

JERZY WÓJCIK – LUBLIN

PSALM 50 IN (EARLY) MODERN ENGLISH CATHOLIC DEVOTIONAL BOOKS: APPLYING TEXT SIMILARITY MEASURES TO THE ANALYSIS OF TEXTUAL TRADITION

1. Introduction

The history of Bible translations into English is a turbulent and complex one, marked by religious, political, and social upheavals. While the circumstances of the English Reformation and the resulting proliferation of English Bible translations in the early modern period are well-studied and understood (see, for example, Daniel 2003 or Norton 2005), the history of Catholic Bible translations into English in the 16th and 17th centuries is less well-researched and less known. For one thing, the policy of the Catholic Church going back to the Arundel's Constitutions¹ of 1409 made the translation of biblical material into English virtually illegal. Secondly, the Revolution of 1688 and the exile of the English Catholic King James II Stuart and the prevailing anti-Catholic sentiment meant that, once the Catholic

DR JERZY WÓJCIK – adiunkt w Katedrze Językoznawstwa porównawczego Angielsko-Polskiego Wydziału Nauk Humanistycznych Katolickiego Uniwersytetu Lubelskiego Jana Pawła II. ORCID: 0000-0001-5283-9017. Kontakt: jwojcik@kul.pl.

1. Arundel's Constitutions, also known as the Constitutions of Oxford, were published by the Archbishop of Canterbury, Thomas Arundel in 1409 as a reaction to Wycliffe's translation of the Bible and the spread of the Lollard movement. Wycliffe's writings and teachings were banned and any unauthorised Bible translation was deemed heretical (Hamlin 2024: 4).

translations of biblical texts did appear, they were first illegal, and subsequently largely neglected and did not receive a lot of scholarly attention.

The sole example of the Catholic translation of the Bible in the Early Modern period is the translation known as the Douay-Rheims Bible, which was the work of English Catholic exiles who set up a College in Douay in 1568. The Douay-Rheims translation was meant to be 'the reply of the Catholic Church to the popular and tendencious Protestant versions' (Isaacs 1954: 190) and an attempt to address the Catholic concerns that the Protestant versions were 'the corruptions whereby the heretics have so long lamentably deluded almost the whole of our countrymen' (Bobrick 2001: 189). The New Testament of this translation was published in Rheims in 1582, while the Old Testament, due to financial difficulties experienced by the Douay College, was printed in two volumes in Douay much later – in 1609 (volume I) and in 1610 (volume II), with the Psalter opening the second volume (Charzyńska-Wójcik in prep).

The Douay-Rheims translation of the Bible was made from the Vulgate². As explained by the translators in the *Preface* to the 1609 edition, this followed from adopting the position expressed by the Council of Trent (1545-1563) which declared the Vulgate to be better than all other Latin translations and also better than the Greek text itself. The *Preface* further expands that the ancient Hebrew and Greek texts which Protestants prefer were 'impure' and 'foully corrupted by Jews and Heretics' (Edgar 1889: 238-239). As mentioned by Charzyńska-Wójcik (in prep), the Douay-Rheims New Testament was reprinted in 1600, 1621, 1630, and 1633, while the Old Testament received only one reprint in 1635. The Douay-Rheims translation functioned as the only Catholic version of the Bible in English for more than a hundred and fifty years when it received a thorough revision prepared by Richard Challoner. His version of the New Testament appeared in 1749, while the entire Bible (Old Testament and the second edition of the New Testament) was published in 1750 (Burton 1909: 286). Challoner's revision of the Douay-Rheims Bible remained the standard version of the Bible for all English-speaking Catholics well into the 20th century (Burton 1909: 288).

As indicated above, the Douay-Rheims Bible was meant to provide English-speaking Catholics with an authoritative, Catholic version of the Bible that would be free from 'the corruptions of the heretics'. At the same time, as stated by Burton (1909: 286-7), 'frequent complaint had been made of the inconvenient bulk of the

2. Note that the Douay-Rheims translation was prepared before the authoritative Catholic version of the Vulgate, known as *Vulgata Sixto-Clementina*, appeared in 1590/1592. This meant that, as stated in the *Preface* to the Douay-Rheims Bible, the translation of the Old Testament had to be revised in accordance with *Sixto-Clementina* before appearing in print in 1609/1610.

folio and quarto volumes,' which made the Douay-Rheims Bible both awkward and expensive to use for an average believer. However, starting in 1599, when Richard Verstegen published the first English translation of the post-Tridentine *Officium Beatae Mariae Virginis* (referred to as the primer when translated into English), the English Roman Catholic community acquired the book that would become one of the most popular works of Catholic private devotion for more than two centuries (Bloom 1979: 1). During the period 1599 to 1800 more than forty editions of the English primer were published. In the course of these two centuries, the primer was 'revised many times and new translations were made of the material it contained' (Bloom 1979: 1). Primers, like their medieval predecessors (i.e. books of hours), contained a collection of psalms, prayers, and readings from the Bible. In particular, when it comes to Biblical texts, they contained 60 different psalms (Charzyńska-Wójcik 2023: 32) – and among these 60 psalms the Gradual and Penitential Psalms were always present. It is important to remember, however, that the primer was not the only book of private devotion available to English-speaking Catholics. As noted by Blom (1979: 13), apart from the primer, the manual, defined as a post-Reformation compilation of prayers, instructions, and devotions for laymen, was among other very popular prayer books used by English Catholics for the purpose of private devotions. The major difference between the primer and the manual was that the primer contained English translation of the *Officium Beatae Mariae Virginis* authorised by the Vatican, while the manual was 'best described as variations on a particular theme. The several editions [of the manual] differ in the amount of matter they contain, and in the number of prayers for each particular occasion; they also differ in the importance each edition assigns to either instructions or devotions' (Blom 1979: 113). According to Blom (1979: 114), the manual originated from the Latin prayerbook by Simon Verepaeus, entitled *Precationes Liturgicae in Dies Septem Digestae*, which was printed for the first time at Antwerp in 1574. Nevertheless, the manual was not a straightforward translation of the Latin original and the successive editions added also material from other sources, reflecting its less rigid composition. Quite importantly, some of these additions involved the text of selected psalms as indicated, for example, in the editor's preface to the 1614 Rouen edition of the manual, which states that the editor added the Seven Penitential Psalms taken out of the primer (Blom 1979: 121).

