



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Western hagiography in topic modelling – applying data analytics and natural language processing to the collection *Acta Sanctorum* (1643 - 1925)

verfasst von / submitted by

Christos Bintsis

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 647

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Digital Humanities UG2002

Betreut von / Supervisor:

PD Dr. Thomas Wallnig, MAS

I would like to give my warmest thanks to my supervisor, Dr. Thomas Wallnig, whose guidance and knowledge steered me through this academic journey.

Special thanks to the faculty members and the University of Vienna for giving me this opportunity.

Sincere gratitude to my parents, Stavros and Alexandra, and my brothers, Nikolaos and Philippos, for their unwavering support.

“Then why do you want to know?”

“Because learning does not consist only of knowing what we must or we can do, but also of knowing what we could do and perhaps should not do.”

Umberto Eco, *The Name of the Rose*

Abstract

The aim of this master's thesis is to apply data analytics and natural language processing methods to *Acta Sanctorum*, a collection of hagiographical texts published between 1643 and 1925, in order to gain insights which are impossible to be detected only by using qualitative methods. Although there are many studies about *Acta Sanctorum*, which are based on traditional methods, the examination of the work in light of the new methods introduced by Digital Humanities is still at an early stage. The thesis starts from a survey of hagiographical literature which highlights the chronological, geographical and historical conditions that led to the formation of hagiography as a literary genre through the centuries. Then, a detailed presentation of *Acta Sanctorum* is included as well as a presentation of the dataset, the data analytics and natural language processing methods applied during the research using the Python programming language. Finally, the results of the analysis are presented with the help of tables and visualizations. These results provide valuable insights that point to chronological, geographical and topical diversity of hagiographical literature, describe the specific factors that led to a unique grouping and supply helpful information about the processing of Latin unstructured data with computational methods, which can be useful for researchers in the field.

Abstract (Deutsch)

Das Ziel dieser Masterarbeit besteht darin, Methoden der Datenanalyse und der NLP auf die *Acta Sanctorum*, eine Sammlung hagiographischer Texte, die zwischen 1643 und 1925 veröffentlicht wurden, anzuwenden, um Erkenntnisse zu gewinnen, die allein durch die Verwendung qualitativer Methoden nicht erfasst werden können. Obwohl es viele Studien zu den *Acta Sanctorum* gibt, die auf traditionellen Methoden basieren, befindet sich die Auseinandersetzung mit den neuen Methoden der Digital Humanities noch in einem frühen Stadium. Die Arbeit beginnt mit einem Überblick über die hagiographische Literatur, der die chronologischen, geografischen und historischen Bedingungen hervorhebt, die im Laufe der Jahrhunderte zur Entstehung der Hagiographie als literarisches Genre geführt haben. Danach erfolgt eine ausführliche Darstellung der *Acta Sanctorum* sowie eine Präsentation des Datensatzes, der Datenanalyse und der Methoden zur Verarbeitung natürlicher Sprache, die während der Forschung mithilfe der Programmiersprache Python angewendet wurden. Abschließend werden die Ergebnisse der Analyse mit Hilfe von Tabellen und Visualisierungen dargestellt. Diese Ergebnisse liefern wertvolle Einblicke in die chronologische, geographische und thematische Vielfalt hagiographischer Literatur, beschreiben die spezifischen Faktoren, die zu einer einzigartigen Gruppierung geführt haben, und bieten hilfreiche Informationen über die Verarbeitung lateinischer unstrukturierter Daten mit Computer-gestützten Methoden, die für Forscher:innen auf diesem Gebiet nützlich sein können.

Table of Contents

<u>Abstract</u>	iii
<u>Abstract (Deutsch)</u>	iv
<u>Chapter 1 – Introduction</u>	1
<u>Chapter 2 – Hagiography in the Middle Ages</u>	4
<u>2.1 Christian Hagiography: From its Appearance to the Fall of the Western Roman Empire (476)</u>	4
<u>2.2 Latin Hagiography in the Middle Ages</u>	8
<u>Chapter 3 – Humanist Latin Christian Hagiography in the Renaissance</u>	15
<u>Chapter 4 – Acta Sanctorum</u>	23
<u>Chapter 5 – Data Analytics</u>	36
<u>5.1 Introduction</u>	36
<u>5.2 Database and Data</u>	39
<u>5.3 Exploratory Data Analysis</u>	41
<u>Chapter 6 – Natural Language Processing</u>	52
<u>6.1 Introduction</u>	52
<u>6.2 Topic Modelling</u>	58
<u>6.3 Results</u>	64
<u>Chapter 7 – Conclusions</u>	77
<u>Bibliography</u>	81

Chapter 1 – Introduction

The appearance of Christian hagiographical texts, i.e. texts that describe aspects of the lives of Christian saints, dates back to Late Antiquity. Later, during the Middle Ages and the Renaissance, they were one of the most widespread literary genres. Although their boundaries are not always distinct, hagiographical texts are divided into categories based on their content, such as the *Vitae* (Lives), *Passiones* (Martyrdoms) and *Miracula* (collections of miracles). Apart from their religious content, hagiographical texts are also valuable sources of history, traditions, morals, and customs. Despite the fact that the hagiographers followed a specific pattern without much room for deviation when writing the hagiographical texts, the different time periods, the different geographical areas and of course the respective needs and conditions left their mark on them.

Acta Sanctorum (Deeds of the Saints) is a multivolume edition of Christian hagiographical texts in Latin published over a period of ca. 300 years, beginning in 1643 in Antwerp. The first volumes were published by the Jesuit scholar Jean Bolland and continued by the *Société des Bollandistes*, an association of scholars named after its founder. The hagiographical texts included in *Acta Sanctorum* cover a time period from Late Antiquity to the Renaissance, while the geographical distribution of the places of action of the saints concerns the entire Christian world with an emphasis on Central and Western Europe especially from the early Middle Ages onwards. The texts are classified according to the feast day of each saint while the work remains still incomplete as the last day of the published calendar year is November 10th. The particularity of the work is found in the fact that it is based on an unprecedented, for its time, historical – critical method as it contains versions from all known sources for each text as well as the necessary research introductions and comments. The main goal was to separate the myth from the real events, thus reflecting the conditions and needs of the time that led to the birth of *Acta Sanctorum*.

This work explores the *Acta Sanctorum* through computational methods. On the one hand, the aim is to give through data analytics methods a complete and detailed overview of the data, the

data analytics process as well as the challenges that arose. On the other hand, with the help of Natural Language Processing, the goal is to identify the most frequent topics in *Acta Sanctorum* as well as to track changes in the topics over time and analyze how they relate to historical events and cultural trends.

Chapter 2 attempts to give a brief presentation of the hagiographical texts from their appearance until the Late Middle Ages, illuminating the various conditions which led to changes in the texts. These changes include the reinterpretation of existing texts through their partial or total rewriting depending on the needs that arose. These needs were sometimes due to purely stylistic reasons, such as the rewriting in a richer literary style and in a “cleaner” Latin language, while sometimes they were due to the needs of the respective era. A typical example are the *Passiones*, texts that chronicle the martyrdoms of the Christian saints, which after the cessation of the persecutions began to decline and give way to texts that described the lives of holy bishops, who were models for the consolidation of the Christian church that had just begun.

Chapter 3 provides a description of the new conditions that prevailed in the writing of hagiographical texts during the Renaissance. On the one hand, new technological discoveries, such as printing, and on the other, Humanism, the intellectual movement with a focus on cultural and educational reform, left their indelible imprint on the writing of hagiographical texts. In addition, important political and religious events such as the Reformation and the Counter-Reformation marked hagiography in their own way.

Chapter 4 provides a full overview of *Acta Sanctorum* as a historical – critical edition. The idea behind the work, the beginning, the key figures that contributed to its implementation and of course the entire journey up to the publication of the last volume. Also, a special focus is given to the particular circumstances in which *Acta Sanctorum* and the *Société des Bollandistes* were born.

The import, the cleaning, and the normalization of the data in the programming language Python are detailed in the first part of Chapter 5. The second part includes an exploratory data analysis,

which explores the text types, compares the length and complexity of the texts, and provides statistics on these, attempting to give a complete overview of the data. Data analytics challenges and the way to deal with them are also part of this chapter.

Natural Language Processing methods, especially topic modelling and its results, are presented in Chapter 6. The most common words and phrases of *Acta Sanctorum*, the most common topics and their relationship with historical events and time periods cover the main part of the chapter as well as the interpretation of the results. The challenges of applying Natural Language Processing methods to Latin texts and the ways of dealing with them are also contained in the same chapter.

Chapter 2 – Hagiography in the Middle Ages

2.1 Christian Hagiography: From its Appearance to the Fall of the Western Roman Empire (476)

Hagiography was one of the most preferred types of Christian literature during the Middle Ages. Etymologically, the word derives from the Greek words ἅγιος (Saint) and γράφειν (write) and constitutes “the body of literature describing the lives and veneration of the Christian saints”. (Encyclopedia Britannica, 2023: *hagiography*) The appearance of this type of writing already dates back to the Patristic Era (c. 100 to 451, Council of Chalcedon) when the first Church Fathers lived and wrote. (McGrath, 1998: 17 – 23) According to the Belgian Jesuit and church historian Hippolyte Delehaye (1859 – 1941), the two main Christian hagiographical types of writing are the liturgical and the literary. (Kosinski, 2016: 13) The former includes sources such as Synaxaria (lists of the saints arranged in the order of their anniversaries) and Menaia (biographies of saints read in sermons for specific dates of the calendar year divided by month) while the latter could be divided into six categories:

- i. The Life of a Saint (lat. *Vita*) is a narrative of the deeds of a saint during his/her lifetime. It shows a conventional structure throughout the Middle Ages: it usually begins with the birth of the saint, occurred with some miraculous events which foreshadow his/her sanctity, while references to his/her childhood include portents of his/her sanctity. At some point in his/her adult life, after renouncing worldly things, he/she will spend a period of ascetic training near one or more spiritual fathers. From then on, he/she follows a model Christian life where through his/her wise teachings and miracles he/she reveals his/her closeness to God and his/her fame is slowly spread among the faithful people. After his/her death, which includes some miraculous events as well, he/she is not to be forgotten by the faithful and thus through the miracles and relics his/her cult is established.

- ii. The Martyrdom (lat. *Passio*) shows many similarities in its structure with the *Vita*. Unlike the *Vita*, the *Passio* focuses primarily on the martyrdom of the Saint. Despite the horrific torture, the Saint refuses to renounce his faith in God and this leads to his execution. His martyrdom is usually likened to a battlefield or bears the characteristics of an athletic contest in order to refer to the battle against demons. (Louth, 2008: 359) His brilliant example as well as his miracles established his worship by the faithful. *Vita* and *Passio* are the two most widespread genres of hagiography. (Sauer, 2011: 1798)
- iii. “The *Inventio* or *Revelatio*, the story of how a new saint or more often a saint’s bodily remains were discovered.
- iv. The *Translatio*, the story of how a saint’s relics were brought to a church or moved to a new shrine.
- v. The *Visio*, the story of how a saint appeared to someone in a vision.
- vi. The *Miraculum*, the story of how a miracle was performed on the saint’s behalf by God”.
(Hayward, n.d.)

Nevertheless, the distinction between these six species is not always distinct and most of the time there is a combination of two or more types between them. Research so far has shown that the origins of hagiography are largely based on the Greco-Roman literary tradition. In the first hagiographical texts of Late Antiquity, influences from literary genres such as the Apotheosis/Deificatio, Encomia and ancient biographies of philosophers can be discerned. (Sauer, 2011: 1799) Nevertheless, the primary source was the Bible, from which examples of moral behavior were drawn as models for the composition of the figure of the Saint who would gain credibility through his resemblance to the famous saints of the Scripture. (Kosinski, 2011: 14) The promotion of the saints’ worship and many times by popular demand, as well as the establishment and teaching of faith decisively contributed to the writing of hagiographical texts. (Pontius et al., 2001: 9) The appearance of these texts already dates back to the time of the persecutions of Christians in the Roman Empire. The first surviving Christian

biography was that of the Carthaginian Bishop Cyprian (d. 258). It was written by the deacon Pontius and chronicles Saint Cyprian's life from the moment he was baptized a Christian until his martyrdom. Another example is the Martyrdom of Saint Polycarp (d. 155), who was Bishop of Smyrna in Asia Minor. The hagiographical text does not stop at Polycarp's martyrdom but also tells how the faithful collected his relics and established his cult. The two cities, Carthage and Smyrna raised up the memory of the martyrs from their own ranks. The Roman Persecutions inadvertently provided one of the primary means through which Christian communities created a sense of identity and historical consciousness. (Head, 1999: 2) Similar phenomena also occurred in bigger cities such as Rome, where the Apostle Peter emerged as the patron saint of the city, but also in smaller cities such as Nola near Naples, where the Martyr Felix was established as a beloved local saint. The cult of Martyrs was, at least in its early stages, localized with a focus on the tomb where worshipers hoped to be benefited from the deceased martyr's intercession in the afterlife. (Rapp, 2004: 1249)

Already from the end of the third century and mainly after the cessation of the Persecutions at the beginning of the fourth century, the religious movement of monasticism and especially in the form of Eremitism saw a great flourishing mainly in Egypt, a fact that is also reflected in the hagiographical texts of the time. Undoubtedly the best known of these are the Life of Saint Anthony, written by the bishop of Alexandria Athanasios and the Life of Saint Paul the Hermit, written by Jerome. In addition, at about the same time, cenobitic monasticism also appeared in Egypt, when Saint Pachomius founded the first monastery. Unlike the eremitic monks, who lived in cells in the desert, the cenobitic monks lived in groups in monastic communities. The general flourishing of monasticism led to the writing of more and more Lives of the Desert Fathers; that is, of the monks who abandoned the worldly and spent their lives in the desert.

The emerging Christian cult of the saints during Late Antiquity was expressed through the hagiographical texts, which were used in the church services of that time and aimed not only to consider the Saint and his good deeds as an example to be imitated but also to be "seen much more as one who, in his earthly life, demonstrated his closeness to God, not only by his godly life, but by his

ready access to God in prayer, and the divine power he was thus able to wield, and who now, as a friend of God in the heavenly court, is able to intercede with God for those for whom he is concerned – in short, a figure of power that can be drawn on by those who cultivate a relationship with him”. (Louth, 2008: 359) The Martyrs and Fathers of the Desert are basically the central figures of the hagiographical texts up to the middle of fifth century and undoubtedly reflect the currents of that time. On the one hand, the vivid descriptions of the Persecutions and the Martyrs in the hagiographical texts contributed to the forging of the identity of Christians both at the universal and local level, while on the other the Lives of the monks brought the faithful into contact with the ascetic life and its desert teachings. As it was previously mentioned, the hagiographical texts of Late Antiquity display motifs from Greco-Roman Tradition as well as from the Bible.

Nevertheless, they introduce new patterns that respond to the needs of the creation of the Christian identity and worldview. For example, the Desert Fathers are often seen meeting animals in the desert (mainly lions), who are their companions or have been healed with their help and show them their gratitude. The animals seem to understand the monk’s holiness, and since nature is now subordinate to God, interacting with them can show people what holiness means and how worshipping saints bring them closer to God. (Gauly, 2020: 180) A new pattern is also the incorporation of pagan traditions and reinterpretation within the framework of the Christian religion through hagiography. Research has shown that the introduction of the cult of Santa Mariña to the region of Galicia in Spain was the driving force behind the Christianization of a landscape previously used as a site of Celtic worship. (González-García & Quintela, 2016: 447) The shaping of the Christian identity through hagiography, in addition to the universal character, also reflected the local specificities as well as the purposes of the hagiographers.

Although the majority of hagiographical texts was anonymous there are cases where the authors are known. Athanasius of Alexandria, in the Life of Saint Anthony, associates the Saint almost entirely with the ascetic movement, completely isolating him from the heretical world and making him the hero of

Orthodoxy and a follower of radical ascetic virtues. (Felix, 2008: 1) In this way, Athanasios leaves no room to claim the sacred heritage of Saint Anthony to the Arians, who cost him his seat as Bishop of Alexandria. The way of presentation of the Life of Saint Martin of Tours by Sulpicius Severus is also interesting. Saint Martin was a monk and then bishop of Tours. Sulpicius Severus portrayed Saint Martin as a man who sought solitude, but later in his role as a bishop in a world full of corruption he became the perfect model of a holy man. (Felix, 2008: 1) It is worth noting that Sulpicius Severus was on the one hand a follower of strict asceticism and on the other a heavy critic of the Gallic clergy.

So, it seems that during the establishment of the Christian religion, local needs also dictated different standards of writing. Regarding the language, although the stories about saints and their miracles during Late Antiquity were spread orally in various languages such as Greek, Latin, Coptic, Aramaic, the majority of written hagiographical texts are in Greek. Greek was not only the language of classical and biblical culture but also the language of monastic and ascetic culture. (Rapp, 2004: 1248) This linguistic influence from the basically Greek-speaking East led to the existence of a large number of Lives of Eastern Saints in the Latin Hagiographical collections of the time even though they did not always directly derive information from Greek sources. (Rapp, 2004: 1248) However, translations of both written sources and oral traditions were also commonplace.

2.2 Latin Hagiography in the Middle Ages

Social, economic, and political changes in the Roman Empire during the fifth century affected both the role of spiritual authority in the state and the internal church system itself. The upper Christian clergy, into which the Roman aristocracy had begun to infiltrate, began to play a more active role in public life and administration. This phenomenon did not leave hagiography unaffected. The Martyrs and Desert Fathers as the main characters of the hagiography until then will give their place to the bishops of the cities. The local bishops, in addition to their ecclesiastical duties, began to occupy an intermediate position between the poles of secular and religious power where they carried out embassies, engaged in the release of captives as well as in the defense of their interests and the interests of their parishes

in dialogue with the state authorities. (Babina, 2011: 73) The general weakening of the secular political and administrative structures on the one hand and the active participation of the educated Roman aristocracy in the upper ranks of the clergy on the other (Saints Paternus, Hilary, Germanus, etc.), contributed to a large extent to the creation of the new role as patrons in relations between their flock and the outside world. Even geographically, the writing of the hagiographical texts shifts mainly further north to the kingdom of the Merovingians. A typical example of this new situation is the Life of Saint Germanus bishop of Auxerre, written in the second half of the fifth century by a cleric named Constantius of Lyons and sets forth a new compromise model of interaction between the saint and secular power. (Babina, 2011: 82 – 83) Through his education, prayer and miracles, Saint Germanus manages to win the respect of secular authorities, participate in diplomatic missions, heal the sick, preach Christianity, obtain tax exemptions for his flock, making it clear that he obeys only one authority, the divine one.

However, the interactions were not always peaceful as it was in the case of the negotiation with King Goar of the Alans. Goar tried to push back Saint Germanus, but the Saint stopped him by grabbing his horse by the reins and as a result the Alans were led to negotiations, which brought peace. Unlike earlier Lives of bishops, such as the aforementioned Saint Martin of Tours, the Lives from the second half of the fifth century (*Vita Ambrosii*, *Vita Epiphani*, *Vita Germani*, etc.) present the bishops as having to some extent worldly heroic elements and yet to come and in conflict with the secular authorities in order to establish the dominance of the church. (Babina, 2011: 92)

Another key element of the hagiographical narratives is the reflection and reinforcement of spiritual and institutional changes within the Church under the Franks in the early Middle Ages. The infrastructure of rural churches, the process of establishing parishes in the countryside, the strengthening of the cult of local saints, the efforts of the Merovingian clergy to make Sunday a day of rest as well as a stream of miraculous stories (particularly healing feats) appear increasingly prominent in hagiography of the era. (Mériaux, 2019: 144 – 145) The Life of Saint Eligius, written by his close

friend and companion in the royal court and the Church, Audoenus, is another example. He began his professional career as a goldsmith at the court of Clotharius II and later as chief adviser to Dagobert I. After the death of Dagobert I he became bishop of Noyon and Tournai in Northern Gaul. (Mériaux, 2019: 145) Following the models of Columbanus' (~540 – 615) monasticism¹, Eligius foreshadows the more explicitly missionary models of Episcopal sanctity, for example in describing at length the visit of the bishop to a small village near Noyon where he tried to forbid pagan dances and sharply focuses on concern for the cult of martyrs and saints from northern Gaul. The efficiency with which the bishops served both God and the Merovingian kingdom is another new element found in the Life of Saint Eligius and also in other hagiographical texts of the period (Desiderius of Cahors, Paul of Verdun, Sulpicius of Bourges, Audoenus of Rouen).

The renewal experienced by the monastic institutions during the seventh century and especially the monastic reform during the Carolingian period also left its mark on the hagiographical texts. In the Concilium Germanicum of 742–743, which was organized under the auspices of the Carolingians, it was stipulated that monks and nuns should live according to the rule of Saint Benedict and govern their lives accordingly, which reflects among other things the desire to establish some kind of order in the Carolingian kingdom. (Kramer, 2020: 434) The *Vita Benedicti Anianensis*, written by the monk Ardo, moves in this direction. Benedict is described as one of the most important reformers of the kingdom through the emperor's delegation where he emphasizes the importance of cooperation as opposed to conflict and power games. (Kramer, 2020: 442) Nevertheless, the hagiography also captures the objections against the reform that existed, as in the collective biography of the founding saints of Sithiu Abbey (later Saint – Bertin) in northern Gaul. Influences from other monastic traditions do not necessarily indicate outright opposition to the reform, but nevertheless reveal a will to preserve certain monastic traditions that differed from the Benedictine rule. (Mériaux, 2019: 151)

¹The *Regula Monachorum Columbans* was stricter than the *Regula Benedicti*. It included individual commitments to study and work based on community services, regular choral prayer, many days of fasting and periods of silence.

