



Open Library of Humanities



Part of the Ubiquity
Partner Network



Research

How to Cite: Edlich-Muth, Christian, and Miriam Edlich-Muth. 2019. "A Computational Approach to Source Adaptation in Thomas Malory's *Morte Darthur*." *Digital Medievalist* 12(1): 5, pp. 1–26. DOI: <https://doi.org/10.16995/dm.86>

Published: 30 December 2019

Peer Review:

This is a peer-reviewed article in *Digital Medievalist*, a journal published by the Open Library of Humanities.

Copyright:

© 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Digital Medievalist is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

RESEARCH

A Computational Approach to Source Adaptation in Thomas Malory's *Morte Darthur*

Christian Edlich-Muth¹ and Miriam Edlich-Muth²

¹ Max Planck Institute for Biology of Ageing, Cologne, DE

² Heinrich Heine University Düsseldorf, DE

Corresponding author: Miriam Edlich-Muth (muth.miriam@gmail.com)

The many source texts that late medieval adaptors like Thomas Malory worked with constitute a potential wealth of information concerning the genesis of the linguistic and stylistic features shaping their works. As a result, their texts provide a useful framework for further developing and testing methods of stylometric analysis in the context of adaptation as a collaborative form of authorship. Our interdisciplinary team has undertaken a stylometric analysis of the eight different sections of Thomas Malory's *Morte Darthur*, in order to identify differences between these sections and how they correspond to the language of the Old French and Middle English sources that Malory is known to have worked with in the different sections of his work. Our findings provide a basis for addressing unresolved scholarly questions concerning Malory's *Morte*, such as the nature of the source used for his "Tale of Sir Gareth" and whether Malory himself was responsible for the differences between the two surviving versions of his "Roman War" episode in Book II of the *Morte*. They further shed light on ongoing controversies concerning the overall textual unity of the *Morte* and the process by which Malory created the different sections of his work—an issue that lies at the heart of broader debates concerning where Malory and other late medieval adaptors are to be situated on the continuum between "faithful translator" and "original author."

Keywords: Medieval; Source Adaptation; Digital Stylometry; Malory; Authorial Originality

1. Introduction

§1 Recent years have seen a widespread application of stylometric analysis for the purpose of authorial attribution in medieval and early modern texts (Grzybek 2014; Kestemont 2012; Binongo 2003; Van Dalen-Oskam and van Zundert 2007; Love 2002;

Souvay and Pierrel 2009; Koppel et al. 2009; Juola 2008). However, authorship can also be a question of degree. When texts arise out of processes of rewriting and adaptation, the dividing line between what is “original” authorship and what is primarily translation or adaptation can become blurred. Several major literary works from the late medieval period, including Thomas Malory's *Morte Darthur* and Geoffrey Chaucer's *Canterbury Tales* are largely adaptations of earlier texts and therefore provide a useful framework for applying quantitative stylistic analysis to the process of source adaptation (Kelly 1978, 291; Kestemont 2012; Kestemont 2015). In particular, the linguistically varied source texts late medieval adaptors worked with constitute a potential wealth of information concerning the genesis of the linguistic and stylistic features shaping the literary texts at hand (Mairey 2006; Norris 2008; Davidson 2008). In this respect, the re-writing and adaptation of source texts allows for greater insights into the act of literary creation than can be afforded by the black box of “original” literary works.

§2 Applying stylometry to medieval texts presents a specific set of challenges and opportunities. On the one hand, the pre-processing required for digitally analyzing these texts is laborious because the corpus needs to be normalized in order to remedy the spelling variation that is characteristic of the period (Piotrowski 2012; Kestemont et al. 2010). On the other hand, many late medieval texts are the product of idiosyncratic processes of source adaptation, as well as collaborations between authors and scribal amanuenses, which have been of particular interest to scholars in the wake of the New Philology (Thompson 2016, 14; Benedict 2004, viiiix; Kestemont et al. 2015). These more complex forms of authorship present a promising field for further developing and testing methods of stylometric analysis (Reynolds et al. 2012). Mike Kestemont et al.'s study of the twelfth-century Latin works of Hildegard von Bingen and her amanuensis Guibert of Gembloux demonstrates how the stylometric techniques employed for authorial attribution studies can be adapted to the task of distinguishing between different authorial styles at work in collaborative medieval texts (Kestemont et al. 2015). Their methodology and results are also highly significant for the study of source adaptation, where the stylistic features of the source texts intersect with the stylistic features of the adaptation.

