
<td>40</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Religious agency</td>
<td>Resurrected</td>
<td>60</td>
<td>3.1</td>
<td>Anointed</td>
<td>80</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heals</td>
<td>326</td>
<td>2.7</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

Table 4b. Distribution of analytical categories in sermons by men. The table presents verbs representative of one of the four categories in relational cases.
Table 5a. Distribution of analytical categories in sermons by women. The table presents verbs representative of one of the four categories in single cases.

<table>
<thead>
<tr>
<th>Category</th>
<th>Verb</th>
<th>Freq.</th>
<th>PMI</th>
<th>Verb</th>
<th>Freq.</th>
<th>PMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single case: Male subject</td>
<td></td>
<td></td>
<td></td>
<td>Single case: Female subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active-Competent agency</td>
<td>Undertook</td>
<td>21</td>
<td>3.2</td>
<td>Gave birth</td>
<td>72</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Promise</td>
<td>388</td>
<td>2.9</td>
<td>Involves</td>
<td>20</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Overcame</td>
<td>57</td>
<td>2.9</td>
<td>Managed</td>
<td>37</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Direct</td>
<td>82</td>
<td>2.8</td>
<td>Achieve</td>
<td>39</td>
<td>3.3</td>
</tr>
<tr>
<td>Cogitative-Perceptive agency</td>
<td>Sensed</td>
<td>29</td>
<td>3.0</td>
<td>Misunderstand</td>
<td>28</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Regrets</td>
<td>63</td>
<td>2.8</td>
<td>Dreamt</td>
<td>82</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>Decides</td>
<td>87</td>
<td>2.9</td>
<td>Imagined</td>
<td>42</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td></td>
<td>Recognizes</td>
<td>201</td>
<td>3.3</td>
</tr>
<tr>
<td>Communicative-Expressive agency</td>
<td>Discussed</td>
<td>32</td>
<td>3.1</td>
<td>Greeted</td>
<td>22</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Propagate</td>
<td>32</td>
<td>2.9</td>
<td>Kisses</td>
<td>72</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Declare</td>
<td>32</td>
<td>3.1</td>
<td>Cries</td>
<td>366</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Asserted</td>
<td>27</td>
<td>3.2</td>
<td>Expressed</td>
<td>83</td>
<td>3.3</td>
</tr>
<tr>
<td>Religious agency</td>
<td>Sermonized</td>
<td>129</td>
<td>2.8</td>
<td>Anointed</td>
<td>110</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Heals</td>
<td>213</td>
<td>2.8</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Preached</td>
<td>101</td>
<td>2.8</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Revealed</td>
<td>101</td>
<td>2.8</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Verb</th>
<th>Freq.</th>
<th>PMI</th>
<th>Verb</th>
<th>Freq.</th>
<th>PMI</th>
<th>PMI</th>
<th>PMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational case: Male subject and Female object</td>
<td>Protects</td>
<td>84</td>
<td>1.6</td>
<td>Protects</td>
<td>84</td>
<td>1.6</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Cogitative-Perceptive agency</td>
<td>Treasures</td>
<td>23</td>
<td>2.2</td>
<td>Treasures</td>
<td>23</td>
<td>2.2</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Observes</td>
<td>26</td>
<td>1.3</td>
<td>Observes</td>
<td>26</td>
<td>1.3</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Overlooked</td>
<td>25</td>
<td>2.4</td>
<td>Overlooked</td>
<td>25</td>
<td>2.4</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Accepted</td>
<td>32</td>
<td>2.5</td>
<td>Accepted</td>
<td>32</td>
<td>2.5</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Communicative-Expressive agency</td>
<td>Declare</td>
<td>32</td>
<td>3.1</td>
<td>Declare</td>
<td>32</td>
<td>3.1</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Shoved</td>
<td>21</td>
<td>2.2</td>
<td>Shoved</td>
<td>21</td>
<td>2.2</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Enhances</td>
<td>108</td>
<td>2.7</td>
<td>Enhances</td>
<td>108</td>
<td>2.7</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Reveals</td>
<td>47</td>
<td>1.8</td>
<td>Reveals</td>
<td>47</td>
<td>1.8</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Religious agency</td>
<td>Preached</td>
<td>101</td>
<td>2.8</td>
<td>Preached</td>
<td>101</td>
<td>2.8</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Revealed</td>
<td>67</td>
<td>2.8</td>
<td>Revealed</td>
<td>67</td>
<td>2.8</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Washed</td>
<td>69</td>
<td>2.7</td>
<td>Washed</td>
<td>69</td>
<td>2.7</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

Table 5c. Distribution of analytical categories in sermons by men. The table presents verbs representative of one of the four categories in single 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).66 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 sociolinguistic 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 communities\(^\text{67}\).

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.\(^\text{68}\) 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,
genedered 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.
Chapter 7

- Unveiling the character gallery of sermons – A social network analysis of 11,955 Danish sermons (Article 2)

Abstract
In this article, we examine the character gallery in a digitised corpus of 11,955 Danish sermons from the Evangelical-Lutheran Church in Denmark written between 2011 and 2016. We study these sermons as a collective text production, where the characters represented illustrate aspects of Christian tradition and cultural history. We depart from three basic questions: Which characters populate Danish sermons? How do pastors adapt characters in their sermons? And what do the encounters between characters tell us about the pastors’ representations of bible and time? In line with Bakhtinian thought, we approach the sermonic character gallery as a polyphony of voices, and informed by the Deleuzian idea of the rhizome, we understand this character gallery as a network, where characters through their connections to other characters form thematic narrative clusters. We have represented this network through a social network analysis, using computational tools, and analysed closely which and how characters connect in sub-groups of the network. We find that biblical figures in particular enhance stories of Jesus as saviour, teacher or caretaker, while political figures tend to be dissociated from biblical figures and representing narratives of atrocities in history. Surrounding these subgroups, a large group of anonymised characters prevail.

Introduction
At the centre of narratives are agents. Danish literary scholar Jørn Vosmar has argued that four components are constitutive in how writers structure literary worlds: an I, time, space and surroundings. This means that literary worlds centre on characters situated in temporal and spatial contexts, in which other subjects and objects exist (Vosmar 1969, 89-101). But narratives are not just innate to the literary field. Within narrative theory, scholars such as David Herman and Monika Fludernik has argued that humans on a very fundamental level understand the world through narratives (Herman 2004; Fludernik 1996). People populate their world with agents and story lines in order to provide a meaning of the world (Nielsen 2010, 173). American psychology scholar, Donald Polkinghorne argues that narratives do not explain the world on the same terms as scientific statements, but instead, narratives make connections between events. As such, they can establish coherence between past and present (Polkinghorne 1988, 21). Thus,
agents do not only form a character gallery in literary texts, but also – albeit differently – in other types of texts, including news, social media, speeches and religious texts. In line with this perspective, we investigate how the character gallery unfolds in contemporary Danish sermons.

Sermons do, of course, differ from literary texts. Sermons are theological and religious texts produced by local pastors for the weekly Sunday service to address the congregation. Elucidating a biblical text (a *pericope*) is at the heart of the interaction between preacher and listener, and this entails inviting a range of biblical figures into a sermon. However, sermons are mediated in a “here and now” in church, which potentially invites an entirely different range of characters. Our basic tenet is therefore that pastors populate their sermons with a broad variety of biblical and non-biblical characters. We can therefore acquire important insight into contemporary preaching by examining who these characters are and how they interrelate. Studying the literary world of a single pastor – or a few selected pastors – would provide insight into *their* particular narrative world. However, as argued by literary scholar Franco Moretti with respect to literature (Moretti 2007, 3-4), we see preaching as a vast field that cannot be understood by stitching together single cases. Therefore, sermons must also be understood as a collective practice fostered by theological and pastoral training, ecclesial tradition as well as cultural history. To learn more about the collective practice of preaching rather than single cases, we must study a large sample of sermons. Based on a text corpus of 11,955 sermons written by 95 pastors in the Evangelical-Lutheran Church of Denmark (ELCD) in the period 2011-2016 (Agersnap et al. 2020), we examine sermons as a collective text production representing contemporary preaching in the ELCD. We depart from three basic questions:

Which characters populate Danish sermons? How do pastors adapt characters in their sermons? And what do the encounters between characters tell us about the pastors’ representations of bible and time?

**The Evangelical-Lutheran Church in Denmark**

The Evangelical-Lutheran Church in Denmark is a majority church with high membership (74.3% of the entire population of Denmark as of 1 January 2020). However, membership is in stable decline primarily due to demographic changes (down from 80.9% of the entire population in January 2010). Membership levels vary significantly across
the country with the diocese of Copenhagen having the lowest membership (56.4%) and the diocese of Viborg having the highest (83.9%). Despite high membership, regular church attendance is low, with approximately 11% of the population attending church at least once a month (Andersen et al. 2019). Although membership levels are declining in general, it is still common to engage in church ceremonies for important life events: in 2018, 58.7% of all children born in Denmark were baptised; in 2019, 68.3% of all young people had a confirmation; and, in 2019, 82.7% of all those who died received a church funeral. The ELCD is affiliated to the state of Denmark through the constitution §4, which states that “The Evangelical-Lutheran Church is the Danish people’s church and as such supported by the state”. The church is divided into ten dioceses with ten bishops, which are further divided into deaneries and geographically determined parishes. There is no national synod and the local congregational councils are relatively free to determine the life of the local church.

The Sunday service follows a liturgical order approved by the ten bishops and authorised by the Queen of Denmark. However, the liturgical order describes a number of variations that the local parishes can choose between. The service must contain two or three bible readings: one from the Old Testament and/or Epistles and one from the Gospels. These readings are determined by two lectionaries – one for even and one for uneven liturgical years. The pastor always has to preach on the gospel pericope of the given Sunday.

Sermons: A polyphony of voices

In recent decades, within homiletics, there has been an increased interest in sermons as dialogues. In particular, there has been a growing recognition of the listeners as active contributors to the meaning of preaching. This role of the listener was acknowledged and developed in the idea of the roundtable pulpit, originally promoted by American theologian John McClure (McClure 1995), in which pastors are encouraged to invite congregants to a round table on the biblical text as part of their preaching preparations. Empirical research has shown that listeners do not necessarily limit their understanding

---

69 All statistics can be found at: https://www.km.dk/folkekirken/kirkestatistik/
71 For further information on the Evangelical-Lutheran Church in Denmark, see Nielsen & Kühle 2011 or Christoffersen 2010.
to what the preacher expressly says or means. Rather, listeners creatively co-create and further develop the meaning of a given sermon depending on their own context and reflection. This has been illustrated in a Danish context by (now) bishop Marianne Gaarden and theologian Marlene Ringgaard Lorensen (Gaarden & Lorensen 2013; Gaarden 2014) but similar results can be found in other contexts. Based on empirical insights such as these, homiletical interest has increasingly shifted towards creating sermons that are open to a variety of individual interpretations. It is possible to create such sermons through direct interactions with potential listeners (and thereby co-authors) following the tradition of the roundtable pulpit (Lorensen 2013, 55). However, this also requires a revaluation of what a sermon is and what the pastor ought to achieve when preaching. Inspired by Russian literary theorist Mikhail Bakhtin, Lorensen strongly distances herself from the sermon as a transmission that attempts to convey a meaning from the preacher to the listeners, since this communication model posits the listeners as passive recipients. Lorensen instead follows Bakhtin in understanding language as dialogical, which refers both to a way of practicing language and an ethical standard of recognising the other as subject. The works of Russian author Fyodor Dostoevsky play a key role in the Bakhtinian understanding of language and communication as something that is – or should be – dialogical and deeply oriented towards the other as itself a subject. For Bakhtin, Dostojevsky’s multi-voiced authorship becomes an ideal for literary art, insofar as it allows every voice to be authoritative – this does not imply each voice is truthful but that it must be taken seriously. Lorensen thus introduces the Bakthinian concept of polyphony to indicate how an author can renounce monological privileges and instead seek to establish polyphonic interactions between autonomous characters in the text (Lorensen 2013, 66). For the communication taking place, this means that:

…the addressee does not just hear one voice, that of the author, but many voices, voices belonging to ‘personalities’ with whom the reader, via the act of reading, is drawn into a dialogue. In this polyphonic encounter the border between discourse and reality is transcended. No one has the monopoly of interpretation; neither author nor characters or readers possess the truth because the truth cannot be born and kept within the mind of an individual. Instead it is born between people who seek the truth in dialogical interaction. (Lorensen 2013, 66).
This ideal of open dialogue – a polyphony of voices that invites the recipients of the novel or sermon to participate as active interpreters and co-creators – was developed by Lorensen and has recently gained empirical support (Lorensen 2016; Lorensen & Buch-Hansen 2018). However, in this article, rather than concentrating on the recipients or the co-creative process of listening, we will instead focus on the characters who embody the polyphony of voices.

In fictional genres such as the novel, characters and their voices are predominantly literary constructions. Although they may eventually assume a life of their own for the readers, at the outset, they are under the control of the author. However, sermons are not fictional texts and the characters included are always known beforehand. This pertains to both biblical and non-biblical characters. Although the pastor is in charge of inviting voices into the sermon and constructing their roles, the characters themselves evoke other stories and contexts in which they occur, along with previous encounters between these characters and the listeners. From a cognitive-linguistic perspective, we can say that these characters act as reference points (Langacker 2009, 83-85). These reference points might provide access to richly textured background knowledge, meaning that the mere mention of certain characters could result in a wide range of connotations for certain listeners (Semino 1997, 119-159). The characters are therefore not floating signifiers to which the pastors assign meaning; they already carry meaning and it is as such that they enter sermons.

Rhizomatic networks

To uncover not only who the characters in the sermons are but also how they are interrelated and connected, we deploy the Deleuzean concept of rhizomes (Delueze & Guattari 1980). Gilles Deluze and Félix Guattari differentiate between “the tree” and “the rhizome” as systems of thinking. The tree system of thinking, arborescent thinking, is linear and vertical: the tree can only grow in one predefined direction and all its many branches have developed from the one same stem. This arborescent mode thus represents coherence and unity between phenomena. In particular, Deleuze and Guattari find the arborescent principle in Freudian psychoanalysis, where one’s identity can be explained in terms of past occurrences. The rhizome system of thinking adheres to completely different principles. A rhizome is an open system that establishes connections between
elements in all directions and has no beginning or end point. Rhizomes are naturally occurring systems. A good example of a rhizome is an animal burrow, with underground tunnels that can connect chambers in a number of different ways. The development of the burrow depends on the conditions that the soil presents for the animal and so the animal responds to its surroundings in the here and now. Unlike the tree that grows in a linear, predefined direction, the rhizome is changeable and develops horizontally owing to momentary interactions between elements. From a Deleuzean perspective, human activity can also be understood rhizomatically, in that human beings are also affected by their surroundings. The environment and encounters human beings experience affect them in unforeseeable ways and produce new emotions and thoughts in the mind (Deleuze & Guattari 1983, 1-58).

Sermons may also be approached rhizomatically. They are affected by given surroundings such as genre, church holiday, the time of the year, and the specific congregation; and the characters included in the sermons connect to a range of contexts, whether biblical figures, politicians, radio hosts, musicians or authors. A sermon facilitates an encounter between characters who carry with them different cultural connotations. In this sense, sermons as texts are an ongoing, changeable production that establishes and dissolves interactions between different spheres of culture through the characters that pastors choose to include or exclude.

To approach sermons as a rhizome, we read our sermon corpus not as a chronological text collection but as a network, as a momentary and multidimensional snapshot of preaching in the ELCD from 2011 to 2016. The sermon characters are therefore not bound to the individual sermon but are instead singular phenomena understood through their connections to other characters. Each character stems from a particular context, a cultural domain. However, in the sermons, the characters are configured into new contexts through encounters with other characters.

**Methods**

The character gallery we study is derived from a corpus of 11,955 Danish ELCD sermons from 95 different pastors. The corpus was sampled directly from pastors and is accompanied by metadata concerning the context of each sermon (the date, holiday, parish size and diocese as well as the pastor’s pseudonym, gender, age and place of
Collective Testimonies to Christianity and Time

education). Approximately 96% of the sermons in the corpus were written between 2011 and 2016, whilst a limited number were written on either side of this period. There is also a small number of sermons for which we have no date (approximately 0.02% of the entire corpus). For the present study, we chose to draw on the full range of data available.

Representing characters: From singular text components to an interactive network

To establish a character gallery, we developed a pipeline for extracting and identifying characters in the corpus and for representing and interpreting the relations between these characters in a rhizomatic network. This involved combining computational tools with manual analyses. We used a computational method known as Named Entity Recognition (NER) (Jurafsky & Martin 2009) to extract all the references to specific individuals in the corpus (NER automatically extracts entities with proper names, such as people, places, and institutions, from a text). In this study, we focus only on people. Following this automatic extraction, we manually coded the extracted entities by labelling them with metadata regarding their originating contexts. Overall, we distinguished between whether or not characters originated from biblical sources. We further sub-coded these characters to differentiate between Old and New Testament characters as well as between the cultural contexts of non-biblical characters. This approach gave us a thorough insight into the characters populating the corpus.

To understand how these characters from different contexts were related to each other in the sermons, we performed a Social Network Analysis or a Network Analysis to visualise the relational patterns of the characters throughout the entire collection. In recent years, this method has been applied with great success across a range of topics in the humanities, including the structure of dramatic texts (Stiller et al. 2003; Fischer et al. 2017) and 16th-century protestant letter writing in England (Ahnert & Ahnert 2015). In Network Analysis, each unique entity in a sermon is considered a node in a network. Each of these nodes is joined to other nodes that appear in the same sermon by an edge. In our network

---

72 For an in-depth presentation of this sermon corpus (sampling, annotation and archiving procedures) and basic corpus statistics, see Agersnap et al. 2020.

73 A more detailed overview of the computational methodology can be found at the online repository for this article, located here [INSERT LINK]. For the PhD assessment committee: Contact supervisor Kirstine Helboe Johansen (kp@cas.au.dk) for access to this information.
analysis, we used a community detection algorithm to more clearly show the internal structure of the network (Blondel et al. 2008). Put simply, this involves partitioning the network into smaller sub-units or communities, which are understood as clusters of highly interconnected nodes. Characters in the network are assigned a modularity class (MC) that indicates the particular sub-group they belong to and, hence, other characters with whom they are most strongly connected. The modularity class parameter thus serves as a bottom-up grouping of characters that we initially categorised in a top-down manner. By using it, we were able to understand whether pastors reconfigured certain characters and their connotations into new domains in their sermons.