The Episcopal and monastic models of holiness bishops. In the Merovingian period were adopted and further developed in the Carolingian period as well, during which almost no new saint-bishops were created but the reforms introduced as early as the end of the eighth century led to the rewriting of a large number of biographies of existing holy bishops. (Mériaux, 2019: 146) The rewriting of hagiographical texts began to be an increasingly frequent phenomenon. According to Monique Goullet, rewriting could be divided into three types:

- i. Purely quantitative aspects. This part includes the techniques of reduction (removing parts of a text), augmentation of the text (adding parts of a text) and substitution, which is a mixed technique where someone deletes a part of a text and replaces it with another.
- ii. Non-quantitative formal aspects. Translation from one language to another, versification (prose into verse), prosification (verse into prose), change in the way of narration (dramatization and narrativization), etc. are some examples of non-quantitative formal aspects.
- iii. Semantic or conceptual dimension. As a person's path to holiness is already planned before the beginning of time it becomes difficult for his action to be modified by the hagiographer. Nevertheless, semantic changes usually concern modifications for the subject of the action to adapt to a new historical context or concern transformations of a historiographical type that aim to stabilize the story more clearly in its context. Another category of semantic formation has to do with the transformation of a minor character into a main character, for example when a collective *Passio* is divided in order to be dedicated separately to each saint. (Goullet, 2003: 109 – 113)

There is no doubt that most of the time a rewrite implements these three categories simultaneously. (Goullet, 2003: 109) Research has shown that the phenomenon of rewriting hagiographical texts (especially *inventio* and *miracula*) was much frequent and concerns mostly local rather than global saints. (Snijders, 2013: 26) Spiritual and political reasons as well as changing demands and practical concerns often led to the rewriting of a hagiographical text. For example, monasteries that possessed

the relics of a saint often wished to have more than one scenario for finding the relics of that saint. (Snijders, 2013: 18) The reasons could of course also be stylistic: the Carolingian scholars considered that the Latin grammar of the Merovingian era needed correction, which led them to systematically rewrite hagiographical texts. (Mériaux, 2019: 146)

The Christianization of Anglo-Saxon England from the seventh century and later of the Scandinavian countries influenced in turn the production of hagiographical texts by imparting new narrative patterns. An example could be, Bede's *Historia Ecclesiastica Gentis Anglorum*, animated "by an underlying moral purpose which notably includes both the earthly contributions of English holy people and miracles attesting to their supra-secular presence". (Sanok, 2019: 423)

On the other hand, the post-Conquest hagiography introduces new standards as there was a "production of texts that presented patron saints of monastic institutions to a new audience of Norman prelates, administrators and monks". (Sanok, 2019: 426) Studies have shown that during the Christianization of Britain many local pagan and folklore legends were mentioned and incorporated into the hagiographical texts. (Powell, 2010) The same fact applies to the Scandinavian countries. (Lindow, 200) The *Clemens Saga*, the legend of the Bishop of Rome Clement, describes pagan practices as it recounts how the Saint through his wisdom managed to teach the pagans the Christian faith. Although Saint Clement died in the year 96 and his Life presents with names of gods from Ancient Greek and Roman religion, the eleventh century Icelandic hagiographical text replaced the names with their counterparts from Norse mythology. (Carron, 2005: 8 – 9)

The growing hostility towards the Islamic world during the period of the Crusades was also reflected in the hagiography. Greek hagiographical texts of the ninth and tenth centuries, narrating the encounters of saints with Muslims, were discovered and reused in the West during the Crusades after being translated into Latin. (Jotischky, 2022: 3) Peter the Venerable compiled a dossier of anti-Islamic polemics, forging a "Cluniac identity" in which the monastery was the intellectual and spiritual engine room of Christendom's defense against the challenges posed by heresy and infidelity while in one of

his letters he pointed out that the Arab conquests in Asia, Africa and the Iberian Peninsula mimicked the “infection” of the same regions by heresy. (Jotischky, 2022: 4) The heresies during the same era brought forth saints, who emerged as models of orthodox piety that contradicted the heretical beliefs and practices of not only by their example but also through their words and their preaching. (Kienzle & Stevens, 2014: 33)

Already from the High Middle Ages, hagiographical texts began to be (re)written more and more frequently in languages such as Middle English, Old French, Old Norse etc. in order to be understood by the laity, who had poor or no knowledge of Latin. This resulted in the influences of the flourishing vernacular literature penetrating the Latin *Vitae* to a great extent and imparting new content and form. Legends of older saints often underwent a transformation enriched by many time-related factors in order to survive alongside them, and new saint cults flourished, with changed forms of piety and a changed ideal of holiness. (Werner, 2014: 523) The spread and influence of the religious orders and their sermons brought in turn new patterns of hagiography. Such a work was the *Legenda Aurea* written in Latin between 1260 – 1275 by Jacobus de Voragine, archbishop of Genoa and member of the Dominican Order. (Sauer, 2011: 1800)

This work is a collection of Lives of Saints distributed according to their feast day. In addition to the *Vitae*, it also contains the etymology of the names of the saints as well as references from the sources from which Jacobus de Voragine drew his information. The *Legenda Aurea* enjoyed great appeal in Europe during the Late Middle Ages, a fact that later led to its translation into other languages. (Sauer, 2011: 1800 – 1801) From the thirteenth century, collections of saints’ Legends, Lives, and Miracles became much more mass-produced literary goods. A new feature is the diversity of fauna found in these texts as it includes legendary beasts such as griffins, unicorns and cynocephali, “exotic” animals as well as more familiar species indigenous to Europe. The representation of animals in these texts deviates from the lived experience of their readership, whether for the benefit of religious edification or dramatic pleasure-seeking spectacle, this literature offers its audience a glimpse of an unknown

world. (Salter, 2018: 84) Another thirteenth century hagiographical text that reflected the ideas of its time is the *Life of Saint Elisabeth of Thuringia*, written by the Dominican Monk Dietrich von Apolda. Saint Elisabeth was considered one of the most popular saints of the late Middle Ages, and her *Vita* was one of the most widely circulated saint *Lives* in Central Europe in the fourteenth and fifteenth centuries. (Werner, 2014: 525) Saint Elisabeth (1207 – 1231) was the daughter of the Hungarian king Andrew II and Gertrude of Merania, daughter of Bertold Duke of Merania. After marrying at the age of 14 to Louis IV the Saint, widowed at 20, giving up her vast wealth, becoming a symbol of Christian grace, and finally dying very early at just 24.

The new and different about this *Vita* compared to its templates and predecessors was Dietrich's attempt to combine the scattered tradition of Saint Elisabeth, which had hitherto had little literary form, into a portrayal of the life of this famous saint that was appropriate in terms of content to the importance of her cult. A second and more decisive factor ensured resounding success and hagiographical tradition forming power: It was the first linking of the previously purely hagiographically processed Elisabeth material with the extensive "profane" traditional property, which revolved around the Thuringian sphere of Elisabeth's life. Dietrich managed to combine both sets of traditions with great literary skill into a unified, coherent work. The realization of this project leads to a depiction of the outer life of Saint Elisabeth, a biography that captures the historical individuality and which, one might further assume, for precisely this reason also lends the saint "human traits" and takes an interest in her private fortunes. (Werner, 2014: 540)

Chapter 3 – Humanist Latin Christian Hagiography in the Renaissance

The French term Renaissance (English: Rebirth) describes the European period of transition from the Middle Ages to the Modern Age and mainly concerns the 15th and 16th centuries, although its beginnings already exist from the 14th century. The term Renaissance was established by the French Historian Jules Michelet (1798 – 1874) in the middle of the 19th century, who presented in the 7th volume of his *Histoire de France*, the Renaissance as a distinct era, which he considered as the beginning of modern thought. (Garin, 1964: 431) The Renaissance was characterized not only by a multitude of geographical discoveries (New World) but also technological inventions (printing), which, in combination with social and political transformations such as the retreat of feudalism and the development of trade, radically changed the European status quo. On the other hand, Renaissance art was characterized by a focus on realism and naturalism, the use of classical forms and motifs as well as an emphasis on secular themes, while literature by the revival of classical learning and wisdom, logical thinking, and reasoning in a larger base of readers. More generally, during the Renaissance, there was a flourishing of the study of the classical heritage, an intensive search for manuscripts from the classical Greco-Roman world as well as the thorough study of the Latin and ancient Greek languages. At the same time, Renaissance Humanism appeared in Northern Italy, a strong intellectual current that later spread to the rest of Western Europe and aimed at the cultural and educational reform of the model that prevailed until then. The whole educational program of the Renaissance is known by the Latin term *Studia Humanitatis*, a term that was already used from the 14th century and from which later (16th century) the Humanists, the representatives of this intellectual current, took their name. The Humanists provided a basis for the study of ancient texts and the assimilation of the values of the classical world at all levels, from art to the study of practical and natural sciences. As language was seen as a means of self-expression of human rationality and people's ability to convey meaning, a basis was given to the cultivation of expressive ability mainly through the teaching of rhetoric, grammar, and poetry.

The new conditions of the Renaissance could only influence the writing of hagiographical texts and give them new characteristics. The “cleansing” of the Latin language from barbarous medieval corruptions, the refinement of style and the disdain of other languages by the Humanists led to a widespread writing of Latin hagiographical texts as classical Latin was the mark of the superiority of the new intellectual movement over the medieval past. The few new Lives of Saints that appeared during the Renaissance indicate that the hagiographers did not have the Popes and the Roman curia in their minds as a principal audience but the local readership, as the new hagiographical texts were composed as a part of the canonization petitions (Collins, 2006: 517). In fact, there has been a great wave of rewriting of already existing texts, which can elucidate the social and intellectual contexts of their composition more directly than the great literary works of the Renaissance can (Frazier Knowls, 2005: 13). Alison Knowles Frazier in her book *Possible Lives: Authors and Saints in Renaissance Italy*, divides the structural model used by humanist authors into five categories:

1. Classical *Vitae*: The main influences came from the Gospels, from well-known authors such as Sallust and Suetonius but also from the hagiographical texts of the Late Medieval era.
2. Patristic *Vitae* and early Christian *Passiones*: By translating and rewriting patristic and early Christian hagiographical texts, the humanists contributed to the spiritual life and literature of the West in a way that paralleled the impact of their translations of the Greek classics.
3. In contrast to rigid medieval hagiographical canonization models that contained a specific order of events, humanist hagiographers rarely repeated the mass collection of miracles entailed by canonization.
4. Mendicant encyclopedic model: This is a model that appeared during the Late Middle Ages (e.g. *Legenda Aurea*) and was preserved during the Renaissance. It largely contains an authoritative and entertaining abridgment of hagiographical narratives.
5. Liturgical: The hagiographical texts during their rewriting were arranged according to the liturgical year and were read during the Divine Liturgy and the Sacred Mysteries. They were generally shorter and divided into sections and could also contain excerpts from paternal

sermons and passages from the Bible. As this genre required elements of contemplative intelligence and rhetoric, the Humanists applied them to the abilities of the audience. (Frazier Knowls, 2005: 20 – 26)

Another new feature introduced in the Latin hagiographical texts of the Renaissance is the reduction of miracle accounts and the increase of the part presenting the life of a saint. This contrasted with contemporary vernacular narratives but also with earlier medieval practice as Humanist hagiographers placed more weight on the biographical narratives. As far as the central figures of the Renaissance hagiographical texts are concerned, the Humanists dealt little with the lives of saints related to mysticism and extreme asceticism without missing some examples such as the Life of Saint Nikolaus von der Flühe. On the contrary, the hermits, the martyrs and the bishops had a central role.

On the one hand, elements such as prayerfulness, humility and moderate asceticism were expressed through the hermits and the Martyrs while on the other, the bishops were examples, whose main responsibilities were not within the monastic walls but in the outside world (Collins, 2006: 535). The Humanists were primarily concerned with the holy bishops and focused on their characteristics such as secular spirituality and the reformation of religious life outside the monasteries, qualities which were in harmony with the *Studia Humanitatis*. The preparation of an exemplary diocese as the goal of spiritual reform is reflected through the lives of the bishops, which now acquire an idealized socio-political role and are the model of Episcopal action. These concerns led to the emergence of a new phenomenon: the emphasis on the last words of the bishop before his death, which functioned as a testament. This phenomenon appeared to a small extent in the Late Middle Ages but during the Renaissance it received large dimensions. The deathbed oration was a moment of public teaching that combined the spiritual and temporal concerns of the time and imparted to the bishops the concerns of the Humanists, such as conscientious attention to the care of souls, the peace and prosperity of the city through charity, justice and education, qualities that are frequent motifs in hagiographical texts. For example, Pietro Barozzi (d. 1507), bishop of the Italian city of Belluno, translated a Greek Life of Basil the Great. The text closed with the admission that a good bishop was like a good teacher and that

a good account of the life of that Bishop made the lesson exciting. The humanist Francesco da Castiglione (ca 1420 – 1484) in his *Life of the Dominican Antonino Pierozzi*, Archbishop of Florence, argued that the more the bishop resembled a humanist educator, the better. It is clear that the humanists projected and called for imitation of their ideals through the hagiographical texts they produced.

Research has shown that Humanists primarily rewrote Lives from Late Antiquity and the Late Middle Ages (Frazier Knowls, 2005). Although the neglect of Medieval hagiographical texts from the 6th to the 12th century could mean contempt due to the view of barbarism of the time, the reasons have mainly to do with the Humanists' preoccupation with specific concerns coinciding with Late Antiquity (e.g. revival of classical learning) and the Late Middle Ages (e.g. the patron saints of their audiences derived mainly from that era).

Of course, many hagiographical texts of the Renaissance also reflect purposes which were in sharp contrast with the geographical and political environment where the saints lived (Collins, 2006: 529 – 533). One such example is the work of the Westphalian Humanist Johannes Cincinnius (ca 1485 – 1555) entitled *De vita et sancta conversatione beatae Idae*. Saint Ida of Herzfeld lived during the 9th century and was married to a Saxon Duke while her ancestry was linked to the Carolingians. After the death of her husband, she devoted her life to the poor and lived the rest of her life close to a group of nuns. Johannes Cincinnius placed her Life within another context, strongly influenced by the work *Chronicon Westfaliae* of the monk Werner Rolevinck (1425–1502). His goal was “to inspire the Westphalians with descriptions of the great deeds of their ancestors, the beauty of the Westphalian landscape, and the significance of their political and social institutions and customs, and so to mobilize them to make their present situation as glory as their historical one”. (Collins, 2006: 531)

Another example is the Thuringian Legend of Saint Boniface, a work written in the 16th century in Erfurt. The peculiarity lies in the description of political and ecclesiastical matters of Thuringia, which were largely unrelated to the life of St. Boniface. At the same time, the anonymous author presented

very briefly the Martyrdom of the Saint, characteristic of humanistic hagiographical texts, even though it was a popular narrative during the Middle Ages. Finally, special mention should be made of a new factor of that era which shaped new conditions in hagiography as well, the printing. In addition to the easier production of copies, new features, such as poems and epigrams at the beginning and at the end of the hagiographical texts, started to adorn them with new features.

Although the Humanist spiritual movement originated in Northern Italy as early as the 14th century, it spread throughout Western Europe during the 15th and 16th centuries and set its own rules and patterns for the writing of Hagiography. An exception is in a way England, as there was a slow spread of humanism which resulted in the rarity of hagiographical texts (Henry, 2013: 536). Nevertheless, the English examples of hagiography are primarily written in Latin, unlike their predecessors in the Late Middle Ages, which were mostly in Old English.

October 31, 1517, when Martin Luther nailed his 95 Theses to the door of All Saints' Church in Wittenberg, Saxony, is considered the beginning of the Reformation, a movement that was a turning point in history, as it gave rise to Protestantism, the third largest branch of Christianity. The motivation for the Reformation was primarily theological, as Luther's criticism was directed against the (later so-called) Catholic Church, the sale of indulgences and the doctrines of Purgatory, the Last Judgment, the worship of the Virgin and the saints, the most sacraments, the compulsory celibacy of the clergy, the monasticism, and the authority of the Pope. However, there were other causes such as the rising nationalism, the widespread corruption in the Vatican, as well as the influence of humanist philosophy and Renaissance thinking, which put human at the center. Luther's publicized displeasure quickly gained traction, bringing to the surface a widespread sense of discontent. The nascent reform movement made widespread use of pamphlets, the distribution of which was made possible by the mechanized printing press introduced a little earlier. The Reformation movement spread to Germany along with other reform movements arose independently of Luther's. The largest groups that worked to spread the Reformation were the Lutherans and the Calvinists. Lutheran churches took root

especially in Germany, the Baltic countries and Scandinavia, while Calvinism spread especially in Switzerland, France, the Netherlands, and Scotland. The Protestant movement also had a significant influence on the Anglican Church, which was introduced to England in the 1530s.

According to Luther, valid church tradition simply had to be based on the Bible and his goal was to be able to communicate the “rediscovery of the Gospel” to people (Brooks, 1984: 144). His views on salvation were also at odds with the Catholic Church: Luther argued that God secures salvation through faith and not in relation to human works. It is noticeable that within such a doctrinal framework, not only some of the Holy Mysteries were rejected, but also the worship of the Mother of God and the saints. The collections of indulgences and venerating relics or going on pilgrimage were considered acts equivalent to blasphemy and idolatry as they distracted the faithful from the search for faith for Christ. Luther was convinced that modelling a life on a Saint not only obscured the redemptive act of Christ, but invariably led to a fatal dependence on the righteousness of good works (Brooks, 1984: 143). As for the hagiographical texts, they were heavily criticized not only for the reasons mentioned above but also for containing false evidence or even being entirely legends and false stories, which cultivated and strengthened the Vatican’s propaganda. The saints in the hagiographical texts were presented as examples to be imitated and many times their lives contained the concept of infallibility, something that was usurped by the respective prior of the Holy See. Moreover, the Catholic clergy surrendered the authority of the Saints to support the sacraments of the Canon, which Luther denounced as completely contrary to the Word of God. In short, the Lives of the Saints were not to be valued in the same way as the miracle stories of the Gospels.

As early as 1520, Pope Leo the 10th condemned 41 of the 95 positions as heretical, while in 1521 Luther was excommunicated. In the same year, Luther was called to an apology before the Diet of Worms where he refused to retract his positions. The Catholic Church’s official reaction to the Reformation is known as the Counter-Reformation. This reaction was characterized by the acceptance of some demands for ecclesiastical renewal, which had also been emphasized by Luther, Calvin and the other

reformers such as the fight against immorality in the clergy, but also by a warlike and defensive reaction to Protestantism, to prevent other Catholics from passing to the evangelical Churches but also with the violent repression of evangelical Christianity in the context of a collaboration between the Papacy and the “Catholic States”, mainly Spain (through the control of the press, the role of the Roman Inquisition, etc.).

Hagiographical texts after 1517 reflect this polemic against the Reformation, such as the poem *De vita et gestis divi Hyacinthi* of Nicolas of Hussow, which was printed in Krakow in 1525 and captures the struggle between the reform movement and its opponents in Poland within a political context linked to the Prussian Homage (Ledzińska, 2015). The Counter-Reformation period began with the Council of Trent (1545 – 1548) and ended conventionally with the Peace of Westphalia in 1648, which ended the Thirty Years’ War (1618–1648) and added Protestantism to the list of acceptable faiths.

The Council of Trent was held from 1545 to 1563 under the presidency of three successive popes: Paul III, Julius III, and Pius IV. Doctrinally, the council clearly formulated the Catholic doctrines of salvation, the sacraments, and the biblical canon, and standardized worship throughout the church, prohibiting local variations and clearly distinguishing the religious theory and practice of Catholicism from that of Protestantism. It strengthened the centralization of the church and the power of the bishops, subordinate only to the Pope. The Council of Trent also emphasized the need for increased religious and moral education in contrast to Protestant propaganda/teaching.

The leaders of the Catholic Church countered Protestant criticism and questioning of the veneration of the saints by arguing that popular manifestations and practices of worship were not only unfamiliar to the Christian faith but also brought believers closer to piety. Research has shown that areas where the cult of saints was highly developed effectively resisted Protestantism in contrast to areas with few saints, which more readily adopted the Reformation (Pfaff, 2013). This fact gave the Catholic Church the opportunity to further promote the traditional worship of the Saints, to encourage the worship of new saints but also the worship of relics, the promotion of pilgrimages and the celebration of miracles,

especially if they could interpret them in a way that neutralized the Protestant faith (Luria, 2018: 114 – 117). The new model of holiness of the Counter-Reformation had to be a real person, preferably a priest or member of a religious order. During his/her career on earth, the saint ideally had either founded a religious order, conducted missionary work, given charity, administered a diocese, or performed feats of mysticism (Nalle, 1992: 33). Typical examples of Counter-Reformation saints are Teresa de Ávila, Ignatius Loyola, and Diego de Alcalá. It is perceived that in the period after the Council of Trent the inhabitants of the Spanish monarchy appeared on the scene with great power, following the vast reach of the Spanish monarchy itself, and played an important role in the formulation and dissemination of goals and norms.

From the late 16th century, as the Protestant impact began to recede, the Catholic Church increasingly viewed cults with a contempt, born of humanist education and a fear of disorder, which encouraged repression in cases of disobedience among the laity (Luria, 2018: 115). The new conditions necessitated both a reexamination of the veneration of the saints and the codification of canonization rules, which became evident with the establishment of the Sacred Congregation of Rites in 1587, which was entrusted with the task of overseeing the entire process of canonization while Urban's reforms VIII in 1634 completed the process of centralization. The participation of the faithful in the canonization of the new saints decreased and the hagiographical texts of the old saints had to be critically examined in order to verify their authenticity and to separate the legends from the real facts. New conditions imposed new standards and rules as Western Europe moved into the Age of Enlightenment.