§3 In the following we will present the results of a basic stylometric analysis of the language used in the different sections of Thomas Malory's fifteenth-century Arthurian collection *Morte Darthur* (*Morte*) and discuss how our findings contribute to a number of scholarly debates concerning the *Morte*.

§4 As an interlinked collection of eight “tales” or “books” based on a range of source texts, the *Morte* exemplifies the variety and scope of late medieval adaptations (Norris 2008, 164; Archibald 2013, 176). The current state of the art regarding which sources Malory used in which sections of his work has been defined by Ralph Norris' seminal monograph on *Malory's Library* (Norris 2008), which devotes a chapter to each section, or “book”, of Malory's Arthuriad and discusses in detail which sections of which sources Malory uses where. Norris identifies eight major source texts, two in Middle English and six in Old French, which were used for different sections of the *Morte* (Norris 2008, 164, see **Table 1**). He synthesizes and expands upon seven decades of preceding research concerning Malory's sources (Simko 1957; Bennett

Table 1: Sections of the *Morte Darthur* listed alongside their major sources as identified by Norris (2008, 163–4). For the purposes of this study, the *Morte Darthur* has been divided into eight books in accordance with the findings of Eugène Vinaver (1947), each corresponding to a “tale”. The differing chapter subdivisions contained in William Caxton's edition of the *Morte Darthur* are given for information only. The “source language” column describes the major sources only.

| Book | Tale | Caxton Chapter | Source Language | Major Source |
|------|----------------------|----------------|-------------------------------|---|
| I | Arthur | 1–4 | Old French | Prose <i>Merlin</i> , Post-Vulgate <i>Suite du Merlin</i> |
| II | Roman War | 5 | Middle English | Alliterative <i>Morte Arthure</i> |
| III | Lancelot | 6 | Old French | Prose <i>Lancelot</i> |
| IV | Gareth | 7 | ? | ? |
| V | Tristram | 8–12 | Old French | Prose <i>Tristan</i> |
| VI | Sankgreal | 13–17 | Old French | Vulgate <i>Queste del Saint Graal</i> |
| VII | Lancelot & Guinevere | 18–19 | Middle English/ Old French | Stanzaic <i>Morte Arthur</i> / Vulgate <i>La Mort Artu</i> |
| VIII | Death of Arthur | 20–21 | Middle English | Stanzaic <i>Morte Arthur</i> |

and Oakeshott 1963; Lumiansky 1964; Batt 2002), and his findings can be regarded as resembling “a complete list of Malory’s sources” (Norris 2008, 163).

§5 While the overall question concerning which sources Malory used has largely been answered, medievalists have continued to focus on four questions that remain unresolved. In the following we will outline the state of the art concerning each of these open questions.

§6 One such question concerns the source Malory used for his Book IV, “The Tale of Sir Gareth”, which is the only tale for which neither Norris nor any other Malorians have convincingly identified a major source. Norris suggests that the source was a now-lost romance written in Old French (Norris 2008, 159), meanwhile D. Thomas Hanks Jr suggests an oral folktale source (Hanks Jr 2003, 52) and P.J.C. Field has posited that Malory was working with a lost Middle English romance (Field 1995, 255). Arnold Sanders, on the other hand, argues that there was no lost major source and that Malory’s tale represents an original reworking of the Gawain romances and the trope of the “Fair Unknown” (Sanders 2006, 34). The controversy concerning whether or not Malory worked with a major source for this section of his work and, if so, what the genre and language of that source were remains a point of particular interest to Malory scholars.

§7 A second and more fundamental unresolved question regarding the *Morte* concerns the differences between the two surviving versions of the text: that contained in the fifteenth-century Winchester manuscript and that contained in William Caxton’s 1485 print edition. The question of how these two versions are related to each other first arose following W.F. Oakeshott’s discovery of the Winchester manuscript in 1936 and Eugène Vinaver’s revelation that the manuscript version differed significantly from the Caxton print edition, which had hitherto been the only surviving version of the *Morte* (Gordon and Vinaver 1937; Norris 2006, 68). Vinaver claims that the Winchester manuscript presents the Arthurian story in eight separate sections, referred to as “tales” or “books” (Gordon and Vinaver 1937). Based on this evidence, Vinaver concludes that Malory did not intend the *Morte* to be read as one work and that his text was instead a collection of eight separate tales (see Vinaver’s edition of *The Works of Sir Thomas Malory*, 1947). This “unity” debate

remains unresolved, and, while several recent editions have combined different aspects of the Winchester and Caxton versions and present the *Morte* as one work (Field 2013; Shepherd 2003), the degree to which the tales are unified remains a point of debate (Clark 2014, 92–3).