**Entity extraction and labelling**

The NER identified 12,357 unique personal entities in the corpus. However, in some cases, the NER incorrectly tagged certain linguistic features as referring to people. We wished to concentrate on the most common characters and therefore only included entities that occurred ten or more times in the corpus. This excluded many of the incorrect tags just mentioned and also removed uncommon entities. With this criterion, we reduced our data considerably to 966 entities.

Many of these 966 entities actually referred to the same characters appearing with different spellings, misspellings and possessive forms. We therefore collapsed these variations into one unique character. For example, we collapsed the extracted entities *Hitler*, *Hitler’s* and *Adolf Hitler* in to one character under the name *Adolf Hitler*. For the names *Paulus (Paul)* and *Saulus (Saul)*, we collapsed the entities in to one character, *Paulus/Saulus*, but we maintained both names in the final representation of this entity in order to acknowledge possible reasons for pastors choosing one name over the other in the data. If a character’s names are separated with a forward slash (/), this indicates that a character can appear in the sermons under two names. By concatenating entities and discarding linguistic phenomena that the NER tagger had incorrectly identified as personal entities, we ended up with a final dataset of 600 unique characters.
<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>B</td>
<td>Biblical character</td>
</tr>
<tr>
<td></td>
<td>B_a</td>
<td>Ambiguous biblical character</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Non-biblical character</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Proper nouns</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Social role</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>Male character</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Female character</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Character unidentifiable as male or female</td>
</tr>
<tr>
<td>Time</td>
<td>OT</td>
<td>Old testament character</td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>New testament character</td>
</tr>
<tr>
<td></td>
<td>ONT</td>
<td>Character or name in the old and new testament</td>
</tr>
<tr>
<td></td>
<td>CON</td>
<td>Character who was alive in 2011-2016</td>
</tr>
<tr>
<td></td>
<td>PAST</td>
<td>Character who was not alive from 2011-2016</td>
</tr>
<tr>
<td>Nationality</td>
<td>DK</td>
<td>Danish character</td>
</tr>
<tr>
<td></td>
<td>NOR</td>
<td>Icelandic, Finnish, Swedish, Norwegian, Faroese or Greenlandic character</td>
</tr>
<tr>
<td></td>
<td>EUR</td>
<td>European character</td>
</tr>
<tr>
<td></td>
<td>NAM</td>
<td>US or Canadian character</td>
</tr>
<tr>
<td></td>
<td>GLO</td>
<td>Non-European and non-North American character</td>
</tr>
<tr>
<td>Domain</td>
<td>POL</td>
<td>Politician</td>
</tr>
<tr>
<td></td>
<td>LIT</td>
<td>Fictional writer</td>
</tr>
<tr>
<td></td>
<td>MUS</td>
<td>Musician</td>
</tr>
<tr>
<td></td>
<td>ART</td>
<td>Artist within the visual arts</td>
</tr>
<tr>
<td></td>
<td>THE</td>
<td>Theologian</td>
</tr>
<tr>
<td></td>
<td>ACT</td>
<td>Actor</td>
</tr>
<tr>
<td></td>
<td>ROY</td>
<td>Royal character</td>
</tr>
<tr>
<td></td>
<td>REA</td>
<td>Character from reality show</td>
</tr>
<tr>
<td></td>
<td>TV</td>
<td>TV personality (show host, anchor etc.)</td>
</tr>
<tr>
<td></td>
<td>RAD</td>
<td>Radio personality (show host, anchor etc.)</td>
</tr>
<tr>
<td></td>
<td>SOC</td>
<td>Social media personality (bloggers, youtubers etc.)</td>
</tr>
<tr>
<td></td>
<td>SPO</td>
<td>Athlete</td>
</tr>
<tr>
<td></td>
<td>BUS</td>
<td>Personality associated with business communities</td>
</tr>
<tr>
<td></td>
<td>ACA</td>
<td>Academic</td>
</tr>
<tr>
<td></td>
<td>DEB</td>
<td>Public voices or debaters</td>
</tr>
<tr>
<td></td>
<td>FIC</td>
<td>Fictional character ex. from literature, movies, tv shows etc.</td>
</tr>
<tr>
<td></td>
<td>LEG</td>
<td>Legendary character</td>
</tr>
<tr>
<td></td>
<td>OUT</td>
<td>Unidentifiable character</td>
</tr>
</tbody>
</table>

*Table 1. Coding manual*
We constructed a coding manual for labelling the characters, which is shown in Table 1 above. We coded each character as one of four types. B characters are characters from biblical sources; N characters are non-biblical but specific characters (for example, C.S. Lewis); P characters are non-specific characters with generic proper names (for example, Anders); and R characters are named according to a social role (for example, Enken (the widow) or Kongen (the King). B_a characters and N_a characters are entities that most likely refer to biblical characters or non-biblical specific characters due to their high frequency scores but that are also common Danish names – like Maria (Mary) or Johannes (John). All character types are gender coded, B characters and N characters are time coded, and N characters are further nationality and domain coded. Domain coding provides information about the cultural domain in which N characters originate.74

Social network analysis

Having completed the NER, we were left with a structured dataset consisting of named individuals and their character type, all of which were linked to the sermons in which they appeared. The next step was to join these characters together as a network, which entailed constructing a data set in which every connection between two nodes was represented. Thus, for sermon pr1000_76 in Table 2 below, the nodes are Eva, Adam, and Jesus/Jesus Kristus.

---

74 For the PhD assessment committee: Contact supervisor Kirstine Helboe Johansen (kp@cas.au.dk) for access to an appendix holding all characters including labels.
Table 2. Example of nodes in individual sermons

These nodes can be joined together to form an edgelist, showing every possible combination of the pairs as follows:

| Eva   | -- -- | Adam          |
| Eva   | -- -- | Jesus/Jesus Kristus |
| Adam  | -- -- | Jesus/Jesus Kristus |

We automated the process of creating an edgelist using a short computer program.⁷⁵ Across all the sermons, the top five pairings between characters can be seen below in Table 3.

---

⁷⁵ Most of the network analysis was performed using the open-source software Gephi (Bastian et al 2009). As with the NER process, additional information can be found online [INSERT LINK]. For the PhD
Table 3. Example of an edgelist

<table>
<thead>
<tr>
<th>Character A</th>
<th>Type</th>
<th>Character B</th>
<th>Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jesus/Jesus Kristus</td>
<td>undirected</td>
<td>Johannes</td>
<td>1938</td>
</tr>
<tr>
<td>Jesus/Jesus Kristus</td>
<td>undirected</td>
<td>Paulus/Saulus</td>
<td>1794</td>
</tr>
<tr>
<td>Jesus/Jesus Kristus</td>
<td>undirected</td>
<td>Kristus</td>
<td>1495</td>
</tr>
<tr>
<td>Jesus/Jesus Kristus</td>
<td>undirected</td>
<td>Maria</td>
<td>1343</td>
</tr>
<tr>
<td>Jesus/Jesus Kristus</td>
<td>undirected</td>
<td>Peter/Simon Peter</td>
<td>1283</td>
</tr>
</tbody>
</table>

The first and third columns show the names of the individual nodes. The fourth column shows the number of times these entities appear together in a sermon, and the second column indicates that this relationship is undirected. The full edgelist shows not only a long list of binary relations but can be conceptualised as a complex network of overlapping edges and nodes. The number of connections each node has is known as degree, which indicates how well connected a character is in the network. For the purposes of our current data, this means that the algorithm shows specific communities of characters in the sermons based on their patterns of co-occurrence. We experimented and fine-tuned the parameters to return the most coherent groupings from a human perspective, which finally resulted in ten distinct modularity classes. However, it remains unclear how to interpret these mathematically extracted clusters in a coherent and human-friendly way. In the following section, we present some of the main characteristics of each modularity class in the network, identifying certain clear patterns and tendencies in the sub-groupings.

assessment committee: Contact supervisor Kirstine Helboe Johansen (kp@cas.au.dk) for access to this information.

76 All edges between nodes in this analysis are undirected. An example of a directed edge is a situation in which we could say, for example, that Jesus was talking to Paul. This directionality is not a feature of our current data.
The social worlds of sermons

Our methodological pipeline provided us with insights into how the social worlds of sermons are structured – both in terms of the distribution of character types and the connections between character types. In Figure 1 and Figure 2 below, we see two representations of biblical (B/B_a) characters, non-biblical (N/N_a) characters, proper names (P characters) and social roles (R characters) in our data.

Figure 1. Total number of biblical characters, non-biblical characters, proper nouns and social roles in the corpus

Figure 2. Number of unique biblical characters, non-biblical characters, proper nouns and social roles in the corpus
Figure 1 shows that, according to raw frequency, biblical characters by far outnumber non-biblical characters in the corpus; whereas Figure 2 shows that the group of non-biblical characters contains many more unique characters than the group of biblical characters. Moving closer to the varying class of non-biblical characters as represented in Figure 3 below, we find a distinct majority of theological characters. The character gallery of Danish sermons is thus primarily populated by biblical figures and characters who actively engage or have engaged with the biblical sources. Such a finding is unsurprising for this type of corpus, since pastors are expected to interpret biblical sources in their sermons. From the remaining cultural domains, we find that pastors seem to prioritise what might be called “high culture” by focusing primarily on literature (LIT), politics (POL) and art (ART). In comparison, characters from more popular cultural platforms, such as radio (RAD), television (TV) or social media (SOC), are represented to a much lesser degree.

![Figure 3. Distribution of cultural domains among non-biblical characters – counts of unique characters](image)

77 Some labels, such as RAD, REA or SOC, are not plotted in Figure 3, due to limited representations in the data.
Table 4. Sample of seven modularity classes

<table>
<thead>
<tr>
<th>Modularity Class</th>
<th>Nodes</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC0</td>
<td>10</td>
<td>Main characters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jesus/Jesus Kristus (Jesus/Jesus Christ), Kristus (Christ), Paulus/Saulus (Paul/Saul), Johannes (John), Peter/Simon Peter, Jakob (Jacob), Josef, Maria (Mary), Stéphane Hessel, Adriana</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Herren (the Lord), Martin Luther, N.F.S. Grundtvig, Søren Kierkegaard, Esajas (Isaiah), Gud (God), David/Kong David (David/King David), Herodes (Herod), Faderen/Fader (the Father/Father), Augustus/Kejser Augustus (Augustus/Augustus the emperor)</td>
</tr>
<tr>
<td>MC1</td>
<td>54</td>
<td>Nativity and theology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eva (Eve), Kain (Cain), Djevelen (the Devil), Adam, Abel, Erik Hagens, Niels Bohr, Felix, Primo Levi, Randi</td>
</tr>
<tr>
<td>MC2</td>
<td>10</td>
<td>The fall of man</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moses, Hans Anker-Jørgensen, Judas, Markus (Mark), Stefanus (Stephanus), Messiah (Messiah), Isak, Henrik, Jens, Andreas (Andrew)</td>
</tr>
<tr>
<td>MC3</td>
<td>402</td>
<td>Versatile characters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.C. Andersen/Andersen, Christian/Kristian, Martin A. Hansen, Maria Magdalene (Mary Magdalena), Thomas, Niels/Nils, Storm P/Storm, Søren Ulrik Thomsen, Søren, Benny Andersen</td>
</tr>
<tr>
<td>MC4</td>
<td>84</td>
<td>Scandinavian literature and narratives of otherness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Matthaus (Matthew), Lukas (Luke), Johannes Doberen (John the Baptist), Simon, Abraham, Martha, Lazarus, Samaritaneren (the Samaritan), Faust, Nemo</td>
</tr>
<tr>
<td>MC5</td>
<td>11</td>
<td>Other biblical stories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Matthaus (Matthew), Lukas (Luke), Johannes Doberen (John the Baptist), Simon, Abraham, Martha, Lazarus, Samaritaneren (the Samaritan), Faust, Nemo</td>
</tr>
<tr>
<td>MC7</td>
<td>17</td>
<td>Politics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adolf Hitler, Nelson Mandela, Desmond Tutu, Anders Breivik, Barack Obama, Dietrich Bonhoeffer, Mao Zedong, Karl Ove Knausgaard, Stalin, Osama Bin Laden</td>
</tr>
</tbody>
</table>

Our network data allows us to see how the individual characters relate to each other, which gives us a more nuanced understanding of the characters’ roles in the sermons. The community detection algorithm returned ten distinct classes. Of these, three classes consisted of only two nodes with relatively few edges, which we excluded from our analysis. The remaining classes are summarised and named in Table 4 above\(^\text{78}\).

According to the network analysis, the most central characters, both in terms of frequency scores and degree, belong to MC0. We have therefore named this community “Main characters”. Apart from two characters – Stéphane Hessel and Adriana, who have a low degree – all the characters in this community are New Testament figures.

The modularity class “Nativity and theology”, MC1, consists primarily of biblical figures and then of theologians. Many of the New Testament characters appear in the nativity narratives. In parallel, the Old Testament characters are mainly prophets or characters known in the Christian tradition for forecasting the birth of Jesus Christ – for example,

---

\(^\text{78}\) For the PhD assessment committee: Contact supervisor Kirstine Helboe Johansen (kp@cas.au.dk) for access to the full network data, including the modularity classes in full.
Esajas (Isaiah) and David/Kong David (David/King David). The theologians in this modularity class are mainly past authoritative figures and/or prominent Danish theologians.

In the modularity class “The Fall of man”, MC2, the most central characters are a coherent group of Genesis characters from the same mythological context. This modularity class is also connected to a World War II/Post-World War II context through the author and concentration camp prisoner Primo Levi, the Danish World War II liberation fighter Erik Hagens, and the Danish physicist Niels Bohr, who were involved in the US Manhattan project.

Modularity class MC3, “Versatile characters”, is the largest modularity class of 408 characters, encompassing the majority of unique characters in the data. It is worth noting that the vast majority of P characters, R characters and Old Testament figures in the corpus are found in this class. Relative to the size of MC3, only few characters have high degrees (such as Moses and the Danish theologian and hymn writer Hans Anker Jørgensen). Primarily, however, this group consists of characters with only a few connections to other characters in the sermons.

In MC4, the main character type is N characters (41), which mostly consists of authors (LIT: 11) mainly from Scandinavia. Other considerable N characters are legendary figures (LEG: 6) and fictional characters (FIC: 4). Taken together, these three types indicate that narrative contexts are an important aspect of MC4. Only eight B characters appear in this group: Maria Magdalene (Mary Magdalena), Levi/Levi tolderen (Levi/Levi the publican), Thomas, Mikael/Michael, Tolderne (The publicans), Talsmanden (The Comforter), Predikeren (The Preacher) and Tiggeren (the Begger). We have named this group “Scandinavian literature and narratives of otherness”.

We have named MC5 “Other biblical stories” as this is a small group with eight B characters and three N characters with very low degrees. The biblical characters are figures that primarily relate to each other during the Trinity season in the Danish lectionaries.

MC7 is the only modularity class that does not include any biblical figures. It instead contains fourteen N characters, including political figures (POL: 7), theologians (THE: 2), and terrorists coded as criminals (CRI: 2). The remaining N characters are the Norwegian author Karl Ove Knausgaard, the psychoanalyst Freud, and the former chair
of the Danish Atheist Society, Anders Stjernholm. We call this cluster of characters “Politics”, as, although they may not be coded POL, the majority of them seem to be politically motivated or engaged in some way.

The network analysis shows that the characters we assigned to domains top-down seem to be reconfigured into new domains bottom-up. These reconfigurations follow a Deleuzian logic that previously known phenomena can create new understandings based on their connections to other phenomena. From our subsequent manual analysis, we find that biblical, literary and political figures in particular reconfigure into distinct domains in interesting ways, while a large number of characters are assigned to one group, MC3.

**Narrative and thematic dynamics in sermons**

*Biblical narrative worlds*

The network analysis showed that biblical figures partake in different modularity classes, indicating that this character class is not a coherent group. Their roles vary depending on the other characters with whom they appear. However, we find that biblical characters stand out as both the most frequently mentioned and the most central character type in the corpus, since they tend to form the nodes of largest degrees in the network. This is particularly true for the group “Main characters”, in which the characters do not denote specific biblical contexts but are well connected to many other characters in the corpus and highly likely to occur with them. An interesting dynamic for New Testament characters occurs in the corpus. The groups “Nativity and theology”, “Scandinavian literature and narratives of otherness” and “Other biblical stories” seem to represent biblical narrative worlds in which Jesus, who belongs to “Main characters”, enters as the protagonist.

In “Nativity and theology”, the biblical character gallery represents the prophecies and the nativity story in which Jesus/Jesus Kristus, Maria and Josef – who belong to the group “Main characters” – are central figures. It is worth noticing that all God representations – Gud, Herren and Fader/Faderen – are assigned to this narrative context and are not a part of “Main characters” with the New Testament protagonists. As an agent, God seems to be part of the narrative background that legitimises Jesus’ centrality in the sermons, in company with nativity characters and Old Testament prophets. This is a narrative setting in which God is the initiative agent. Overall, this modularity class seems to facilitate a
distinct encounter between Old Testament, New Testament and authoritative theologians, most likely mediated by God as a central agent.

In contrast to “Nativity and theology”, the groups “Scandinavian literature and narratives of otherness” and “Other biblical stories” represent narratives from Jesus’ adult life. In the first of these, we find a small group of characters from the gospels personifying societal outcasts, whom Jesus cares for: Maria Magdalene, Levi/Levi tolderen, Tolderne and Tiggeren. Along with these characters, we find a large group of authors who are known for writing about and enhancing what is aberrant, odd or complex. For example, Benny Andersen writes existential and humorous poetry about peculiarities in everyday life; Astrid Lindgren writes about children being different from the norm (Pippi Longstocking) or vulnerable and self-sacrificing (The Brothers Lionheart); H.C. Andersen writes about the boy Kaj, who was tempted and corrupted but saved by Gerda (The Snow Queen); and Fyodor Dostoevsky writes about morality, doubt and sacrifice (The Brothers Karamazow and Crime & Punishment). This modularity class thus points to biblical narratives being told through and in support of literary narratives. In contrast, the group “Other biblical stories” is primarily defined by biblical characters. These stories represent biblical narratives that comprise gospel sources for sermons written in the Danish Church Calendar’s period of Trinity:79 the parable of the good Samaritan (Luke 10,25-37), the story of Martha’s and Mary’s encounter with Jesus (Luke 10,38-42), the parable of the rich man and Lazarus in the arms of Abraham (Luke 16,19-31), and the story of the death and resurrection of Lazarus, where Martha and Mary also feature (John 11,19-45). The liturgical texts in this period are known for representing many of Jesus’ teachings for living a Christian life, and several of them are from the gospel of Luke, who also appears in this modularity class.