Chapter 4 – *Acta Sanctorum*

On August 15, 1534, the Society of Jesus, best known as the Jesuit Order (Latin: *Societas Iesu*) was founded by the Basque nobleman and ex-military Ignatius Loyola and his associates in Paris. The foundation of the order was officially confirmed on September 27, 1540, in Rome by Pope Paul III and the following year Loyola was elected as the first head of the Society of Jesus. It was an essentially new Christian religious order of the Roman Catholic Church, whose members were called Jesuits.

Initially, Loyola and his successors were not elected for a fixed term, as was customary, but for life, while each newly inducted member had to take the vow of poverty, chastity, and obedience, as well as go through the “spiritual exercises” developed by Loyola. In the 16th century, the “exercises” were criticized, as many Catholics saw in them a special form of mysticism, far from the accepted rules of orthodoxy, which is why Loyola had to defend them several times. In addition, they had to take a fourth vow, which was the personal vow of submission to the Pope. This fourth vow significantly distinguished the Jesuits from other religious orders.

Among the main tasks of the Jesuits was the preaching of the Christian religion to the New World, as well as fighting the forces of the Reformation in Europe. Through the numerous schools and universities, they founded, the Jesuits provided Catholic education to large numbers of people in Europe and especially to the aristocracy. Their educational institutions were structured around a comprehensive curriculum called *Ratio Studiorum*, and they taught Latin, Greek, classical literature and poetry, philosophy, languages, sciences, and arts. (Schwickerath, 1911) They secured from the Holy See a large number of privileges, including the right to grant degrees from their universities, no local authority (secular or spiritual) having jurisdiction over them, while they could also unmolested enter the territory of any existing monastery. Two of the members of the order were among the main leaders of the Council of Trent, and the order as a whole contributed greatly to the victory of Catholicism in Poland. Thanks to Loyola’s leadership, the good organization and training of the Order’s members, the Catholic Church gave rise to the Counter-Reformation and managed not only to survive

but also to go on the offensive. Both the Society of Jesus and the other religious orders played an important role in the spread of Christianity outside of Europe, and as early as the beginning of the 17th century, a spiritual renaissance in hagiography appeared, which mainly concerned new saints who taught and/or were martyred in the colonies. Hagiographical texts about pious lay men and women, who were not only European but also native or mixed-race missionaries, appeared more and more frequently during the 17th and 18th centuries. (Leavitt – Alcántara, 2014: 717)

Nevertheless, in Europe after the Council of Trent, the canonizations of saints became increasingly rare as the Catholic Church proceeded to a centralization of the process, leaving a smaller percentage of participation to the laity. The post-Tridentine Catholicism, challenged by Protestantism, tried to recreate its historical genealogy, and renew its interests in its traditions in the light of rationality that stemmed from the humanistic education. The standardization of canonization rules, the revision of the canonization framework as well as the reconsideration of the popular veneration of the saints also necessitated the editing of the existing hagiographical texts. One of the main elements of this treatment was the separation of legends from real events. Moreover, it was the verification of their authenticity as well as their reinterpretation as for example the “assembly” of many local stories into a universal frame, which in turn gave a uniformity to the worship of the saints. This not only strengthened the centralization of the Catholic Church but also answered the Protestant accusations of blasphemy. Within this context, the *Acta Sanctorum* was “born”.

In 1607 a small volume entitled *Fasti Sanctorum quorum vitae in belgicis Bibliothecis manuscriptae* was published in Antwerp. Its author was the Jesuit Heribert Rosweyde (1569– 1629), a philosophy professor at the Jesuit college in Douai. Rosweyde’s published volume was only a detailed plan for the work he intended to produce a few years earlier. In 1603, he had obtained permission from Olivier Mannaerts (1523 – 1614), founder of the Society of Jesus in South Holland, to collect and publish the hagiographical manuscripts of the libraries of Belgium. Altogether, the manuscripts numbered 1300

and Rosweyde procured copies of most of them. The plan included 18 volumes, which were divided as follows:

- “De vita Christi et festis eius,
- De vita beatae Mariae et festis eitis,
- De sanctorum festis diebus publicis solemnibus,
- 12 volumes with Lives of Saints divided by month in chronological order,
- 1 volume of Martyrologies²,
- 1 volume with commentary on the texts (8 books in total),
- 1 volume with 13 tables”. (Delehaye, 1920: 10 – 11)

Notes	Tables
1. Authors of the Lives of saints	1. Alphabetical table of the saints
2. Torments of martyrs	2. Table of the saints with indication of the country of origin, of the condition, of the quality, of the time, of the place of birth, of the author, of the Life
3. Images of saints	3. Table of saints by state (religious, virgins, widows, married persons)
4. Ecclesiastical rites mentioned in the Lives	4. Table by functions and dignities (apostles, bishops, etc.)
5. Profane rites	5. Table by countries and provinces
6. Chronological issues	6. Table by localities where the saints are honored as patrons

²No details are provided about the contents of this volume.

7. Geographic issues	7. Table by order of patronage in certain illnesses
8. Glossary of obscure terms	8. Table by order of patronage of various trades
	9. Proper names of persons and names of places
	10. Texts of Scripture
	11. Index for controversy
	12. Index for catechisms
	13. Alphabetical index of subjects and words

Table 1: An overview of the “*Fasti Sanctorum quorum vitae in belgicis Bibliothecis manuscriptae*”

(Delehay, 1920: 11 – 12)

Rosweyde’s aim was to restore the texts to their integrity and for this reason all controversial points that could not be verified were removed. The first edition came in 1616 under the title *Vitae Patrum* and contained Lives of the Desert Fathers. The popularity of these Lives had led to widespread reproduction of manuscripts and later printed copies, resulting in inaccuracy and confusion. Rosweyde managed to compare them, classify them, and finally create a volume suitable for scientific use. The *Vitae Patrum* consisted of a total of 10 chapters and an appendix:

Vitae Patrum
1. Collection of saints’ Lives (<i>Vitae virorum</i> and <i>Vitae mulierum</i>)
2. <i>Historia monachorum</i> translated by Rufinus of Aquileia
3. <i>Verba seniorum</i> translated by Rufinus of Aquileia

4. A collection of saints' Lives, drawn from the writings of Sulpicius Severus and John Cassian
5. <i>Verba seniorum</i> translated by Pelagius
6. <i>Verba seniorum</i> translated by John the Subdeacon
7. <i>Verba seniorum</i> translated by the deacon Paschasius of Rome
8. <i>Paradise of Heraclides</i> , which is actually the <i>Lausiac History</i> by Palladius of Galatia
9. <i>De Vitis Patrum</i> by Theodoret of Cyrus
10. <i>Pratum spirituale</i> by John Moschus

Table 2: An overview of the “*Vita Patrum*” (Delehaye, 1920: 11 – 12)

Rosweyde added prologues at the beginning, with topics about the authors, translators, original language, etc. He also provided an introduction to each chapter, while, at the end of the volume, there is a glossary of rare words, a table of contents, tables of names and locations, and comments.

On October 5, 1629, Rosweyde died, leaving behind a considerable amount of work, but in the state of raw material, while just a year later, the Belgian Jesuit Jean Bolland (1596 – 1665) was called to Antwerp to examine it. According to the Jesuit Hippolyte Delehaye, Heribert Rosweyde’s *Vitae Patrum* “was the cornerstone of the *Acta Sanctorum* (*Deeds of the Saints*)”. (Delehaye, 1920: 16)

Jean Bolland was born in 1596 in the village of Julémont, southeast of the city of Liège. In 1612, he became a member of the Society of Jesus and in 1620 he was sent for theological studies at the University of Leuven. Earlier, he had been a professor of humanities in various cities (Brussels, Antwerp, Roermond, Den Bosch). He was ordained in 1625. In 1630, he was invited to Antwerp by the superior of the Belgian province of the Society of Jesus to scrutinize the manuscripts left by Rosweyde, which he accepted with great joy. Bolland found the work sketchy, and the plan flawed but

nevertheless agreed to complete it on two basic conditions: first, he asked for freedom of movement in order to modify Rosweyde's plan as he saw fit, and second, he asked for a place of his own where he could work collecting and studying the available manuscripts (De Smedt, 1907). Until then, the library of the Professed House³ in Antwerp was the place of collection and study of the manuscripts.

Bolland's terms were accepted, and he began his work, leaving his position at the college of Mechlin. Bolland's new modifications could be divided into two categories: Firstly, he extended the provenance of the manuscripts geographically. Unlike Rosweyde, who had collected only the manuscripts of the Belgian region, Bolland approached his colleagues in various countries of Europe to send him relevant copies. Secondly, Bolland decided to broaden the genre of his sources. Even if there were no Lives or Passions of saints available, their veneration was already established, and the historical sources should be cited. Bolland's aim was to present all the available evidence for each saint, noting at the same time, the type of each source from which the information was drawn. Furthermore, he rejected Rosweyde's distinction between texts and elucidations for each saint. Bolland's plan provided a folder for each saint, which would be a complete dossier of the available hagiographical texts and would contain an introduction, summaries and appropriate commentary followed by the deeds of the saint. Finally, appropriate tables would be appended to each volume.

The main concerns that Bolland faced in his work were first of all, whether the events mentioned in the sources actually took place and secondly, whether the events and visions were divinely inspired or demonic in origin (*discretio spirituum*). (Machielsen, 2012: 125) In this way, he managed to create a historical core for the saint that could not be disputed and that could strengthen the saint's authenticity. The great difference with Rosweyde's *Vitae Patrum* was that the latter "did not question the truth of the original sources; only their interpretation as concern for truth had been transplanted by a concern for orthodoxy". (Machielsen, 2012: 124)

³Houses in which Jesuit priests lived. As they had no fixed income, the houses were supported by the generosity of benefactors.

After 5 years, Bolland admitted that the work was beyond his individual powers due to the increased material that had to be processed and so after receiving financial support from the abbot of Liessies, Antoine de Wynghe (1562 – 1637), he hired Godefrid Henschen (1600 – 1682) as his assistant. Henschen had studied humanities at the Jesuit College of Bois-le-Duc and was a professor at various colleges in Belgium. He was a member of the Society of Jesus as early as 1619, while in 1636 he was ordained a priest in Antwerp.

Henschen's help proved invaluable. When he arrived in Antwerp, he undertook the editing of the manuscripts for the first February volume as Bolland had already prepared the January volumes and was preparing to print them. Henschen's knowledge and zeal resulted in a new, more detailed method of annotation, which led Bolland to cancel the planned printing in order to make the appropriate additions. Bolland's preliminary comments of the January saints were essentially limited to the manuscript (where the texts were found), to a commentary section and a list of variants while Henschen's comments attempted to solve any problems, regarding the matter of chronological, geographical, historical, and philological interpretation (De Smedt, 1907).

Finally, in 1643 the first two volumes of the *Acta Sanctorum* for the month of January were printed, causing excitement both in the scientific world and in the Catholic Church as it was the first time that a systematic attempt was made to classify and critically publish such a large number of hagiographical texts. The same enthusiasm prevailed in 1658 when the three volumes of February were exhibited, which showed a marked improvement in commentaries over the January volumes. In 1659, Bolland and Henschen acquired a new partner, Daniel Papebroch. Papebroch (1628 – 1714), already having Bolland as a spiritual guide from a young age, focused on learning foreign languages and literary composition and additionally studied philosophy at Douai. In 1658, he was ordained a Catholic priest, having already been a member of the Society of Jesus in Belgium for years.

Correspondence between Bolland and his collaborators with libraries across Europe was an important part of their work as it enriched their collection of manuscripts. The success of the January and

February volumes made it possible to further expand the sources for the following months as Pope Alexander VII invited Bolland to Rome in 1660 not only to congratulate him but also to give him access to manuscripts in the Vatican Library. Bolland was unable to travel due to his advanced age but nevertheless sent Henschen and Papebroch in his place. Their staying in Rome lasted about a year, during which they had the opportunity to copy a large number of manuscripts. Both on their journey to Rome and on their return, Henschen and Papebroch received warm hospitality from their colleagues in various cities because of the European fame the *Acta Sanctorum* had acquired. This gave them the opportunity to acquire additional copies. In summary, they visited a large number of cities in Germany (Cologne, Frankfurt, Nuremberg, Munich), Italy (Florence, Bologna, Venice) and France (Grenoble, Lyon, Cluny) and collected fourteen hundred manuscripts as well as a large number of notes.

On their return to Antwerp in 1662, they began the preparation of the three March volumes, which were published in 1668. In these volumes only the names of Henschen and Papebroch appeared as Bolland had already died three years earlier. Bolland's colleagues as well as his successors were given the honorary nickname Bollandists and their group was called the Society of the Bollandists (French: *Société des Bollandistes*), which exists to this day. In 1675, the 3 April volumes were published, followed by the May volumes: I, II, III in 1680, IV, V in 1685 and VI, VII in 1688. At the same time, P. Jean Ravesteyn (from 1670 to 1675) and P. Daniel Cardon (from 1675 to 1678) became members of the Society of the Bollandists. In 1679, Conrad Jannick was hired as an assistant. He was extremely studious, and, two years later, after the intervention of Papebroch, he was sent to Rome to complete his studies. P. Baertius d'Ypres was his replacement and contributed decisively to the publication of volumes I and II of June (1695 and 1698 respectively). However, he was replaced in 1698 by P. François Verhoeven, who died just two years later. The death of Henschen in 1682 and of P. François Verhoeven in 1701 in conjunction with the blindness of Papebroch contributed to the escalation of the crisis of the Society of the Bollandists, which had begun a few years earlier.

The edition of Volume III in March 1668 and Volume I in April 1675 contained the *Acta* of St. Berthold and St. Albert respectively. On the one hand, Saint Berthold practiced asceticism on Mount Carmel in Israel and a group of ascetics gathered around him from which it is believed that the Carmelite Order arose. On the other hand, Saint Albert, Patriarch of Jerusalem, contributed the Carmelite Rule to the newly founded Carmelite Order. According to the tradition, the origin of the Carmelite Order goes back to the Prophet Elias, who is considered to be its founder. Papebroch had presented in a preliminary comment these theories as insufficiently substantiated. In addition, in volume II of April he included a section entitled *De praetensa quorundam Carmeliticorum conventuum antiquitate*. (Badea, 2012: 384) These actions resulted in the wrath of the Carmelites. (De Smedt, 1907)

In the following years, a large number of pamphlets were circulated, including abusive language against Papebroch, while the Carmelites tried to take the controversy to Rome in order to have the particular volumes condemned. Their effort was not made possible as Papebroch refuted the accusations point by point with arguments. The Carmelites' next move was to appeal to the court of the Spanish Inquisition, which in turn issued a decree in 1695 condemning the *Acta Sanctorum* in its entirety. (Badea, 2012: 386) The work was considered to contain erroneous and heretical propositions on matters of faith, which degraded saints and offended several popes. Papebroch's efforts to secure the revocation of the decree proved fruitless despite his assurances that he was prepared to correct or even retract the disputed points. For this purpose, Jannick was sent to Rome from 1697 to 1700 where he managed to obtain assurances that the volumes of the *Acta Sanctorum* would not be condemned. Furthermore, the Bollandists accepted the decision of the Roman Congregation of 22 December 1700, to place "on the Index of chronological and historical Essay on the Popes, published in the Propylaeum Maii, a decree issued, on account of the sections bearing on certain conclaves and requiring merely the correction of the passages in question". (De Smedt, 1907) The decree of the Spanish Inquisition was not fully revoked until 1715, a year after Papebroch's death.

During the 18th century other members were added to the Society of the Bollandists and the publication of the volumes progressed further. In 1702, Johannes Baptista Sollerius (1669 – 1740) undertook the revision and finalization of the essay on the Patriarchs of Alexandria. Other members contributing their work to the publication of the volumes were Johannes Pinus (? – 1749), Pierre Van den Bossche (? – 1736), Urbain Stickerus (? – 1753), Johannes Stilting (1703 – 1762), Constantin Suyskens (1714 – 1771), Jacques De Bue (1728 – 1808), Cornelius de Bye (1727 – 1801) and others. By 1773, seven volumes of June, seven of July, six of August, eight of September and the first three volumes of October had been published.

On July 21, 1773, Pope Clement XIV issued in Rome the *Dominus ac Redemptor Noster*, a decree which, among other things, provided for the cessation of the activity of the Society of Jesus and, in effect, its abolition. Already in the middle of the 18th century, suppressions had begun in the territories of Portugal, France, and Spain. The reasons that led to these suppressions could be briefly attributed to the ever-increasing financial power of the Society of Jesus on the one hand and on the other hand to the effort to centralize and secularize power from the point of view of monarchies as the autonomy of the Jesuits within the monarchies and the proximity to the pope were an obstacle to their plans. Only two months after the papal decree, in September 1773, the Bollandists were expelled from the Professed House of Antwerp, which was sealed. Nevertheless, they were allowed to settle in Brussels, where they remained until 1789. There Constantine Suyskens, Cornelius De Bye, John De Bue, Joseph Ghesquière (1731 – 1802) and Ignatius Hubens (1737 – 1782) managed to publish volumes IV and V of October. In 1789, the Imperial Government of Vienna came to an agreement with the abbey of the Premonstratensians of Tongerlo and the library of the Bollandists was moved there.

Despite the difficulties caused by the Brabantine Revolution (1789–1790), the Bollandists managed to publish Volume VI of October in 1794. The invasion of Belgium by French troops in the same year effectively suppressed the work of the Bollandists and dissolved their Society. The extant manuscripts were either hidden in peasant houses or transported to Westphalia. Many of them were destroyed

while a large part of those that remained were put up for sale and ended up mainly in the libraries of The Hague, Brussels and Burgundy. After many attempts from both France and Belgium, the Society of Bollandists was only re-established in 1837 in Brussels. This was followed by the publication of volumes VII to XIII of October (1845 – 1875) and then the Propylaeum and 4 volumes of November between 1887 and 1931. The last publication was made in 1940 and was the Propylaeum of December. The work remains still incomplete as the last published feast day of the year is November 10th.

It is worth noting that the Bollandists at the end of the 19th century and the beginning of the 20th century also edited three auxiliary catalogues, which contain hagiographical texts (Vitae, Passiones, Miracula etc.) sorted alphabetically by saint:

- i. *Biblioteca Hagiographica Latina (BHL)*: It contains hagiographical texts in Latin.
- ii. *Biblioteca Hagiographica Graeca (BHG)*: It contains Greek hagiographical texts in Greek.
- iii. *Biblioteca Hagiographica Orientalis (BHO)*: It contains hagiographical texts in Arabic, Armenian, Coptic, Ethiopian and Syrian.

Snijders divides this classification of these catalogues into 3 categories citing the example of Saints Eucharis, Valerius and Maternus:

- i. Dossier: It includes all available hagiographical texts written about a saint or group of saints.
- ii. Distinction between works for each saint. The works are classified, numbered, and an introduction and epilogue are also added. The enumeration bears the initials of the corresponding bookshelf (e.g. BHL) and a number.
- iii. Scripta “that can be said to ‘contain’, ‘express’, or ‘perform’ these works”. (Snijders, 2012: 7)

Dossier	EUCHARIUS, VALERIUS EN MATERNUS	
Work	BHL 2655 (Vita)	BHL 2655 (Vita) & BHL 2656 (Additamentum)
Scripta	1. Douai BM 838 2. Brussels KBR 9289 3. Namur MA 15	1. Brussels KBR II 976 2. Douai BM 842 3. Douai BM 865 4. Valenciennes BM 514

Figure 1: The extant manuscripts of saints Eucharis, Valerius and Maternus (Snijders, 2012: 8)

In summary, *Acta Sanctorum* was the first large-scale effort to classify and critically publish hagiographical texts from the Middle Ages and the Renaissance. To understand the plan of Bollandus, one must have before his/her eyes the proper object, which is the written monuments of the history and worship of the saints. In the early Middle Ages, the veneration of the saints was clearly limited to the church of their origin, where one could find the necessary evidence to create a dossier for each. Soon, there was a tendency to cross these narrow borders. By the relics, which in a way multiplied the sepulchral abode of the saints, and by the written records which carried the celebrity of the saints far and wide, conditions were modified. (Delehay, 1920: 55) This resulted in each church having, in theory at least, documents for its specific saints, as well as for the relics of those it borrows from other churches.

A multitude of reasons, as described in the previous two chapters, led to the mass production of hagiographical texts which were scattered throughout Europe. In order to have a critical edition, first of all, as many texts as possible had to be collected. Extensive correspondence with libraries and monasteries, as well as travels and visits for the purpose of copying manuscripts, occupied much of the Bollandists' time. Of course, it was not possible to acquire them all, let alone include them all in the published volumes. It is important to note that the goal of the Bollandists was not to reconstruct the archetype text but to classify and critically edit it. Thus, the result led to a single text but included the variations in the form of references that were not rejected during the critical process. Nevertheless, the publication of the text was only one part of the whole undertaking. The other part was the classification, the interpretation, and the commentaries. It included interpretations related to "literary history such as origin, author, period, tradition of the document, genre to which it belongs and then, the interpretation of obscure passages, the clarification of linguistic, geographical, and historical ambiguities". (Delehay, 1920: 99)

In conclusion, *Acta Sanctorum* was established as a project in a time of transition from the late Renaissance to the Baroque. The new conditions that arose forced a treatment of the sources in a way

different from the previous ones. On the one hand, the simple quotation of the texts and the restoration of the archetype pave the way to the critical approach and interpretation. It is clear that Bolland tried to separate fact from myth. On the other hand, the clear belligerent tones of earlier works against Protestantism smooth the way to a milder treatment. Bolland responded through his work to the Protestants, who mocked the worship of saints, with the aim of not defeating them but of converting them. (Machielsen, 2012: 128)

Chapter 5 – Data Analytics

5.1 Introduction

The massive digitization of archives and libraries in recent years, the further development of computer software and hardware as well as interdisciplinary study programs are just some of the factors that have led to an increase of the use of computational methods and to a further development of Digital Humanities. As David Berry notes “the key to understanding Digital Humanities is to reject the idea that digital technology is invading the academy, as computers were used for humanist ends from very early on in their history, and not only, as one might expect, as mere storage for large libraries of text”. (Berry, 2019) Transcribing manuscripts for digital scholarly editions, coding data for qualitative and distant reading methods, data visualization, analysis of large archives, creating databases and applications, 3D virtual reconstruction etc. are only a small sample of the techniques and methods used in the Digital Humanities.