What transpires from the above is that selected biblical material, including a number of psalms, circulated in Catholic primers and manuals used by English-speaking Catholics starting from the end of the 16th century. This raises the question concerning the source of the biblical material contained in these primers and manuals, especially in view of the fact that the Douay-Rheims Old Testament, being the first English translation officially recognised by the Catholic Church,

appeared in 1609 and the second volume containing the Book of Psalms in 1610. At the same time, we know that beginning in 1599, when Richard Verstegen published the English version of the primer, an English translation of the psalms began to circulate in numerous editions of both the primer and the manual. The question, then, is which translation of the psalms appeared in these Catholic devotional books printed prior to the publication of the Douay-Rheims Bible, and whether its appearance influenced the versions of the psalms found in later primers and manuals. A related question concerns the relationship between Challoner's 1750 revision of the Douay-Rheims Bible and the biblical texts found in the primers and manuals printed during the 17th and 18th centuries.

In what follows, I will offer some tentative answers to the questions posed above. Due to space limitations, my focus will be on Psalm 50. As one of the Seven Penitential Psalms, it appears in all editions of the primer as well as in those editions of the manual that include the penitential psalms. While conclusions based on the examination of a single psalm must necessarily remain tentative, they nonetheless provide valuable insight into the textual transmission of the biblical text contained in the sources under consideration. In the ensuing discussion, I will employ a methodological approach that has been shown effective for examining such textual relationships, as developed in a number of recent studies, for example, Charzyńska-Wójcik (2021), Charzyńska-Wójcik and Wójcik (2022), Wójcik (2023, 2024), Lis and Wójcik (2023), where different versions of historical texts have been compared and their mutual dependencies studied using cosine similarity measurements. The paper is organised as follows: Section 2 provides an overview of text comparison methodology using cosine similarity between texts, Section 3 presents the results of similarity calculations, and finally Section 4 offers some conclusions.

2. Cosine similarity measurements

The cosine measure is among the most widely used methods for computing similarity between documents (Steinbach et al. 2000: 5) and is well established across a broad range of textual datamining³ applications. As noted above, its application to the study of textual traditions, particularly in the context of the dissemination of different versions and translations of the same original, is a relatively recent development. As Charzyńska-Wójcik and Wójcik (2022: 213) observe, the principal advantage of cosine similarity measurements lies in their capacity

3. For an overview of text data-mining techniques see, for example, Manning et al. (2008), Han et al. (2012), and Aggarwal (2015).

to analyse large datasets (the number of text versions that can be compared is theoretically unlimited) and, more importantly, to express the results in precise numerical terms, thereby eliminating the scope for subjective judgement. This represents a significant improvement over traditional philological approaches to comparing different text versions, which have too often relied on inherently subjective and impressionistic labels. For instance, researchers examining the same biblical translation might variously describe it as a *revision*, a *deep revision*, or *practically a new translation*. Likewise, some texts are described by certain scholars as identical, while others regard them as the same text with only minor modifications (CharzyńskaWójcik and Wójcik 2022: 212).

Performing cosine similarity measurements⁴ involves, first of all, representing the text in a way which makes the computation of similarity possible. Typically, a document is viewed as a collection of words (also referred to as *terms*) accompanied by their frequency (the number of their occurrences) in a text. This is known in the literature as the bag-of-words model, in which the exact ordering of the terms in a document is ignored but the number of occurrences of each term is material (Manning et al. 2008: 107 and Aggarwal 2015: 430). The resulting representation is referred to as the vector space representation of text⁵. The vector space representation of a document is a multidimensional vector that contains a frequency associated with each word (dimension) in the document (Aggarwal 2015: 430). The overall dimensionality of the collection of documents is equal to the number of distinct words in the documents in question. The words present in the collection of documents that are absent from a given document are assigned a frequency of 0. In effect, the set of documents in a collection may be viewed as a set of vectors in a vector space.

As indicated above, an important feature of the bag-of-words model representation is that word-order information is not preserved. This is potentially problematic as the aim of the analysis performed here is to discover a potentially complex set of mutual textual dependencies and disregarding word-order information means leaving an important aspect of a text completely out of sight. The situation can

4. A detailed presentation of the methodology involving cosine similarity measurements can be found in Wójcik (2024).

5. In text-mining, turning texts into vectors involves some text modification steps, e.g. stop-word removal or stemming, in order to strip the document of words that are deemed semantically irrelevant (Aggarwal 2015: 431). While these steps are necessary in text-mining as they reduce the computational load without impacting the semantic content of a document, neither stop-word removal, nor stemming are performed here because that would eliminate relevant linguistic features of the analysed texts. Also, function words, singular and plural forms or the choice of grammatical tenses, disregarded in typical text-mining applications, are all retained here as important aspects of textual comparison of different psalm versions.

be remedied, however, by modifying the basic unit of textual representation (i.e. the terms of a term-frequency vectors) so that word-order information is still taken into account in the analysis. This is done by utilising *n-grams*. Vinson et al. (2016: 910) define an *n-gram* as a sequence of *n* consecutive items from a given text. In the case of *n*=1, we are dealing with a 1-gram (unigram); if *n*=2, a 2-gram (or bigram) is created; if *n*=3, we have a trigram, and so on⁶. Crucially, when we take the basic textual unit to be an *n-gram* (where *n* > 1), word-order information will be preserved (Russel and Norvig 2021) since word ordering is directly encoded in an *n-gram*. All calculations undertaken for the purposes of this study were based on bigrams.

The final point to note is that working with historical data requires a systematic treatment of spelling variation characteristic of texts created before spelling standardisation. This is achieved by means of spelling normalisation, which removes irrelevant spelling differences and ensures that words (or *n-grams*) that show up in compared texts with more than one spelling are spelled consistently in one way only.