The modularity classes “Nativity and theology”, “The fall of man”, “Scandinavian literature and narratives of otherness” and “Other biblical stories” all enhance distinct biblical narratives. In this structure, the group “Main characters” seems to constitute a narrative centre of the network, whereas the majority of the biblical modularity classes present small narrative worlds that revolve around Jesus as the forecasted saviour, the compassionate or the teacher.

79 The liturgical year of the ELCD is ordered according to different periods, such as Advent, Lent and Easter. Trinity follows immediately after Pentecost and lasts from the first Sunday after Pentecost until the last Sunday in the church year (at the end of November).
In contrast to the other biblical modularity classes, the group “The fall of man” represents a distinct Old Testament narrative context – and the only Old Testament narrative that is so distinctly demarcated and highlighted in the network. Since the Old Testament characters are in company with a few characters associated with the time around World War II (Erik Hagens, Primo Levi and Niels Bohr), “The fall of man” establishes a narrative about evil and corruption told through Old Testament sources and recent 20th-century historical sources.

The characters in all of the above-mentioned modularity classes appear in thematically delimited sub-networks constituting small worlds, which indicates that there is a consensus among pastors that these biblical narratives are particularly important to disseminate and engage with. This would indicate a pastoral consensus on how to frame these narratives. These narratives are sometimes told through biblical sources and sometimes told through contributing characters.

**Politics and morality**

Among our modularity classes, “Politics” stands out as the only sub-group in which we find primarily N characters and no biblical figures. Most of the characters in this modularity class are political figures who seem to come from two different contexts. Adolf Hitler, who is the most central node in this cluster, Mussolini, Stalin and Mao Zedong were dictatorial political leaders adhering to totalitarian ideologies around the time of World War II, whereas Barack Obama, Nelson Mandela and Saddam have been central figures in more recent global politics. Both of these political contexts pose a distance to the context of Danish preaching in 2011-2016 – either temporally (the World War II motif) or spatially (recent global politics). Even though the label POL is a rather common character type among N characters in the corpus, these characters rarely represent national political issues. This forms a contrast with the groups of authors and theologians, who are mainly from Danish or Nordic contexts. In this regard, “Politics” is noteworthy for including only two Danish figures, Omar and Anders Stjernholm.

Another group of characters in this modularity class are terrorists – Anders Breivik, Osama Bin Laden and Omar80 – whose terrorist attacks have been ideologically framed.

---

80 Initially, we labelled Omar as a P character, but, due the other characters in this modularity class, we infer from the network analysis that he is most likely the terrorist who attacked the synagogue and the cultural centre “Krudttønden” in Copenhagen and killed two people in 2015.
Combined with the dictatorial leaders, politics in this modularity class thus connotes violence, strife and distress. The terrorist motif also emphasises a societal agenda that has been prominent in a global context since the 9/11 attacks in New York and remained prominent in the 2011-2016 period. In this regard, the motif also resonates with the political figures representing contemporary global politics.

A remaining group in this modularity class seems to be distinctly associated with the two political contexts. The archbishop Desmond Tutu and Nelson Mandela share a context as activists for human rights in South Africa. These two characters thereby pose a remarkable contrast to the politically motivated terrorists and dictators, and the political domain in the sermons thus seems to represent a polarised ethical domain about good and bad. The other theologian in the modularity class is Dietrich Bonhoeffer, who was also part of a resistance movement against oppression – in his case, against Hitler’s dictatorship in Germany. Karl Ove Knausgaard has explored, in his writing, Hitler’s life and psyche in an attempt to understand the dictator and his dispositions for violent actions (Knausgaard 2013). Knausgaard pondered the personality of Anders Breivik in a similar way after the Norway massacres in 2011 (Knausgaard 2014), and Knausgaard therefore seems to be part of the cluster due to his literary engagement with distinct characters and contexts in it.

It is interesting that the chair of the Danish Atheist Society (DAS) in 2016, Anders Stjernholm, also appears in this group. Stjernholm was not publicly known until March 2016, when DAS began a campaign encouraging people to renounce their membership of the ELCD. His atheist agenda seems to be strongly problematised given the other characters in this modularity class. Along with Knausgaard, Breivik and Omar, he adheres to a group of characters who gained public attention between 2011 and 2016. This observation suggests that this modularity class provides a context in which problematic issues in the eyes of ELCD pastors are framed.

Taken together, the figures in “Politics” seem to form a distinct domain in the sermons in which mainly political contexts are reconfigured into raising this-worldly narratives about good and evil – narratives that are represented without distinct anchoring in biblical sources. The moral dimension, which becomes a part of the political domain, is not unique for “Politics”. In the modularity classes “Scandinavian literature and narratives of otherness” and “Other biblical stories”, the motifs of care and otherness and the focus on
Jesus’ teachings also appear to be represented within a moral frame – about what is right and wrong. However, the most similar class to “Politics” is “The fall of man”. Here, a narrative about evil is told through biblical sources, supplemented by few World War II figures. The history of World War II and stories of evil are thus contexts that are, in some cases, associated with the biblical sources (“The fall of man”) and, in other cases, associated with this-worldly events in recent history (“Politics”). As such, World War II seems to mark an important historical period in the sermons.

The anonymised mass

The small clusters of biblical figures and the group of political figures in particular appear as consensus groups. Pastors seem to agree that the characters in these groups are substantial parts of specific and, in their opinion, important stories. In contrast, the modularity class “Versatile characters” stands out as a large group of several nodes with low degrees and low frequencies in the corpus. This characteristic reduces the distinctiveness of the character gallery within this group. Their main similarity to each other appears to be that they are secluded from the consensus groups; they are peripheral characters in the corpus. Therefore, this group stands in stark contrast to “Main characters”, as the versatile characters are the least likely to take part in the established narrative worlds in the network. The peripheral role of these characters also pertains to the character types, which consist primarily of P characters and R characters. Since we cannot infer specific identities from their names, these characters become anonymous figures in the network, who lack distinct and consistent roles in the corpus. Their anonymity seems to group them together and define this modularity class.

More than half of the Old Testament characters in the data appear in the group “Versatile characters”. In “Nativity and theology”, the Old Testament characters represent narratives used to forecast and legitimise Jesus Christ as saviour. The most central Old Testament characters in the corpus seem to occur in “The fall of man”. In contrast to these more distinct clusters, the many Old Testament characters that appear in “Versatile characters” are part of a group that tend towards periphery and anonymity, a diffuse group of left-over characters. This indicates that Old Testament characters generally play a minor role in the sermons compared with the New Testament characters. Furthermore, this finding
points to different accentuations of the biblical texts in the lectionaries, as pastors tend to highlight gospel narratives and epistles more than Old Testament pericopes.

We also find a large number of contemporary N characters in the class “Versatile characters”. Therefore, contemporary characters rarely seem to gain distinct or central roles in the corpus, and pastors do not appear to reach a consensus about which parts of contemporary culture to engage with – apart from when it comes to contemporary antagonists, i.e. the group “Politics”. In that modularity class, we find that some new and contemporary characters are placed within a clearly delimited group about good and evil. In contrast, “Versatile characters” might represent an underlying and ongoing negotiation among pastors of which ‘new’ elements to integrate into sermons. The N characters in this group are likely to be the most floating characters in the corpus. Some are prospects that might obtain a place in a consensus group in future sermons – and others might disappear altogether and will be replaced by other prospects.

**Discussion: The rhizome revisited**

In *A thousand plateaus*, Deleuze and Guattari write: “A rhizome doesn’t begin and doesn’t end, but is always in the middle, between things, inter-being, intermezzo. The tree is filiation, but the rhizome is alliance, exclusively alliance” (Deleuze & Guatteri 1983, 57).

In our study, we analysed the sermon corpus as a network based on a rhizomatic logic. However, our analyses have re-posed the question of whether sermon characters illustrate rhizomatic alliance or arborescent filiation – whether sermons are production or reproduction. The modularity classes have provided different insights to this question.

The biblical figures who dominate the modularity classes “Nativity and theology”, “The fall of man” and “Other biblical stories” seem to cluster in the modularity classes in terms of familiarity over difference by being part of the same biblical narratives or part of narratives from the same period of the liturgical year. The group “Main characters” does not point to specific narratives but unites characters with roles of narrative or authoritative weight in the New Testament. The co-occurrences of the biblical characters thus follows an arborescent rather than a rhizomatic mode. They seem to form what we might call *trees of consensus*, in that pastors broadly agree that these characters and their narratives are particularly important to reproduce. They seem to be kept stable and predominantly
within a biblical universe. Due to the high frequency of their appearance and the tightly
knit clusters they form, we would likely see a similar picture of biblical characters if we
examined a sermon corpus collected before or after the period we studied. In
“Scandinavian literature and narratives of otherness”, we find a group of biblical
characters who share narrative themes but are in company with characters from
completely different contexts. In our interpretations, we have focused on the literary
context, though this modularity class could be further examined in terms of the legendary
(LEG) and aesthetic (MUS and ART) contexts that are also present. The formation of this
modularity class is rhizomatic in the sense that the biblical characters are unfolded in
dialogue with other cultural contexts, indicating that stories of biblical outcasts in
particular are told through other sources than just biblical sources. The various domains
are thus configured into a new setting in sermons and the stories facilitated anew.

The modularity class “Politics” is particularly interesting because it has no biblical
characters. However, since this modularity class is primarily anchored in the World War
II motif with Hitler as the most frequent and most connected character, it seems to be
reproductive of narratives of evil and strife particularly connected to a specific historical
period. The World War II context seems to have become a rooted narrative, thus
following an arborescent logic. Meanwhile, we see that this modularity class is
particularly capable of encompassing characters who are new on the public stage
provided they are somehow attached to the rooted narrative. In this sense, this modularity
class, we have characterised as a distinct ethical domain, might be understood as a kind
of tree rhizome, since its context is rooted, while it is able to encompass new elements
representative of the period 2011-2016. These new characters might not be awarded a
continuous role in future sermons and could be replaced by new characters who configure
narratives about evil and ethics in the given period.

A significant number of the “new” characters in the corpus can be found in the largest
modularity class “Versatile characters”. Here we find characters of all types and domains
from our manual labelling: B characters, N characters, R characters and, in particular, P
characters. In this sense, characters here clearly operate on rhizomatic terms. They
compose the widest sub-network and their connections seem to be contingent and
momentary, with the majority of the characters appearing infrequently and with a low
number of connections. It is therefore difficult to infer any common features or significant
relationships between these characters, other than the fact they are peripheral. However,
this does not mean that they should be considered obsolete. On the contrary, “Versatile characters” includes characters that have not (yet) obtained consensus among pastors with respect to their roles in the sermons, and it thereby reveals a field of potential movement and renewal. Rhizomatically, it remains to be seen which routes are blind alleys or serve only momentary purposes and which routes lead to stable incorporation in future networks. This modularity class thus appears to represent the creative test lab of the sermons.

**Conclusion: Characters in a chain of memory**

The characters in the sermon corpus seem to have two main functions: they are invited into sermons in order to enhance and illuminate the story about Jesus, or they are invited into sermons in order to expound ethical or existential issues. These main functions unfold in a mixture of arborescent and rhizomatic thinking. The New Testament characters accompanied by selected Old Testament characters and N characters appear to consolidate a number of thematic trees. Around these, a myriad of versatile characters offer and test new hermeneutical and illustrative potentials in rhizomatic movements. From our Deleuzian analysis of their sermon production, the pastors appear focused on both passing on what seems collectively defined as the main stories and on telling these stories in contemporary ways – though perhaps more the former than the latter. Our analyses also demonstrate how preaching, besides unfolding in the dialogue between preacher and congregants, unfolds a collective dialogue between bible and history. The many sermon documents therefore raise awareness of sermons as the pastors’ collective theological testimonies to time – that is, as historical documents that are intertextually linked.

French sociologist Danièle Hervieu-Léger has argued that the primary role of the old majority religions in late modern society was to maintain strong links to the religious past and tradition in order to ensure its continuity in a societal structure where change and development is otherwise valued. She thus describes religion as a chain of memory (Hervieu-Léger 2000, 83-100). One could argue that adapting tradition to the present age has always been an innate feature of preaching and is not limited to a late modern societal context: every Sunday, pastors are expected to expound and interpret a biblical text for the assembled congregations. As such, an external motivation encourages pastors to
incorporate biblical narratives, which explain their centrality in the sermon corpus. After all, interpreting the bible is at the core of preaching. However, in our study, we have seen that selected distinct biblical narratives receive most attention in the sermons. It therefore seems there is an underlying yet collective consensus among pastors about which narratives are particularly important to enhance and thereby which parts of tradition are particularly important to preserve. Similarly, the World War II motif in the sermons demonstrates that pastors not only preserve and reproduce a religious tradition but also incorporate a chain of memory anchored in a historical event – a memory chain composed of other links than the links to biblical narratives and themes. In these sermons, then, religion is not just a matter of reproducing religious tradition but also of reproducing a narrative anchored in history and society. Following this line of thought, we might interpret the “Versatile characters” as disparate parts without links attaching them to a chain. However, even though these figures are not currently discernable as strong links, this picture might change in future sermon productions.

References

2020   Sermons as data: Introducing a corpus of 11,955 Danish sermons. – *Journal of Cultural Analytics* 12, 1-27. DOI: 10.22148/001c.18238

Ahnert, R. & S. Ahnert  

Andersen, Peter B., Jakob Erkmen & Peter Gundelach  

Bastian, Mathieu, Sebastien Heymann & Mathieu Jacomy  
Blondel, V.D., J. Guillaume, R. Lambiotte and E. Lefebvre  

Christoffersen, Lisbet  

Deleuze, Gilles & Félix Guatteri  

Fischer, F., G. Dazord, M. Göbel, C. Kittel and P. Trilcke  

Fludernik, Monika  

Gaarden, Marianne & Marlene Ringgaard Lorensen  
2013  Listeners as Authors in Preaching: Empirical and Theoretical Perspectives. – *Homiletic (Online)* 38(1), 28-45.

Gaarden, Marianne  

Herman, David  
Hervieu-Léger, Danièle

Jurafsky, D. & J.H. Martin
2009  *Speech and Language Processing, 2nd ed.* Prentice-Hall: Upper Saddle River, NJ.

Knausgaard, Karl Ove

Langacker, R.W.

Lorensen, Marlene Ringgaard

Lorensen, Marlene & Gitte Buch-Hansen
2018  Listening to the voices: refugees as co-authors of practical theology. – *Practical Theology* 11(1), 29-41. DOI: 10.1080/1756073X.2017.1415577

McClure, John

Moretti, Franco
Nielsen, Marie Vejrup

Nielsen, Marie Vejrup & Lene Kühle

Polkinghorne, Donald

Semino, E.

Stiller, J., D. Nettle and R.I.M. Dunbar
2003 The Small World of Shakespeare's Plays. – Human Nature 14(4), 397-408. DOI: 10.1007/s12110-003-1013-1

Vosmar, Jørn
Chapter 8

- “What is love...?” – A study of thematic constructions in 11,955 Danish sermons (Article 3)

Abstract
Writing sermons in the Evangelical-Lutheran Church in Denmark (ELCD) is a shared practice, which means that pastors collectively produce a comprehensive text material, when they prepare their weekly sermons. This article studies a corpus of 11,955 ELCD sermons as a collective text production, as we investigate the role of theological discourses in emergent thematic fields in the corpus. With the aid of computational tools from the field of digital humanities, we explore the overall thematic framework of the corpus, the construction of specific theological concepts (love and sin) and the relationship between thematic constructions in the corpus and thematic content in the liturgical texts that pastors expound in sermons.

Introduction
When they preach on church holidays, pastors perform a significant interpretative task of making Christian narratives and teachings pertinent for assembled congregations in the present. As a consequence of this practice, pastors synchronously produce a large number of sermons on a weekly basis. In a Danish context, the vast majority of pastors write a full sermon manuscript, and these documents provide a unique insight into how pastors collectively on behalf of their church represent symbolic language and theological discourse. In this article, we explore the semantic fields of a collection of 11,955 sermons from the Evangelical-Lutheran Church in Denmark, primarily in the period 2011-2016, by attending to thematic constructions and the role of theological discourses. Our approach is informed by the digital humanities, and in our study, we make use of three computational tools: word count analysis, semantic network analysis and topic modelling. Our interest in sermons stems from the sociological study of religion, but we consider our study to be in line with the empirical turn in homiletics, which has paved the way for descriptive analyses of preaching. So far, this turn has in particular been interested in the sermon as a component in a communicative situation, whether the focus has been on the
speaker or more recently the listener\textsuperscript{81}. In our approach, we do not attend to the communicative intentions or consequences of preaching. Instead, we perceive the production of sermon documents as a collective testimony to Christianity and time.

The Evangelical-Lutheran Church in Denmark

The Evangelical-Lutheran Church in Denmark (ELCD) constitutes the majority religion in Denmark, and is affiliated with the state through section 4 in the Danish constitution, declaring: “The Evangelical-Lutheran Church is the Danish people’s church and as such supported by the state”.\textsuperscript{82} The church is organised with ten dioceses divided geographically and supervised by ten bishops. Each diocese is divided into deaneries, and each deanery consists of local parishes. Membership in the ELCD is high, but in steady decline primarily due to demographic changes. On 1 January 2020, 74.3\% of the Danish population were members compared to 80.4\% on 1 January 2010. Despite high membership levels, regular attendance at church services is comparably low – about 11\% of the population attend ELCD services at least once a month\textsuperscript{83}. Meanwhile, the church has higher attendance rates in connection with ceremonies framing life events: in 2018, 58.7\% of all newborns were baptised; in 2019, 68.3\% of all youth in the valid age range had their confirmation; and of all the people who died in 2019, 82.7\% had a church funeral\textsuperscript{84}. The theology of the church is generally liberal: women pastors have been ordained since 1948, three out of ten bishops are women, and the performance of same-sex weddings has been allowed since 2012.

The current liturgy used at regular church services in the ELCD was approved in 1992 by the bishops at the time and authorised by the Danish Queen. Pastors are expected to follow this liturgy, but they can choose from a number of variations. The service includes two or three bible readings: from the Old Testament and/or Epistles and from the Gospels. The readings for each holiday are prescribed in two different lectionaries, one for odd and one for even liturgical years. Combined these texts provide a thematic basis for each holiday.

\textsuperscript{82} Translation from Danish: §4. Den evangelisk-lutherske kirke er den danske folkekirke og understøttet som sådan af staten.
\textsuperscript{84} All statistics can be found at: https://www.km.dk/folkekirken/kirkestatistik/
holiday. Although the two lectionaries are not identical, the order of texts represents the narrative of salvation throughout the course of a liturgical year\textsuperscript{85}.