A Digital Humanities project example is *The French Book Trade in Enlightenment Europe (FBTEE)* project. It is a “project of international significance mapping the production, marketing, dissemination, policing, and reception of books (and hence ideas) in the late eighteenth century that aims to bring together and make interoperable and publicly available in a single digital resource multiple historical bibliometric database”. (Burrows et al., 2014) Essentially it is a tracing of the activities of the Société Typographique de Neuchâtel (STN), a publishing house in Switzerland that operated between 1769 and 1794. The database provides information on the time of transaction, the author, the title, and the type of books as well as the customer details, such as name, address, occupation, etc. The users of the database are given the possibility of a specialized search through a multitude of series of option menus and filters. Finally, the authors of the database provide a number of other supporting materials for further exploration of the data, such as videos, maps, and footnotes.

The present chapter of the master’s thesis is an attempt to explore the data obtained from the *Acta Sanctorum* database using a technique called exploratory data analysis. Exploratory data analysis (EDA)

is an approach for data analysis, and its formal methodology was introduced by the mathematician and statistician John Turkey (1915 – 2000) in 1977 through his work entitled *Exploratory Data Analysis*. This approach is a fundamental stage after the collection and pre-processing of the data and aims to help in the exploration and evaluation of their quality. It provides tools for generating hypotheses by visualizing and understanding data usually through graphical representations. Usually, these graphical representations used in the exploratory data analysis are simple such as plotting the raw data (histograms, scatter plots), simple statistics (box plots, mean plots) and “positioning such plots so as to maximize our pattern recognition abilities”. (Natrella, 2010) It is a useful way to approach data sets without assumptions as it employs a “variety of techniques that maximize insight into a data set, uncover underlying structure, extract important variables, detect outliers and anomalies, test underlying assumptions, develop parsimonious models and determine optimal factor settings”. (Natrella, 2010) Many times, exploratory data analysis is confused with statistical graphics, which is a collection of techniques used in the field of statistics for data visualization. In contrast, exploratory data analysis is an approach to data analytics that allows “the data itself to reveal its underlying structure” (Natrella, 2010) and pattern as well as a guideline as to how an analysis is performed on a data set. As for the visualization techniques used, they largely depend on the type of data. For example, histogram, scatterplots, and box-and-whisker plots are used for outlier detection while descriptive statistics are used for categorical data. (Komorowski et al., 2016: 187)

Exploratory data analysis can be used in various domains and sciences, such as statistics, medicine, geography, business intelligence etc. In recent years the increasing use of quantitative methods resulted in similar processes in the humanities, such as history, archaeology, and literature studies. Karsdorp et al. in their book *Humanities Data Analysis: Case Studies with Python* attempted an exploratory data analysis using data from US cookbooks. Specifically, the data came from seventy-six cookbooks covering a chronological range between the end of the eighteenth century and the beginning of the twentieth century. (Karsdorp et al., 2021) These books are just a small sample of a total of seven thousand cookbooks that were published in that time period. The dataset is called *The*

Feeding America: The Historic American Cookbook dataset and includes 48.032 recipes encoded in the xml format, which contain comments on recipe type, ingredients, measurements, cooking tools etc.

The two main research questions were the following:

1. “Which ingredients have fallen out of fashion, and which have become popular in the nineteenth century?

2. Can we observe the influence of immigration waves in the Feeding America cookbook collection?”

(Karsdorp et al., 2021)

An interesting observation about the first question concerns the most favorite category of recipes, which corresponds to bread and sweets. The use of yeast began to decline from the early nineteenth century and was later replaced by baking powder in American kitchens as it proved to be better and faster acting. Another trend discovered through exploratory data analysis was the ups and downs of tomato use. As it was considered poisonous it had little presence in recipes until the creation of a new breed and its transformation into a commercial crop at the end of the nineteenth century. Regarding the second question, exploratory data analysis showed that the introduction of specific ingredients coincided with the arrival of specific migration waves. For example, soybeans began to appear in the US cookbooks after the immigration wave from China, while ingredients such as garlic and olive oil began to appear after the immigration wave from the Mediterranean countries of Europe. (Karsdorp et al., 2021)

For the exploratory data analysis and the topic modelling in this master’s thesis, Python has been chosen. Python is a programming language, which belongs to the category of imperative programming languages and supports both procedural programming and object-oriented programming. (*General Python FAQ*, n.d.) It was created by Dutchman Guido van Rossum at the Centrum Wiskunde & Informatica research center in the Netherlands in 1989 and first released in 1991. Apart from programming, Python is also used in other fields such as data analysis, data science, machine learning etc. The high readability for new data sources, the simple syntax, the great open-source community

and of course the extensive range of libraries for several tasks are only some of the many advantages of Python in the field of data analytics and machine learning. For the data analytics methods described in this chapter the following packages have mainly been used:

- *Pandas* is “a package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive”. (*Pandas 0.25 Documentation*, 2019) *Pandas* can handle data stored in Excel and flat files such as csv, tsv, etc. and convert them to data frames. Merging, joining, reshaping, grouping, slicing and subsetting data as well as handling of missing data are some of *Pandas*’ main characteristics.
- *NumPy (Numerical Python)* is a Python package suitable for working with numerical data. “It contains multidimensional array and matrix data structures that can be used to perform a wide variety of mathematical operations”. (*NumPy: The Absolute Basics for Beginners*, n.d.)
- *Matplotlib* “is a library for creating static, animated, and interactive visualizations in Python”. (*Matplotlib: Visualization with Python*, n.d.)
- Other libraries that have been used are *os*, for data import, *string*, for common string operations and *re* for regular expression operations.

5.2 Database and Data

The data for this work were obtained from ProQuest through the Vienna University Library. ProQuest is a global information-content and technology company based in Michigan, USA. According to the company’s own website, “ProQuest is a key partner for content holders of all types, preserving and enabling access to their rich and varied information. Those partnerships have built a growing content collection that now encompasses 90,000 authoritative sources, 6 billion digital pages and spans six centuries. It includes the world’s largest collection of dissertations and theses; 20 million pages and three centuries of global, national, regional and specialty newspapers; more than 450,000 eBooks; rich aggregated collections of the world’s most important scholarly journals and periodicals; and unique vaults of digitized historical collections from great libraries and museums, as well as organizations as

varied as the Royal Archives, the Associated Press and the National Association for the Advancement of Colored People”. (*About – Who We Are*, n.d.)

The *Acta Sanctorum* Database “is an electronic version of the complete print edition published in sixty-eight volumes by the *Société des Bollandistes*. It is a collection of documents examining the lives of saints, organized according to each saint’s feast day. It runs from the two January volumes published in 1643 to the Propylaeum to December published in 1940. The entire *Acta Sanctorum* is available, including all prefatory material, original texts, critical apparatus, and indices along with *Bibliotheca Hagiographica Latina* reference numbers”. (*Acta Sanctorum*, n.d.) The data were obtained in accordance with paragraph 1.6.a. of the website’s Terms and Conditions, as the main and only purpose is research and analysis of *Acta Sanctorum* in the context of writing this master’s thesis.

The criteria selected for the search in the database include all texts of the category “Latin Hagiographical Texts Pre-1500” and “Latin Hagiographical Texts Post-1500”. As the aim of this work is the study of the hagiographical texts, all the duplicate documents as well as calendars, martyrologies, introductions, dissertations, commentaries and miscellaneous were excluded (see Chapter 4). The search results lead to 4665 unique hagiographical texts, which have been downloaded and stored locally in .txt files only for research purposes. The data consists of forty-seven .txt files where each one contains one hundred texts while the last one contains only the remaining sixty-five of them. Their total size is 197 MB. The character encoding of the data is UTF-8, which is the default option provided by the database.

The structure of all the texts shows common patterns, which could be divided as follows:

- i. First part: This is an introduction which contains metadata such as the title, the URLs linking the text to the database, the name of the saint, the feast day and the BHL number (*Bibliotheca Hagiographica Latina*).

- ii. Second part: In this part there is the main text which is divided into chapters. Each chapter has its own title and at the end there are footnotes with comments and metadata. The footnotes always begin with the words “Side Note” or “End Note”.
- iii. Third part: This is standard in every text and includes all the following metadata: saint, saint gender, feast date, religious work type, title, BHL reference, publication title, pages, publication year, publisher, place of publication, source type, language of publication, document type, ProQuest document ID, document URL, last updated and the name of the database.

5.3 Exploratory Data Analysis

With the help of the *os* library, all the data were uploaded into the Jupyter Notebook⁴ in a list form. The next step was to turn them into a data frame, a two-dimensional table of rows and columns, in order to carry out the (pre)processing part. The library *pandas* was used for this purpose. To do this, the straight line separating each text from the next one in the txt files was used as a separator pattern. The generated data frame was divided into rows where each row included a unique text. Finally, all the empty rows were filtered out (figure 2).

	texts
0	
1	\n\nVita S. Fulgentii A Quodam Eius Discipulo ...
2	\n\nVita Ex Vitis PP. Collecta\n\nhttps://uacc...
3	\n\nDe Eodem Sancto Martyre Ex Ruffino, et MSS...
4	\n\nVita\n\nhttps://uaccess.univie.ac.at/login...
5	\n\nVita SS. Severini Et Victorini\n\nhttps://...
6	\n\nVita Auctore Anonymo\n\nhttps://uaccess.un...
7	\n\nVita S. Guilielmi Auctore Glabro Rodulpho\...

	texts
1	\n\nVita S. Fulgentii A Quodam Eius Discipulo ...
2	\n\nVita Ex Vitis PP. Collecta\n\nhttps://uacc...
3	\n\nDe Eodem Sancto Martyre Ex Ruffino, et MSS...
4	\n\nVita\n\nhttps://uaccess.univie.ac.at/login...
5	\n\nVita SS. Severini Et Victorini\n\nhttps://...
6	\n\nVita Auctore Anonymo\n\nhttps://uaccess.un...
7	\n\nVita S. Guilielmi Auctore Glabro Rodulpho\...
8	\n\nElevatio S. Odilonis\n\nhttps://uaccess.un...

Figure 2: The texts column of the data frame after removing the empty lines

⁴ “Jupyter Notebooks are an open document format that contain a complete record of the user's sessions and include code, narrative text, equations, and rich output”. (*Project Jupyter*, n.d.)

The metadata was then used to extract information about each one of the texts. With the help of the function `str.extract()`, five new columns were added to the data frame:

1. Saint
2. Gender
3. Place of Publication
4. Publication Year
5. Type (it includes the different types of texts such as *Vita*, *Translatio* etc.)

	texts	Saint	Gender	Place of publication	Publication Year	type
1	BHL Number: 3208 Per eivs Discip. PROLOGVS AV...	Fulgentius, Episcopus Ruspensis in Africa (S.)	Male	Antwerp	1643	Vita S. Fulgentii
2	BHL Number: 5096 Ex vitis PP. CAPVT I. S. Mac...	Macarius Alexandrinus, Abbas in Aegypto (S.)	Male	Antwerp	1643	Vita Ex Vitis
3	BHL Number: 5007 Ex variis. [1] [S. Lucianus...	Lucianus, Presbyter Antiochenus Martyr (S.)	Male	Antwerp	1643	De Eodem Sancto
4	BHL Number: 5905 Ex Ioan. Capgr. [1] Dvm in ...	Melorus, Martyr in Anglia (S.)	Male	Antwerp	1643	Vita https://uaccess.univie.ac.at/login?url=ht...
5	BHL Number: 7659 Avct. Anonymo. ex mss. [1] ...	Seuerinus, Episcopus Septempedanus in Italia ...	Male	Antwerp	1643	Vita SS. Severini

Figure 3: The newly generated columns of the data frame

The next step was to clean all metadata from the original text column. Thus, all URLs were replaced with a single whitespace and the above-mentioned parts *i* and *iii* were deleted completely. Deleting the metadata of the part *ii* was challenging and for this purpose, the upcoming pattern was followed. The words “Side Note” and “End Note” were set as the beginning of the pattern as they are the beginning of all footnotes. Considering that the footnotes are always followed by the title of the next chapter, which is always enclosed in square brackets, the “[” symbol was defined as the end of the pattern. In this way all metadata and footnotes were deleted and only the original hagiographical text of the manuscripts remained in the data frame. Given that the footnotes only contain critical notes and comments while their deletion doesn’t affect the results. The few remaining metadata words were added to the stopwords list and later removed from the texts (*Side*, *End*, *Note*, *BHL*, *Number*). Nevertheless, within the texts of the data frame there were also some other elements that had to be

removed before going further with the analysis. Such elements were all Arabic numerals from 0 to 9 as well as all parentheses and square brackets together with their contents. This had the effect of deleting not only the titles of each chapter but also all the explanations and comments that were between these symbols and were irrelevant for this work. In addition, all the punctuation was removed while 48 letters and combinations of letters were replaced with simpler forms in order to achieve uniformity in the texts. (see figure 4) The remaining special characters, such as Greek letters, and symbols (cross, percentage symbol, etc.) were added to a list called *special_characters_new*. Finally, all the big whitespaces (everything bigger than one) were replaced with a single one. The last step before the exploratory data analysis was to find all non-Latin words and store them in a list called *non_latin_words*. The total number of the list elements was 157.858 and almost all of them were Greek words.

```
#Replace letters with their diphthong
df_filtered["texts"] = df_filtered['texts'].str.replace('æ', 'ae')
df_filtered["texts"] = df_filtered['texts'].str.replace('Æ', 'AE')
df_filtered["texts"] = df_filtered['texts'].str.replace('œ', 'oe')
df_filtered["texts"] = df_filtered['texts'].str.replace('Œ', 'OE')
df_filtered["texts"] = df_filtered['texts'].str.replace('ā', 'a')
df_filtered["texts"] = df_filtered['texts'].str.replace('ē', 'e')
df_filtered["texts"] = df_filtered['texts'].str.replace('ō', 'oe')
df_filtered["texts"] = df_filtered['texts'].str.replace('ó', 'o')
df_filtered["texts"] = df_filtered['texts'].str.replace('ū', 'u')
df_filtered["texts"] = df_filtered['texts'].str.replace('ī', 'i')
df_filtered["texts"] = df_filtered['texts'].str.replace('ü', 'u')
df_filtered["texts"] = df_filtered['texts'].str.replace('ñ', 'n')
df_filtered["texts"] = df_filtered['texts'].str.replace('ß', 'ss')
```

Figure 4: An overview of the letters that have been replaced

Starting from the analysis of new generated columns, the results (figure 5) showed that the available texts were published in Antwerp (3856), Brussels (578), Paris (194) and Tongerlo (34).

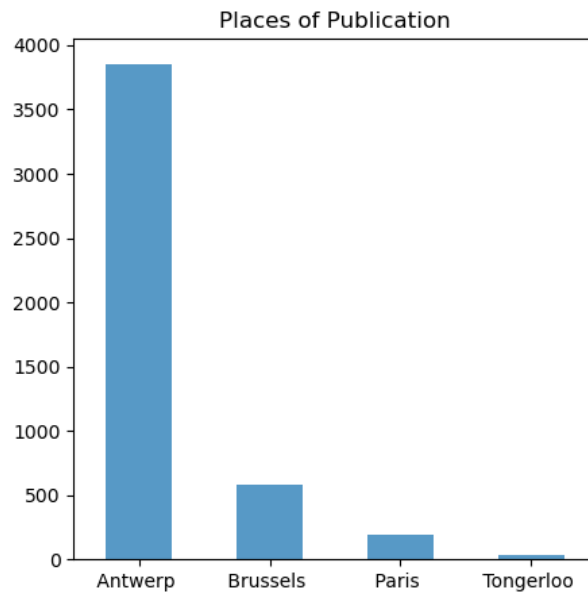


Figure 5: *The publication places of the volumes*

Regarding gender, most hagiographical texts contained only male saints (3.700) compared to those containing only female saints (724), while 225 texts contained both male and female saints (figure 6).

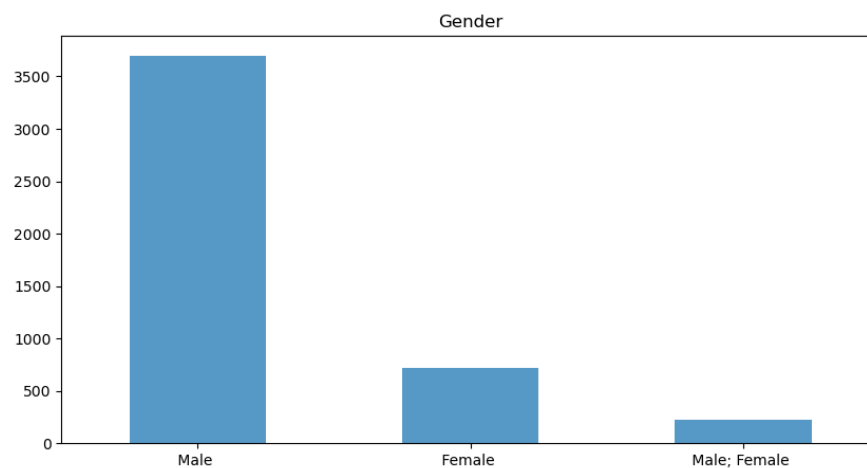


Figure 6: *The gender of the saints*

The publication dates of the texts were merged and visualized by century as shown in figure 7. Most of them were published during the 17th century and followed a downward trend in publications (see chapter 4).

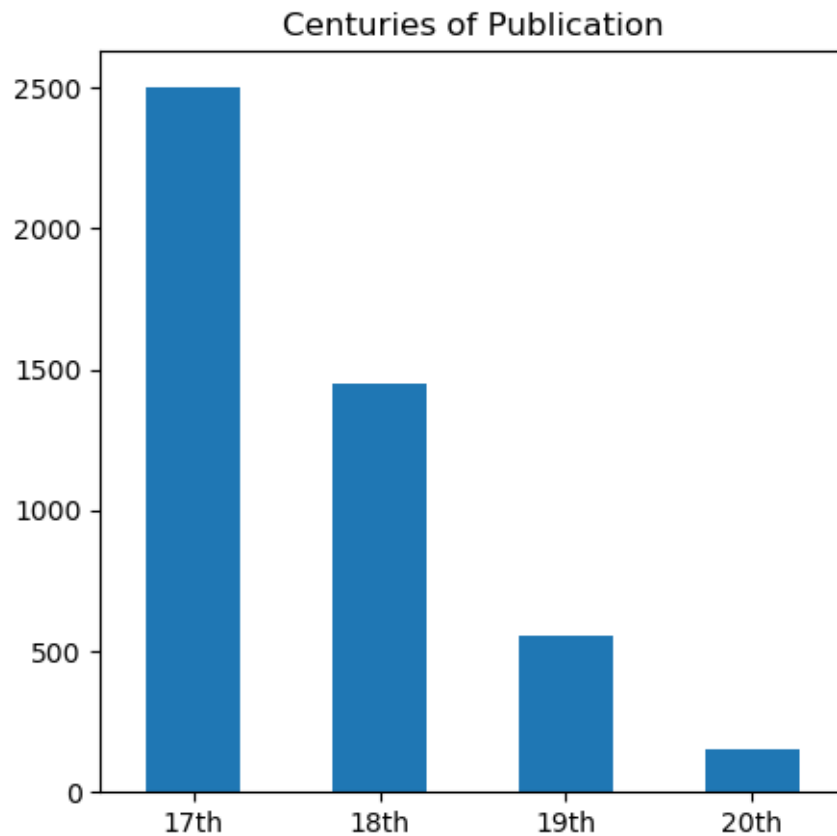


Figure 7: The centuries of publication of the texts

The next step was the analysis of hagiographical texts that appear more than once in the data frame, but that are not duplicates (they were excluded during the database search). These texts are mainly different variants of *Vitae* and *Passiones* as well as multiple collections of *Miracula*. Generally, 2061 of the entries in column “Saint” are duplicates. Most hagiographical texts with different variants concern local bishops and martyrs of Late Antiquity and the Early Middle Ages, who came from all over Europe and the Near East (Asia Minor, Egypt, etc.). On average there are 3–5 different texts for each of these saints. Figure 8 shows the seven saints for whom there are more than ten different hagiographical texts in the data frame. In the first places are two saints from the region of Belgium, St. Amandus

(c.584 – 679) and St. Vedestus (c.453 – c.540). Both were bishops during the early Middle Ages with intense missionary activity and organization of churches. Many miracles were attributed to both of them, and their cult spread not only in Belgium and the wider region but also in England. The area of their activity and their later cult could justify the existence of a large number of hagiographical texts as it is the same as the seat of the Bollandists which gave them greater and easier access to manuscripts. In third place is St. Thomas Aquinas (1225 – 1274). He belonged to the Dominican Order and is considered one of the most important philosophers and theologians of the Catholic Church. Two of his works were *Summa Theologica* and *Summa Contra Gentiles*. Thomas Aquinas breathed his last at the Cistercian abbey of Fossanova and his funeral was organized by the Cistercians, who claimed ownership of his body from the Dominicans. This resulted in his body being moved several times over the next few years as well as the removal of his head and right hand. The canonization process of Thomas Aquinas began in 1303 and was completed twenty years later, in 1323. During this time, a profile of the saint was drawn up and research was carried out with the aim of examining all the testimonies of miracles, which in the meantime were increasing. Various versions of his *Vita* and his *Miracula* are included in the *Acta Sanctorum*. It was the Bollandists who noticed that the hand of Thomas Aquinas still appeared unchanged in “color, intact, with skin, flesh and nails and moreover gave off a very sweet smell”. (Riccardi, 2022) In fourth place is St. Magnus Erlendsson (1080 – 1117). He was earl of Orkney of Scotland and was noted for his religious zeal. Magnus and his cousin Haakon jointly ruled the county from 1105 to 1114 but later came into conflict. The climax of this conflict led to Magnus being executed by Haakon on the island of Egilsay. Narratives of Magnus’ martyrdom spread rapidly, and this led to the production of *Sagas*, *Vitae*, and *Hymns*. Research showed that the Latin Magnus *Hymns* emphasize the laid on Magnus’ ascetically pure marital life and they were composed for singing by monastic communities, at canonical services etc. (Beveridge, 1939: 360) The Bollandists included in the *Acta Sanctorum* mainly *Hymns* as well as *Vitae*. The next three saints (Berignus, Theodorus and Georgius) came from the Near East and all of them were martyred in Late Antiquity. Berignus (third century) was born in Smyrna in Asia Minor but went to France to teach Christianity. He

was martyred in Dijon, of which he is the patron saint. Theodorus (281 – 319) and Georgius (280 – 303) were also born in Asia Minor where both followed military careers. They did not hesitate to confess their Christian faith in front of the local rulers which led to their martyrdom. Their worship spread quickly and was widespread throughout the Christian world, especially in the Byzantine Empire.