§8 The discovery of the Winchester manuscript raised a third controversial issue concerning the genesis of the *Morte*. Vinaver draws particular attention to Book II of the *Morte*, known as the “Roman War” section, as the part of the text where the greatest differences between the Winchester and Caxton versions of the story can be found (Gordon and Vinaver 1937, 81). Following the discovery of the Winchester Manuscript, Jan Simko produced a parallel-text analysis of the version of the “Roman War” episode contained in the Winchester manuscript, that contained in the Caxton edition, and the Alliterative *Morte Arthure*, which is the major source of the “Roman War” section (Simko 1957). He concludes that differences in word order between the Caxton and the Winchester versions were dictated by functional, grammatical, rhythmic and stylistic factors (Simko 1957, vii). Simko also argues that the differences between the two versions should be attributed to Caxton’s editorial intervention (Simko 1957, ix). This position has been opposed by William Matthews, who draws on internal stylistic evidence from the Caxton and Winchester versions of the “Roman War” episode to argue that Malory himself made the revisions to his text that appear in the Caxton edition (Matthews 1997; Moorman 1995, 25). There is currently no scholarly consensus on who was responsible for the differences between the two surviving versions of the “Roman War” episode (Moorman 1995, 24–5).

§9 One of the reasons that scholars have been so interested in the “Roman War” episode is because of the possibility that the two different versions represent different stages of Malory’s revision process (Matthews 1997; Moorman 1995, 28–9). If it were possible to establish that Malory was himself responsible for both versions of this section, researchers would gain new insights into Malory’s creative process as a rewriter.

§10 This issue feeds into the fourth and final open question regarding the *Morte*. Scholars have disagreed on Malory’s role in creating the *Morte*, and in particular on where he should be placed on the continuum between “faithful translator” and

“original author” (Lewis 1963; Lynch 2006; Davidson 2008). This debate has also hinged on the question of the extent to which Malory re-shapes the linguistic and stylistic features of his source texts in ways that go beyond straightforward translation. Malory was long regarded as a simplistic translator who “has no style of his own” (Lewis 1963, 23). However, since the 1990s there has been a growing interest in Malory's written style and reshaping of his narrative material. These arguments have been reviewed in Andrew Lynch's article on “A Tale of ‘Simple’ Malory” (Lynch 2006). Thus, for example, Jeremy Smith argues that Malory's paratactic composition of the text is “intensely audience-centred” and indicative of stylistic experimentation (Smith 1996, 104). Similarly, Ingrid Tieken-Boon van Ostade, Shunichi Noguchi, and Toshiyuki Takamiya have undertaken targeted analyses of specific linguistic and stylistic features of Malory's writing (Tieken-Boon van Ostade 1995; Noguchi 1995; Takamiya 1993). More recently, Roberta Davidson has drawn on Catherine Batt to argue for a view of Malory's adaptation process that gives due recognition to his role at the interface between being a “translator as writer” and a “translator as reader” (Batt 1989, 143–7; Davidson 2008, 133–4).

§11 Overall, scholarship on the *Morte* has so far failed to reach a consensus on the four questions discussed above. On the one hand, there have been numerous studies of the *Morte*, which have focused on small scale qualitative study of individual sections of the work in order to determine Malory's treatment of his sources in individual tales. On the other hand, larger scale studies of Malory's language have focused on analyzing linguistic and stylistic usage throughout the *Morte* without fully taking into account the role of Malory's many different source texts in shaping his writing.

§12 What these studies leave open is the question of the extent to which the eight sections of Malory's *Morte* differ from each other linguistically and how these differences may reflect the linguistic and stylistic features of the corresponding Middle English and Old French source texts he was working with. Our study is based on the hypothesis that there is linguistic variation between the different sections of the *Morte* and that these differences reflect the influence of the various source

texts Malory was working with for each of his tales. It has been undertaken on the understanding that identifying these influences has the potential to shed light on several unresolved questions concerning the overall unity of the *Morte*, Malory's adaptation process, the possible source used for Malory's "Tale of Sir Gareth" and the genesis of the two surviving versions of Malory's "Roman War" episode.