\textbf{Sermons: Collective testimonies to Christianity and time}

The empirical turn within homiletics has paved the way for analysing preaching inductively. This development has largely entailed a shift from a kerygmatic to a communicative understanding of preaching\textsuperscript{86}. Instead of regarding preaching firstly as an event of divine proclamation uttered by the preacher, the empirical turn has increasingly focused on the listeners’ experience of sermons. Danish theologians Marlene Ringgaard Lorensen and Kirstine Helboe Johansen have outlined the status of empirical research within the field of homiletics in the Nordic countries\textsuperscript{87}. Here, the strongest tendencies draw inspiration from the newer North American tradition known as \textit{Other-wise Preaching}, which has been developed in particular in the works of John McClure\textsuperscript{88}. In Other-wise preaching, focus is on the asymmetrical relationship between preacher and listener, meaning that the preacher cannot assume what congregants learn from the sermon, as the congregants come with backgrounds and perspectives different from those of the preacher. In a Nordic context, this approach to sermons has informed studies with an increasing focus on what congregants hear and interpret from the sermon. In Denmark, homiletician and now bishop Marianne Gaarden has studied this phenomenon in her doctoral thesis, arguing that listening to a sermon creates a third room between speaker and listener\textsuperscript{89}. Similarly, in Norway, a collective of researchers has studied how listeners at different ages experience sermons\textsuperscript{90}, and in Sweden, homiletician Jonny Karlsson has examined how different congregations have been addressed in the sermon collection of theologian Gustaf Wingren\textsuperscript{91}. With such studies, the homiletical discipline has become

\textsuperscript{85} Henning Thomsen, Prædikenens plads og indhold, in: Pastoralteologi, eds. Helle Christiansen and Henning Thomsen, Copenhagen 2007, 260-269.
\textsuperscript{87} Marlene Ringgard Lorensen & Kirstine Helboe Johansen, Homiletics in the Nordic Countries, in: Praktische Theologie 55.2 (2020), 88-93.
\textsuperscript{89} Marianne Gaarden, The Third Room of Preaching: The Sermon, the Listener, and the Creation of Meaning, Louisville 2017.
\textsuperscript{90} See the project web-page: \texttt{https://www.mf.no/kom/forskning/foss-forkynnelse-sma-store}
\textsuperscript{91} Jonny Karlsson, Predikans samtal: en studie av lyssnarens roll i predikan hos Gustaf Wingren utifrån Michail Bachtincs teori om dialogicitet, Linköping University 2000.
subject to empirical research with an interest in not only what preaching ought to be, but also how preaching practices actually unfold.

The empirical turn in the Nordic countries has primarily entailed studies of sermons with regard to the communicative situation in church between preacher and congregant. Yet, sermons are not merely subjective realities – not just the intentional proclamation of the preacher and the experienced perception of the listener. Sermons take part in a collective voice, which is documented by the thousands of sermons pastors write every week.92 Literary scholar Franco Moretti uses the term collective to describe the relations between literary works of the same historical period. His description illustrates a shift from approaching texts as unique individual works to understanding them as parts of a larger literary field. Moretti argues: “a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn’t the sum of individual cases: it’s a collective system, that should be grasped as such, as a whole”93. Therefore, a literary work is not just the product of an author’s creative mind, but also a product conceived in a specific cultural and temporal context in history. As such, it gives access to a collective consciousness that may bypass the author’s own intention or awareness – a collective consciousness that we can only study if we examine the individual work as intertextually related to the larger production of works sharing historical context. This perspective in itself is not unique to the field of literary history, but Moretti radicalises the approach: even though literary historians may study hundreds of works of a given period, these collections tend to comprise canonised works, which means that the same “few” literary works are scrutinised closely over and over again by different scholars.94 Instead of reading only a few works, Moretti’s concept of distant reading implies adopting a wider view on literature, by attending to a larger collection of works from a given period in time.95 The idea of distant reading has become important for the field of digital humanities, which since then has contributed with a range of computational tools for large-scale text studies.

94 Ibid., 3-4.
Moretti’s perspective on literature as embedded in a collective system provides new ways of embarking on empirical research of sermons. In parallel to literary works, sermons are conceived in a particular temporal and cultural context, where certain discourses and ideas prevail, and their shared contemporaneity situates the sermons in an intertextual relationship. Furthermore, even more so than for literary works, sermons are embedded in a concrete system due to the liturgical years, whereby sermons are produced in parallel to each other and on a continuous, recurring basis. In a Danish context, the biblical passages prescribed for each holiday in fact provide a tangible intertextual basis for sermons.

The interest in studying sermons in the intersection between theological and contemporary public discourses is not unique in a Nordic context. For example, a research collective including Lorensen and Norwegian homiletician Tone Stangeland Kaufman have examined the influence of the international refugee crisis in 2015 on preaching in Denmark, Norway and Sweden. The study was done on 104 sermons delivered over the course of six months – an impressive amount of text for a close-reading study. However, with Moretti’s more radical understanding of collective text productions, we can broaden the scope of study even more through the aid of a large digital archive of sermons and computational methods. By approaching sermons as a large collective production of texts by pastors, we leave out the communicative purpose and impact of individual orally delivered sermons. What is to be gained, however, is an insight into sermons as a collective testimony to Christianity and time. As such, we are able to investigate prevalent themes, and how theological vocabulary and discourse unfold in a temporally coherent collection of sermons.

Transformational patterns in contemporary Christianity

Our approach to conducting semantic investigations of thematic content and theological dynamics in sermons is inspired by sociologist Linda Woodhead’s concept of the

---

97 Moreover, we are aware that pastors may not have delivered the sermons in church as written in the sermon manuscripts. The document content, however, is still an attestation of pastors’ theological work today, whether pastors’ oral presentation of the sermon reflects the content as stated in the document or not. The sermon document is a concrete imprint that has been made before the service, and a concrete imprint that remains after it.
displacement of theology. She uses the concept to explain how theological expertise has been devalued in public life in European societies, in particular since the 1960s. Due to new power structures, theology as an academic discipline lost both social power and cultural influence. In society, the influence of Christianity thus seems to have shifted from a ‘power from above’ (confessional theology and authoritative dogmas) to ‘a power from below’ (lay peoples’ demands to religion)\(^98\). Recent studies have investigated what these developments mean for the practices in the ELCD. Marie Vejrup Nielsen and Kirstine Helboe Johansen demonstrate how new types of church services, such as Halloween and Valentine’s Day services, are situated in a church context while clearly drawing on mainstream cultural codes in framing them\(^99\). Another study conducted by Nielsen focusing on the profile and motivations of the users of the ELCD shows that a majority of the informants are occasional users, who are interested in attending specific activities in their church, for example baby hymn singing for parents and babies, rather than regular attendance at church services\(^100\). These studies indicate that theological displacements occur in the ELCD, as churches provide new activities that tend to meet the demands, needs or interests of its members rather than providing a setting for intellectual theological discussion. Since these activities are supplementary to the official church calendar and not bound to the liturgical order of church services, they allow a higher degree of diversion from intellectual or doctrinal theological discourse in order to reach specific target groups. For the weekly and authorised church services, the situation is different. The sermon for the official church holidays is a genre that has traditionally been a distinct domain of theological expertise. Where Woodhead uses the term theological displacement to describe a societal dynamic, and where Nielsen and Johansen look at dynamics on an organisational level by attending to supplementary church activities, we investigate which theological dynamics occur in one of the church’s core practices shared by all pastors in the ELCD. To explore this, we look for thematic content inductively in a comprehensive sermon corpus. This approach entails paying close attention to pastors’ shared vocabulary, the configurations of the vocabulary and the role of theological language in these configurations. Through these patterns, we are able to explore the


collective voice of ELCD pastors, and whether they tend to share in theological 
preservations or displacements in their sermons.

Methods and observations

Data and computational tools

The corpus we analyse consists of 11,955 sermons from the ELCD. Each sermon is 
annotated with metadata, including pastor pseudonym, date, holiday, parish size, diocese, 
pastors’ gender, pastors’ year of birth and pastors’ place of education. The sermons were 
submitted by 95 pastors, and they are predominantly from the period 2011-2016 (roughly 
96% of the corpus), while a few sermons are from before or after this period or are lacking 
information about the date.

In our approach to studying themes in sermons, we attend to word representations in the 
text collection and explore the semantic contexts they produce and are part of. In this 
endeavour, we model lexical semantics in terms of the co-occurrence and associative 
structure of the corpus at the level of individual sentences, documents and the corpus as 
a whole. For the analysis, we are less interested in the concrete syntactic context of words 
in individual sermons; instead we focus on semantics in terms of lexical centrality, 
repetition and relations across the entire corpus.

We have applied three tools for our semantic investigations: word count analysis, topic 
modelling and seeded semantic network analysis. Word count analyses provide an 
overview of the vocabulary of the corpus in terms of word frequencies. We thus consider 
a high frequency of some words rather than others as an indicator of central status in the 
collection. As we attend to the high frequency words in relation to each other, we get an 
immediate sense of some of the important themes in the collection. Topic modelling is a 
tool to automatically represent latent themes that run through a large collection of texts. 
A topic model makes the assumption that a document is a distribution over a finite set of 
latent topics, and that a latent topic is a distribution over the corpus lexicon. Furthermore, 
it makes the assumption that the topics within a text have been chosen by the author 
before writing the text, and that the words within a text are realisations of these pre-

101 For a thorough presentation of the sermon corpus (sampling, metadata and basic corpus statistics), see: 
Anne Agersnap, Ross Deans Kristensen-McLachlan et al., Sermons as data: Introducing a corpus of 11,955 
existing topics. With topic modelling, you try to reverse this process and trace the words back to the topics they came from, and consider the documents in terms of their distribution of topics. This can be done using a probabilistic approach, where an algorithm computes the probability that every word within every document belongs to any of the number of topics you are looking for. Thus, in topic modelling, you define how many topics it is reasonable to look for within a given text collection before conducting the analysis. Topic modelling regards words to be central to a given theme by considering how each word in the text collection occurs within every document of the collection. This approach implies relating documents to each other based on their distribution of topics. Whereas word count analyses and topic modelling provide bottom-up approaches for uncovering central terms and their semantic context, a seeded semantic network queries the associative structure of concrete seed terms in the corpus. This technique uses simple neural word embeddings to represent the corpus lexicon and builds an associative graph by using words as nodes and distance in word context as edges. By iteratively applying the algorithm with threshold for graph size, the technique builds a hierarchical graph, where the primary associative level consists of the $m$ most similar terms to the seed term, and the secondary associative level consists of the $n$ most similar terms to the primary associative level. Finally, the edges for all terms of the graph are computed with a distance threshold, and the Louvain method for community detection is used to extract semantic clusters. In our terminology, the repetitive or relational structures that these tools uncover establish semantic fields that can be scrutinised closely. The words in the sermon collection are thus understood as text components facilitating links between the texts. These links exhibit latent patterns that would not be possible to uncover with the naked eye and without the aid of computational tools.

Pipeline: Word count

For our word count analysis, we tokenised and casefolded the entire corpus. We then calculated how frequently identical tokens, i.e. words, appear in the text collection. There are in total 14,381,998 tokens in the corpus, and 125,311 unique tokens, which we call terms. This calculation provides a list of the most frequently mentioned words. The list shows that the most frequent words (MFW) in particular are stopwords, which are generic or function words – such as “and”, “that”, “to”, “we”, “there”. This type of words are innate terms for constructing most written texts despite genre. We filtered out such stopwords, leaving only nouns and proper nouns on the list, in order to represent terms that carry semantically significant content in this text collection. These terms are presented in table 1 below.
<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency rank</th>
<th>Frequency score</th>
</tr>
</thead>
<tbody>
<tr>
<td>god (gud)</td>
<td>26</td>
<td>96,278</td>
</tr>
<tr>
<td>jesus (jesus)</td>
<td>27</td>
<td>88,555</td>
</tr>
<tr>
<td>god’s (guds)</td>
<td>45</td>
<td>45,269</td>
</tr>
<tr>
<td>life (liv)</td>
<td>48</td>
<td>43,227</td>
</tr>
<tr>
<td>day (dag)</td>
<td>56</td>
<td>32,523</td>
</tr>
<tr>
<td>human beings (mennesker)</td>
<td>57</td>
<td>32,307</td>
</tr>
<tr>
<td>the life (livet)</td>
<td>72</td>
<td>26,047</td>
</tr>
<tr>
<td>world (verden)</td>
<td>74</td>
<td>25,732</td>
</tr>
<tr>
<td>word/words (ord)</td>
<td>78</td>
<td>24,699</td>
</tr>
<tr>
<td>faith (tro)</td>
<td>82</td>
<td>23,906</td>
</tr>
<tr>
<td>love (kærlighed)</td>
<td>98</td>
<td>19,769</td>
</tr>
<tr>
<td>human being (menneske)</td>
<td>108</td>
<td>17,273</td>
</tr>
<tr>
<td>time (tid)</td>
<td>119</td>
<td>15,511</td>
</tr>
<tr>
<td>jesu (jesu)</td>
<td>120</td>
<td>15,507</td>
</tr>
<tr>
<td>christ (kristus)</td>
<td>132</td>
<td>13,401</td>
</tr>
<tr>
<td>year (år)</td>
<td>142</td>
<td>12,401</td>
</tr>
<tr>
<td>son (søn)</td>
<td>151</td>
<td>11,364</td>
</tr>
<tr>
<td>children (børn)</td>
<td>155</td>
<td>10,916</td>
</tr>
<tr>
<td>joy (glæde)</td>
<td>163</td>
<td>10,428</td>
</tr>
<tr>
<td>light (lys)</td>
<td>170</td>
<td>9,659</td>
</tr>
<tr>
<td>thing (ting)</td>
<td>172</td>
<td>9,491</td>
</tr>
<tr>
<td>realm/rich (rige)</td>
<td>173</td>
<td>9,358</td>
</tr>
<tr>
<td>path (vej)</td>
<td>174</td>
<td>9,310</td>
</tr>
<tr>
<td>place (sted)</td>
<td>176</td>
<td>9,138</td>
</tr>
<tr>
<td>death/dead (død)</td>
<td>177</td>
<td>9,112</td>
</tr>
<tr>
<td>child (barn)</td>
<td>178</td>
<td>9,081</td>
</tr>
<tr>
<td>man (mand)</td>
<td>180</td>
<td>8,955</td>
</tr>
<tr>
<td>the faith (troen)</td>
<td>186</td>
<td>8,448</td>
</tr>
<tr>
<td>people (folk)</td>
<td>188</td>
<td>8,383</td>
</tr>
<tr>
<td>hope (håb)</td>
<td>189</td>
<td>8,349</td>
</tr>
<tr>
<td>earth (jorden)</td>
<td>190</td>
<td>8,283</td>
</tr>
<tr>
<td>the church (kirken)</td>
<td>191</td>
<td>8,245</td>
</tr>
<tr>
<td>the death (doden)</td>
<td>192</td>
<td>8,221</td>
</tr>
</tbody>
</table>

Table 1: List of the most frequent content words in the sermon corpus. The table describes the terms’ frequency rank and frequency scores. Terms with ambiguous translations are reported with a dash (-). Terms in parenthesis are the Danish term.
Pipeline: Seeded semantic network analysis

We use a seeded semantic network to trace the associative structure of two core theological concepts, love and sin, throughout the entire corpus. “Love” is high on the MFW list (the 98th most frequent word), while “sin” is absent from the top of the list (the 492nd most frequent word). With the semantic network analysis, we analyse and visually inspect the semantic constructions that these terms tend to be part of, and how the terms relate to each other. We thereby have a tool to examine the role of innate theological concepts regardless of their frequency in the corpus.

To build a seeded semantic graph, we initially train simple neural word embeddings using the CBOW model architecture\(^{105}\) to learn distributed lexical representations. To compensate for the relatively small corpus size, the model is initialised on pre-trained weights for Danish. In order to build the graph, the algorithm first computes the distance between a set of seed terms, love and sin, and the corpus lexicon using the inverse trigonometric arccosine function. For each seed, the algorithm then extracts the \(m\) terms with the shortest distance, the primary associations, and then the \(n\) terms with the shortest distance, the secondary associations, to the primary associations. The distance between all terms (i.e., seeds, primary and secondary associations) are computed and terms are connected based on their distance under a given threshold (thresholds are estimated from the distance variance structure). Finally, semantic clusters are extracted using the Louvain method, and the graph is represented visually with terms as nodes and thresholded distances as edges. For visualisation purposes, seed and primary associations are represented with UPPERCASE and node colour indicates cluster\(^{106}\).

We represent the associative structures of “love” and “sin” in three graphs. In Figure 1 and Figure 2 below, we have the semantic network of “love” and “sin” individually.


Figure 1. Semantic network of the seed term “love” and primary associations.

Figure 2. Semantic network of the seed term “sin” and primary associations.
In Figure 3 below, we have represented the semantic networks in relation to each other, where the grey nodes represent terms associated with “love”, and the pink nodes are terms associated with “sin”. Each seed in Figure 3 appears with ten primary associations, and each primary association has three secondary associations.
Collective Testimonies to Christianity and Time

**Pipeline: Topic modelling**

We designed our topic modelling analysis as a hypothesis-deductive test of whether holidays in the liturgical year explain thematic content in sermons, asking: do sermon documents from the same holidays share the same themes? For each of the 95 pastors, we chose to concatenate their sermons according to one of ten holiday periods in the church year, and we did therefore not run the topic model on 11,955 individual documents. Before the concatenation, we divided the holidays of the church year into ten sub-groups, which we defined. Our criteria for this division were that holidays named according to same period in the liturgical year should belong to the same sub-group; that the holidays in the same sub-groups should follow each other consecutively according to the liturgical year; and that the sub-groups should be as even in size as possible, while complying with the first two criteria. See the holiday sub-groups in Appendix II.

As the time of Trinity extends over five months in the liturgical year, we divided this season into three different sub-groups. Similarly, the holidays after Easter are divided between two groups, “Post-paske1” and “Post-paske2”, even though these groups contain holidays that actually represent a delimited holiday period themselves (for example Pentecost). However, due to our criteria, they do not constitute their own group in our division. Based on the holiday metadata in our corpus, we concatenated the pastors’ sermon productions according to the ten holiday periods in Appendix II. Our new corpus thus holds up to ten concatenated documents from each pastor, but in practice, we ended up with a corpus of 871 documents. Therefore, some pastors are represented with fewer documents than ten. The documents are named according to holiday sub-group and the random pseudonym the pastors have in the metadata.