The 7 most common texts are:

Amandus, Ep. Traiectensis, Elnone in Belgio (S.)	15
Vedastus, Ep. Atrebatensis in Belgio (S.)	14
Thomas Aquinas, Doctor Angelicus Ordinis Prædicatorum (S.)	12
Magnus, Comes insularum Orcadum in Scotia (S.)	12
Benignus, martyr Divione in Gallia (S.)	12
Theodorus martyr Euchaïtis (S.)	11
Georgius Megalo-Martyr Nicomediensis, Lyddæ seu Diospoli in Palæstina (S.)	11

Figure 8: The saints that have been most often object of hagiographical writing.

Most of the hagiographical texts in the data frame basically belong to the following types: *Vitæ* (1983), *Acta* (713), *Miracula* (337), *Translatio* (207) and *Passio* (184). Other types, which are not included in figure 9, are rarer such as *Encomium*, *Epistola*, *De Eodem*, *Honor*, *Gloria*, etc or texts just talking about “the same” saint (*De Eodem*), her/his honour (*Honor*) or her/his glory (*Gloria*). A very small number of hagiographical texts do not include any reference to type. As mentioned in the first chapter, the distinction between the hagiographical text types is not always clear and most of the time, there is a combination of two or more types under an umbrella type such as *Vita* and *Passio*.

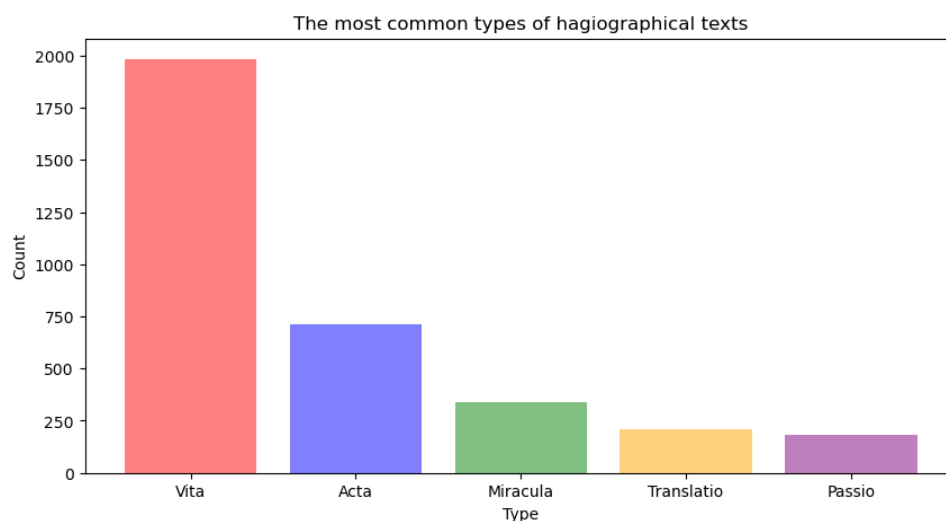


Figure 9: The most common types of hagiographical texts

The next part of the exploratory data analysis is divided into two sections and includes a statistical analysis of the texts with stopwords and one without them. First, the punctuation has been completely removed from the texts and all the words have been lowercased; neither intervention seriously concerns distinctive features of early modern Latin. Then a new column was created with all the tokenized texts called *texts_with_stopwords*.

To analyze the texts without stopwords the same procedure was followed with an additional step, the removal of a list of words. The list contained 4.001 elements of which only 255 were symbols and, Latin and Arabic numerals while all the rest were Latin words and abbreviations; the list, named *stopwords_latin.txt*, was downloaded from the GitHub repository of A. Berra (Berra, n.d.). The new column is named *texts_without_stopwords*.

	texts_with_stopwords	texts_without_stopwords
Maximum number of words in the dataset	170.018	91.395
Minimum number of words in the dataset	34	26
Average number of words in the dataset	4.696	3033
Total number of words in the dataset	21.907.227	14.148.693
Number of texts with more than 10.000 words	464	257
Number of texts with less than 10.000 words	4201	4408

Table 3: An overview of the statistical analysis

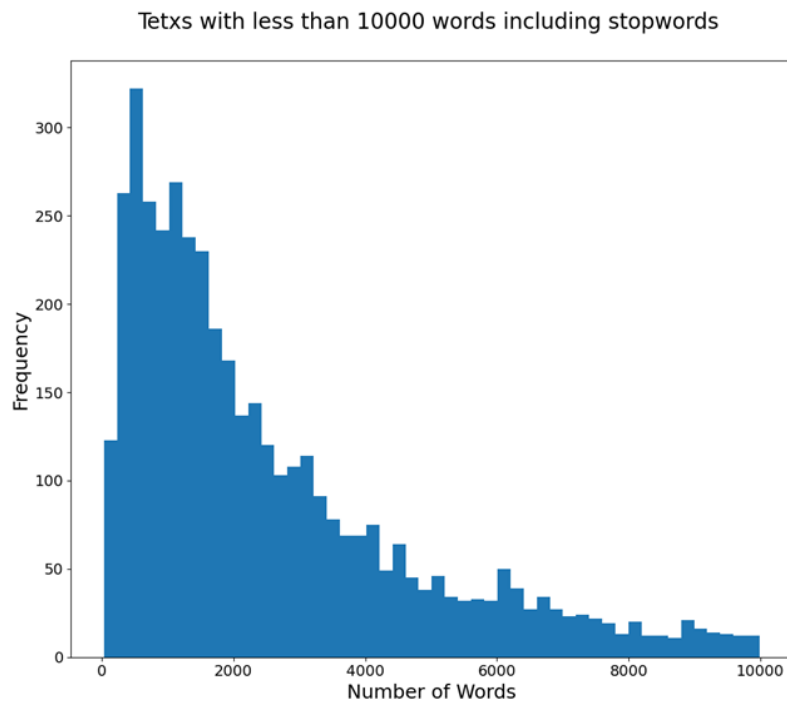


Figure 10: Number of texts with less than 10.000 words including stopwords

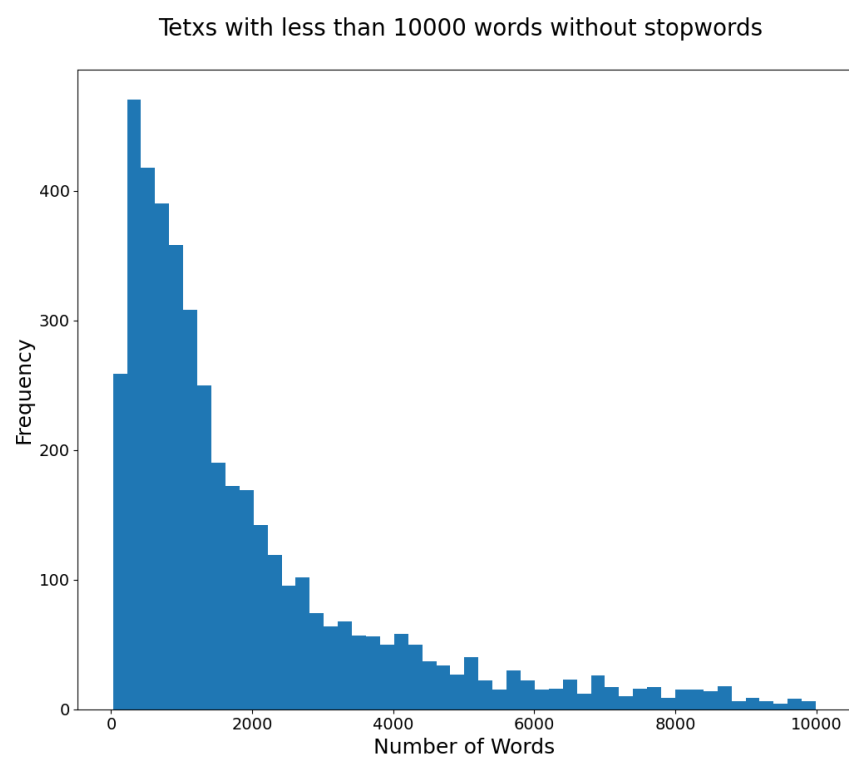


Figure 11: Number of texts with less than 10.000 words without stopwords

As can be seen in figures 12 and 13, the average number of letters in the hagiographical texts with stopwords is 6 - 6.25, while without stopwords it is 7.25 - 7.5.

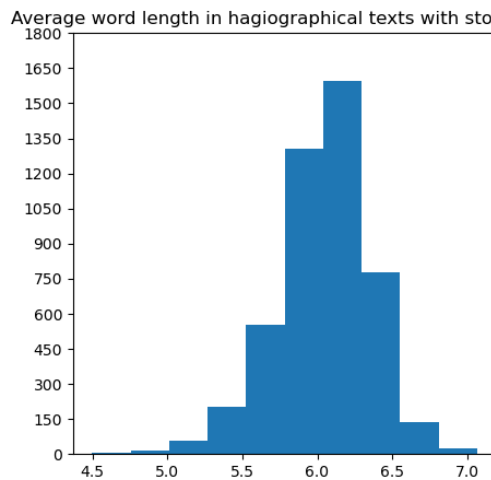


Figure 12: Average length of words including stopwords

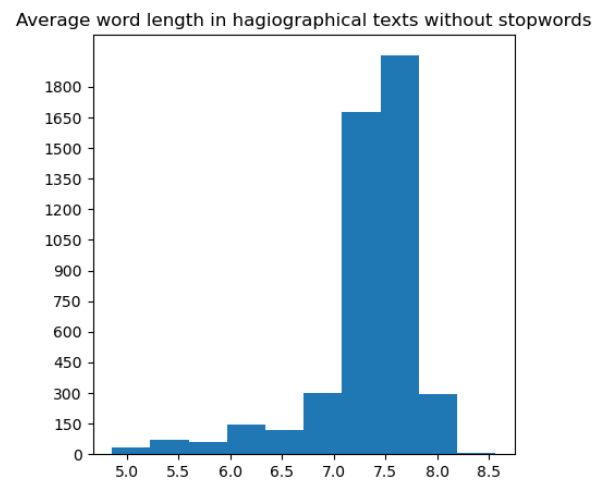


Figure 13: Average length of words without stopwords

The last part of the exploratory data analysis is about the Greek texts of the data frame. In total there are 352 texts which contain Greek words but only 121 of them are at least 50% in Greek. The highest percentage is 76% and it is a text entitled *Pelagia Virgo, Martyr Tarsi in Cilicia*, published in 1680 in Antwerp. More generally, these hagiographical texts appear either entirely in Greek with an introduction in Latin or contain the Greek version with the corresponding Latin translation. All of them concern saints of Late Antiquity and the Early Middle Ages and most of them were bishops, martyrs, hermits, and monks. The origin of these saints was from the Near East, Greece, Rome and Southern Italy, and their worship was widespread mainly in the Byzantine Empire. As mentioned in the first chapter, Greek was basically the language of Christianity during the first Christian centuries. The hagiographical texts about the monks, the Desert Fathers, the martyrs of the Roman persecutions and the bishops were mainly written in Greek, especially in the Eastern part of the Roman Empire, a fact that can also be proved from the Greek texts of the *Acta Sanctorum*. Indicatively, of the 121 texts that contain mostly Greek words, 75 of them are about martyrs, 17 of them about bishops and 13 of them

about monks. As for the texts of the data frame with a small percentage of Greek words, they only contain some Greek terms as explanations and comments in the mainly Latin corpus. One of the reasons that led to the use of individual Greek terms in Latin texts is probably due to the way of learning Greek in the Western part of the Roman Empire, as this way was based on lists of words organized alphabetically or by subject. This had the result that many Latin authors sprinkled their works with occasional Greek words relying on such lists rather than an active command of the Greek language. (Rapp, 2004: 1244 – 1245)

After the analysis of the Greek texts, all the symbols of the *special_characters_new* list created above as well as all Greek words were removed from the *texts_without_stopwords* column leaving a total of 13.132.114 Latin words to be used in the next part of the script about topic modelling.

Chapter 6 – Natural Language Processing

6.1 Introduction

Natural language processing (NLP) is “a branch of computer science and more specifically, a branch of artificial intelligence or AI”. (*What is Natural Language Processing?*, n.d.) Understanding and producing natural language, that is, trying to enable computers to understand human linguistic data and extract the full meaning from them, are the core challenges of NLP. (*What is Natural Language Processing?*, n.d.) NLP has been a research field since the 1950s. Characteristic is the paper by the English mathematician Alan Turing (1912 – 1954) entitled “Computing Machinery and Intelligence”, with which he introduced the imitation game (now known as the Turing Test), a method to determine if a machine can exhibit human-like intelligence. Also worth mentioning is the Georgetown-IBM experiment in 1954, a machine translation of about sixty sentences from Russian into English. Until the 1970s, most NLP systems were based on complex handwritten rules, while from the 1980s onwards, the increase in computing power brought machine learning algorithms to the fore for natural language processing. The focus was on statistical models, which over time became more and more complex. In addition, the introduction of unsupervised learning algorithms and the widespread use of the internet in turn brought a new revolution in NLP. Syntactic analysis (parsing, sentence breaking), lexical semantics (named entity recognition, sentiment analysis), machine translation etc. are some of the modern research areas of NLP. (McGillivray et al., 2020; Schofield, 2020; Johnson et al., 2021)

Regarding the use of NLP methods in the humanities, there is a particular interest in the development of new computational techniques for data modelling as the ever-increasing amount of digital data makes it impossible to process it with traditional methods and tools. Handwritten text recognition (HTR), historical NLP pipes, authorship attribution etc. are just a few examples. The collaboration of the humanities with other sciences, such as computer science, has made it possible to ask and investigate new questions that until now have been impossible to answer due to limitations in methodological tools. Nevertheless, the new methods also introduced new problems that were

previously unknown. For example, NLP methods can introduce bias into the research, which can in turn affect the performance of the model. (Fokkens et al., 2014; Henriot, 2020; Schofield, 2022) The complexity of data in the Humanities combined with the limitations of existing NLP methods, highlight a still existing issue: the lack of communication between the NLP and (Digital) Humanities communities. (McGillivray et al., 2020) Existing tools lack high-tech capacity for humanities data processing resulting in research limitations.

An approach that is part of NLP is topic modelling. Specifically, it is an unsupervised machine learning approach that is used to determine the topics of a set of documents based on their content. In other words, topic modelling is a type of statistical model, which not only creates so-called “topics” but also determines which topic each document belongs to as well as which words each topic consists of. That is, a model understands that a document refers to a specific topic as certain words appear more often than others. For example, the words “ball” and “field” appear frequently in documents about sports, while the words “bread” and “water” frequently in documents about food. Words like “and” and “is” appear in both topics. Usually, a document covers many topics in different proportions, so a document in which 10% of the topic is food and 90% of the topic is sports can be considered to have nine times as many words about sports. Topic modelling reflects this proportion of topics in a statistical structure that allows one, based on the study of a collection of documents and the study of the frequency characteristics of the words in each document, to conclude that each document contains a certain number of topics. (Boyd-Graber et al., 2017; Saxton, 2018; Zosa & Pivovarov 2022)

Marketing, human resources, operations, strategy, and finance are only some of the domains where topic modelling is used in order to analyze consumer profiling, buying patterns, purchase predictions, and discovering online communities and topics. Generally, the use of topic modelling in social sciences and humanities has the basic goal of a better understanding of certain phenomena through the text that people have written. However, research has shown that two main barriers arise: a) accessibility, e.g. technical barriers that make text processing and topic modelling difficult, and b) trust, e.g. the

exploration and validation of a model by researchers. (Ramage et al., 2009: 1) Regarding the trust in topic model output, the research mentions four main challenges – observations during the evaluation:

- i. “Characterizing topics is hard
- ii. Naming topics is hard
- iii. Topics mean different things in different contexts
- iv. Topic models must find what we know is there”. (Ramage et al., 2009: 2 – 4)

An evaluation method of a probabilistic topic model could be human judgment. Visualizations of the topics, e.g. wordclouds, could provide an overview of the most dominant words in each topic that could lead to their labeling based on human interpretation. Stanford University researchers introduced the *Termite* visualization which is described as a “term-topic distribution produced by topic models”. (Chuang et al., 2012: 74) The specificity of *Termite* visualization is found in the two calculations it introduces: the *seriation* method, which discloses clusters of terms and helps identification of coherence, and the *saliency* measure, which extracts words that are more meaningful and informative beyond just the frequency of their occurrence. (Chuang et al., 2012: 74 – 76) *Word intrusion* and *topic intrusion* are also two evaluation methods based on human judgment. In *word intrusion* the topics are presented with groups of six words, where five of them belong to a given topic and one is an intruder word. The poorer the coherence, the harder it is to recognize the intruder word. (Chang et al., 2009: 3) *Topic intrusion* is similar but instead of words, the title and a piece from a document are shown along with four topics. “Only three of those topics are the highest probability topics assigned to that document while the remaining intruder topic is chosen randomly from the other low – probability topics in the model”. (Chang et al., 2009: 4) Regarding the quantitative methods of topic modelling evaluation, most of the evaluations have “focused on statistical measures of perplexity or likelihood of test data”. (Newman et al., 2010: 215) Splitting a dataset into two parts, a training set and a test set, is the first step to calculate the perplexity. After it, the model is trained on the training set and then tested on the test set, which contains previously unseen documents. Nevertheless, “the perplexity measure does not reflect the semantic coherence of individual topics learned by a topic model”

(Newman et al., 2010: 215), while research has also shown that it often could contradict human judgement. (Chang et al., 2009: 3) On the other hand, the semantic similarity between high probability words inside a topic is measured by the coherence. The higher the coherence score, the more interpretable and coherent the topic is considered to be. The coherence pipeline is a four-stage process used to calculate the coherence of topics generated by topic modelling algorithms. It consists of four stages:

- i. “Segmentation: The corpus is segmented into smaller parts, such as sentences or paragraphs.
- ii. Probability Estimation: The probability of each word in the corpus is estimated using a statistical model.
- iii. Confirmation Measure: A measure of the degree of similarity between words in a topic is calculated using the estimated probabilities.
- iv. Aggregation: The individual topic coherence scores are aggregated to obtain an overall coherence score for the entire model”. (Röder et al., 2015: 402)

In Digital Humanities, topic modelling could be a kind of synecdoche of these stages as it is a form of distant reading, which focuses on the processing not of individual texts but of text corpora resulting in seductive yet obscure outputs in the form of easily interpreted “topics”. (Meeks & Weingarten, 2012) As Blei mentions “the goal is for scholars and scientists to creatively design models with an intuitive language of components, and then for computer programs to derive and execute the corresponding inference algorithms with real data”. (Blei, 2012) However, models are intended only to aid in the interpretation and understanding of texts while the actual interpretation and understanding is still the work of the scholar. (Blei, 2012)

One of the most popular methods for performing topic modelling, also in Digital Humanities, is *Latent Dirichlet Allocation* (LDA). (Meeks & Weingarten, 2012) Blei et al. introduced it in 2003 and described it as a “generative probabilistic model of a corpus where the basic idea is that the documents are represented as random mixtures over latent topics and where each topic is characterized by a

distribution over words". (Blei et al., 2003: 993) As shown in figure 14, LDA is a three-level hierarchical Bayesian model and the "parameters α and β are corpus – level parameters, the variables θ_d are document – level variables and the variables z_{dn} and w_{dn} are word – level variables". (Blei et al., 2003: 997)

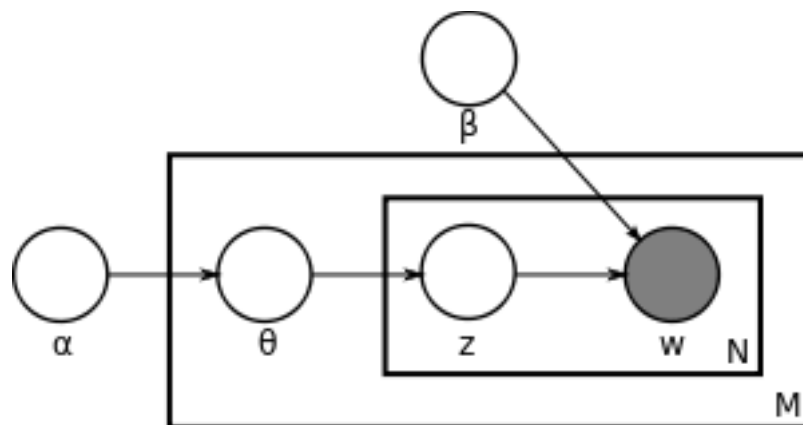


Figure 14: Graphical model representation of LDA (Blei et al., 2003: 997)

The key in LDA is that the words follow a bag-of-words hypothesis, (their order doesn't matter), where the use of a word is to be part of a topic and it communicates the same information no matter where it is in the document. In LDA, each document can be viewed as a collection of different topics. Unlike other methods, LDA also assumes that the topic distribution has Dirichlet distributions as a priori. Practically, this leads to a more rational mix of categories in a document.