Once the texts are represented as vectors, similarity calculations can be performed. As noted by Feldman and Sanger (2007: 85), while calculating similarity between texts (represented as vectors), the similarity function is usually based on the distance between the vectors in some metric (Feldman and Sanger 2007: 85). The standard measure of quantifying the similarity between vectors is the cosine similarity⁷, which calculates the cosine of the angle between vectors (see, for example, Manning et al. 2008: 111; Han et al. 2012: 78; Aggarwal 2015: 76; Cichosz 2015: 323). Since the vectors' dimensions are expressed by non-negative numbers (recall that this follows from the fact that these dimensions correspond to word/*n-gram* frequencies that can only be non-negative), the cosine of the angle between two vectors representing texts will be expressed by a value between 0 and 1, where 1 indicates complete identity of the two texts, while 0 results from the two texts sharing no features.

6. *N-grams* are typically classified into two categories: character-based, when individual letters are used to build an *n-gram* and word-based, when words are used in the construction of an *n-gram* (Mohan et al. 2010: 2).

7. Cosine similarity is calculated using the following formula:

$$\text{similarity}(x, y) = \cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

where *x* and *y* are *n*-dimensional vectors (representing documents), θ is an angle between vectors *x* and *y*, while *A_i* and *B_i* are components (frequencies of terms) of vectors *x* and *y*.

The texts selected for analysis comprise primers and manuals published between 1599 and 1732. The list⁸ begins with the first English primer published by Verstegen in 1599 and includes primers from 1604, 1650, 1669, 1672, 1684, 1685, 1687, 1699, and 1732, yielding a total of ten primers for analysis. In addition, the list includes four manuals from 1658, 1686, 1692, and 1696. The text of Psalm 50 in these Catholic devotional publications was compared with that found in the two Douay-Reims Bible editions from 1610 and 1635 and in Richard Challoner's 1750 revision. For additional context, the analysis also includes the 1611 King James Version, enabling a comparison with a Protestant Bible. The inclusion of the King James Version will allow the verification of the claims made in the literature (e.g. Newton 1940: 218) that in Challoner's translation 'The elements of his text break up into three categories: one being the Douay version, another the King James, a third Challoner himself'. The results of the cosine similarity calculations performed on this collection of texts are presented in section 3 below.

3. Results

Table 1 presents the raw results of the cosine similarity calculations⁹. The grey cells indicate the similarity of each text with itself, which is naturally 1 in every case. The cells above and below the grey diagonal are mirror images of one another, as the similarity between text A and text B is identical to that between text B and text A.

-
8. Full bibliographic details of the analysed texts are provided in the *Primary Sources* section of the References.
 9. All calculations were performed using RStudio (RStudio Team 2023), an integrated development environment for the R programming language (R Core Team 2023). Cosine similarities were computed using the *quanteda* package (Benoit et al. 2018), hierarchical trees were generated with the base R function *hclust()*, and dendrograms were visualised using the *factoextra* package (Kassambara and Mundt 2020).

Table 1. Results of cosine similarity calculations

	Psalm_50_Primer-1732	0,88	0,863	0,858	0,532	0,909	0,882	0,882	0,882	0,904	0,9	0,904
	Psalm_50_Primer-1699	0,889	0,865	0,861	0,538	0,9	0,896	0,896	0,896	0,902	0,898	0,902
	Psalm_50_Primer-1687	0,884	0,816	0,811	0,539	0,878	0,994	0,994	0,994	0,873	0,869	0,873
	Psalm_50_Primer-1685	0,833	0,967	0,962	0,505	0,892	0,823	0,823	0,823	0,904	0,899	0,904
	Psalm_50_Primer-1684	0,857	0,877	0,872	0,526	0,949	0,873	0,873	0,873	0,958	0,954	0,958
	Psalm_50_Primer-1672	0,833	0,967	0,962	0,505	0,892	0,823	0,823	0,823	0,904	0,899	0,904
	Psalm_50_Primer-1669	0,857	0,877	0,872	0,526	0,954	0,873	0,873	0,873	0,963	0,958	0,963
	Psalm_50_Primer-1650	0,863	0,895	0,89	0,52	0,979	0,879	0,879	0,879	1	0,995	1
	Psalm_50_Primer-1604	0,859	0,89	0,886	0,52	0,974	0,875	0,875	0,875	0,995	1	0,995
	Psalm_50_Primer-1599	0,863	0,895	0,89	0,52	0,979	0,879	0,879	0,879	1	0,995	1
	Psalm_50_Manual-1696	0,889	0,822	0,817	0,538	0,884	1	1	1	0,879	0,875	0,879
	Psalm_50_Manual-1692	0,889	0,822	0,817	0,538	0,884	1	1	1	0,879	0,875	0,879
	Psalm_50_Manual-1686	0,889	0,822	0,817	0,538	0,884	1	1	1	0,879	0,875	0,879
	Psalm_50_Manual-1658	0,861	0,893	0,888	0,511	1	0,884	0,884	0,884	0,979	0,974	0,979
	Psalm_50_KJV-1611	0,555	0,496	0,496	1	0,511	0,538	0,538	0,538	0,52	0,52	0,52
	Psalm_50_D-R-Bible-1635	0,838	0,995	1	0,496	0,888	0,817	0,817	0,817	0,89	0,886	0,89
	Psalm_50_D-R-Bible-1610	0,843	1	0,995	0,496	0,893	0,822	0,822	0,822	0,895	0,89	0,895
	Psalm_50_Challoner's-Bible-1750	1	0,843	0,838	0,555	0,861	0,889	0,889	0,889	0,863	0,859	0,863
Psalm_50_Challoner's-Bible-1750												
Psalm_50_D-R-Bible-1610												
Psalm_50_D-R-Bible-1635												
Psalm_50_KJV-1611												
Psalm_50_Manual-1658												
Psalm_50_Manual-1686												
Psalm_50_Manual-1692												
Psalm_50_Manual-1696												
Psalm_50_Primer-1599												
Psalm_50_Primer-1604												
Psalm_50_Primer-1650												