We used DARIAH’s topics explorer, a user-friendly application for performing topic modelling, to analyse our new corpus. The application tokenised and lowercased every word in the corpus and removed stopwords. We used a stopword list that removed 262

---

107 Our reasons for concatenating sermons were as follows: It would be too immense to survey this many documents in relation to all 69 holidays in the liturgical year; the documents in the sermon corpus differ in length, and some documents would therefore not provide enough data for the model to work adequately; the pastors had not provided sermons for every holiday in the liturgical year throughout 2011-2016, and we did therefore not have an even amount of data from each pastor.

108 For example, the document “13_Advent” contains the concatenated advent sermons from pastor number 13.

109 The application is available here: [https://dariah-de.github.io/TopicsExplorer/](https://dariah-de.github.io/TopicsExplorer/)
function words from the corpus. We did not lemmatise the corpus, because we did not want to remove possible nuances that different forms of a word can provide in the topics we extracted. We tested the application with 100 iterations, asking for, respectively, 10, 20, 50 and 100 topics. We report our results based on findings from 20 topics, as the tendencies from each test follow a similar pattern in terms of how documents are grouped based on the topics. See Table 2 below for a distribution of topics in relation to holidays.

\footnotesize
\begin{itemize}
\item \textsuperscript{110} The stopword list is available here: https://gist.github.com/berteltorp/0cf8a0c7afea7f25ed754f24cfc2467b
\item \textsuperscript{111} For the PhD assessment committee: Contact supervisor Kirstine Helboe Johansen (kp@cas.au.dk) for access to all 20 topics.
\end{itemize}
<table>
<thead>
<tr>
<th>Holiday signal</th>
<th>Topic</th>
<th>Top 15 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVENT</td>
<td>Topic 17</td>
<td>john (johannes), jesus, day (dag), time (tid), light (lys), god’s (guds), advent (advent), christ (kristus), zechariah (zakarias), god (gud), joy (glæde), path (vej), awaits (venter), people (folk), isaiah (esajas)</td>
</tr>
<tr>
<td>JUL-NYTAR</td>
<td>Topic 16</td>
<td>god (gud), christmas (julen), child (barn), god’s (guds), world (verden), christmas (jul), year (år), light (lys), born (født), the child (barnet), peace (fred), human beings (mennesker), joy (glæde), bethlehem, joseph (josef)</td>
</tr>
<tr>
<td>HELIGTREKONGER</td>
<td>Topic 9</td>
<td>small (små), old (gamle), men (mænd), children (børn), water (vand), place (sted), child (barn), wise/show (vise), become (blevet), adults (voksne), exactly (netop), at the time (dengang), narrative (fortælling), heaven (himlen), earth (jorden)</td>
</tr>
<tr>
<td>FASTE</td>
<td>Topic 1</td>
<td>jesus, god (gud), god’s (guds), belief (tro), mary (maria), rich/realm (rige), day (dag), human beings (mennesker), son (son), bread/broke (brod), jesu, human being (menneske), power (magt), evil (onde), believe (tror)</td>
</tr>
<tr>
<td>PASKE</td>
<td>Topic 6</td>
<td>jesus, dead/death (død), jesu, “the” death (døden), resurrection (opstandelse), life (liv), jerusalem, died/the dead (dode), the tomb (graven), the disciples (disciplene), christ (kristus), easter morning (påskemorgen), good friday (langfredag), the cross (korset), son (son)</td>
</tr>
<tr>
<td>POST-PASKE1</td>
<td>Topic 4</td>
<td>jesus, peter, god (gud), shepherd (hyrde), life (liv), thomas, the truth (sandheden), good (gode), say (siger), freedom (frihed), day (dag), the path (vejen), the disciples (disciplene), recognise (kender), time (tid)</td>
</tr>
<tr>
<td>POST-PASKE2</td>
<td>Topic 8</td>
<td>god (gud), god’s (guds), jesus, word/words (ord), jesus, christ (kristus), the holy spirit (helligånden), life (liv), spirit (ånd), world (verden), human beings (mennesker), prayer (bøn), christian (kristne), community (fællesskab), christi (kristi)</td>
</tr>
<tr>
<td>TRINITATIS1</td>
<td>Topic 3</td>
<td>jesus, god (gud), rich/realm (rige), god’s (guds), life (liv), human beings (mennesker), peter, love (kærlighed), say (siger), son (son), jesu, to love (elske), the parable (lignelsen), good (gode), money (penge)</td>
</tr>
<tr>
<td>TRINITATIS1</td>
<td>Topic 10</td>
<td>say (siger), right (ret), received (fået), guilt (skyld), said (sagde), place (plads), go (gå), well (vel), to live (lev), day (dag), give (give), act (handler), whole (hele), children (børn), willingly (gerne)</td>
</tr>
<tr>
<td>TRINITATIS2</td>
<td>Topic 11</td>
<td>jesus, god (gud), life (liv), word/words (ord), god’s (guds), martha, day (dag), mary (maria), say (siger), thanks (tak), jesu, human beings (mennesker), lazarus, death (døden), good (gode)</td>
</tr>
<tr>
<td>TRINITATIS3</td>
<td>Topic 18</td>
<td>god (gud), jesus, god’s (guds), life (livet), human beings (mennesker), belief (tro), day (dag), give (give), death (døden), to give (give), each other (hinder), good (gode), jesu, human being (menneske)</td>
</tr>
</tbody>
</table>

Table 2: Relationship between holidays and topic. The table exhibits the holiday signals found from the generated topics and the 15 most related words within each topic. The original Danish terms are in parentheses, and the dash (−) indicates ambiguous meaning in the translation from the Danish terms. Topic numbers are assigned randomly to the topics by the topic model.
Our analysis shows that the holiday signal of the sermons is most significant for how topics are distributed in the documents: among the twenty topics, eleven indicate a strong holiday signal, with eight or more documents adhering to the same holiday period. Thus, each of our ten holiday periods is represented at least once in our topic modelling, while Trinitatis1 words are represented in relation to two topics (topic 3 and 10). In comparison, we find five topics where the pastor signal is strongest with eight or more documents originating from the same pastors. Among the 20 topics, topics 2, 5, 7, 13 and 14 are most frequently represented in the sermons of individual pastors. Topics 0, 12, 15 and 19 reveal neither a strong holiday nor pastor signal112.

Our findings indicate that the structure of the liturgical years influences thematic content in sermons. We thus have a model for exploring how themes and concepts unfold in relation to the liturgical year.

**Thematic contexts in sermons**

*The large framework: Solicitude for human life*

Our MFW list of content words, Table 1, indicates a rather coherent thematic framework for the sermon corpus. “God” and “Jesus” are the most frequent content words, which convincingly points out them as the main figures. The succeeding content words would emphasise a focus on a this-worldly life (“life”, “the world”, “earth”, “death”) among people (“human beings”, “children”, “man”, “people”) living here and now (“day”, “time”, “year”) from an optimistic perspective (“belief”, “love”, “joy”, “light”, “hope”). The conceptual framework found from the MFW list seems to illustrate a general theme about care for human life in the here-and-now, through a vocabulary that includes many terms that do not only have Christian connotations. It is striking that many concepts pointing to dogmas and cosmology are absent from the MFW list. The frequency rank (FR) of such terms is considerably lower – for example “heaven” (FR: 283), “the holy spirit” (FR: 378), “forgiveness” (FR: 387), “resurrection” (FR: 398), “mercy” (FR: 400), “redemption” (FR: 440), “the cross” (FR: 460), “sin” (FR: 492), “judgement” (FR: 672), “the devil” (FR: 742), “evil” (FR: 888), “hell” (FR: 1259). These concepts belonging more innately to doctrinal vocabulary or salvation history have comparably lower

---

112 For the PhD assessment committee: Contact supervisor Kirstine Helboe Johansen (kp@cas.au.dk) for access to all 20 topics.
frequency ranks, and, similarly, concepts that connote hardship or strife are far away from the top of the list. This tendency is also found among the concepts more innately representing a Christian symbolic system, which we have listed: “heaven” is represented far more frequently than “hell”. “the holy spirit”, “forgiveness” “resurrection” and “mercy” are comparably more represented than “sin”, “judgement”, “satan” and “evil”. There is thus a clear preference in the corpus for a recognisable (i.e. not exclusively or distinct theological) vocabulary that in general exhibits a positive attitude; moreover, there is a clear preference for theological concepts that connote optimistic and hopeful perspectives on existence. This indicates a solicitude theme about caring for human life with God and Jesus at the centre.

The word count analyses are good examples of – in Moretti’s terms – a distant reading approach, as we read and analyse words outside of the immediate context of the texts they derive from. To some, these findings may come off as crude observations in an otherwise highly complex material. It can easily be argued that pastors may provide theological understandings of “resurrection”, “creation”, “sin”, “evil” and “salvation” without using these particular terms – it can be argued that these concepts are so abstract that it takes more words and complex sentence structures to adequately represent the content of those concepts. The word count analyses cannot account for the various contexts in which the words in the corpus can take part, but they can show us a clear preference for some terms over others, which adds up to a seemingly coherent thematic framework. In the following, we will initially maintain a distant perspective on our text collection, but we will use tools that are more complex in order to get more intimate knowledge about the semantic structures in which the concepts are embedded throughout the sermon corpus.

Romance and ethics

We have applied the seeded semantic network analysis to focus on the two theological concepts, love and sin, and their semantic embedding in the corpus.

In the primary associations to “love” in Figure 1, we find two word clusters. In one cluster, “mercy” (barmhjertighed), “forgiveness” (tilgivelse) and “grace” (nåde) belong together\(^{113}\). These terms all denote aspects of love that are God’s to give, and thereby

\(^{113}\) The terms in parentheses in this paragraph refer to the Danish terms in figures 1, 2 and 3.
indicate an implicit reference to God as loving and caring. The terms in the other cluster—“gentleness” (*mildhed*), “generosity” (*gavmildhed*), “devotion” (*hengivenhed*), “unlimited” (*grænseløs*)—contextualise love as something safe, kind and altruistic, but not through concepts necessarily associated with God. Overall, the terms in Figure 1 are mostly abstract terms represented through nouns, thus appearing as predicates to love. In comparison, the primary associations to “sin” in Figure 2 are nouns as well as verbs. This network is not as tightly connected as Figure 1, but there is a tendency towards nouns and verbs belonging to different clusters. We see that “sin” in particular shares the same contexts as “original sin” (*arvesynd*) and “evil” (*ondskab*), indicating that sin is a lasting and basic condition. The small cluster of yellow nodes inscribes sin in a context of movement through the terms “trespassing” (*overtrædelse*), “sinner” (*synder*) and “owe” (*skylde*), signifying that sin is process-related and understood in terms of actions. The grey nodes constitute the most close-knit cluster of the three, which includes the terms “repent” (*angre*), “atone” (*sone*), “forgive” (*tilgive*) and “punishment” (*straf*). Instead of contextualising sin in terms of direct actions, these words rather contextualise it in terms of reactions and frame sin as something that can be dealt with through deeds. Even though the network around “love” suggests different contexts, “love” overall appears as something lasting and positive in a rather close-knit network; there are no verbs indicating that love can be lost or is temporary. In comparison, “sin” seems to exhibit more nuances, as it at the one end of the spectrum is associated with stable conditions, while it at the other end is associated with deeds that can overcome or at least deal with sin. Sin thus appears as a more flexible concept in the corpus.

In Figure 3, however, we can interpret both seed terms through more complex structures. In this graph, we have included the secondary associations to the seed terms, and we may observe how the semantic fields around each seed term relate to each other. Even though some terms mostly related to “sin” also relate to “love”, there are no direct connections between “love” and “sin”. When we expand the associative structure around love, we find that the altruistic conceptualisation of love becomes more consolidated through terms such as “unconditional” (*betingelsesløs*), “compassion” (*mefølelse*), “kindness” (*venlighed*), “tenderness” (*ømhed*) and “to love” (*elske*). Furthermore, with these types of terms, the altruistic love seems to denote a romanticised, intimate love. These terms dominate the top of the graph, where they are very closely related. The terms are moreover quite unambiguous – none of them indicate downsides or challenges associated
with love, and none of them refer to love as wild or passionate. Instead, the romanticised representation of love seems rather uncomplicated, carefree and safe. Meanwhile, the conceptualisation of love as “God’s love” still seems to be in its own little cluster separate from the close-knit romanticised love and is thus less predominant in the semantic network around “love”. In the expanded network around “sin”, there are more connections between the terms than in Figure 2, but they are still not as closely connected as in the expanded network around “love”. However, the terms added to the network around “sin” accentuate the action/reaction aspect of sin rather than the stability aspect. Furthermore, with terms such as “revenge” (hævne), “regret” (fortryde), “guilt” (skyld) and “injustice” (uretfærdighed), the semantic field around sin is increasingly becoming an ethical domain.

We thus find semantic fields of core theological content through the semantic network analyses. However, we also see that the term “love” transgresses a delimited narrative of salvation history and instead takes part in a more mainstream understanding of love through a recognisable vocabulary. This insight supports our word count analyses, which illustrated that the traditional authoritative theological concepts are not necessarily used to represent the main themes in the sermon corpus. The representation of a romanticised love would emphasise the solicitude theme, where the close relations in the here-and-now seem enhanced. The semantic network analyses further support the finding that there is an overall positive attitude to be found in the corpus: love does not seem to have downsides, and the action/reaction pattern indicates that sin – while it may be associated with harsh conditions – can be coped with.

**Thematic allegiance to the liturgical year**

Although our word count analyses and semantic network analysis of love show that pastors seem to present thematic content through a recognisable and relatable vocabulary, our topic modelling indicates that pastors also exhibit strong allegiance to the narrative structure of the liturgical year.

Table 2 shows that the terms “god”, “jesus”, “human beings”, “day” and “life” from the top of the MFW list (Table 1) appear in most of the eleven topics, which confirms that these terms are central to the thematic content in the corpus. Meanwhile, Table 2 also demonstrates that these central terms take part in different word configurations in the
topics. A large part of the vocabulary that unfolds in relation to the MFW terms in the topics refers to narrative content from biblical sources, which inscribes the MFW terms in different semantic contexts. Below, we characterise the thematic content that appears from these contexts in topics relating to holiday seasons. We focus specifically on Christmas, Easter, Pentecost and the first part of the Trinity season in order to observe variations throughout the course of the liturgical year.

The topic most related to the JUL-NYTAR sermons clearly exhibits a Christmas theme, which is emphasised through the term itself, “christmas”. The nativity motif of the gospels is also distinctly apparent: “god”, “child”, “born”, “the child”, “bethlehem” and “joseph”. Jesus is not at the centre of this topic with references to “jesus”, “christ”, “saviour” or “messiah”, but as a child. His role in the Christmas sermons is thus framed in terms of his human nature and not his transcendent. A large part of the remaining terms in the topic are high on the MFW list: “world”, “year”, “light”, “human beings” and “joy”. Along with the topic-specific term “peace”, the message at the centre of the Christmas theme seems to be a wish for world harmony – a framing that resembles the optimistic attitude found in the solicitude theme from our word count analyses. However, here it unfolds in a clear context of Jesus entering the world at Christmas.

In the topic related to the PASKE sermons from the Easter tide, we also see a strong allegiance to the holiday theme. We find in this topic that a few terms from the MFW list – “jesus”, “death/dead”, ”the death”, “life” – interrelate with a larger vocabulary that clearly represents the Easter narratives as found in the gospels: “resurrection”, “jerusalem”, “died”, “the tomb”, “the disciples”, “easter morning”, “good friday” and “the cross”. We thus see that the thematic content in Easter sermons resonates less with the larger solicitude framework of the collection than the Christmas topic does. We also find that a more explicit theological vocabulary, as opposed to the solicitude vocabulary, permeates the Easter topic. For example, “resurrection” emerges as a distinctly central term in this topic, while it is absent in the overall solicitude theme. This observation indicates that the occurrence of the term is closely linked with its biblical narrative context, whereas terms such as “light” and “joy” are significant for, but not limited to, the nativity motif.

“the holy spirit” seems to follow a similar pattern as “resurrection” in the liturgical year, as it only appears to be central in the topic related to POST-PASKE2 sermons. We have
previously described this sub-group of sermons as a bit artificial, as the holidays in this
group are not as strongly related as for example the PASKE sermons. However, through
the terms in the POST-PASKE2 topic in Table 3, we see a distinct Pentecost narrative:
The Holy Spirit entering the world into the lives of human beings after Christ’s ascension
to heaven. Yet outside the contexts of Easter and Pentecost in the liturgical year, the terms
“resurrection” and “holy spirit” are not emphasised as central theological components.
Instead, they mostly seem to appear in the corpus in terms of their biblical narrative
contexts.

In the ELCD, the gospel pericopes of the Trinity season include motifs emphasising what
a righteous life of Christian faith entails. Many of these in the form of parables,
disseminating the teachings of Jesus. These stories are, among others, the narratives of
the prodigal son (Luke 15:11-32), the camel and the eye of a needle (Matt 19:16-26), the
house and the storm (Matt 7:22-29), the feast of the rich man (Luke 14:16-24), and the
conversion of Zacchaeus (Luke 19:1-10). From our topic modelling, we found two
different topics most related to sermons from the first part of the Trinity season, and they
seem to emphasise the motif of righteousness and faith from different perspectives. In
one topic, topic 10, there is a focus on ethical demands with a dichotomy of right and
wrong. This framing includes active verbs and shares similarities with the dynamics in
the semantic modelling of “sin”. The other topic, topic 3, does not seem to emphasise
judgment and guilt, but rather a message of God’s love for human beings – especially the
story of the prodigal son appears to belong to this context. In this topic, the semantic
context of “love” promotes an understanding of care for human life, but it is not
constructed as a romanticised love similar to the construction found in the semantic
network analysis. Rather, “love” seems to be associated with the biblical sources, thus
behaving similarly to the terms “resurrection” and “holy spirit”. Furthermore, in topic 3,
Peter appears as a central figure. In the pericopes from this part of the Trinity season,
Jesus’ appointment of Peter as the rock of faith (Matt 16:13-26) and Jesus’ proclamation
of Peter as a fisherman appointed to catch people instead of fish (Luke 5:1-11) occur.
Therefore, there is also a focus on the foundation of the Christian congregations, which
is rooted in the stories of Peter.