Already since 2010, an ever – increasing use of the model has been observed in Digital Humanities. Matthew L. Jockers' *Who's Your DH Blog Mate: Match-Making the Day of DH Bloggers with Topic Modelling* and Cameron Blevins' *Topic Modelling Martha Ballard's Diary* are two of the most popular "blog posts first introduced topic modelling to the broader Digital Humanities community". (Meeks & Weingarten, 2012) On the one hand, Jockers describes how he used topic modelling in the programming language R to analyze the content of Digital Humanities blogs and match bloggers who were interested in similar topics. He also suggested how topic modelling can be used to identify trends and patterns in the content of Digital Humanities blogs. (Jockers, 2010) The result was a matrix of

dimension 117×10 (117 blogs and 10 topics). On the other hand, Blevin used the diary of Martha Ballard (1735 – 1812), an American midwife. For a period of 27 years, she recorded her daily work and domestic life. Blevin used *MALLET (MAchine Learning for Language ToolkiT)*, developed by Andrew McCallum at Umass as “a Java-based package for statistical natural language processing, document classification, clustering, topic modelling, information extraction, and other machine learning applications to text”. (Blevin, 2010) The result showed thirty topics consisting of twenty words each that were classified and titled afterwards by Blevin.

Programming languages and existing special software (e.g. *MALLET*) provide the possibility to perform topic modelling not only using data in English, such as the two aforementioned examples, but also in several other modern languages. Schöch’s *Topic Modelling Genre: An Exploration of French Classical and Enlightenment Drama* contains data in French (Schöch, 2016), while Zosa’s and Pivovarova’s *Multilingual and Multimodal Topic Modelling with Pretrained Embeddings*, contains data in both English and German. (Zosa & Pivovarova, 2022) But what happens in cases like the *Acta Sanctorum* where the data is in Latin? Is it possible to process data and perform topic modelling in ancient languages that are not in use today?

Wishart and Prokopidis performed topic modelling experiments using Hellenistic Greek corpora. The corpora contained texts from ancient Greek authors as well as parts from the New Testament and chronologically spanned from the 8th century BC. until the 2nd century AD. Their aim was “1) to describe the development of a part-of-speech (POS) tagger and a lemmatizer trained on annotated texts written in Hellenistic Greek, and 2) to represent the lemmatized products as topic models in order to examine the effects of automatically processing the texts, and semi-automatically correcting the output of the lemmatizer on tokens occurring frequently in Hellenistic Greek corpora”. (Wishart & Prokopidis, 2017: 39) The generated topic models have been used “in order to test and exemplify the improvements made through manual correction of the lemmatized data”. (Wishart & Prokopidis, 2017: 46) In other

words, Wishart and Prokopidis' paper illustrated the importance of targeting machine learning tools (including topic modelling) toward data from a classical language. (Wishart & Prokopidis, 2017: 47)

In 2018 a group of Carleton College students presented the results of their capstone project where they performed topic modelling in Latin texts. Specifically, the text they used was the *Historia ecclesiastica gentis Anglorum* (*Ecclesiastical History of the English People*), a work written around 751 by Saint Bede (672/3 – 675), an English monk and author. The work is one of the most important sources for Anglo-Saxon history and is rich in material on the history of the Anglo-Saxon conversion to Christianity, while also containing martyrdoms and lives of English saints. The team wrote their project in Python and implemented the LDA algorithm with some specific tweaks to make it perform well on a Latin text. (Durrett n.d.) Then, in order to allow the users to navigate the results of the algorithm easily and intuitively, they connected the LDA topic modelling algorithm with a visualization tool. The interface of this new application, called *Loquela*, has been written in HTML, CSS, and Javascript and has four parts:

- i. "Metadata about the inputted model,
- ii. A re-creation of the text with each word annotated with its topic,
- iii. Heat maps of topics, and
- iv. Word clouds of topics". (Durrett, n.d.)

The biggest challenges the project team had to face were lemmatization, as Latin is a highly inflected language, as well as the interpretation of the results, as LDA is an unsupervised method. The two basic questions that were asked were on the one hand if the topics have different word distributions and on the other hand if the documents have different topic distributions.

6.2 Topic Modelling the *Acta Sanctorum*

The next part of the Python code accompanying this master's thesis contains the data preparation for topic modelling, and its performing. The existing column of the dataframe named

texts_without_stopwords was converted to type *string*, which was named *text*, while at the same time all non – alphanumeric characters⁵ were replaced. The next step contained a function which cleaned the string of the remaining symbols and lowercased all the words, preparing it for lemmatization. As Latin is a highly inflected language, as mentioned above, a specific package was needed that could perform the lemmatization. For this purpose, the use of the library *cltk* (*Classical Language Toolkit*) was chosen. According to the website of the library, *cltk* “offers natural language processing (NLP) support for the languages of Ancient, Classical, and Medieval Eurasia. Greek, Latin, Akkadian, and the Germanic languages are currently most complete. The goals of the *cltk* are to:

- compile analysis-friendly corpora;
- collect and generate linguistic data;
- act as a free and open platform for generating scientific research” (Johnson et al., 2021)

After installing the *cltk* library, the code of the *LatinBackoffLemmatizer* package was saved by the website in an .ipynb folder and stored manually in the local PC memory as it was not one of the default packages of the library. After that, the string *test* was converted to a list named *list_lemma*, to which the *LatinBackoffLemmatizer* was applied. The result was the enrichment of the list with the corresponding lemmatized words as in figure 15. Then the list was cleaned of the remaining numbers and only the lemmatized words were kept.

```
[('de', 'de'), ('pelagia', 'pelagium'), ('proprium', 'proprius'), ('de', 'de'), ('pelagia', 'pelagium'), ('ex', 'ex'), ('menaeis', 'menaeus2'), ('excusis', 'excusis'), ('maii', 'maii'), ('dies', 'dies'), ('proprium', 'proprius'), ('de', 'de'), ('pelagia', 'pelagium'), ('pelagia', 'pelagium'), ('virgo', 'virgo'), ('martyr', 'martyr'), ('tarsi', 'tarsi'), ('cilicia', 'cilicium'), ('enlarge', 'enlarge'), ('this', 'this'), ('image', 'image'), ('hic', 'hic'), ('notatur', 'noto'), ('canendus', 'caneo'), ('hirmus', 'hirmus'), ('initium', 'initium'), ('contacion', 'contacion'), ('festi', 'festus1'), ('praescribuntur', 'praescribo'), ('legenda', 'lego2'), ('synaxario', 'synaxario'), ('prosequitur', 'prosequor'), ('chorus', 'chorus'), ('coeptum', 'coepio'), ('canonem', 'canon'), ('sancta', 'sancio'), ('tum', 'tum'), ('canitur', 'cano'), ('hirmus', 'hirmus'), ('oratio', 'oratio'), ('michaele', 'michaelis'), ('martyre', 'martyre'), ('oratio', 'oratio'), ('de', 'de'), ('michaele', 'michaelis'), ('martyre', 'martyre'), ('theodoro', 'theodoro'), ('metochita', 'metochita'), ('nov', 'nov'), ('iv', 'iv'), ('appendix', 'appendix'), ('oratio', 'oratio'), ('michaele', 'michaelis'), ('martyre', 'martyre'), ('michael', 'michael'), ('martyr', 'martyr'), ('alexandriae', 'alexandriae'), ('oratio', 'oratio'), ('codice', 'co-dico2'), ('vindobonensi', 'vindobonensi'), ('philosophico', 'philosophico'), ('graeco', 'graecus'), ('cf', 'cf'), ('comm', 'comm'), ('praev', 'praev'), ('psalm', 'psalm'), ('ephes', 'ephes'), ('luc', 'luc'), ('ephes', 'ephes'), ('matth', 'matth'), ('rom', 'rus'), ('ephes', 'ephes'), ('gen', 'gen'), ('luc', 'luc'), ('cor', 'cor'), ('tim', 'tim'), ('rom', 'rus'), ('matth', 'matth'), ('matth', 'matth'), ('tim', 'tim'), ('encomium', 'encomium'), ('georgii', 'georgii'), ('mart', 'mart'), ('ex', 'ex'), ('ms', 'ms'), ('codice', 'co-dico2'), ('bibliothecae', 'bibliotheca'), ('vaticanae', 'vaticanae'), ('apr', 'apr'), ('iii', 'iii'), ('ad', 'ad'), ('diem', 'dies'), ('xxiii', 'xxiii'), ('encomium', 'encomium'), ('georgii', 'georgii'), ('mart', 'mart'), ('georgius', 'georgius'), ('megalomartyr', 'megalomartyr'), ('nicomediensis', 'nicomediensis')]
```

Figure 15: The original words with their corresponding lemmatized versions

⁵ All characters that are not numbers (0 - 9) or alphabetic characters (a-z and A-Z).

Before writing the code for topic modelling, an exploration of the corpus was done. More specifically, it was a visualization of the most frequently occurring words, bigrams, and trigrams of the corpus. In order to do this, the *nlTK* (*Natural Language Toolkit*) package has been imported. It is a package which “is a leading platform for building Python programs to work with human language data”. (*Natural Language Toolkit*, n.d.) The first visualization contains the twenty words with the highest frequency in the text corpus (figure 16).

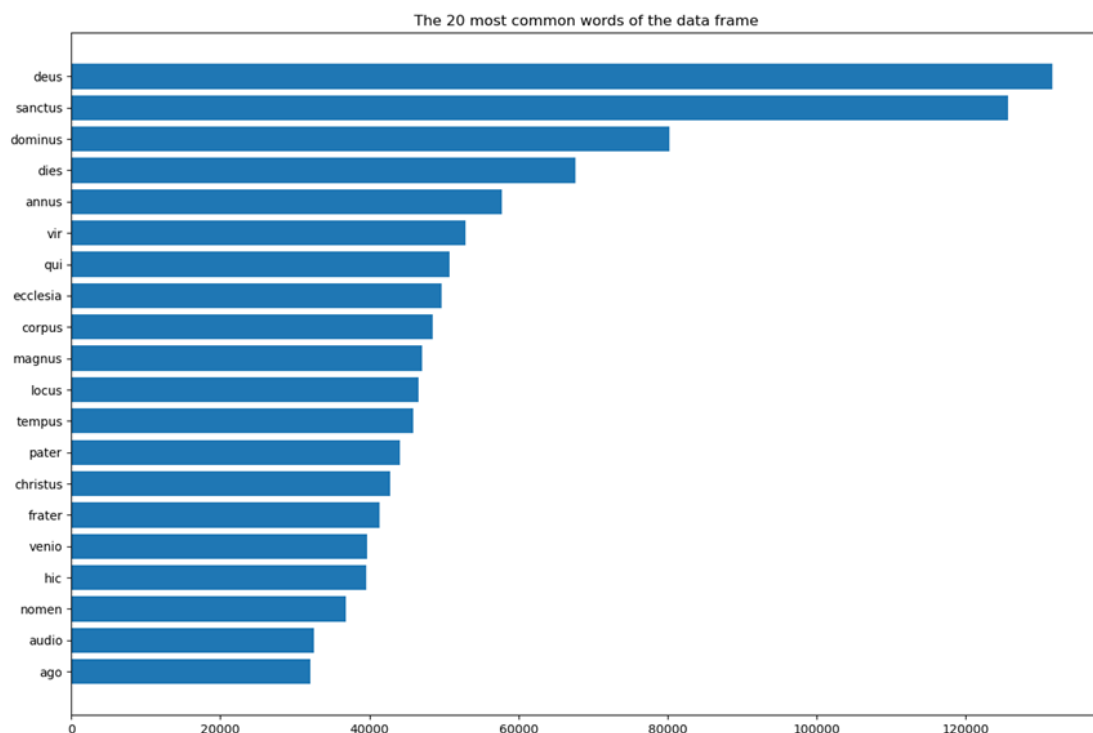


Figure 16: *The 20 most common words of the data frame*

As could be expected, words whose meaning refers directly to the Christian religion, such as *deus*, *sanctus*, *ecclesia*, *dominus*, have a strong presence in the corpus. Although the words are lowercased and lemmatized, it seems that the removal of the stopwords did not work completely as within this visualization there are still the words *qui* and *hic*, which should have been filtered out. At a first glance, the twenty most common words move around a religious framework which, however, addresses more theological as well as more institutional aspects.

A two-word sequence that has a distinct meaning when used together is called bigram while a three-word sequence is called a trigram. As William Mattingly notices, “understanding bigrams and trigrams is essential because in order for a computer to truly understand language the way a human does, it must be able to understand the nuances of a single word and how a word’s meaning not only shifts in context but shifts in meaning when used in conjunction with other word”. (Mattingly, 2022) The second visualization contains the twenty most common bigrams (figure 17).

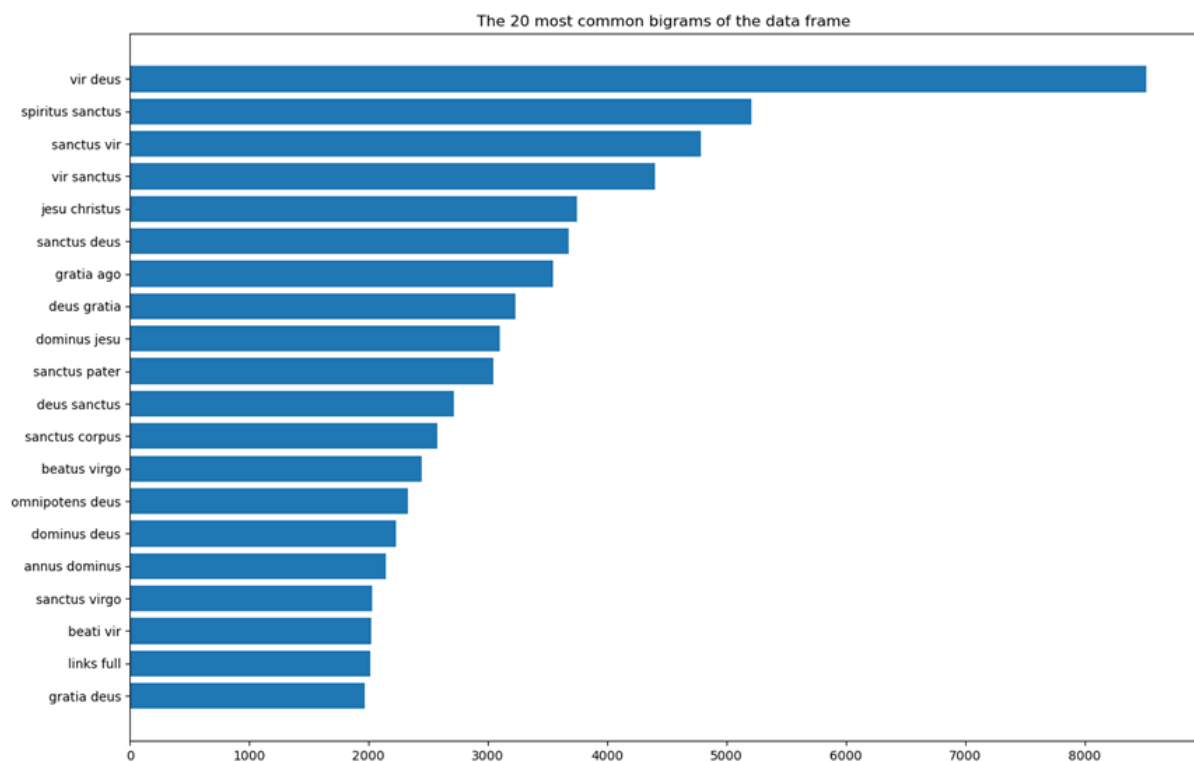


Figure 17: The 20 most common bigrams of the data frame

The bigrams of the corpus reflect more the descriptions of the saints: *vir deus*, *vir sanctus* and *sanctus virgo* are just some of the examples. Of course, there are also references to the Triune God, such as *iesus christus* and *spiritus sanctus*. The visualization of the trigrams contains a greater variety of general liturgical and religious phrases which are not specific to hagiography such as (*secula seculorum amen*, *deus gratia ago*, *dominus iesum christum*, *pater filius spiritus*, etc).

The visualization of the most frequent trigrams also highlighted two issues:

- i. Imperfect lemmatization: For example, in the trigram *apostolus petri paulus*, the name *Petrus* is referred to in genitive case instead of nominative case. Another example is the trigram *secula seculorum amen*, in which neither word has been lemmatized (*amen* is uninflected).
- ii. Spelling: Despite the normalization done in the first part of the code, there are still spelling issues that are not easy to solve. A typical example is the following: *frater ioannis bonus* and *frater ioannes bonus*. Although it is the same word, there are two different versions. Another example, which is directly related to the issue of the incomplete lemmatization is the following: *secula seculorum amen* and *saeculum saeculum amen*. The spelling variations led the computer not only to understand the same trigram as two different ones but also to an incomplete lemmatization in the simple version without the diphthong *ae*.

The next part of the code is topic modelling, which consists of four cells in a Jupyter Notebook, whose first three are taken from William Mattingly's website. (Mattingly, 2022) The library used for topic modelling was *gensim*, a popular open source NLP library used for unsupervised topic modelling. Some of the main advantages of *gensim* are that compared to other libraries it provides more convenient facilities for text processing and doesn't require annotations or hand tagging of documents because it uses unsupervised models. (*Gensim Introduction* n.d.) The first of the three steps introduced the *gensim* library and then a function which converted the list of lemmatized words to a list of lowercased tokens. In the second step, all bigrams and trigrams were removed from the list of lowercased tokens. The third step contained the extraction of *Term Frequency-Inverse Document Frequency* or *TF-IDF* for short. It is a statistical measurement that makes it possible to evaluate the importance of a word contained in a document, relative to a collection of documents. The relevancy of a word to a document in a collection of documents is calculated by multiplying two metrics: the frequency with which a word appears in a document (*Term Frequency – TF*) and the inverse document frequency of the word across a collection of documents (*Inverse Document Frequency – IDF*). (Mattingly, 2022) The aim in this case is to remove frequently occurring words that do not necessarily mean anything important for the

topics that are going to be generated. Finally, the fourth and last part contains the code that creates the dictionary and the corpus and then builds the LDA model.

In order to select the parameters and hyperparameters that have led to the final result, the model was run several times with different values, which each time were evaluated based not only on quantitative methods but also on human judgment. The quantitative methods used were the measurement of c_v coherence and perplexity. On the one hand, c_v coherence is based on a Boolean sliding window of size 110 and “an indirect confirmation measure that uses normalized pointwise mutual information (NPMI)⁶ and the cosine similarity”. (Singh, 2023: 29) “The coherence measure retrieves co-occurrence counts for the given words” (Singh, 2023: 29) which “are used to calculate the NPMI of every top word, to every other top word resulting in a set of vectors (one for every top word)”. (Singh, 2023: 29) “The segmentation of the top words leads to the calculation of the similarity between every top word vector and the sum of all top word vectors where the cosine similarity is used as a measurement unit”. (Singh, 2023: 29) Finally, the coherence score is the average of all these cosine similarities, and the closer this score is to one, the better the coherence in the model. On the other hand, perplexity is a “commonly used measurement in information theory to evaluate how well a statistical model describes a dataset, with lower perplexity denoting a better probabilistic model”. (Zhao et al, 2015: 2)

As far as it concerns the current work, the final parameters and hyperparameters chosen for the LDA model are as follows:

- num_topics = 5: The number of latent topics to be extracted from the training corpus.
- random_state = 42: A random state seed which is useful for reproducibility.
- chunksize = 5000: The number of documents to be used in each training chunk.
- passes = 100: Number of passes through the corpus during the training process.
- alpha = 1: Scalar for a symmetric prior over document-topic distribution.

⁶ In statistics, the point mutual information (PMI) is a measure of correlation and dependence. This criterion compares the probability of two events occurring at the same time with the probability of their occurrence if they are independent.

- $\eta = 1$: Scalar for a symmetric prior over topic-word distribution. (Řehůřek, n.d.)

The hyperparameters α and η are used to control the sparsity of the document – topic and topic – word distributions, respectively. A higher value of α results in more topics per document, while a lower value of α results in fewer topics per document. In the same way, a higher value of η results in more words per topic, while a lower value of η results in fewer words per topic. In this case, both α and η values are considered high.

6.3 Results

The result of the LDA model had a c_v coherence score of 0.84 and a perplexity score of -9.20. The model consists of five topics, which were visualized with the help of the *pyLDavis* library, a library designed to help users interpret the topics in a topic model. (*pyLDavis*, n.d.) It generates an interactive visualization, where the user has the possibility to select each topic individually and also observe the thirty most salient terms of the model (figure 18). Furthermore, each circle of the visualization represents a topic while the size shows the percentage of tokens belonging to the topic (figure 19).