Psalm_50_Pri- mer-1732	0,907	0,867	0,905	0,867	0,876	0,973	1
Psalm_50_Pri- mer-1699	0,917	0,876	0,917	0,876	0,89	1	0,973
Psalm_50_Pri- mer-1687	0,867	0,817	0,867	0,817	1	0,89	0,876
Psalm_50_Pri- mer-1685	0,888	1	0,888	1	0,817	0,876	0,867
Psalm_50_Pri- mer-1684	0,995	0,888	1	0,888	0,867	0,917	0,905
Psalm_50_Pri- mer-1672	0,888	1	0,888	1	0,817	0,876	0,867
Psalm_50_Pri- mer-1669	1	0,888	0,995	0,888	0,867	0,917	0,907
Psalm_50_Pri- mer-1650	0,963	0,904	0,958	0,904	0,873	0,902	0,904
Psalm_50_Pri- mer-1604	0,958	0,899	0,954	0,899	0,869	0,898	0,9
Psalm_50_Pri- mer-1599	0,963	0,904	0,958	0,904	0,873	0,902	0,904
Psalm_50_Man- ual-1696	0,873	0,823	0,873	0,823	0,994	0,896	0,882
Psalm_50_Man- ual-1692	0,873	0,823	0,873	0,823	0,994	0,896	0,882
Psalm_50_Man- ual-1686	0,873	0,823	0,873	0,823	0,994	0,896	0,882
Psalm_50_Man- ual-1658	0,954	0,892	0,949	0,892	0,878	0,9	0,909
Psalm_50_KJV-1611	0,526	0,505	0,526	0,505	0,539	0,538	0,532
Psalm_50_D-R- -Bible-1635	0,872	0,962	0,872	0,962	0,811	0,861	0,858
Psalm_50_D-R- -Bible-1610	0,877	0,967	0,877	0,967	0,816	0,865	0,863
Psalm_50_ Challoner's-Bible-1750	0,857	0,833	0,857	0,833	0,884	0,889	0,88
Psalm_50_Pri- mer-1669							
Psalm_50_Pri- mer-1672							
Psalm_50_Pri- mer-1684							
Psalm_50_Pri- mer-1685							
Psalm_50_Pri- mer-1687							
Psalm_50_Pri- mer-1699							
Psalm_50_Pri- mer-1732							

Examination of the data in Table 1 allows one to make several interesting observations. First, the text most distinct from all others in the collection is the King James Version from 1611, with similarity scores between it and the remaining texts in the set ranging from 0,496 to 0,555. This result is unsurprising, as the text of the Protestant version based on a different source would be expected to differ markedly from that found in Catholic publications. The second point to note is the level of similarity between the 1610 DouayRheims Bible and its 1635 reprint, which stands at 0.995 – a result that is somewhat unexpected. Since the 1635 edition is a reprint, complete identity would normally be anticipated. Nevertheless, the high score of 0.995 indicates a near complete correspondence between the two. The sole difference lies in the presence of *words* (in 50.6) in the 1610 edition and *word* in the 1635 edition. The Latin text reads *sermonibus* at this point, which is a plural form, suggesting that the variation is likely a printer's error.

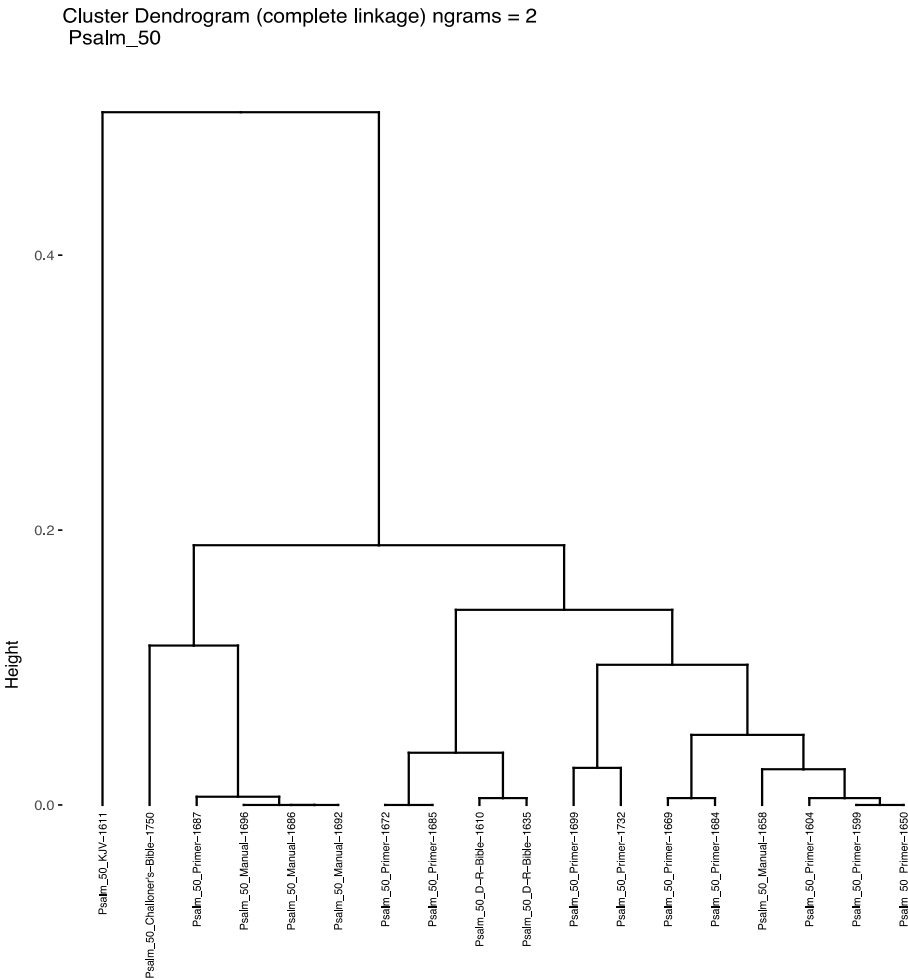
The next observation concerns the similarity between the DouayRheims text and the versions found in the examined primers and manuals. It ranges from a low of 0,811 (in the case of 1687 primer) to a high of 0,967 (in the case of 1672 and 1685 primers, which print the same text). While the range of similarity scores (0,811 – 0,967) indicates an overall close resemblance between the text of Psalm 50 in the primers and manuals and that of the DouayRheims Bible, the number of texts involved makes it difficult to observe and analyse finer distinctions. This is where cluster analysis¹⁰ (or clustering) may prove useful as it partitions a set of data objects into subsets (i.e. clusters) such that objects within a cluster exhibit high internal similarity¹¹, while being dissimilar from those in other clusters. Similarities and dissimilarities are assessed based on attribute values describing the objects and often involve similarity measures (Han *et al.* 2012: 443). The clustering method applied here is agglomerative hierarchical clustering (Moisl 2015: 213). Hierarchical clustering groups data objects – in this case, the text of Psalm 50 from various sources – into a hierarchy of clusters and the agglomerative type uses a bottom-up strategy, initially treating each data object as a cluster of its own and iteratively merging the most similar clusters into larger ones until a single cluster containing all objects is formed. Various algorithms were proposed in the literature that can be used to perform this merging process; in