The topic modelling analysis broadens our understanding of how theological concepts
and themes are represented in the sermons. When we look at themes in relation to the
church year, we find distinct theological discourses that unfold in close relation to the

146
biblical narratives that structure the church year. So even though this discursive content is not correspondingly apparent from our MFW list, it does not mean that traditional authoritative theological concepts are not pertinent in this corpus. Rather, these concepts seem to unfold and become particularly central in expounding the biblical narratives.

**Discussion: Theological dynamics in sermons**

The three approaches applied unveil different thematic aspects of the sermon corpus. This unveiling informs us that the corpus covers a complex semantic field of diverse dynamics. In these dynamics, the role of theological discourse varies. In Woodhead’s explanation, theology is defined as an academic discipline where a confessional basis and doctrinal discourse prevail, as she associates theological expertise with a power dynamic from above. If this dynamic is devalued, theological displacements occur. According to this terminology, the solicitude theme from the larger thematic framework of the sermons indicates that innate theological concepts are displaced. Here, a doctrinal vocabulary seems to be downplayed in favour of a vocabulary emphasising existential themes and care for human life. This framing points to an orientation towards inter-personal relations. These findings would support sociologist Robert Bellah’s early characterisation of symbolisation in modern Christianity. In 1964, he described the symbolic universe as a flexible and dynamic system that adapted to the situation of modern individualised human beings, where its transcendent and dualistic aspects would not be central anymore. Instead, there would be an emphasis on the human situation here and now\(^{114}\). However, when we turn to our topic modelling results, we find a remarkable theological structure below the most frequently used words. We see that many of the terms on the MFW list are part of topics closely related to the liturgical year. The topics distinctly represent biblical narratives, as they very precisely combine agents, locations, liturgical tide and concepts of narrative significance. When the sermon corpus was built, we made sure to remove the pericopes which the sermons are to expound\(^{115}\). Therefore, the biblical content found in the analyses is incorporated by the pastors themselves and not copy-pasted directly from the bible pericopes to the sermon manuscripts. Furthermore, even though we have pre-structured the corpus for the topic modelling analyses by


concatenating the sermons, the design allowed the analyses to show an arbitrary relationship between sermons, if there had in fact not been a strong affiliation between sermons from the same liturgical seasons. Thus, pastors are in general very loyal to the theological premises of the church year.

Meanwhile, the seeded semantic network analysis demonstrates that an innate theological concept such as “love” does not belong solely to a doctrinal discourse, but is also embedded in a more popular cultural idea of love. In Nielsen and Johansen’s study of the new Valentine’s Day services, they found that parishes tend to frame this event as a celebration of romantic love in couples’ lives\textsuperscript{116}. Combined, these observations indicate that a concept such as love provides a link between an original theological context and a broader mainstream cultural awareness – the concept seems to have an ability to be transposed to a new context. However, not all components in the symbolic system seem suitable for such a linking according to our analyses. “sin” does not provide a direct link between cultural contexts, but it does display a flexibility between semantic contexts, and it exhibits an optimism, where sin as an alterable condition resembles the overall positive and hopeful attitude found in the MFW list. Even more so than with “sin”, we are able to observe a remarkable elasticity in the semantic universe of sermons through the term “love”, as it is part of thematic contexts in all three analyses. It is one of the most frequently used terms in the corpus, thus providing content to the solicitude theme. This motif of inter-personal relationships is even more unfolded through the term’s embedding in the semantic network analysis. However, it also obtains a liturgical pertinence in the first part of the Trinity season. Here, love is not romanticised, but seems to denote God’s love for human beings similarly to a father’s love for his son as told through parables.

The analyses closely examining terms and vocabulary challenge a distinction where theological discourse per se belongs to a dynamic from above strictly bound to doctrinal teachings. Rather than something only from above, the analyses indicate that theological terminology can also interact with other cultural codes. In the sermons, theological discourse can be observed from above as well as from below – the latter as a discourse where theological content is the outcome of negotiations with contemporary culture and the everyday lives of churchgoers. Therefore, even though this dynamic might be deemed

a displacement of theology in Woodhead’s terminology, there would be no grounds for claiming that sermons are no longer a distinct domain of theological expertise. Rather, different dynamics are prevalent simultaneously in the sermons, which indicates that sermons are compounded phenomena where theological discourse is emergent – whether in relation to the church year or in relation to contemporary cultural codes.

**Conclusion**

The distant reading of sermons unveils collective patterns. It is a method that shows tendencies in the corpus and, in our approach, dynamics that can be interpreted closely. However, our study also demonstrates the value and necessity in applying different computational tools in order to observe the dynamics that occur simultaneously in the corpus – relying on just one of the three tools would not have made the thematic complexity in the corpus obvious.

Our approach to sermons allows us to address an aspect of preaching different from the oral event. We do not know for sure that the sermons in the corpus have been uttered in church precisely as they were written, and we have no insights into pastors’ individual performative styles. Instead, we focus on the fact that sermon documents exist and that they unfold an interpretive practice. Therefore, even though our approach does not analyse the oral communicative context of sermons, we are still interested in how content is communicated – however, not in terms of individual style, but in terms of implicit dynamics in the collective production of sermons. Therefore, when we call sermons a collective testimony to Christianity and time, we allude to an understanding of the documents as a theological reality that only reveals itself as a collective phenomenon. Sermons are an ongoing documentation of Christianity and history, where pastors bear witness to tradition and the cultural context in which they expound tradition.
Chapter 9

- Conclusion

I have experienced throughout my project period different attitudes towards this PhD project. Some have been sceptical about the quantitative approaches to an otherwise qualitative material, but in general, the amount of data has fascinated most people, leading to their proposals of a variety of specific analyses they found the sermon corpus calls for. I agree that the scope of this material opens for all sorts of research directions: it is interesting as a general compilation of historical religious texts, as a specific corpus of Christian theological documents, as a collection of sources to contemporary Danish history, as a linguistic collection of language or as a resource for rhetorical studies. I have not chosen one clearly defined direction to follow, but instead gathered perspectives from different research arenas in order to grasp the nature of this material before diving into specific research cases. This means that the horizon of inquiry is broad, but I have sought to keep the sermons as a text collection at the centre throughout every stage of the research process. In this way, the material has aided me in finding which paths to follow and which to leave.

Access to Danish sermons

The text material is at the centre of this project directly, since an important outcome is the sermon corpus itself. The corpus is therefore included as a research product introduced in chapter 4, and as a research resource displayed publicly in the article for *Journal of Cultural Analytics* included in chapter 6. The sampling procedure as outlined in chapters 3 and 6 was based on assembling sermons as personal documents provided directly by pastors. In order to resolve the issues of accessibility, I asked bishops and rural deans to help me get in contact with pastors under their supervision, and I sought to make the sampling procedure accommodating to pastors by allowing them to send in sermons in the amount they found feasible. Overall, I used an opportunistic sampling approach, which means I allowed for a considerable amount of variance in the material I received. However, since sermons are the outcome of a living practice and composed in terms of different employment situations, it would be difficult to acquire a material
adhering to external representativity standards even with a probabilistic sampling approach. Meanwhile, I was able to balance some of the variances in the final material by employing a manual procedure of cleaning and annotating the corpus as outlined in chapter 6 and in-depth in chapter 4. The approach to assembling the corpus was influenced by corpus linguistic principles with the purpose of standardising the sermons and metadata in a consistent structure, while maintaining the original sermons as written by pastors. The hands-on cleaning process ensured that documents not containing sermons were discarded, and that any kind of supplementary text to specific sermons in the documents was removed. The metadata annotations I supplied consisted of genre specific labels, such as the holiday labels, and general demographic labels retrieved from two online resources. Combined, the procedures of sampling and assembling sermons ensured that the sermons in the corpus are all part of the same practice, namely pastors’ composition of sermon manuscripts prior to church services for ELCD congregations.

**Interactions with intertextual systems**

Following the corpus linguistics perspectives, the sermon corpus by itself constitutes a large representation of language. This awareness has been important in my endeavour of studying the sermons’ interactions with different intertextual systems, even though my project is not a linguistic study. Language provides a basic access to latent and complex structures in the sermons. The tools used in my analyses are generic and in principle transposable to other genres – to other collections of language. However, by applying them to this particular corpus, we can take sermons seriously as a genre that unveils linguistic patterns in a genre specific way. In each of the article-chapters, chapters 6, 7 and 8, I have observed different interactions with religious and cultural codes in the corpus, illustrating that different dynamics occur in sermons.

Chapter 6 includes a study of the ways in which pastors construct social worlds with a specific focus on gender representations. We attended to gender categories on two levels – in terms of gender categories in language (personal pronouns) and in terms of gender categories in the metadata (gender of pastors). We found that male and female pastors alike represent gendered pronouns rather similarly in terms of frequency. With a clear overrepresentation of male characters, pastors further seem to follow the representation of gendered characters in the lectionary texts that they expound in their sermons. The
agency patterns we observed, meanwhile, suggest that male and female pastors represent gender differently. Male pastors tend to represent gender roles asymmetrically, meaning that male and female characters seem assigned to different agential spaces, as male characters tend to be active or directive agents, while female characters tend to be cogitative or communicative agents. In contrast, female pastors tend to assign active, cogitative and communicative agency to female and male characters alike although with different accentuations, while religious agency are assigned mainly to male characters in the singular cases. In the relational cases (suggestive interactions between male and female characters) female pastors, however, assign religious agency to female agents as well. The observations indicated that male pastors’ construction of social worlds are more in line with classic socio-linguistic theories, where feminine speech communities are associated with communicative speech acts and masculine speech communities with instrumental speech acts that demonstrate competency and independence. Female pastors do not represent these speech communities as gender dependent, since female and male characters alike in their sermons can take part in them. However, when we considered this in terms of the gender profile of the pastors, female pastors seem to display speech acts mostly corresponding to feminine speech communities, since the similarity in agency among the agents comes to represent a social world defined in terms of symmetry over asymmetry. This study as a whole observed a dynamic between sermon practice and cultural discourse focusing on how discursive components used in society at large unfold more specifically in the sermon genre. The study demonstrated that in sermons, this use of everyday language categories are not exclusive compared to language use in other contexts, since the language use follows other socio-linguistic observations. However, with a further subdivision of the corpus in terms of for example age, these patterns might begin to look differently.

Chapter 7 is also a study of social worlds in sermons, though not in terms of gender identity, but in terms of the cultural and religious domains characters represent in the corpus. This means that we took more diverse character identities into account. We observed three main tendencies that describe the ways in which characters connect to each other. The first one concerns the biblical narratives represented in the corpus, where we found different clusters illustrating stories about Jesus, each accentuating different characteristics associated with him: Jesus as saviour, Jesus as teacher and Jesus as compassionate. The second tendency concerns the role of recent history in sermons,
which is particularly anchored in political motifs – most significantly is the World War II motif centred on Hitler. The other motifs in this cluster concerns terrorism and world peace, which means that the political domain in sermons seems to be a highly ethical domain about right and wrong. The third tendency relates to the large group of characters in the corpus who are not inscribed in discernible contexts like the biblical or political figures. These remaining characters tend to be figures that occur sporadically in sermons, and in this way unable to connect to distinct thematic character clusters. We emphasised these characters as representing an ongoing collective negotiation between pastors, where some characters may obtain more significance over time, while others may be written out entirely. This study approached sermons as a rather open genre – open for all sorts of interactions between cultural and biblical narratives. However, what we found, in particular, was separate character clusters, where biblical characters and historical figures respectively belong together, but do not intermix. The most diverse and interrelating characters are the characters least represented in the corpus, but their vastness and low representativity make them a large, anonymised group. Nonetheless, they seem to represent a room of possibility – an agential space from which acceleration and novelty may grow.

Chapter 8 focuses specifically on sermons as a theological genre through a study of how thematic content connects to intertextual systems. The three approaches we applied in the study illustrated different thematic dynamics in the corpus. The point of departure is in the most frequently used words in the corpus, which suggests an overall thematic framework centred on compassion for human life in the present with an emphasis on close relations. In this framework, there are only few representations of core theological concepts and of transcendent elements. The other tools we applied, semantic network analysis and topic modelling, allowed us to look for complex structures below this overall thematic pattern. Utilising topic modelling, we explored the relationship between liturgical themes and sermon themes, and we found that pastors are extremely loyal to the narrative of salvation that the lectionary texts represent. In contrast to the word count analysis that suggested that theological concepts are not prevalent as a strong structure, the topic model showed that many authoritative concepts are closely related to the biblical narratives structuring the church year. The semantic network analyses focused on authoritative theological concepts and how they are semantically constructed in the corpus. Our analysis of “sin” showed different ontological framings of sin as a basic
condition versus sin as an alterable condition. With our expansion of the network, we saw nuances that consolidate the latter more than the former, and that sin seems to create an ethical domain – a reference from which to discuss and interpret right and wrong. The semantic network analysis of “love” did not show comparable nuances, as the concept is mostly visible as a romanticised concept that is rather one-dimensionally constructed – love is safe, kind and uncomplicated. Meanwhile, love is a key concept in all three analyses, where we find different accentuations of the concept. In the semantic network analysis, love resembles a popular cultural idea about romance, while it seems to operate specifically as a theological accentuation of God’s love to humankind in relation to the liturgical themes. Moreover, love is also one of the theological concepts ranging highest among the most frequent words in the corpus and as such inscribed in the overall solicitude theme. The three analyses in chapter 8 showed a diversity of thematic dynamics in the corpus. There is a strong theological dynamic “from above” through pastors’ allegiance to the liturgical year, but also a theological dynamic “from below” where a traditional theological concept such as love can connect to other cultural codes. Similarly, the accentuation of the close relations in the here and now in the solicitude theme seems to demonstrate a general orientation towards the everyday life of congregants.

These three chapters combined uncover different dynamics that place sermons in diverse intertextual interactions with Christian tradition and cultural fields. In chapter 6, we investigated how cultural codes move into sermons and embed structures in the genre, while we in chapter 8 observed how the vocabulary of the church itself unfold structures, which in content are oriented towards the lives of congregants as well as towards the lectionary texts. The design in chapter 7 looked for narratives represented through characters, and in this design, we explored the trajectories that the sermon corpus seems to suggest “by itself” towards biblical and cultural narratives. These different dynamics show different movements and directions between religious and cultural fields in sermons – movements and directions that do not preclude one and other, but co-exist. In this way, the dynamics expose some latent, but core structures in the sermon corpus regarding pastors’ collective interaction with intertextual systems.
In between biblical and cultural contexts

The theoretical framework of this PhD project draws up a rather wide space of possible interactions between culture and religion in sermons – there are no pathways to or connections between religious and cultural domains that are disregarded by default. Similarly, I have not based my investigations on conceptual definitions of culture and religion. With the lived religion background, these concepts are sensitising concepts that gain substance by concrete observations in the sermons. With this approach the practical space of interaction with culture and religion is considerably narrower than the wide theoretical one.

Biblical structures in sermons

All of my analytical studies demonstrate that biblical structures are very apparent in the corpus. On a very general level, the frequency representations of pronouns in chapter 6 indicate that the character gallery in regards to gender identities resemble the character galleries in the lectionary texts. The analyses in chapter 7 and 8 further unveil specific biblical narratives from the lectionaries. The topic modelling of the sermons unveil core narratives from the different holiday tides, and the network analysis group the biblical characters, who constitute a rather small group in terms of unique entities in the study, into six of the seven modularity classes. In these groups, besides the group of “Versatile characters”, the biblical figures tend to co-occur with characters they share narratives or thematic contexts with. One of the diverging aspects between the analyses in chapter 7 and 8 is the appearance of Jesus and God. Jesus is a central component in both studies – in the network, he is the most central character in terms of frequency and connections, and in the word count analysis he is one of the most frequently mentioned content terms and apparent in most of the retrieved topics. God is comparably apparent in the topics and in fact higher than Jesus on the list of most frequent words in the corpus. Meanwhile, the network analysis did not assign God to the group of main characters, and this analysis found a considerably lower frequency for God than the word count analyses in chapter 8 did. The reason for this discrepancy would be the NER-tagger used in the network analysis, since it only extracts entities it recognises as agents based on their syntactic positions in the corpus. Therefore, God is the most frequent character in the corpus, but as a character, God does not seem to be endowed with a lot of agency. The reason why
God in the network analysis is grouped with nativity characters is quite likely that God directly initiates the narrative of Jesus’ birth. Therefore, in the holiday topics, God is probably not included as an acting agent like Jesus. The analyses in this project will not be able to interpret more thoroughly the role or agency of God in the sermons, but it would be an evident objective to explore further in this corpus. At this point, it is merely obvious that God and Jesus are significant components in the biblical structures in the corpus in quite different ways.

The biblical narratives that chapter 7 and 8 both seem to enhance in the corpus are the stories from the Christmas and Trinity tides in the liturgical year. In the network, the group named “Other biblical stories” represent in particular stories from the Trinity tide, which is significant, considering the many other Gospel pericopes that could be represented in principle. Comparably, the topic model shows that the first part of the Trinity season is the only holiday period that have not only one, but two themes in the data. These studies in combination indicate that pastors engage in particular with the Gospel texts in this part of the church year. Christmas is another holiday that emerges in the network data through the many characters – from Old and New Testament – clearly related to this narrative context. Meanwhile, the network cluster with the nativity motif is one of the considerably larger ones in the network data, which suggests that this cluster is more connectable than the cluster with Trinity stories. From the topic model, we found that the thematic vocabulary in the Christmas topic shares many connotations with the overall solicitude theme. These two holiday contexts are interesting in comparison to each other, as the Trinity tide constitutes a very long coherent period of up to 25 Sundays that, except for All Hallows Eve, does not include any of the extraordinary holidays in the liturgical year. This means that the churchgoers in this season will typically be regular and local ones. Christmas Eve, in comparison, is the holiday where Danes at large visit the church as part of their Christmas celebration. The above observations could mean that when pastors target a more regular group of congregants, they are particularly focused upon the biblical narrative worlds in their sermons. At Christmas, the nativity narrative is also significantly present in sermons, but it seems more open for associations to other intertextual contexts and more framed in terms of the solicitude theme’s focus on close relations in the present.

The analyses in general demonstrate a significant lectionary structure in the corpus, and that the religious codes in the corpus are biblical ones in particular. This structure
indicates a general consensus among pastors, where the biblical narratives are important not just in principle but very much in practice, too. The lectionaries as a concrete intertextual system is also considerably smaller than the vast cultural systems pastors have the opportunity to engage with. However, even from this narrow intertextual source, it is still evident that the pastors are in consensus about some biblical narratives over others, which the network analyses illustrate with the open design.