2. Methods

§13 Our first step was to download Caxton's edition of the *Morte* from the Corpus of Middle English Text and Verse hosted by the University of Michigan (<http://name.umdl.umich.edu/MaloryWks2>). We then removed all HTML tags, notes and chapter titles and merged the 21 "chapters" into eight "books" (see **Table 1**). These books are aligned with the divisions in the Winchester manuscript identified by Vinaver as separating Malory's work into eight books (Gordon and Vinaver 1937 *passim*). We normalized the text to unify spelling variants (**Table 2**), adopting the procedure outlined by Kestemont et al. (2015). Thus all non-standard Latin characters were replaced (e.g. "þ" with "th") and "i/j/y" were treated as the same letter, as were "u/v/w". Double consonants were replaced with single consonants. The raw and the normalized corpus were made publicly available under the document identifier DOI: 10.5281/zenodo.2639708.

§14 After normalizing the spelling, we applied three tokenization methods, which we term "simple stemming", "Porter like stemming" and "dictionary lemmatization". Our "simple stemming" involved removing suffixes such as "-e" and "-es" (with a few exceptions, such as "yes") and harmonizing the spelling of a small

Table 2: Lemmatization. Letter patterns were substituted with the replacements indicated. "{}" indicates a zero-length character string. Furthermore, every word was stemmed by shortening it to four letters.

| Pattern Replacement | |
|---------------------|----|
| -e, -es, -est, -eth | {} |
| i,j | y |
| u,w | v |
| -oo | o |

number of high frequency words. Finally, we shortened all tokens to a maximum of four letters. We then manually inspected our list of approximately 200 high-frequency words and carried out further tokenization where necessary, for example “hit” was always categorized as representing the same token as “it”. Some words representing content rather than function were removed to prevent plot features biasing the statistical analysis. Thus, we excluded proper names (“Arthur”, “Launcelot”) and nouns and adjectives referring to the main protagonists (“quen”, “fayr”) from our further analysis. The full list of excluded words is given in the caption of **Table 3**. Following this process of manual curation, we were left with a list of 135 words, which we have used as the basis for our further analysis (**Table 3**).

§15 For comparison we also undertook “Porter-like stemming”, where we applied the same work-flow as for “simple stemming”, however, we also implemented

Table 3: High frequency one-word tokens in the *Morte*. Words were shortened to four letters, suffixes such as “e” and “es” were removed and spelling was simplified (see Table 2), e.g. the letters “i”, “j” and “y” are all represented as “y”. Only words that occurred reproducibly in 80% of the samples were selected. The following words, which we regard as being indicative of content, were removed from the list: “syr”, “knyg”, “kyng”, “lavn”, “arth”, “gava”, “lady”, “qven”, “lord”, “fayr”, “nobl”, “damo”, “bors”.

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| a | afte | agey | al | am | an | and | anon | as | at |
| bata | be | ben | both | brod | bvt | by | cam | cast | com |
| covn | covr | day | ded | depa | do | don | dovn | dyd | ever |
| for | from | god | good | gret | had | hand | hath | hav | he |
| her | hors | hov | hym | hys | knov | let | lov | mad | mak |
| man | many | may | me | men | moch | mor | my | mygh | nam |
| neve | no | non | not | nov | nygh | of | on | ony | or |
| othe | over | ovt | pass | pray | rod | rygh | sav | say | sayd |
| see | self | she | shal | shel | shol | slay | smot | so | sper |
| stro | svch | sver | tak | the | that | them | then | ther | they |
| thov | thr | thvs | thy | thys | to | told | took | tvo | tym |
| vas | ve | vel | vent | ver | vet | vhan | vhat | vher | vhyc |
| vnto | vold | vors | vpon | vs | vyl | vyth | yes | ye | yf |
| yn | yov | yovr | ys | yt | | | | | |

many of the rules of the Porter stemmer: thus suffixes such as “-ly” and “-yng” were removed in addition to the ones listed in **Table 2**. Finally, we tested a form of “dictionary lemmatization”: we created a dictionary for all high-frequency words and manually assigned them to a lemma, for example “hors”, “horses”, “horsbak”, “horsed”, “horsback”, “horse”, “horsemen”, “horsbere”, “horsfeet”, and “horseman” were all assigned to the lemma “hors”. The final dictionary contained 160 lemmata. The overall results of our “Porter-like stemming” and our “dictionary lemmatization” can be viewed in the supplementary material (Supplementary Figure S1 and S2). The results of the “dictionary lemmatization” were also used for the MDS bi-plot showing the word loadings in **Figure 5**.