10. For a detailed discussion of different clustering methods see Han *et al.* (2012: Ch. 10 and Ch.11), Moisl (2015: 156ff). Comprehensive accounts of cluster analysis can be found in numerous handbooks on the subject, e.g. Duda *et al.* (2001) or Everitt *et al.* (2001).

11. The internal similarity of a cluster is typically calculated based on the distance between data points. When using cosine similarity, the relationship between cosine similarity and cosine distance is expressed by the formula: *cosine distance* = 1 – *cosine similarity*. This means that identical texts (cosine similarity = 1) will have a cosine distance of 0.

this study, the complete linkage¹² method was employed. The output of hierarchical clustering is a tree-like structure known as a *dendrogram*, which visually represents the hierarchical relationships among the clustered texts. The results are presented in Figure 1.

Figure 1. Cluster Dendrogram based on cosine similarity measures for the text of Psalm 50



12. A thorough discussion of the various cluster agglomeration methods (also known as linkage methods) used at successive steps in a tree-building process can be found in, for example, Moisl (2015: 208ff).

In the hierarchical dendrogram shown in Figure 1, the distance (or dissimilarity) between texts (or between clusters of texts) is represented by the length of the vertical lines connecting them. In other words, the height at which two elements are joined in the dendrogram reflects their degree of difference. A connection at height 0, for example, indicates that the grouped texts are identical, with a *dissimilarity* score of 0.

As noted above, dendrogram visualisation enables the identification of finer relationships among the analysed texts. To begin the discussion of what can be inferred from the dendrogram in Figure 1, it is useful to repeat the three questions posed above. What was the source of the biblical material in Catholic devotional books? Which translation of the psalms appeared in Catholic devotional works published prior to the printing of the DouayRheims Bible? And to what extent does Challoner's revision diverge from the DouayRheims Bible, and how does it relate to the biblical texts found in the primers and manuals?

In relation to the first two questions, which are interconnected, an inspection of the relations depicted in Figure 1 shows clearly that all the analysed Catholic texts belong to the same tradition based on the Douay-Rheims translation. The only outlier that is most unlike the rest of the texts is the Protestant KJV version, a conclusion that we were able to infer quite easily from the raw data. All the remaining texts, although they do display some differences that I will return to below, show high levels of similarity. The horizontal level at which all the Catholic texts are joined is slightly below 0,2. This means that even the most dissimilar texts (among the Catholic ones) display a high level of similarity, i.e. 0,811, which is the score between primer from 1687 and the Douay-Rheims Bible. Although the two texts are most dissimilar, the primer from 1687 clearly represents the Douay-Rheims translation. To quote just two examples of differences between these texts which nicely illustrate the general closeness between these versions: 1687 *because I know my iniquity* vs. 1635 *because I do know my iniquity*, or 1687 *and my sin is always against me* vs. 1635 *and my sin is before me always*.

A more intriguing question is the one regarding the biblical text version which appeared in Catholic books of devotion prior to the publication of the Douay-Rheims Bible. Recall that the first Catholic English primer was published by Richard Verstegen in 1599 – eleven years before the volume of the Douay-Rheims Bible containing the Book of Psalms appeared. The scholarly views on the source of biblical translations in Verstegen's primer are ambiguous. In a recent publication (Hamlin 2024: 356), it is claimed that the author of the English psalm translation in the first Catholic primer from 1599 is Richard Verstegen himself. The claim is repeated several pages later (Hamlin 2024: 360), where the author states that 'Verstegen's Psalm translations were published along with other paraphrases of devotional texts' clearly attributing the translated psalm texts to Verstegen. This

is in stark contrast to what is claimed by Blom (1979: 16), where it is maintained that ‘the primer of 1599 does contain the Douay translation, albeit with some insignificant variants’. As was noted above, the translators of the Douay-Rheims Bible published the New Testament in 1582, while the Old Testament was published only in 1609/1610 due to financial difficulties, although the translation was ready. As further explained by Blom (1979: 16), ‘[t]he editor of the primer must have had at his disposal Martin’s [i.e. the translator of Douay-Rheims Bible] manuscript translation and fortunately the biography of the man responsible for the primer edition yields abundant evidence of his connection with Gregory Martin and the people who assisted him with the Bible translation’. The analysis of the textual dependencies between versions of Psalm 50 depicted in the dendrogram in Figure 1 makes it possible to establish that Verstegen’s version is almost identical to the one contained in the Douay-Rheims translation, thus corroborating Blom’s views on the matter.

The dendrogram further shows the first Catholic English primer from 1599 is identical or almost identical to the primers from 1650 and 1604, and the manual from 1658. This near perfect equivalence between Verstegen’s primer and the other texts in this group can be explained if we consider their place of publication. All were published in Antwerp. The 1604 edition was issued by Arnold Conings, the printer of Verstegen’s original 1599 primer and it is in fact a reprint of the first English Catholic primer, while the 1650 and 1658 editions – although produced by different Antwerp printers – clearly draw on the same text as that found in Verstegen’s 1599 publication. The two almost identical¹³ primers from 1684 and 1669 were published in Rouen in France and their level of similarity to the previous four publications shows clearly that they essentially contain the same text of Verstegen’s publication. The same can be said about primers 1699 (place of publication unknown – likely London as Catholic publications issued in England often lacked publication details) and 1732 (with a London imprint). Note further that the text of the Douay-Rheims Bible in its two editions is repeated almost without change in two primers from 1672 and 1685 forming together a coherent group of twelve publications of Psalm 50 that contain essentially the same text. It is quite clear, then, that the first Catholic primer by Richard Verstegen (and a succession of primers and manuals that followed it) is based (in its biblical material) on the translation associated with the Douay-Rheims Bible.