*Representations of culture and time in sermons*

Symptoms of culture, to use Marjorie Garber’s concept, is clearly detectable in sermons, but not necessarily closely related to the religious codes or corresponding to the visibility of the religious codes. However, one of the cultural structures that match the biblical ones in visibility is the gender patterns observed in chapter 6. While the representation of gendered characters in terms of frequency resemble the social worlds in the lectionaries, the discursive constructions of these agents in sermons are indicative of cultural discourses. Another more general cultural pattern is the solicitude theme found in chapter 8 from the most frequently used words in the corpus, which suggests that the vocabulary used in sermons occur as relatable for congregants. These findings demonstrate that sermons are conceived in cultural contexts and are, in this way, products of religion as well as culture, since the sermons are not based in an exclusive religious language or discourse. The general linguistic and thematic patterns are supplemented by distinct cultural narratives found in the network analysis. The narrative of good and evil in the world represented through political characters and rooted in the World War II motif indicate that a writing of specific aspects of history takes place in the collective production of sermons. Moreover, we see that literary narratives are represented at large in the sermons and that this kind of cultural narratives interact with the biblical ones the most. This enhancement of literature, politics and history also indicate an intellectual core in the sermons. Meanwhile, the most distinct cultural narratives originate back in time, since contemporary figures are not discernible in the same way except for the ones associated with the World War II motif. These characters called “Versatile characters” do, however, display a large space for new elements in sermons, but these characters do not engage in strong connections corresponding to biblical, historical or political figures. Therefore, contemporary cultural codes exist in the corpus, but the very open design of
the network analysis may not be fit to show them as parts of text structures. It seems that distinct contemporary codes need to be searched out more explicitly, such as the semantic network analysis allows, because when we turn to a specific concept such as “love” and its semantic construction, we observe clear traces of popular culture. In fact, the semantic network unveils love as a rather one-dimensionally romanticised concept.

The observations of contemporary cultural codes demonstrate an interesting variation. They seem very apparent on a general level, as the gender analyses and the word count analysis illustrate – a level that concerns an overall use of language. On the other hand, as the semantic network analysis suggests, contemporary cultural codes are particularly evident, when we zoom in on specific terms in language. The analyses therefore suggest that cultural discourse shapes sermons on a very general and fundamental level, while new specific elements seem to occur sporadically, as the studies combined do not discover any new cultural trends that pastors collectively embed in sermons in the same way they embed biblical narratives.

*Dialogues between domains*

The purpose of this project has not been to observe religious and cultural content necessarily as separate from each other in sermons. With the theoretical framework presented in chapter 2, it has been just as much an aim to observe intersections between religious and cultural domains in sermons. I have treated the dynamics between religious and cultural discourses as separate from each other in the observations above, but there are also concrete findings, where they intersect. While I have just enhanced “love” as a concept that displays a particular cultural code, it furthermore illustrates a significant flexibility in the corpus, since it is also closely related to the biblical narratives from the topic modelling analyses. “Love” therefore demonstrates a theological and interpretative negotiation between scripture and cultural discourse. We were only able to observe the popular cultural framing by looking specifically for the semantic construction of “love”. This approach allows for discovering concepts that may hold the same kind of dynamic by querying other theological concepts enhanced in the topics from the topic modelling in the same way.

A theme emerging in chapters 7 and 8 alike is compassion. In chapter 8, I described it as the solicitude theme that comprises a general semantic framework of the corpus. In the
group called “Scandinavian narratives and stories of otherness” in chapter 7, we found some of the social outcasts in the bible, whom Jesus cares for, thus enhancing a framing of him as the compassionate. This group, however, largely includes literary and fictional characters, representing literary narratives of otherness and moral. This narrative cluster constitutes the group of characters that are most connected across biblical and cultural domains, indicating that the biblical narratives in this cluster are more connectable to a broader cultural context. Similarly, the solicitude theme also resembles a broader framework through the representation of everyday language oriented towards the lives of congregants. Compassion or solicitude as themes therefore illustrate an intersection of cultural and religious codes.

Another example of intersections between biblical and cultural codes occurs in the gender analyses in chapter 6. The overall tendency was that the verbal content we found mostly involved verbs that could appear in other contexts than in sermons, but we also found a narrow category called “religious agency” that distinctly refers to biblical vocabulary. I have not focused in-depth on this category since it is not as prevalent as the others, but it resembles in a way the same dynamic observed in chapter 8, where the biblical structure is revealed below the general thematic framework. In chapter 6, we see that the biblical structure emerges within the overall gender structure.

By comparing findings in chapters 7 and 8 again, we can observe different types of ethical domains. The cluster of political characters illustrates this domain, by representing a tension between world peace and historical events of strife and violence. Biblical characters are not included and only two theologians closely related to the concrete historical events in the cluster occur (Dietrich Bonhoeffer and Desmond Tutu). As such, the ethical domain is framed as a particular this-worldly tension. The semantic network of “sin” in chapter 8 unveils the ethical domain through traditional Christian vocabulary, such as “atone”, “repent” and “eternal sin”, and the ethical dimension of the concept is emphasised with focus on active actions that seem to cope with or maybe even resolve the condition of sin. The ethical aspects in the two chapters arise from different contexts and semantic fields due to the different designs of the studies. It would therefore be interesting to include a broad range of theological concepts in the character network, where associations to text components are not defined within a sentence as in the semantic network analysis, but defined within a sermon to see how theological concepts would relate to the thematic clusters represented by personal entities.
Collective testimonies to Christianity and time

The dynamics in the sermon corpus and the core structures they unveil show that pastors in the ELCD bear various kinds of testimonies to Christianity and time. The term testimony has a range of religious connotations related to Christian communities in particular, and is often understood as a performance of one’s personal experience of being or becoming a Christian. However, that is not the understanding I am aiming for. Theologian Anna Carter Florence has proposed the term testimony for understanding the content and performance of preaching:

(...). I am drawing on the classical definition of testimony as both a narration of events and a confession of belief: we tell what we have seen and heard, and we confess what we believe about it. A sermon in the testimony tradition is not an autobiography but a very particular kind of proclamation: the preacher tells what she has seen and heard in the biblical text and in life, and then confesses what she believes about it (Florence 2007, xiii).

Florence’s emphasis on sermons as a narration of biblical texts and life is relatable to my point of departure for this project, given that my purpose was to study sermons in the intersection of Christianity and cultural fields. However, I have not intended to study how pastors’ beliefs arise from their narrations, but rather how they practice the narrations in themselves, which are unveiled in the form of varied dynamics. In my portrayal of sermons as testimonies, I combine the narration aspect with the aspect of the term “testimony” that connotes attestation and documentation. In this, I enhance sermons as text phenomena that are not objective documentations of the world and religious traditions, but a type of documentation in which pastors witness the world and religious tradition by narrating them. This narration is however just as much an implicit, unconscious narrating as it is explicit, which the latent patterns found in the sermon corpus disclose.

I have only found these latent patterns by approaching sermons as a collective text production. This approach implies that we can understand the observed patterns as common features in sermons and as structures that presuppose a consensus among pastors. Since the collective understanding of sermons is a premise in this PhD project,
the common features are discovered, because I have sought to find them. This means that I have not intended to take individual variations in the corpus into account. Instead, I have conducted all analyses on the entire collection. In the gender analyses in chapter 6, we segmented the corpus into corpora from male and female pastors respectively, but we still approached the pastors as two collective groups. Meanwhile, in our design of the topic modelling analyses we contrast collective factors with individual factors, as we concatenated the sermons according to holiday as well as pastor pseudonym, thereby allowing the model to show individual topics. Although some of the topics do relate to individual pastors, the overall analysis shows a collective practice of representing the holiday pericopes.

Just as the studies do not focus on individual variations, nor do they attempt to infer an average way of composing sermons. The many patterns I have observed do not add up to a uniform genre that can be generalised to represent pastors overall. Even the pastors in the corpus do not represent every structure detected – rather, the structures are co-occurring tendencies that reveal different aspects of sermons. With the understanding of sermons as parts of a collective text system, we are able to look for variations in the system without focusing on individual cases, and we are able to look for tendencies without claiming a uniform voice in the system. The collective text system is therefore not a closed system, but an open and interactive one that is upheld through a shared practice. This understanding reveals a wide space of possible interactions between culture and religion, and my endeavour has been to frame and access this semantic and intertextual space by observing the dynamics that occur in practice. These dynamics reveal a range of core structures that in their diversity outline variations on the role of cultural and religious codes in sermons.
Bibliography


English summary

Collective Testimonies to Christianity and Time – A collection and large-scale text study of 11,955 Danish sermons from 2011-2016

Pastors in the Evangelical-Lutheran Church in Denmark (ELCD) continuously and collectively produce an enormous religio-cultural text material, when they prepare their sermons for the weekly church service. This type of text material thus contains valuable and comprehensive knowledge about how Danish pastors today collectively represent, interpret and engage with biblical, historical and contemporary contexts. However, access to sermons in the form of written manuscripts are restricted, since the documents remain the private property of pastors after the church service, for which they are composed.

This dissertation holds the project thesis that these sermon manuscripts are outcomes of a collective practice, and that they can be studied as such; as collective phenomena. It is further the argument of the thesis that pastors through this practice facilitate an encounter between religion and contemporaneity. The PhD project has had two main objectives: firstly, to gain access to ELCD sermons in their collective and textual form and provide a lasting research resource of sermon manuscripts; secondly, to explore sermons as a collective and intertextual text production in which religion and cultural codes intersect.

The first objective has resulted in a sermon corpus of 11,955 Danish ELCD sermons written mainly in the period from 2011 to 2016 by 95 pastors. The sermon corpus is included in the dissertation as a research product. The second objective concerns the analytical outcome of the PhD project, which is presented in the form of three research articles that are included as chapters in the dissertation. In the first article, I explore how cultural codes in society unfold in sermons by attending to the presence of gender discourses in the corpus. In the second study, I attend to the character gallery that appears from the sermon production, in order to investigate which cultural and religious contexts pastors seem to interact with in particular. The last article is a study of theological dynamics in the corpus, where I examine emergent themes as related to liturgical texts and contemporary cultural codes. In combination, the articles scrutinise the dynamics in three different relations between Christianity and cultural fields: 1) How everyday use of language embed cultural codes in sermons; 2) How cultural and religious contexts interact; and 3) How theological themes relate to contemporary cultural codes.
The project thesis poses distinct epistemological and methodological demands to this PhD project. The project is guided by currents from three research fields: *Lived religion*, as I argue that sermons as religious phenomena cannot be grasped isolated of surrounding culture; *the empirical turn in homiletics*, as this development has opened the homiletical discipline for new descriptive and compelling approaches for studying preaching; and *digital humanities*, as I argue that new technology and large-scale text studies facilitate a new perspective and approach to sermons, where they are not understood as individual documents, but as parts of a collective production and consciousness.

The analyses are conducted through applications of computational tools. The methodological ambition of this project is to study the material on its own terms as far as possible. In this endeavour, I investigate how sermons’ dialogue with each other in the text collection reveals a collective dialogue with Christianity and cultural codes in society. Central analytical concepts in this effort are *intertextuality, distant reading* and *close reading*. I have based on these concepts sought out text components in the sermon collection that are carriers of cultural information, and as such establish cultural and religious representations in sermons. With this method, I have strived to approach cultural and religious codes as anchored in concrete textual components.

The first part of the dissertation is dedicated to presenting the theoretical and methodological framework of the PhD project as a whole, while the latter part includes a presentation of the sermon corpus that holds the empirical outcome of the project, and the articles that hold the analytical outcomes of the project. While the articles are composed under consideration of the overall framework of the project, they appear as independent studies, as they target different journals and audiences.

I present the corpus of 11,955 sermons as a research product in a product report in *chapter 4*. I outline and discuss the process of transforming sampled sermons into a coherent and well-structured text corpus. This process has been conducted manually in order to ensure consistency in the material. The approach to assembling the corpus is influenced by *corpus linguistic* principles with the purpose of standardising the sermons and metadata in a consistent structure, while maintaining the original sermons as written by pastors. The procedures of sampling and assembling sermons ensured that the sermons in the corpus are all part of the same practice, namely pastors’ composition of sermon manuscripts prior to church services for ELCD congregations.
Article 1 is entitled “Sermons as data: Introducing a corpus of 11,955 Danish sermons” and is enclosed in chapter 6. The article introduces the ELCD as a public actor and sermons as access to the public voice of the church. The first part of the article is devoted to presenting the corpus in terms of sampling, metadata and corpus statistics. The second part is a case study of gender representations in the corpus, where I investigate how female and male pastors respectively construct social worlds in sermons. This study observes a dynamic between sermon practice and cultural discourse focusing on how discursive components used in society at large unfold more specifically in the sermon genre. The study demonstrates that in sermons, this use of everyday language categories are not exclusive compared to language use in other contexts, since the language use resembles other socio-linguistic observations.

Article 2 is enclosed in chapter 7 called “Unveiling the character gallery of sermons – A social network analysis of 11,955 Danish sermons”. It studies the representation of named entities – characters – in the corpus, as I regard these entities as access to uncovering a broad scope of the biblical and cultural domains pastors navigate between, when they compose their sermons. Through a social network analysis, I explore how various domains meet and reconfigure in the sermon corpus. Three main tendencies describe the ways in which characters connect to each other. The first one concerns the biblical narratives represented in the corpus with different character clusters illustrating stories about Jesus, each accentuating different characteristics associated with him: Jesus as saviour, Jesus as teacher and Jesus as compassionate. The second tendency concerns the role of recent history in sermons, which is particularly anchored in political motifs – most significantly a World War II motif. The other motifs in this cluster concerns terrorism and world peace, which means that the political domain in sermons seems to be a highly ethical domain about right and wrong. The third tendency relates to the large group of characters in the corpus who are not inscribed in discernible contexts like the biblical or political figures. These remaining characters tend to be figures that occur sporadically in sermons, and in this way unable to connect to distinct thematic character clusters. These characters seem to represent an ongoing collective negotiation between pastors, where some characters may obtain more significance over time, while others may be written out entirely.

Article 3 is entitled “What is love? – A study of thematic constructions in 11,955 Danish sermons”, which is included in chapter 8. In this article, I investigate the role of
theological discourses in emergent thematic fields in the corpus. The study is based on three computational tools: word count analysis, semantic network analysis and topic modelling. With these tools, I explore the overall thematic framework of the corpus, the construction of specific theological concepts (love and sin) and the relationship between thematic constructions in the corpus and thematic content in the liturgical texts that pastors expound in sermons. The three analyses in chapter 8 show different thematic dynamics in the corpus. There is a strong theological dynamic “from above”, as pastors display a strong allegiance to the themes of the liturgical year, which the topic modelling analysis illustrate. Meanwhile, there is also a theological dynamic “from below”, since a traditional theological concept such as “love” can connect to other popular cultural codes. Similarly, the word count analyses show an accentuation of close relations in the present in the form of a solicitude theme, which seems to demonstrate a general orientation among pastors towards the everyday life of congregants.

The three articles combined disclose some core structures in the corpus in the form of different dynamics between religious and cultural contexts. These core structures do not add up to a uniform genre that can be generalised to represent pastors overall. The structures are co-occurring tendencies that reveal different aspects of sermons. With the understanding of sermons as parts of a collective text system, we are able to look for variations in the system without focusing on pastors individually, and we are able to look for tendencies without claiming a uniform voice in the system. The collective text system is therefore not a closed system, but an open and interactive one that is upheld through a shared practice.
Dansk resumé

Kollektive Vidnesbyrd om Kristendom og Tid – En samling og et large-scale tekststudie af 11.955 danske prædikener fra 2011-2016

Præster i den danske folkekirke producerer på kontinuerlig og kollektiv basis et enormt religio-kulturelt tekstmateriale, når de forbereder deres prædikener til den ugentlige gudstjeneste. Denne type tekstmateriale rummer derfor en værdifuld og omfattende viden om, hvordan danske præster i dag kollektivt repræsenterer, fortolker og interagerer med bibelske, historiske og nutidige kontekster. Disse prædikensmanuskripter er dog ikke uden videre tilgængelige, da disse tekster forbliver præsternes private dokumenter, efter den gudstjeneste, de er udformet med henblik på.

kontekster interagerer; og 3) Hvordan teologiske temaer relaterer sig til nutidige kulturelle koder.

Projektets tese stiller særlige epistemologiske og metodologiske krav til dette ph.d.-projekt, og det er derfor influeret af strømme fra tre forskningsområder: *Lived religion*, da jeg argumenterer for, at prædikener som religiøse fænomener ikke kan forstås isoleret fra den omgivende kultur; *Den empiriske vending i homiletik*, da denne udvikling har åbnet den homiletiske disciplin for nye deskriptive og interessante tilgange til at studere prædikener; og *digital humanities*, idet jeg hævder, at ny teknologi og large-scale tekststudier faciliterer et nyt perspektiv og en ny tilgang til prædikener, hvor de ikke forstås som individuelle dokumenter, men som dele af en kollektiv produktion og bevidsthed.

Analyserne er udført ved hjælp af computationelle værktojer. Den metodiske ambition i dette PhD projekt har været at studere tekstmaterialet på dets egne præmisesr så vidt muligt. I denne bestrebelse undersøger jeg, hvorledes prædikernes dialog med hinanden i tekstsamlingen afslører en kollektiv dialog med kristendom og kulturelle koder i samfundet. De centrale analytiske begreber er *intertekstualitet, distant reading og close reading*. På baggrund af disse begreber har jeg søgt efter tekstkomponenter i prædikenerne, der bærer kulturel information, og som derved etablerer kulturelle og religiøse repræsentationer i korpusset. Med denne metode har jeg stræbt efter at tilgå kulturelle og religiøse koder som forankret i konkrete tekstkomponenter.


Jeg præsenterer prædikenkorpusset bestående af 11.955 prædikener som et forskningsprodukt i form af en produktrapport i kapitel 4. Jeg skitserer og diskuterer processen, hvormed jeg har omdannet indsamlede prædikener til et sammenhængende og velstruktureret tekstkorpus. Denne proces blev udført manuelt for at sikre en ensartethed i materialet. Tilgangen til at samle korpusset er influeret af *korpus lingvistiske* principper, der har haft til formål at standardisere prædikener og metadata i en ensartet struktur,
Collective Testimonies to Christianity and Time

samtidig med at det originale indhold i prædikenerne er blevet bevaret. Denne procedure har sikret, at prædikenerne i korpusset alle er resultat af samme praksis, nemlig præsters udarbejdelse af prædikenmanuskripter forud for gudstjenester for folkekirkemenigheder.