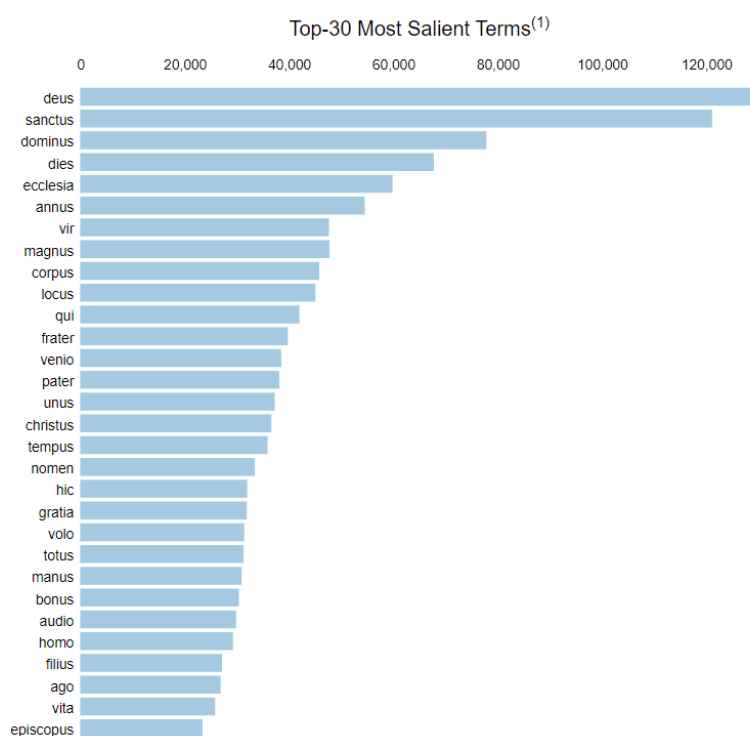


Figure 18: The top-30 most salient terms

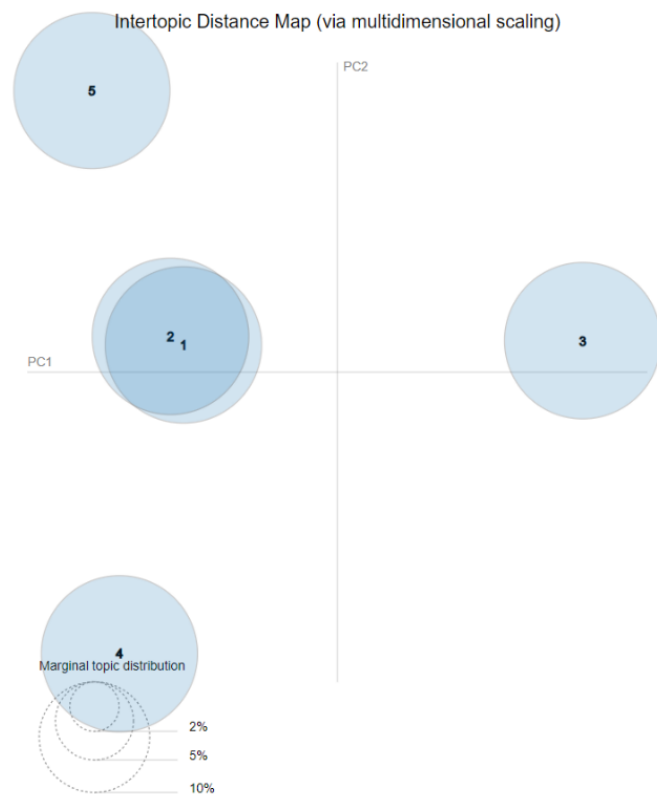


Figure 19: An intertopic distance map of the model

Table 4 presents the ten most relevant terms of each topic. On the one hand, it seems that the lemmatization process worked successfully, as all the words are correctly lemmatized, but on the other hand there are words like *qui* and *hic*, which shows that the removing of stopwords did not work completely. Each topic consists mainly of nouns but there are also some adjectives and verbs. The meaning of the vast majority of words is religious or directly related to religious matters:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Dominus	deus	Corpus	sanctus	dies
Magnus	locus	Pater	ecclesia	vir
Frater	qui	Unus	annus	manus
Nomen	bonus	Tempus	venio	hic
Filius	sanctus	Gratia	christus	volo

Virgo	ago	Coepio	totus	homo
Sequor	beatus	Opus	audio	fides
Pono	episcopus	Mors	vita	populus
Domus	rex	Mereo	virtus	ordo
Os	verbum	Invenio	terra	mater

Table 4: The 10 most relevant terms of each topic

The next step included the extraction of the dominant topic and its percentage for each hagiographical text, which were then stored in two new columns called *dominant_topic* and *percentage* respectively.

The result for each topic is as follows:

Topic 1 contains 981 texts or 21% of all texts. This topic basically includes saints of the High Middle Ages and the Late Medieval Period as well as saints who lived during the early modern period. Apart from the similarities in the chronological period in which they lived and acted, the great majority of the documents concerned are about saints-members of religious orders, many of whom had missionary activity. Worth mentioning is the fact that in this particular topic there is a large number of female saints in contrast to the other topics (about 17%). The chronological period in which the female saints of this topic lived and acted extends throughout the whole Middle Ages. Table 5 presents the five female saints with the highest scores on the topic:

Saint	Century	Region	Comments
Wenefrida	Eighth	England	Nun and Martyr
Wiborada	Tenth	Saint Gallen, Switzerland	Nun and Martyr
Hemma	Eleventh	Gurk, Carinthia	Noblewoman and founder of churches and monasteries

Clara	Thirteenth	Assisi	Founder of <i>Ordo sanctae Clarae</i>
Angela	Thirteenth	Bohemia	Hermitess

Table 5: The female saints with the highest scores in Topic 1

In addition, the way in which the saints are presented in *Acta Sanctorum* in relation to their area of action is of particular interest. Although the majority of saints are associated with their city, as in the Early and High Middle Ages, here we can observe a slight tendency for them to be associated with the name of a country or a wider region. For example, Saint Michael Giedroyć has the nickname Lithuanian and Saint Stanislav the nickname Polonus, although he is also known as Stanislaus of Szczepanów or Stanislaus the Martyr. This phenomenon is more pronounced in female saints (Wenefrida of Anglia, Angela of Bohemia, etc). The following table presents the top ten saints of the topic with the highest scores:

Saint	Century	Region	Comments
Michael Giedroyć	Fifteenth	Lithuania	Member of the <i>Ordo Canonorum Regularium Mendicantium S. Mariae de Metro de Poenitentia Sanctorum Martyrum</i>
Stanislav	Eleventh	Poland	Martyr and Bishop
Martin de Porres	Seventeenth	Peru	Member of the Dominican Order
Raymond Palmarius	Thirteenth	Piacenza	Pilgrim

Antonius Baldinucci	Eighteenth	Pofi, Italy	Member of the Jesuit Order
Johannes Bonus	Thirteenth	Mantua, Italy	Lay Brother
Henricus Baucenensis	Fourteenth	Bozen/Bolzano and Trevizo	Pilgrim
Franciscus de Paula	Sixteenth	Paola/ Plessis-lèz- Tours	Member of the <i>Ordo Minimorum (Paulaner Order)</i>
Bernard of Calvo	Thirteenth	Spain	Member of the Cistercian Order
Johannes Angelus Porro	Sixteenth	Milan	Member of the <i>Ordo Servorum Beatae Mariae Virginis</i>

Table 6: The saints with the highest scores in Topic 1

Topic 2 includes 270 texts which is 5.8% of all texts. In this topic the type of sainthood, the chronological period and the geographical areas are the common characteristics among the saints. These are Martyrs of Late Antiquity, who lived and were martyred in the eastern part of the Roman Empire or in Rome and mostly have Greek names. Specifically, Rome, Anatolia, the Near East, and North Africa are basically the geographical areas where the saints lived, while chronologically they are mainly placed during the time of the Roman persecutions against the Christians (mostly in the third and fourth centuries). The following table lists the ten saints with the highest score in this topic:

Saint	Century	Region	Comments
Galaction and Episteme	Third	Emesa (t. in Syria)	Martyrs
Orestes	Third	Cappadocia	Martyr
Theodorus	Fourth	Euchaita (t. in Turkey)	Martyr
Tryphon Respicius	Third	Rome	Martyrs

Nympha			
Epimachus	Fourth	Alexandria	Martyr
Hieron and 33 Companions	Third	Melitene (t. in Turkey)	Martyr
Eleutherius	Fourth	Tarsus (t. in Turkey)	Martyr
Martyres XXXVII Philippopoli	Fourth	Philippopolis (t. in Bulgaria)	Martyrs
Christophorus	Third	Lycia (t. in Turkey)	Martyr
Eusebius and 9 Companions	Third	Rome	Martyrs

Table 7: *The saints with the highest scores in Topic 2*

It is worth mentioning that all the different versions of the same hagiographical text belong to the same topic. For example, the eleven versions of the Life and Martyrdom of Saint Theodore that exist in the *Acta Sanctorum* belong to Topic 2 as well as the ten versions of the Life and Martyrdom of Saints Tryphon, Respius and Nympha. The same phenomenon is also observed in hagiographical texts with a smaller number of versions, such as for the Saints Galaction and Episteme, for whom there are two versions, both belonging to the same topic.

Topic 3 contains 2.382 texts or 51% of all texts. The hagiographical texts of the topic cover a period of time between the end of the fifth century and the beginning of the eleventh century with an emphasis on the Early Middle Ages. Geographically, they cover Central and Western Europe. In the vast majority of the texts there are saints who held some high degree of priesthood, such as bishop or abbot, while in some cases they also held the status of martyr. The saints of this topic are directly connected to the city in which they lived, acted, and sometimes were martyred. In several cases, these are saints who preached Christianity and founded the first Christian community in a city, such as St. Gerard, the first bishop of Csanád in Hungary, or St. Bruno, also a bishop, who was martyred in Prussia teaching Christianity. Indicatively, the table below contains fifteen saints of this topic, who had the highest scores:

Saint	Century	Region	Comments
Leoluca	Tenth	Sicily	Abbot
Severinus	Fifth	Noricum (t. in Austria)	Bishop
Decorosus	Seventh	Capua, Italy	Bishop
Gerard	Eleventh	Hungary	Bishop
Adomnán	Eighth	Scotland	Abbot
Vedastus	Sixth	Arras, Belgium	Bishop
Bruno	Eleventh	Prussia	Bishop and Martyr
Eneco	Eleventh	Aragon, Spain	Abbot
Wulfram	Eighth	Sens, France	Archbishop
Botolph Adulf	Seventh	Thorney, England	Abbot and Bishop respectively
Cuthbert	Seventh	Lindisfarne, England	Bishop
Benedict	Sixth	Nursia, Italy	Monk, Founder of the <i>Ordo Sancti Benedicti</i>
Gall	Sixth	Clermont, France	Bishop
Wolbodo	Eleventh	Liège	Bishop
Hugh	Eighth	Rouen	Archbishop

Table 8: The saints with the highest scores in Topic 3

Finally, it is worth mentioning that the saints of the same period who lived in the Byzantine Empire (mainly in Constantinople and Anatolia) are also part of this topic. The vast majority of them held the title of bishop or hegumen, a term similar to abbot. Such examples are Saint Hilarion the Younger (ninth century), hegumen of the Dalmatian Monastery in Constantinople, and Saint Nikitas (eighth century), abbot of the Medikion Monastery in Asia Minor. In addition, the patriarchs of Constantinople,

Jerusalem, and Alexandria, who belong to the same chronological period as the rest of the saints of this topic, are also part of this subgroup. Table 9 shows the five patriarchs with the highest scores:

Saint	Century	Region	Comments
Ioannes Eleemosynarius	Seventh	Alexandria	Patriarch
Sophronius	Seventh	Jerusalem	Patriarch
Ignatius	Ninth	Constantinople	Patriarch
Nikephoros	Ninth	Constantinople	Patriarch
Cyrillus	Fifth	Alexandria	Patriarch

Table 9: *The Patriarchs with the highest scores in Topic 3*

Topic 4 includes 336 texts which is 7.2% of all texts. The connecting link of this group is not some common time period or geographical region of action but the status of each saint. The vast majority of the saints in this topic held some political, military, or spiritual authority during their earthly lives. In the first places are the versions of the life of Saint Elesbaan, king of Aksum, followed by Saint Eric IX, king of Sweden, Saint Sigebert III, the Merovingian king of Austrasia and Saint Richard of Wessex, also known as Saint Richard the Pilgrim. In the table below, the fifteen saints who occupy the top positions of the topic are shown indicatively, together with the century they lived, their region, and their status:

Saint	Century	Region	Comments
Elesbaan	Fifth	Aksum (t. in Ethiopia)	King
Eric IX	Twelfth	Sweden	King
Sigebert III	Seventh	Austrasia (t. in France)	King
Richard	Eighth	Wessex (t. in England)	King
Adeodatus I	Seventh	Rome	Pope
Edward	Eleventh	England	King

Gontrand	Sixth	France	King
Ethelbert	Eighth	England	King
Paschalis I	Nineth	Rome	Pope
Theodorus Craterus	Nineth	Amorium (t. in Turkey)	Army General
Ladislaus V	Fifteenth	Hungary	King
Magnus Erlendsson	Twelfth	Orkney (t. in Scotland)	Earl
Charlemagne	Eighth/Ninth	France	Emperor
Gregory I	Seventh	Rome	Pope
Paul I	Eighth	Rome	Pope

Table 10: *The saints with the highest scores in Topic 4*

Observing the table above, it becomes clear that power is the common characteristic of the saints of this topic. Kings, popes, emperors, counts, military leaders, and even nobles occupy most of this topic. In addition, there are saints whose names have been linked to the founding of a monastery or a religious order, such as St. Stephen of Muret, founder of the Abbey of Grandmont and the Order of Grandmont in France (eleventh/twelfth century) and also St. Gilbert of Sempringham, founder of the Gilbertine Order in England (tenth/eleventh century). The geographical area does not seem to play an important role in the grouping of this topic, nor does the century of action as the saints cover the entire Christian world both geographically and chronologically.

Topic 5 includes 696 texts or 14.9% of all texts. This topic contains hagiographical texts of martyrs and bishops of Late Antiquity, who lived in the central and southwestern part of the Roman Empire. In several cases, there is a combination of the status, i.e. a person was already a bishop when he was martyred. The names of the saints are Latin, and their period of activity was from the first to the fifth centuries. The following table contains the names of the ten saints with the highest score in this topic:

Saint	Century	Region	Comments
Severus	Fourth	Barcelona	Martyr
Valentinus Hilarius	Fourth	Viterbo, Italy	Martyrs (they were Bishop and Deacon respectively)
Claudius Lupercus Vittoricus	Third	Spain	Martyrs
Calocerus	Second	Albegna, Italy	Martyr
Processus Martinian	First	Rome	Martyrs
Gavinus Protus Januarius	Fourth	Sardinia	Martyrs
Juvenal	Fourth	Narni, Italy	Bishop
Orientius	Fifth	Auch, France	Martyr and Bishop
Restitutus	Fourth	Tricastinum (t. in France)	Bishop
Erasmus	Fourth	Formia, Italy	Martyr and Bishhop

Table 11: *The saints with the highest scores in Topic 5*

Summing up, the results of the topic modelling clearly reflect the evolution of hagiography as a literary genre from its appearance during Late Antiquity until the early modern period. There is a clear distinction according to the chronological period, the geographical area, and the status of the saints, which of course were directly forged by the prevailing conditions and influenced by the emerging needs. Late Antiquity is strongly represented in two topics (2 & 5), which on the one hand agree

chronologically and based on the status of the saints (martyrs), but their differences are geographical as one topic includes martyrs from the basically Greek-speaking Eastern Mediterranean while the other includes martyrs of the basically Latin-speaking Western Mediterranean. As Claudia Rapp notes, “the ‘parting of way’ that resulted in eroding the cultural unity of the Later Roman Empire and the separate developments of Byzantium and the Medieval West is often measured by the declining knowledge and eventual disappearance of Latin in the East and Greek in the West”. (Rapp, 2004: 1221) It seems that the language as a carrier and means of expression of culture gave its own characteristics to the hagiographical texts of the martyrs of Late Antiquity. Although all *Acta Sanctorum* texts used in this work are in Latin (either the original or their translation), it is clear that the cultural differences between the two parts of the Roman Empire led to the above-mentioned kind of grouping.

The end of Late Antiquity, the rise of Christianity and the establishment of local churches and Christian communities during the Early Middle Ages is captured in Topic 3. The bishops of local parishes began to become the central figures of the hagiographical texts and play an active role both in religious and political life. The hagiographical texts of this era turn more and more towards Central and Western Europe and concern fewer saints who lived in the Byzantine Empire. Moreover, the gradual spread of Christianity is also reflected through the central figures of the texts. For example, the fifth and sixth centuries are more concerned with bishops of today’s Italy, Southern France, and Spain, while as Christianity spread further north and new communities were established, the more saints-bishops of the northernmost regions (England, Scandinavia) come to the fore. The need to create the founding myth of each local diocese, led to the production of more and more hagiographical texts related to the life of bishops in combination, of course, with the prevailing conditions (the strengthening of the cult of local saints, the need to centralize political and religious power, establishing parishes in the countryside, etc.).

Conditions from the middle of the eleventh century changed. The Crusades, the establishment of the religious orders, Humanism, the appearance of Protestantism and the Counter-Reformation as well as

the missions beyond Europe are just some of the factors that changed the way hagiographical texts were written and gave them new characteristics. These changes are reflected in Topic 1, where the standards of holiness turn in a different direction than in the Early Middle Ages. Members of religious orders, priests who fought heresies and at the same time proved their obedience to the Holy See, pilgrims, and missionaries are the central figures of hagiography from the eleventh century until the Enlightenment.

Finally, Topic 4 captures an extraordinary grouping, which is based neither on the common chronological period/common geographical area motive, nor on the religious status of each saint (martyr, bishop, etc.) but it is based on a special characteristic: the power of the saints during their earth life. Regardless of century and region, kings, emperors, counts, army officers and popes are the main figures of the topic with a connecting link of political and/or religious power. Reading the hagiographical texts of this topic, one can understand that there is a special focus on the power and the way it is managed by the people who possess it without, of course, omitting the religious aspects of their lives.

Summarizing, the five topics of the LDA model present coherence and uniformity at a rate of 65 - 70% but, of course, they also contain outliers. It is also worth mentioning the fact that the different versions of hagiographical texts for certain saints belong, in the vast majority, to the same topic. Such examples are the hagiographical texts on Saint Theodorus and Saints Tryphon, Respicus and Nympha, as mentioned above. However, this phenomenon does not only concern Topic 2, but is a more general phenomenon in the LDA model. As far as outliers are concerned, one can come across hagiographical texts about saints of Late Antiquity in the topic with saints of the Late Middle Ages or vice versa. Another example is the members of the religious orders, which are overwhelmingly concentrated in topic 1, but some scattered texts exist in other topics as well. Nevertheless, the existence of outliers does not alter the meaning and coherence of the topics as they are a minority. After all, the goal in unsupervised machine learning ways to organize texts, such as topic modelling, is not to avoid the

existence of outliers but to draw safe conclusions and interpret the results based on both human judgment and qualitative methods of evaluation. In this case it was possible to draw safe conclusions from the results of the LDA model, which were also largely verified by the qualitative measurements.

Chapter 7 – Conclusions

This master's thesis aimed to examine the *Acta Sanctorum* using computational methods and specifically data analytics and topic modelling. The aim was the in-depth exploration of the work of the Bollandists and the quantitative examination of its content looking for aspects which are not perceptible with the use of qualitative analysis only. For this reason, the first part of the research focused on the study of hagiography as a literary genre as well as the presentation of the *Acta Sanctorum* and the context in which the project was established. The second part contains the quantitative analysis methods that were applied in order to answer the research questions raised. Regarding the data of the study, these come entirely from the hagiographical texts of the *Acta Sanctorum*, which have been cleaned and processed with the help of the Python programming language. Then the references, comments and in general the contribution of the Bollandists to the work were removed in order to use for the analysis only the original version of each hagiographical text.

In a nutshell, based on the methods of natural language processing applied, it can be concluded that the hagiographical texts included in the *Acta Sanctorum* can be grouped into different topics, depending on their content and characteristics. Although all these texts follow the same writing pattern and their content shows uniformity in terms of the main purpose, the LDA model showed that there are some secondary features, which play their own special role and lead to the grouping of the different topics. Focusing on the results of the LDA model, the geographical space, the time period that the saints lived, and the status of each saint, as well as the combination of these characteristics, result in the existence of particularities and differences between the various topics. Although there is evidence of these characteristics from historical research, they are difficult to discern using qualitative analysis methods alone. In other words, the spatio-temporal context, and the conditions in which the hagiographical texts of the *Acta Sanctorum* were written, as presented in the first part of the study,

were not only confirmed by the results of topic modelling, but at the same time, previously unknown insights came to the fore.

In addition, the data analytics methods gave a complete and deep overview of the data providing statistics about the complexity of the hagiographical texts. The exploratory data analysis and the corresponding visualizations gave, among other things, information about the gender of the saints, the centuries and places of publication of the volumes as well as the types of hagiographical texts. It is also important to mention that the data analytics methods confirmed that the types of hagiographical texts (*Vita*, *Passio*, *Miracula*, etc.) do not have distinct boundaries and many times there is a combination of them in a single text regardless of how it is titled. The second part of the data analytics methods provides a comparison of the data before and after the removal of stopwords, providing statistical information on the hagiographical texts. Through this analysis it is realized that there is generally a homogeneity in the sizes of the texts and words of the *Acta Sanctorum*. The geographical area, the chronological period and the status of the saints, i.e. the factors that determined the five different topics, do not play any role in the sizes or composition of the texts.