The final point to be discussed concerns the nature of Challoner’s revision of the Douay-Rheims Bible. As indicated above, it is claimed in the literature that

13. The difference in similarity between them is just 0,005. This result is the same as that observed between the two editions of Douay-Rheims Bible, which as was noted above followed from a single word difference – likely a printer’s mistake.

Challoner's revision drew from the original Douay-Rheims version, King James Version from 1611, and contained original Challoner's elements (Newton 1940: 218). Note as well that the primary motivation behind Challoner's work was to provide an updated version of the Douay-Rheims Bible since 'the only Catholic version of the Sacred Scriptures in English was 150 years old, difficult to procure, and hard to understand by reason of its antique style' (Burton 1909: 271). What follows from the inspection of the dendrogram is that Challoner's version exhibits relatively low similarity scores (0,843/0,838) when compared with the original DouayRheims text as printed in the two editions of the Bible (1610 and 1635). These are among the lowest similarity scores between Challoner's version and any of the other texts analysed (with the exception of the King James Version). Notably, the lowest score occurs in comparison with the primers from 1672 and 1685, which are themselves closely aligned with the DouayRheims version. This suggests that while the primers and manuals generally adhered to the DouayRheims text, Challoner's version represents a genuine revision and, as expected, diverges most significantly from the DouayRheims original¹⁴. Challoner's revision is, in fact, most similar to a group of identical texts: manuals from 1686, 1692, and 1696, as well as a primer from 1687. The identical nature of the texts in this group can be explained by considering their place of publication. The manual from 1686 and the primer from 1687 were both printed in London by the prominent Catholic printer Henry Hills, during the reign of the Catholic monarch James II. Although the two manuals (from 1692 and 1696) do not disclose a printer, one of them (i.e. the one from 1692) shows Paris as its place of publication. It is hardly surprising that a version identical to the one printed in London under a Catholic king would appear in Paris in 1692, as the area became the seat of James II's exiled court following his departure from England in 1688. It is clear that the printers of the 1692 Paris edition and the 1696 edition (whose place of publication is unknown) drew directly on Henry Hills's London¹⁵ publication. When it comes to the alleged similarity of Challoner's revision to the King James Bible, it is interesting to note

-
14. There exists an intriguing possibility that Challoner's revision was influenced by a little-known and largely overlooked English translation of the Psalms, produced at Saint-Germain-en-Laye, where King James II established his exiled court following the events of 1688. This translation was published anonymously in 1700 under the title *The Psalms of David, translated from the Vulgat*. For a comprehensive discussion of this version, see Charzyńska-Wójcik (2023). The relationship between this translation and the texts within the Douay-Rheims tradition remains an open and promising subject for future research.
 15. Henry Hills fled England after an anti-Catholic mob attacked his printing house in Blackfriars in December 1688. He took refuge in St. Omer, France, where he died shortly thereafter in 1689, although some Victorian-era sources claim that he continued his printing activities in London in the early 18th century (Caswell 2016: 72).

that the similarity scores between KJV and the remaining texts in the dataset, although generally low (as expected of a distinct textual tradition) seem to reflect that affinity. In particular, the text that KJV is most similar to (or perhaps least dissimilar to) is Challoner's revision, at 0,555.

4. Conclusion

The aim of this paper was to analyse the textual relationships between versions of Psalm 50 found in selected primers and manuals printed between 1599 (the first edition of Verstegen's primer) and 1732 (the latest of the analysed texts), and to compare these texts with the authorised Catholic translation of the Bible, the DouayRheims editions of 1610 and 1635, as well as Challoner's 1750 revision. This study demonstrates the usefulness of a methodological approach based on cosine similarity measures, enabling objective comparison across large sets of historical data. The approach is particularly well suited to the analysis of texts that are inherently similar (translation versions of the same original, for example) where differences are often subtle and difficult to assess by means of traditional methods.

Several observations emerge from the analysis. First, the data demonstrate beyond any doubt that Richard Verstegen's first edition of the English primer (1599), contrary to even very recent claims, such as Hamlin (2024), contains the text of the Douay-Rheims translation. Second, the differences in the text of Psalm 50 found in the primers and manuals, although not pronounced, as they all belong to the Douay-Rheims tradition, follow a clear pattern in which the place of publication determines the specific version used in each edition. Finally, Challoner's 1750 revision shows a striking affinity with versions published in France after 1688 that reproduce the text of London editions associated with the printer Henry Hills. It should be noted, however, that these conclusions are based on the analysis of a single penitential psalm and that further research on the remaining biblical material contained in primers and manuals is needed to confirm the observations presented here.

References

Primary sources in chronological order

1599 = *The primer, or, Office of the Blessed Virgin Marie in Latin and English, according to the reformed Latin, and vvith lyke graces priuileged*. Bibliographic name/number: STC (2nd ed.) / 16094. Antwerp: By Arnold Conings, 1599.