Artikel 1 hedder “Sermons as data: Introducing a corpus of 11,955 Danish sermons” og indgår i kapitel 6. Artiklen introducerer folkekirkens som en offentlig aktør og prædikener som adgang til kirkens offentlige stemme. Den første del af artiklen præsenterer prædikenkorpusset med fokus på dataindsamling, metadata og korpusstatistik. Anden del er et casestudie af kønsrepræsentationer i korpusset, hvor jeg undersøger, hvordan henholdsvis kvindelige og mandlige præster konstruerer sociale verdener i prædikener. Denne undersøgelse observerer en dynamik mellem prædikenpraksis og kulturelle diskurser med fokus på, hvordan diskursive komponenter (kønnede pronomener), der anvendes i samfundet generelt, udfolder sig mere specifikt i prædikenen. Undersøgelsen viser, at i prædikener er denne brug af hverdagslige sprogkategorier ikke markant anderledes end sprogbrug i andre sammenhænge, da fundene afspejler generelle sociolinguistiske observationer af køn og sprog.


Tilsammen afslører de tre artikler nogle dybdestrukturer i prædikenkorpusset i form af forskellige dynamikker, der udholder sig imellem religiøse og kulturelle kontekster. Disse dybdestrukturer udgør ikke tilsammen en ensartet genre, som kan generaliseres ud til præster i bred forstand. Disse strukturer er i stedet sammenfaldende tendenser, der afslører forskellige aspekter ved prædikener. Ved at forstå prædikener som del af et kollektivt system, er det muligt at opdage variationer i systemet uden at fokusere på præster individuelt, og det er muligt at opdage tendenser uden at hævde, at der er en ensartet stemme i systemet. Det kollektive tekstsystem er således ikke et lukket system, men et åbent og interagerende system, som er opretholdt igennem en fælles praksis.
Appendix I

- Invitation addressed to pastors in the ELCD
Kære præst i Folkekirkens

Vi henvender os til dig, da vi har brug for din hjælp i forbindelse med et forskningsprojekt på
religiøs videnskab og teologi ved Aarhus Universitet. Under vejledning af lektor i Praktisk Teologi,
Kirstine Helboe Johansen, undersøger Anne Agersnap i sin ph.d.-projekt centrale indholdsdimensioner
i nutidige prædikener.

Folkekirkens prædikener udgør et vigtigt stykke kulturhistorie i Danmark, som fortjener at blive
belyst og bevaret for eftertiden. Derfor søger vi din og dine kollegers støtte til at opbygge en database
bestående af prædikener i tusindvis fra hele landet. Denne indholdsrige samling vil således kunne give
helt ny indsigt i, hvordan kristendom aktualiseres og formidles i kirken i dag.

Projektets udgangspunkt er, at prædikenen skaber et særligt møde mellem bibeltekst,
menighed og samfund. Ud fra dette perspektiv vil vi ved hjælp af digitale analyseredskaber udforske
mønstre, relationer og forskelle prædikenerne imellem. Herved vil det blive undersøgt, hvordan
søndagens evangelietekster, teologiske begreber og lokale, nationale og globale begivenheder
optræder i samtlige prædikener i databasen. Deres betydning med betydning med og
aktuelt prædikenskab.

Vi håber, at du vil hjælpe med at realisere projektet, og derfor beder vi dig: Send dine prædikener fra
2011-2016 til en mail til anag@cas.au.dk.

Det skal du vide:

- **Send prædikenerne på den måde, det er lettest for dig!** Har du eksemplervis ikke mulighed for at sende
  prædikener fra alle årene, så send så mange, du har mulighed for. Du er også velkommen til at sende
  flere prædikener end de efterspurgte, hvis det er nemmere ift. dit arkiveringssystem.
- **Der er intet krav om, at du redigerer eller ryder op i dine prædikener inden
  indsendelse.** Når vi modtager dine dokumenter, vil vi sørge for, at der ikke inddrages tekstr
  i undersøgelsen, som ikke indgår i en prædiken.
- **Undersøgelsen vil ikke forholde sig til prædikenernes originalitet eller formidlingsmæssige
  og teologiske kvalitet.** Det er således ikke problematisk for undersøgelsen, hvis dine prædikener rummer
  indhold eller inspiration fra andre eller andres prædikener.
- **Der vil ikke blive knyttet forbindelser mellem specifikke prædikener og den konkrete præst.**

**Dataopbevaring og tilgængelighed**

Prædikenerne vil indgå i en forskningsdatabas i de kommende ti år, hvor også ph.d.-vejleder, Kirstine
Helboe Johansen, og bi-vejledere, Uffe Schjødt og Kristoffer Løggaard Nelbo, vil arbejde med prædikenen.
I overensstemmelse med persondataloven vil de involverede præsters og sognes navne blive holdt skjult i
alle former for offentliggørelse. Databasen administreres under Aarhus Universitet og adgang vil kræve
tilladelse gennem universitetet. Efter ti år slettes alle navne og baggrundsoplysninger, men
prædikendarbasen vil blive bevaret som kildermaterial for andre forskningsprojekter.

**Sådan deltager du i projektet:**

- **Åbn en ny mail og tryk ’vedhæft fil’**
- **Vælg mappen med dine prædikener og marker filerne med prædikener fra 2011-2016.** Tryk ’Indsæt’.
- **Kriv evt. i mailen, at du ønsker at komme på en malinglister for at blive holdt opdateret om projektet
  og gjort opmærksom på tilknyttede forskningspublikationer.**
- **Indsæt følgende formulering i mailen for at give samtykke til, at dine prædikener må indgå i
  forskning:**
  “Jeg er indforstået med, at mine prædikener og baggrundsoplysninger indgår i en
  forskningsdatabas. Baggrundsoplysninger slettes efter ti år, mens prædikendarbasen bevares
  for andre forskningsprojekter.”

Aarhus Universitet
- Du vil efter din indsendelse modtage et meget kort spørgeskema angående forskellige baggrundsoplysninger, som vi håber, du desuden vil besvare.

NB: Der kan ca. vedhæftes 60 prædikere i én mail, hvorfor det kan være nødvendigt at fordøle indsendelsen over flere mails. Samtykkeerklæringen skal blot sendes én gang. Det tager ca. 15-25 minutter i alt. Vi indhenter også gerne prædikenerne på anden vis, fx via delinger i Dropbox, Google Drev etc.

Hvis du har lyst til at læse mere om ph.d.-projektet, kan det lade sig gøre via dette link: http://cas.medarbejdere.au.dk/nyhedsbreve/nyhed/artikel/anne-agersnap-ny-phd-ved-afdeling-for-religionsvidenskab/

Skulle du desuden have spørgsmål til projektet eller opleve problemer med indsendelsen af dine prædikere, må du endelig kontakte os, Anne Agersnap (mail: anag@cas.au.dk, tlf.: 87162903 / 29802875) eller Kirstine Helboe Johansen (mail: kp@cas.au.dk).

Med venlig hilsen
Anne Agersnap, Ph.d.-studerende, Religionsvidenskab, Aarhus Universitet
Kirstine Helboe Johansen, Lektor, Teologi, Aarhus Universitet
Kristoffer Laiaaard Nielsen, Lektor, Religionsvidenskab, Aarhus Universitet
Uffe Schjødt, Lektor, Religionsvidenskab, Aarhus Universitet
Appendix II

- Liturgical Holidays
The table below illustrates Danish liturgical holidays assigned to sub-groups. The table includes name of holiday sub-group, concrete holiday and name of holidays from the corpus metadata. Sub-groups are named based on Danish holiday names. Number of holidays in the Epiphany tide and Trinity tide varies depending on when Easter falls in a given year, so there are not always respectively seven and twenty six holidays in the two tides.

<table>
<thead>
<tr>
<th>Holiday sub-group</th>
<th>Holiday (English)</th>
<th>Holiday (Danish)</th>
<th>Metadata annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVENT (Advent)</td>
<td>1st Sunday of Advent</td>
<td>1. søndag i advent</td>
<td>% 1Adv</td>
</tr>
<tr>
<td></td>
<td>2nd Sunday of Advent</td>
<td>2. søndag i advent</td>
<td>% 2Adv</td>
</tr>
<tr>
<td></td>
<td>3rd Sunday of Advent</td>
<td>3. søndag i advent</td>
<td>% 3Adv</td>
</tr>
<tr>
<td></td>
<td>4th Sunday of Advent</td>
<td>4. søndag i advent</td>
<td>% 4Adv</td>
</tr>
<tr>
<td>JUL-NYTAR (Christmas-NewYear)</td>
<td>Christmas Eve</td>
<td>Juleaften</td>
<td>% Jul</td>
</tr>
<tr>
<td></td>
<td>Christmas Day</td>
<td>Juledag</td>
<td>% JuD</td>
</tr>
<tr>
<td></td>
<td>Saint Stephen’s Day/2nd Day</td>
<td>Sankt Stefansdag/2. Juledag</td>
<td>% 2Jul</td>
</tr>
<tr>
<td></td>
<td>of Christmas</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Christmas Sunday</td>
<td>Julesøndag</td>
<td>% JuS</td>
</tr>
<tr>
<td></td>
<td>New Years Day</td>
<td>Nytårsdag</td>
<td>% Ny</td>
</tr>
<tr>
<td></td>
<td>New Years Eve</td>
<td>Nytårsaftensdag</td>
<td>% NyA/S:ER</td>
</tr>
<tr>
<td>HELLIGTREKONGER (Epiphany)</td>
<td>Epiphany Sunday</td>
<td>Helligtrekongers søndag</td>
<td>% Hel</td>
</tr>
<tr>
<td></td>
<td>1st Sunday after Epiphany</td>
<td>1. søndag efter Helligtrekonger</td>
<td>% 1Hel</td>
</tr>
<tr>
<td></td>
<td>2nd Sunday after Epiphany</td>
<td>2. søndag efter Helligtrekonger</td>
<td>% 2Hel</td>
</tr>
<tr>
<td></td>
<td>3rd Sunday after Epiphany</td>
<td>3. søndag efter Helligtrekonger</td>
<td>% 3Hel</td>
</tr>
<tr>
<td></td>
<td>4th Sunday after Epiphany</td>
<td>4. søndag efter Helligtrekonger</td>
<td>% 4Hel</td>
</tr>
<tr>
<td></td>
<td>5th Sunday after Epiphany</td>
<td>5. søndag efter Helligtrekonger</td>
<td>% 5Hel</td>
</tr>
<tr>
<td></td>
<td>Last Sunday after Epiphany</td>
<td>Sidste söndag efter</td>
<td>% SHel</td>
</tr>
<tr>
<td>FASTE (Lent)</td>
<td>Septuagesima Sunday</td>
<td>Søndag Septuagesima</td>
<td>% Sept</td>
</tr>
<tr>
<td></td>
<td>Sexegesima Sunday</td>
<td>Søndag Seksegesima</td>
<td>% Seks</td>
</tr>
<tr>
<td></td>
<td>Lent</td>
<td>Fastelavns søndag</td>
<td>% Fas</td>
</tr>
<tr>
<td></td>
<td>1st Sunday of Lent</td>
<td>1. søndag i Fasten</td>
<td>% 1Fas</td>
</tr>
<tr>
<td></td>
<td>2nd Sunday of Lent</td>
<td>2. søndag i Fasten</td>
<td>% 2Fas</td>
</tr>
<tr>
<td></td>
<td>3rd Sunday of Lent</td>
<td>3. søndag i Fasten</td>
<td>% 3Fas</td>
</tr>
<tr>
<td></td>
<td>Laetare Sunday</td>
<td>Midfaste søndag</td>
<td>% MidF</td>
</tr>
<tr>
<td></td>
<td>Feast of the Annunciation</td>
<td>Marie Bebudelsesdag</td>
<td>% MBeb</td>
</tr>
<tr>
<td>PASKE (Easter)</td>
<td>Palm Sunday</td>
<td>Palmesøndag</td>
<td>% Pal</td>
</tr>
<tr>
<td></td>
<td>Maundy Thursday</td>
<td>Skærtorsdag</td>
<td>% Skær</td>
</tr>
<tr>
<td></td>
<td>Good Friday</td>
<td>Langfredag</td>
<td>% Lang</td>
</tr>
<tr>
<td></td>
<td>Saturday before Easter</td>
<td>Påskelørdag</td>
<td>% PåsL</td>
</tr>
<tr>
<td></td>
<td>Easter Day</td>
<td>Påskedag</td>
<td>% PåsD</td>
</tr>
<tr>
<td></td>
<td>2nd Day of Easter</td>
<td>2. Påskedag</td>
<td>% 2PåsD</td>
</tr>
<tr>
<td>POST-PASKE1 (After-Easter1)</td>
<td>1st Sunday after Easter</td>
<td>1. søndag efter Påske</td>
<td>% 1Pås</td>
</tr>
<tr>
<td>Event</td>
<td>Danish Name</td>
<td>Abbreviation</td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>2nd Sunday after Easter</td>
<td>2. søndag efter Påske</td>
<td>2Pås</td>
<td></td>
</tr>
<tr>
<td>3rd Sunday after Easter</td>
<td>3. søndag efter Påske</td>
<td>3Pås</td>
<td></td>
</tr>
<tr>
<td>Common Prayer Day</td>
<td>Bededag</td>
<td>Bed</td>
<td></td>
</tr>
<tr>
<td>4th Sunday after Easter</td>
<td>4. søndag efter Påske</td>
<td>4Pås</td>
<td></td>
</tr>
<tr>
<td>POST-PASKE2 (After-Easter2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Sunday after Easter</td>
<td>5. søndag efter Påske</td>
<td>5Pås</td>
<td></td>
</tr>
<tr>
<td>Christ’s Ascension</td>
<td>Kristi Himmelfartsdag</td>
<td>KHim</td>
<td></td>
</tr>
<tr>
<td>6th Sunday after Easter</td>
<td>6. søndag efter Påske</td>
<td>6Pås</td>
<td></td>
</tr>
<tr>
<td>Pentecost</td>
<td>Pindedag</td>
<td>Pin</td>
<td></td>
</tr>
<tr>
<td>2nd day of Pentecost</td>
<td>2. Pindedag</td>
<td>2Pin</td>
<td></td>
</tr>
<tr>
<td>TRINITATIS1 (Trinity1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trinity Sunday</td>
<td>Trinitatis søndag</td>
<td>Trin</td>
<td></td>
</tr>
<tr>
<td>1st Sunday after Trinity</td>
<td>1. søndag efter Trinitatis</td>
<td>1Trin</td>
<td></td>
</tr>
<tr>
<td>2nd Sunday after Trinity</td>
<td>2. søndag efter Trinitatis</td>
<td>2Trin</td>
<td></td>
</tr>
<tr>
<td>3rd Sunday after Trinity</td>
<td>3. søndag efter Trinitatis</td>
<td>3Trin</td>
<td></td>
</tr>
<tr>
<td>4th Sunday after Trinity</td>
<td>4. søndag efter Trinitatis</td>
<td>4Trin</td>
<td></td>
</tr>
<tr>
<td>5th Sunday after Trinity</td>
<td>5. søndag efter Trinitatis</td>
<td>5Trin</td>
<td></td>
</tr>
<tr>
<td>6th Sunday after Trinity</td>
<td>6. søndag efter Trinitatis</td>
<td>6Trin</td>
<td></td>
</tr>
<tr>
<td>7th Sunday after Trinity</td>
<td>7. søndag efter Trinitatis</td>
<td>7Trin</td>
<td></td>
</tr>
<tr>
<td>8th Sunday after Trinity</td>
<td>8. søndag efter Trinitatis</td>
<td>8Trin</td>
<td></td>
</tr>
<tr>
<td>TRINITATIS2 (Trinity2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9th Sunday after Trinity</td>
<td>9. søndag efter Trinitatis</td>
<td>9Trin</td>
<td></td>
</tr>
<tr>
<td>10th Sunday after Trinity</td>
<td>10. søndag efter Trinitatis</td>
<td>10Trin</td>
<td></td>
</tr>
<tr>
<td>11th Sunday after Trinity</td>
<td>11. søndag efter Trinitatis</td>
<td>11Trin</td>
<td></td>
</tr>
<tr>
<td>12th Sunday after Trinity</td>
<td>12. søndag efter Trinitatis</td>
<td>12Trin</td>
<td></td>
</tr>
<tr>
<td>13th Sunday after Trinity</td>
<td>13. søndag efter Trinitatis</td>
<td>13Trin</td>
<td></td>
</tr>
<tr>
<td>14th Sunday after Trinity</td>
<td>14. søndag efter Trinitatis</td>
<td>14Trin</td>
<td></td>
</tr>
<tr>
<td>15th Sunday after Trinity</td>
<td>15. søndag efter Trinitatis</td>
<td>15Trin</td>
<td></td>
</tr>
<tr>
<td>16th Sunday after Trinity</td>
<td>16. søndag efter Trinitatis</td>
<td>16Trin</td>
<td></td>
</tr>
<tr>
<td>TRINITATIS3 (Trinity3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17th Sunday after Trinity</td>
<td>17. søndag efter Trinitatis</td>
<td>17Trin</td>
<td></td>
</tr>
<tr>
<td>18th Sunday after Trinity</td>
<td>18. søndag efter Trinitatis</td>
<td>18Trin</td>
<td></td>
</tr>
<tr>
<td>19th Sunday after Trinity</td>
<td>19. søndag efter Trinitatis</td>
<td>19Trin</td>
<td></td>
</tr>
<tr>
<td>20th Sunday after Trinity</td>
<td>20. søndag efter Trinitatis</td>
<td>20Trin</td>
<td></td>
</tr>
<tr>
<td>All Hallow’s Day</td>
<td>Allehelgensdag</td>
<td>AHelg</td>
<td></td>
</tr>
<tr>
<td>21st Sunday after Trinity</td>
<td>21. søndag efter Trinitatis</td>
<td>21Trin</td>
<td></td>
</tr>
<tr>
<td>22nd Sunday after Trinity</td>
<td>22. søndag efter Trinitatis</td>
<td>22Trin</td>
<td></td>
</tr>
<tr>
<td>23rd Sunday after Trinity</td>
<td>23. søndag efter Trinitatis</td>
<td>23Trin</td>
<td></td>
</tr>
<tr>
<td>24th Sunday after Trinity</td>
<td>24. søndag efter Trinitatis</td>
<td>24Trin</td>
<td></td>
</tr>
<tr>
<td>25th Sunday after Trinity</td>
<td>25. søndag efter Trinitatis</td>
<td>25Trin</td>
<td></td>
</tr>
<tr>
<td>Last Sunday of the liturgical year</td>
<td>Sidste søndag i kirkelåret</td>
<td>Sidst</td>
<td></td>
</tr>
</tbody>
</table>