The results of the research also include the highlighting of the particularities that exist in the application of data analytics and natural language processing methods in Digital Humanities, especially when primary sources in Latin are used. These peculiarities have to do basically with issues concerning the changes that have occurred in the Latin language over the centuries, such as the addition of new words, or the nature of the language, such as the inflection in Latin. Although in recent years there has been a development in the processing of classical Latin with the help of computational methods, the processing of Latin of the Middle Ages still remains a great challenge, let alone in cases such as the hagiographical texts of the *Acta Sanctorum*, which show large discrepancies not only due to of the time range of their writing but also because of the wide range of places of their writing which coincides with the geographical spread of Christianity. The main challenges during the analysis had to do with the peculiarities of the Latin language and mainly with the different versions of spelling for certain

words. For this purpose, the process of normalization of the data was deemed necessary, i.e. the replacement of the different spelling versions by a specific one, which in all cases was the dominant one. The next challenge had to do with the fact that Latin is a highly inflected language, which makes certain processes, such as lemmatization, more demanding. In order to address this challenge, a specific *cltk* library package was installed and used, which resulted in the issue being resolved. Last it is worth mentioning the technical challenges encountered during the research process. In addition to the search and installation of libraries necessary for the analysis of the data and the application of natural language processing methods, which at the same time had to be compatible with Python, further solutions had to be sought in order to cope with the RAM memory and the temporary storage area in processing such a large corpus of tokenized words. This resulted in the selection of specific parameters, which lengthened the time it took the LDA model to run but did not interfere with the parameters responsible for the quality of the results. In other words, a slow process - sometimes over the course of several days - was preferred, which did not affect maintaining the quality level of the output. Nevertheless, it is clear that data analytics and topic modelling are powerful tools for analyzing Latin texts as they reveal unknown aspects and demonstrate the value of computational approaches in the humanities.

To sum up, the master's thesis reveals the importance of quantitative methods in the study of Western Christian hagiography. By using new tools and approaches, new insights of *Acta Sanctorum* have been uncovered and aspects that have been overlooked in the past or were impossible to identify, have been brought to light. Of course, this small contribution could be a basis and inspiration for future studies concerning both the *Acta Sanctorum* and Christian hagiography in general. To better understand the implications of these results, future studies could go even deeper in analysis, by including the non Latin texts, the contribution of the Bollandists and even the joint examination of the work with other collections of hagiographical texts. Moreover, another kind of approach through other tools and technical methods could illuminate aspects that were not possible to study within the limits

of a master's thesis. As the field of Digital Humanities continues to evolve, new tools, new research methods, and of course new research questions can add further data and expand studies around hagiography.

Bibliography

1. *About - who we are*. (n.d.). ProQuest. Retrieved October 26, 2023, from <https://about.proquest.com/en/about/who-we-are/>
2. *Acta sanctorum*. (n.d.). ProQuest. Retrieved October 26, 2023, from <https://about.proquest.com/en/products-services/acta/>
3. Babina, A. A. (2011). Sviatye u trona: modeli vzaimodeistviia svyatykh i svetskoi vlasti v Latinskoi agiografii IV—VI vv. [The Saints at the Throne: Models of interaction between saints and secular power in the Latin hagiography of the IV-VI centuries]. *Intelligentsiya i mir. [Intelligence and the World]* 4. 72 – 93.
4. Badea, A. (2012). (Heiligen-)Geschichte als Streitfall. Die Acta sanctorum und Mabillons Epistola de cultu sanctorum ignotorum und die römische Zensur. In T. Wallnig, T. Stockinger, I. Peper & P. Fiska (Ed.), *Europäische Geschichtskulturen um 1700 zwischen Gelehrsamkeit, Politik und Konfession* (pp. 379-404). Berlin, Boston: De Gruyter. doi.org/10.1515/9783110259193.379
5. Berra, A. *Ancient Greek and Latin Stopwords for textual analysis* [Repository]. Retrieved October 26, 2023, from <https://github.com/aurelberra/stopwords>
6. Berry, D. (2019, February 13). What are digital humanities?. *The British Academy*. <https://www.thebritishacademy.ac.uk/blog/what-are-digital-humanities/>
7. Beveridge, J. (1939). Two Scottish Thirteenth-Century Songs. *Music & Letters*, 20(4), 352-364.
8. Blei, D. M. (2012). Topic Modeling and Digital Humanities. *Journal of Digital Humanities*. <https://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/>
9. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(1), 993-1022
10. Blevins, C. (2010, April 1). TOPIC MODELING MARTHA BALLARD'S DIARY. CAMERON BLEVINS. <https://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>
11. Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of Topic Models. *Foundations and Trends in Information Retrieval*, 11(2–3), 143–296. <https://doi.org/10.1561/15000000030>
12. Brooks, P. (1984). A Lily Ungilded? Martin Luther, the Virgin Mary and the Saints. *Journal of Religious History*, 13(2), 136-149. <https://doi.org/10.1111/j.1467-9809.1984.tb00192.x>

13. Burrows, S., Curran, M., Hiribarren, V., Kattau, S., & Merivale, H. (2014). The French Book Trade in Enlightenment Europe Project, 1769-1794 [FBTEE Project]. <http://fbtee.uws.edu.au/stn/>
14. Carron, H. (Ed). (2005). *Clemens saga: The life of St Clement of Rome*. (Trans. H. Carron). Viking Society for Northern Research, University College London.
15. Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems* (pp. 288-296). Association for Computational Linguistics.
16. Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization Techniques for Assessing Textual Topic Models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)* (pp. 74-77). Association for Computing Machinery. <https://doi.org/10.1145/2254556.2254572>
17. Collins, D. J. (2006). Latin Hagiography in Germania, 1450-1550. In G. Philippart (Ed.), *Hagiographies: Histoire internationale de la littérature hagiographique latine et vernaculaire en Occident des origines à 1550* (Vol. 4, pp. 521-582). Brepols.
18. Cruz, A. J., & Perry, M. J. (Eds.). (1991). *Culture and Control in Counter-Reformation Spain*. University of Minnesota Press.
19. De Smedt, C. (1907). The Bollandists. In *The Catholic Encyclopedia*. New York: Robert Appleton Company. <http://www.newadvent.org/cathen/02630a.html>
20. De Voragine, J. (1998). *The Golden Legend: Selections*. Penguin Books.
21. Delehaye, H. (1920). *L'oeuvre des Bollandistes a travers trois siècles, 1615-1915*. Bruxelles: Bureaux de la Société des Bollandistes.
22. Delehaye, H., & Crawford, V. M. (2014). *The Legends Of The Saints: An Introduction To Hagiography*. Literary Licensing, LLC.
23. Ditchfield, S. (2007). Tridentine worship and the cult of saints. In R. Hsia (Ed.), *The Cambridge History of Christianity* (Cambridge History of Christianity, pp. 201-224). Cambridge: Cambridge University Press.
24. Durrett, M. (n.d.). Topic Modeling. <https://marthadurrett.com/selected-work/topic-modeling/>
25. Encyclopedia Britannica. (2023, September 22). Lemma: *hagiography*. Retrieved October 26, 2023, from <https://www.britannica.com/topic/hagiography>

26. Felix, T. (2008). *Propaganda in Hagiography*. All Empires History Forum. <http://www.allempires.com/allempires.com-redirect/article/files/File/Issue%2021%20PDFs/Propaganda%20in%20Hagiography.pdf>
27. Fokkens, A., ter Braake, S., Ockeloen, N., Vossen, P., Legêne, S., & Schreiber, G. (2014). BiographyNet: Methodological Issues when NLP supports historical research. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 3728-3735). Reykjavik, Iceland: European Language Resources Association.
28. Forrestal, A. (2008). The Church in the Tridentine and Early Modern Eras. In G. Mannion & L. S. Mudge (Eds.), *The Routledge Companion to the Christian Church* (pp. 95-112). Routledge.
29. Fouracre, P. (Ed.). (2005). *The New Cambridge Medieval History (The New Cambridge Medieval History)*. Cambridge: Cambridge University Press.
doi:10.1017/CHOL9780521362917
30. Frazier Knowls, A. (2005). *Possible lives: Authors and Saints in Renaissance Italy*. Columbia University Press.
31. Garin, E. (1964). Weltkulturen. Renaissance in Europa. In G. Mann & A. Heuss (Eds.), *Propyläen Weltgeschichte, Band 6: Die Kultur der Renaissance*. Berlin. Propyläen Verlag.
32. Gauly, B. (2020). Von Löwen und Mönchen: Tiere in spätantiker Hagiographie. *Antike und Abendland*, 65–66(1), 174-189.
33. *General Python FAQ*. (n.d.). docs.python. <https://docs.python.org/3/faq/general.html>
34. *Gensim Introduction*. (n.d.) Tutorialspoint. https://www.tutorialspoint.com/gensim/gensim_introduction.html
35. Gibbs, F., & Owens, T. (2012). Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly*, 6. <https://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html>
36. González-García, A. C., & Quintela, M. V. G. (2016). FROM HAGIOGRAPHY TO CELTIC COSMOLOGY: ARCHAEOASTRONOMY AND CHRISTIAN LANDSCAPE IN OURENSE (NW SPAIN). *Zenodo (CERN European Organization for Nuclear Research)*. <https://doi.org/10.5281/zenodo.220969>
37. Goullet, M. (2003). Vers une typologie des réécritures hagiographiques, à partir de quelques exemples du Nord-Est de la France. In M. Goullet & M. Heinzelmann (Eds.), *La réécriture hagiographique dans l'Occident médiéval: Transformations formelles et idéologiques* (pp. 9-28). Thorbecke.
38. Hayward, P. (n.d.). *Medieval Primary Sources—Genre, Rhetoric and Transmission*. Retrieved August 13, 2023, from <https://www.lancaster.ac.uk/staff/haywardp/hist424/index.htm>

39. Head, T. (1990). The ideal of sanctity formation imitation and dissemination. In *Hagiography and the Cult of Saints: The Diocese of Orléans, 800–1200* (Cambridge Studies in Medieval Life and Thought: Fourth Series, pp. 102–134). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511562457.004
40. Head, T. (1999). The Development of Hagiography and the Cult of Saints in Western Christendom to the Year 1000. *The ORB: On-Line Reference Book for Medieval Studies*. Hunter College and the Graduate Center, CUNY. <https://the-orb.arlima.net/encyclop/religion/hagiography/hagindex.html>
41. Henriot, C. (2020). *Rethinking historical research in the age of NLP*. Elites, Networks and Power in Modern China. <https://enepchina.hypotheses.org/3275>
42. Henry, J. (2013). Humanist Hagiography in England, c. 1480 - c. 1520. *Literature Compass*, 10(7), 535–543.
43. Hotson, H., & Wallnig, T. (Eds.). (2019). *Reassembling the Republic of Letters in the Digital Age*. <https://doi.org/10.17875/gup2019-1146>
44. Jockers, M. (2010). Who's Your DH Blog Mate? Match-Making the Day of DH Bloggers with Topic Modeling. *Matthew Jockers*. <https://www.matthewjockers.net/2010/03/19/whos-your-dh-blog-mate-match-making-the-day-of-dh-bloggers-with-topic-modeling/>
45. Johnson, K. P., Burns, P. J., Stewart, J., Cook, T., Besnier, C., & Mattingly, W. J. B. (2021). The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations* (pp. 20–29). <https://doi.org/10.18653/v1/2021.acl-demo.3>
46. Jotischky, A. (2022). Monks and the Muslim Enemy: Conversion, Polemic and Resistance in Monastic Hagiography in the Age of the Crusades, c. 1000–1250. *Transactions of the Royal Historical Society*, 1–18. <https://doi.org/10.1017/S0080440122000159>
47. Karsdorp, F., Kestemont, M., & Riddell, A. (2021). *Humanities Data Analysis: Case Studies with Python*. Princeton University Press. <https://www.humanitiesdataanalysis.org/>
48. Kern, E. (1994). Counter-Reformation Sanctity: The Bollandists' Vita of Blessed Hemma of Gurk. *The Journal of Ecclesiastical History*, 45(3), 412–434. doi:10.1017/S0022046900017061
49. Kienzle, B. M., & Stevens, T. E. (2014). Preaching, Heresy, and the Writing of Female Hagiography. In *Beyond Catholicism: Heresy, Mysticism, and Apocalypse in Italian Culture* (pp. 33–53). Palgrave Macmillan.
50. Knowles, M. D. (1958). Presidential Address: Great Historical Enterprises. I. The Bollandists. *Transactions of the Royal Historical Society*, 8, 147–166. <https://doi.org/10.2307/3678913>

51. Komorowski, M., Marshall, D., Saliccioli, J., & Crutain, Y. (2016). Exploratory data analysis. In *Secondary Analysis of Electronic Health Records*. (pp. 185-203). Oxford University Press. https://doi.org/10.1007/978-3-319-43742-2_15
52. Kosinski, R. (2016). The Holy Man in Late Antiquity. In A. Kaldellis (Ed.), *Holiness and Power: Constantinopolitan Holy Men and Authority in the 5th Century* (pp. 6-20). De Gruyter.
53. Kramer, R. (2020). Monasticism, Reform, and Authority in the Carolingian Era. In A. Beach & I. Cochelin (Eds.), *The Cambridge History of Medieval Monasticism in the Latin West* (pp. 432-449). Cambridge: Cambridge University Press.
54. Leavitt-Alcántara, B. (2014). Holy women and Hagiography in Colonial Spanish America. *History Compass*, 12(9), 717–728. <https://doi.org/10.1111/hic3.12191>
55. Ledzińska, A. (2015, July 6-9). *Hagiography in the Service of Counter-Reformation: De vita et gestis Divi Hyacinthi opusculum of Nicolaus Hussoviensis* [Paper presentation]. International Medieval Congress, Leeds, United Kingdom.
56. Lindow, J. (2001). Norse Mythology and the Lives of the Saints. *Scandinavian Studies*, 73(3), <https://www.jstor.org/stable/40920325>
57. Louth, A. (2004). Hagiography. In M. M. Cameron, B. Ward, & M. Whitby (Eds.), *The Cambridge history of early Christian literature* (pp. 358-361). Cambridge University Press.
58. Luria, K. P. (2018). *Territories of Grace: Cultural Change in the Seventeenth-Century Diocese of Grenoble: Cultural Change in the Seventeenth-Century Diocese of Grenoble*. University of California Press.
59. Lupton, J. R. (1996). *Afterlives of the Saints: Hagiography, Typology, and Renaissance Literature*. Stanford University Press.
60. Machielsen, J. (2012). Heretical Saints and Textual Discernment: The Polemical Origins of the *Acta Sanctorum* (1643-1940). In C. Copeland & J. Machielsen (Eds.), *Angels of Light? Sanctity and the Discernment of Spirits in the Early Modern Period* (pp. 103-141). Brill.
61. *Matplotlib - Visualization with Python*. (n.d.). Matplotlib. <https://matplotlib.org/>
62. Mattingly, W. (2022). *Introduction to Python for Digital Humanities*. python-textbook.pythonhumanities.com
63. McGillivray, B., Poibeau, T., & Fabo, P. R. (2020). Digital Humanities and Natural Language Processing: “Je t’aime. . . Moi non plus.” *HAL (Le Centre Pour La Communication Scientifique Directe)*. <https://hal.science/hal-02970302>
64. McGrath, A. E. (1998). *Historical theology: An Introduction to the History of Christian Thought*. Oxford

65. McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
66. Meeks, E., & Weingart, S. B. (2012). The Digital Humanities Contribution to Topic Modeling. In *Journal of Digital Humanities*, 2(1). <https://journalofdigitalhumanities.org/2-1/dh-contribution-to-topic-modeling/>
67. Mériaux, C. (2019). Bishops, Monks and Priests: Defining Religious Institutions by Writing and Rewriting Saints' Lives (Francia, 6th–11th centuries). In S. Herrick Kahn (Ed.), *Hagiography and the history of Latin Christendom, 500–1500* (Vol. 4, pp. 143-160). Brill.
68. Nalle, S. (1992). A Saint for All Seasons: The Cult of San Julián. In A. J. Cruz & M. E. Perry (Eds.), *Culture and Control in Counter-Reformation Spain* (pp. 25-50). University of Minnesota Press.
69. Natrella, M. (2010). *NIST/SEMATECH e-Handbook of Statistical Methods*. Retrieved August 13, 2023, from <https://itl.nist.gov/div898/handbook/>
70. *Natural Language Toolkit*. (n.d.) NLTK. <https://www.nltk.org/>
71. Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010). Evaluating topic models for digital libraries. In *Proceedings of the ACM International Conference on Digital Libraries* (pp. 215-224).
72. *NumPy: The absolute basics for beginners*. (n.d.). NumPy. https://numpy.org/devdocs/user/absolute_beginners.html
73. *Pandas 0.25 documentation*. (2019, July 18). DevDocs. <https://devdocs.io/pandas~0.25/>
74. Pfaff, S. (2013). The true citizens of the city of God: The cult of saints, the Catholic social order, and the urban Reformation in Germany. *Theory and Society*, 42(2), 189-218.
75. Pontius, S., Paulinus, a. 4. c. c., Possidius, S., Athanasius, S., Jerome, S., Ennodius, M. F., . . . Deferrari, R. J. (2001). *Early Christian biographies: Lives of St. Cyprian*. Catholic University of America Press.
76. Powell, H. (2010). 'Once Upon a Time There Was a saint ...': Re-evaluating Folklore in Anglo-Latin Hagiography. *Folk-Lore*, 121(2), 171-189.
77. *Project Jupyter*. (n.d.). Jupyter. <https://jupyter.org/>
78. *pyLDAvis*. (n.d.). PyPi. <https://pypi.org/project/pyLDAvis/>

79. Ramage, D., Rosen, E., Chuang, J., Manning, C. D., & Mcfarland, D. (2009). *Topic Modeling for the Social Sciences*. Stanford University.
80. Rapp, C. (2014). Hagiography and monastic literature between Greek East and Latin West in late antiquity. In G. Fiaccadori & G. Rota (Eds.), *Cristianità d'Occidente e cristianità d'Oriente* (secoli VI-XI) (pp. 1221-1280). Edizioni di Storia e Letteratura.
81. Rehurek, R. (n.d.). LDA Model. *Gensim*.
<https://radimrehurek.com/gensim/models/ldamodel.html>
82. Rembold, I. (2015). Rewriting the founder: Werden on the Ruhr and the uses of hagiography. *Journal of Medieval History*, 41(4), 363-387.
<https://doi.org/10.1080/03044181.2015.1084352>
83. Riccardi, F. (2022). Non c'è pace per le ossa di San Tommaso. *Alessio Porcu*.
<https://www.alessioporcu.it/articoli/non-ce-pace-per-le-ossa-di-san-tommaso/>
84. Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining* (pp. 399-408). Association for Computing Machinery. <https://doi.org/10.1145/2684822.2685324>
85. Salter, D. (2018). Animals in Late-Medieval Hagiography and Romance. In B. Boehrer, M. Hand, & B. Massumi (Eds.), *Animals, Animality, and Literature* (Cambridge Critical Concepts, pp. 73-87). Cambridge: Cambridge University Press.
86. Sanok, C. (2019). Hagiography. In J. Jahner, E. Steiner, & E. Tyler (Eds.), *Medieval Historical Writing: Britain and Ireland, 500–1500* (pp. 420-436). Cambridge: Cambridge University Press.
87. Sauer, M. (2011). Hagiographical Texts. In A. Classen (Ed.), *Handbook of Medieval Studies: Terms – Methods – Trends. 3 Volumes* (pp. 1798-1807). Berlin, New York: De Gruyter.
<https://doi.org/10.1515/9783110215588.1798>
88. Sawilla, J. (2009). *Antiquarianismus, Hagiographie und Historie im 17. Jahrhundert*. Berlin, New York: Max Niemeyer Verlag. <https://doi.org/10.1515/9783484970823>
89. Saxton, M. (2018). A Gentle Introduction to Topic Modeling Using Python. *Theological Librarianship*, 11(1), 18–27. <https://doi.org/10.31046/tl.v11i1.506>
90. Schöch, C. (2016). *Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama*. <https://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html>
91. Schofield, A. (2022). The Possibilities and Limitations of Natural Language Processing for the Humanities. In J. Schreibman, R. Siemens, & R. Guthrie (Eds.), *The Bloomsbury Handbook to the Digital Humanities* (pp. 169-182). Bloomsbury Academic.

92. Schwickerath, R. (1911). Ratio Studiorum. In *The Catholic Encyclopedia*. New York: Robert Appleton Company. <http://www.newadvent.org/cathen/12654a.html>
93. Singh, M. (2023). Public sentiment and opinion regarding the CARES Act. *Business Economics*, 58(1), 24-33. <https://doi.org/10.1057/s11369-022-00299-8>
94. Snijders, T. (2013). Rewriting hagiography in high medieval monasteries: Towards a quantitative approach. In Y. Desplenter, R. Gabriël, & J. Oosterman (Eds.), *Between stability and transformation: Textual traditions in the medieval Netherlands* (pp. 1-27). Academia Press.
95. Thomson, S. C. (2018). Telling the story: Reshaping Saint Christopher for an Anglo-Saxon Lay Audience. *Open Library of Humanities*, 4(2), 29, 1-31. <https://doi.org/10.16995/olh.306>
96. Tsamis, D. (2010). *Hagiologia tis Orthodoxis Ekklisias*. Pournaras Publications
97. Van Liere, K., Ditchfield, S., & Louthan, H. (Eds.). (2012). *Sacred History: Uses of the Christian Past in the Renaissance World*. Oxford University Press.
98. *What is Natural Language Processing?*. (n.d.). IBM. <https://www.ibm.com/topics/natural-language-processing>
99. Werner, M. (2014). Die Elisabeth-Vita des Dietrich von Apolda als Beispiel spätmittelalterlicher Hagiographie. *Vorträge Und Forschungen*, 31, 523–541. <https://doi.org/10.11588/vuf.1987.0.16163>
100. Wishart, R., & Prokopidis, P. (2017). Topic modelling experiments on Hellenistic corpora. In *Proceedings of the Workshop on Corpora in the Digital Humanities* (pp. 39-47). Association for Computing Machinery.
101. Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*, 16(Suppl 13), S8. <https://doi.org/10.1186/1471-2105-16-S13-S8>
102. Zosa, E., Pivovarov, L. (2022). Multilingual and Multimodal Topic Modelling with Pretrained Embeddings. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 4037–4048). Association for Computing Machinery.