- 1604 = *The primer, or Office of the Blessed Virgin Marie in Latin and English: according to the reformed Latin, and vvith lyke graces priuileged*. Bibliographic name/number: STC (2nd ed.) / 16095. Antwerp: by Arnold Conings, 1604.
- 1610 = *The holie Bible faithfully translated into English, out of the authentical Latin. Diligently conferred with the Hebrew, Greeke, and other editions in diuers languages. With arguments of the bookes, and chapters: annotations. tables: and other helps ... By the English College of Doway*. Bibliographic name/number: Darlow & Moule (Rev. 1968), 300; STC (2nd ed.) / 2207. Douai: By Laurence Kellam, at the signe of the holie Lambe, 1609-1610.
- 1611 = *The Holy Bible conteyning the Old Testament, and the New: newly translated out of the originall tongues: & with the former translations diligently compared and reuised, by his Maiesties speciall co[m]mandement. Appointed to be read in churches*. Bibliographic name/number: Darlow & Moule (Rev. 1968), 309; STC (2nd ed.) / 2216. London: By Robert Barker, printer to the Kings most excellent Maiestie, 1611.
- 1635 = *The Holy Bible faithfully translated into English out of the authentical Latin diligently conferred with the Hebrew, Greek, & other editions in diuers languages: with arguments of the bookes and chapters, annotations, tables, & other helps, for better vnderstanding of the text: for discouerie of corruptions in some late translations, and for clearing some controuersies in religion / by the English Colledge of Dovvay*. Bibliographic name/number: STC (2nd ed.) / 2321. 2 v. Rouen: Printed by Iohn Covstvier, 1635.
- 1650 = *The primer, or, Office of the Blessed Virgin Marie, in Latin and English. According to the reformed Latin and with lyke graces priuileged*. Bibliographic name/number: Wing (2nd ed.) / P3464. Antwerp: By the widow of John Cnobbaert, for James Thompson, 1650.
- 1658 = *A Manual of prayers and litanies: distributed according to the days of the week. With other excellent devotions, fitted for all persons and occasions*. Bibliographic name/number: Wing (2nd ed.) / M544G. Antwerp: Printed by Michael Cnobbart, 1658.
- 1669 = *The primer more ample and in a new order containing the three offices of the B. Virgin Mary in Latin and English, and al[l] offices and devotions which were in former primers*. Bibliographic name/number: Wing / P3461. Rouen: Printed by David Mavrry, 1669.
- 1672 = *The primer or, Office of the Blessed Virgin Mary in English: exactly revised, and the new hymnes and prayers added, according to the Reformation of Pope Urban. 8*. Bibliographic name/number: Wing (2nd ed.) / P3466. Saint-Omer: [s.n.], 1672.
- 1684 = *The primer more ample, and in a new order containing the three offices of the B. Virgin Mary, in Latin and English, and al[l] offices and dévotions, which*

- were in former primers. In this last edition the hymne are in a better verse, and six offices newly added, I. Of the Holy Trinity II. Of the B. Sacrament. III. Of the holy name of Iesus, with a lotany. IV. Of the Immaculate Conception of our B. Lady, with a letany. V, of the angel-guardian. VI. Of S. Joseph. And sundry sweet devotions, and instructions taken out of the Holy Scripture for to live a devout Christian life. A large and short examen of conscience. To the calendar are annexed many English and Irish saints. Bibliographic name/number: Wing (2nd ed.) / P3462. Rouen: by Nicolas le Tourneur, ruë S. Lovis-à-vis la porte du Palais, au tourneur. VVith permission, 1684.
- 1685 = *The primer or, office of the Blessed Virgin Mary in English: exactly revised, and the new hymns and prayers added, according to the Reformation of Pope Urbans*. Bibliographic name/number: Wing (2nd ed.) / P3467. Antwerp: for T.D, 1685.
- 1686 = *A Manual of prayers and other Christian devotions*. Bibliographic name/number: Wing / M544O. Printed by Henry Hills Printer to the King's Most Excellent Majesty for his houshold and chapel, 1686.
- 1687 = *The office of the B.V. Mary in English. To which is added the vespers, or even-song, in Latin and English, as it is sung in the Catholic Church upon all Sundays and principal holy-days throughout the whole year. With the compline, rosary, hymn and prayers that are sung at the benediction of the b. sacrament. The prayers for the King, Queen, &c. The ordinary of the Holy Mass; the sequence, dies iræ, dies illa, that is sung at solemn Mass for the dead, and the libera that is sung after Mass for the dead: all in Latin and English. Together with several other devout prayers in English*. Bibliographic name/number: Wing (2nd ed.) / O149. London: printed by Henry Hills, printer to the King's most Excellent Majesty for his houshold and chappel; and are to be sold at his printing-house on the Ditch side in Black Fryars, 1687.
- 1692 = *Manual of prayers and other Christian devotions*. Bibliographic name/number: Wing / M544P. Paris: [s.n.], 1692.
- 1696 = *A manuel of devout prayers and other Christian devotions fitted for all persons and occasions*. Bibliographic name/number: Wing / M544B. s.l.: s.n., 1696.
- 1699 = *The primer, or, office of the Blessed Virgin Mary*. Bibliographic name/number: Wing (2nd ed.) / P3467A. s.l.: s.n., 1699.
- 1732 = *The primer; or, office of the B. Virgin Mary, with a new and approv'd version of the Church-Hymns. To which are added the remaining hymns of the Roman breviary*. Bibliographic name/number: ESTC Number T125400. [London?]: Printed for T. Meighan in the year, M.DCC.XXXII. [1732].
- 1750 = *The Holy Bible translated from the Latin Vulgat: ... first published by the English College at Doway, Anno 1609. Newly revised, and corrected, according*

to the Clementin edition of the Scriptures. With annotations ... Bibliographic name/number: ESTC Number T107533. [Dublin?]: Printed in the year, 1750.

Secondary sources

- Aggarwal, C. C. 2015. *Data Mining: The Textbook*. Springer Publishing Company.
- Benoit, K., K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo. 2018. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. <https://doi.org/10.21105/joss.00774>.
- Blom, J. M. 1979. *The Post-Tridentine English Primer*. Krips Repro B.V.
- Bobrick, B. 2001. *The Making of the English Bible*. London: Weidenfeld and Nicolson.
- Burton, E. H. 1909. *The Life and Times of Bishop Challoner, 1691-1781*. London and New York: Longmans, Green.
- Caswell, V. 2016. "Meeting Henry Hills: Printer to the King's Most Excellent Majesty." *Elements* 12 (2). <https://doi.org/10.6017/eurj.v12i2.9435>.
- Charzyńska-Wójcik, M. 2021. "Familiarity and Favour: Towards Assessing Psalm Translations." *Linguistica Silesiana* 42: 43–77. <https://doi.org/10.24425/lin-si.2021.137231>.
- Charzyńska-Wójcik, M. 2023. "Too Good Is Bad: On a Forgotten Translation of the Book of Psalms from 1700." *Roczniki Humanistyczne* 71 (6sp): 15–43. <https://doi.org/10.18290/rh237106.2s>.
- Charzyńska-Wójcik, M. in preparation. *Seven Penitential Psalms in a Manuscript of the Exilic Poor Clares: Translation, Tradition, and Transmission*.
- Charzyńska-Wójcik, M. and J. Wójcik. 2022. "Similarity Measurements in Tracing Textual Affinities. A Study of Psalm 129 in 16th-Century Devotional Manuals", *Token* 14, 191–220.
- Cichosz, P. 2015. *Data Mining Algorithms: Explained Using R*. John Wiley & Sons.
- Duda, R., P. Hart, and D. Stork. 2001. *Pattern Classification*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Daniell, D. 2003. *The Bible in English: Its History and Influence*. New Haven, CT: Yale University Press.
- Edgar, A. 1889. *Bibles of England*. London.
- Everitt, B., S. Landau, and M. Leese. 2001. *Cluster Analysis*. 5th ed. Hoboken, NJ: Wiley.
- Feldman, R., and J. Sanger. 2007. *The Text Mining Handbook*. Cambridge University Press.
- Han, J., M. Kamber, and J. Pei. 2012. *Data Mining: Concepts and Techniques*. 3rd ed. Waltham, MA: Morgan Kaufmann Publishers.

- Hamlin, H., ed. 2024. *The Psalms in English 1530–1633*. Tudor and Stuart Translations 19. MHRA. <https://doi.org/10.2307/jj.19341356.8>.
- Isaacs, J. 1954. “The Sixteenth-Century English Versions.” In *The Bible in Its Ancient and English Versions*, edited by H. W. Robinson, 146–95. Oxford: Clarendon Press.
- Kassambara, A., and F. Mundt. 2020. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>.
- Lis, K., and J. Wójcik. 2023. “French and English Texts of the Laws of Oléron: Assessing Proximity between Copies and Editions by Means of Cosine Similarity.” *Bulletin of the John Rylands Library* 99 (2): 103–26. <https://manchester-universitypress.co.uk/9781526178503/>.
- Manning, C., P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Mohan, A., I. M. Baggili, and M. K. Rogers. 2010. “Authorship Attribution of SMS Messages Using an N-Grams Approach.” *Proceedings of CERIAS Tech Report 2010-11*, Center for Education and Research in Information Assurance and Security, Purdue University, USA: 1–12.
- Moisl, H. 2015. *Cluster Analysis for Corpus Linguistics*. Berlin: Mouton de Gruyter.
- Newton, W., L. 1940. “Bishop Challoner’s Revision of the Old Testament.” *The Catholic Biblical Quarterly*, 2 (3): 215–228.
- Norton, D. 2005. *A Textual History of the King James Bible*. Cambridge: Cambridge University Press.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- RStudio Team. 2023. *RStudio: Integrated Development Environment for R*. Version 2023.06. Boston: Posit Software, PBC. <https://www.posit.co/>.
- Russell, S., and P. Norvig. 2021. *Artificial Intelligence: A Modern Approach*. Global ed. Pearson Higher Education.
- Steinbach, M., G. Karypis, and V. Kumar. 2000. “A Comparison of Document Clustering Techniques.” University of Minnesota Digital Conservancy. <https://hdl.handle.net/11299/215421>.
- Vinson, D. W., J. K. Davis, S. Sindi, and R. Dale. 2016. “Efficient N-Gram Analysis in R with Cmscu.” *Behavior Research Methods* 48 (3): 909–21. <https://doi.org/10.3758/s13428-016-0766-5>.
- Wójcik, J. 2023. “Cluster Analysis in Tracing Textual Dependencies – A Case of Psalm 6 in 16th-Century English Devotional Manuals.” *Digital Humanities Quarterly* 17 (3): 1–16. <http://www.digitalhumanities.org/dhq/vol/17/3/000694/000694.html>.

Wójcik, J. 2024. "On Measuring Psalm Similarity: A Case for Word-Level N-Grams." *Roczniki Humanistyczne* 72 (5), special issue PSALMICA 1: 185-208. <https://doi.org/10.18290/rh24725.2>.

Summary

The paper investigates textual affinities among English versions of Psalm 50 found in (Early) Modern Catholic materials. It compares eighteen renditions of Psalm 50 appearing in primers and manuals of devotion with those in the official Catholic Douay-Rheims Bible (1609-10 and 1635) and its revision prepared by Richard Challoner in 1750. Psalm 50 from the King James Bible (1611) is also added to the comparison on account of the claims concerning Challoner's reliance on this version in preparing his revision. Using cosine similarity measures to assess textual similarity, the study uncovers patterns of textual dependence in Catholic publications from the (Early) Modern English period.

Keywords: Douay-Rheims Bible Psalms, Catholic devotional books, text similarity, cosine similarity, digital humanities.

Streszczenie

Artykuł analizuje pokrewieństwa tekstowe pomiędzy angielskimi wersjami Psalmu 50 występującymi we (wczesno) nowoangielskich publikacjach katolickich, porównując osiemnaście wersji Psalmu 50 pochodzących z modlitewników z tekstami zawartymi w oficjalnym katolickim przekładzie Biblii Douay-Rheims (1609-1610 i 1635) oraz w jego rewizji dokonanej przez Richarda Challonera w 1750 roku. Do porównania włączono również Psalm 50 z Biblii Króla Jakuba (1611), z uwagi na pojawiające się w literaturze sugestie dotyczące korzystania z tej wersji przez Challonera podczas przygotowywania rewizji. Wykorzystując miarę podobieństwa cosinusowego, badanie ujawnia zależności tekstowe w katolickich publikacjach z okresu (wczesno) nowoangielskiego.

Słowa kluczowe: Psalmi Biblii Douay-Rheims, katolickie księgi modlitewne, podobieństwo tekstów, podobieństwo cosinusowe, humanistyka cyfrowa.