§16 Our next step was to undertake statistical analysis of two different feature frequency tables: one in which the text was tokenized as 1-word tokens and one in which the lemmatized text was tokenized as consecutive 2-word tokens. After tokenization, 20 independent random samples were drawn by sampling with replacement, of 5,000 or 75,00 tokens for 1-word and 2-word tokens respectively (this process of sampling is discussed by Eder 2015). As the text overall contains fewer high-frequency 2-word tokens, we had to collect more tokens per sample in this dataset, in order to ensure the numerical stability of the analysis.

§17 We retrieved high-frequency tokens by “culling” (Hoover 2004a, 2004b) with a document frequency cutoff of 80% (i.e. by removing all words that occur in less than 80% of all samples). Drawing on the culled frequency table, we then calculated Burrows’ Delta distance (Burrows 2002) between the samples and carried out multi-dimensional scaling (MDS). We repeated the analysis with two alternative distance measures: the Cosine Delta distance (Jannidis et al. 2015) and the Euclidean distance. (It is worth noting that “classical” MDS, based on singular value decomposition and using the Euclidean distance, is equivalent to performing principal component analysis (PCA)). In addition to 1-word and 2-word tokens, we also collected samples of character 3, 4 and 5-grams of normalized but unstemmed words (Stamatatos 2006). After culling with a 95% document frequency cutoff, we performed MDS using the same methodology as before. All computation was undertaken in R (R-Core Team 2018) making heavy use of functions provided in the *stylo* package in R (Eder

et al. 2016). We employed the package `plot3D` (Soetaert 2017) for 3D plotting and the package `car` (Fox and Weisberg 2011) for plotting ellipses using the function `dataEllipse` at confidence intervals of 0.5 and 0.95. We carried out MDS using the function `cmdscale`. MDS eigenvalues were used to calculate the proportion of variance represented in each MDS dimension. Word loadings of an MDS dimension were calculated using Pearson's correlation coefficient.

3. Results

§18 The MDS plot of our results, shown in **Figure 1**, strongly supports our hypothesis that there is linguistic variation between the different sections of the *Morte* and that this can be attributed in part to the influence of the different source texts Malory was working with for his different tales. This is confirmed by the affinity our graphs show between the text sections that are based on major English source texts (Books VII and VIII, marked in light and dark blue respectively). The samples drawn from these sections cluster in the upper right-hand quadrant of **Figure 1**. Meanwhile, the samples drawn from those sections based primarily on major French source texts (Books I, III, V, and VI, marked in orange, red, yellow, and pink respectively) cluster in the lower right hand quadrant of the MDS graph. We further observe that the samples from Book III ("Lancelot") lie at one extreme, and the samples from Books VII and VIII lie at the other extreme of the second MDS dimension (MDS2, corresponding to the vertical axis in **Figure 1**). Since Malory's major source texts are French for the former and English for the latter, it seems feasible to conclude that MDS2 captures stylistic features that reflect the linguistic influence of the language of the source texts. This is confirmed by the fact that the samples from Books I, V and VI, which are also known to be based on French source texts are all positioned closer to samples from Book III in MDS2. It is also particularly striking that the samples drawn from the "Roman War" section (Book II, marked in green), which has attracted so much scholarly attention, also appear as an anomaly in our analysis and stand out from all other sections by clustering in the left-hand half of **Figure 1**.

§19 The separation of the "Roman War" section is visible in the first MDS dimension (MDS1, horizontal axis) of **Figure 1**, not the second (MDS2, vertical axis). The first dimension of an MDS represents the features that constitute the largest

degree of dissimilarity between the samples (i.e. the difference between the books of the *Morte* assessed by token frequencies). On a hermeneutic level the features represented in MDS1 should be regarded as the most salient features of the dataset. As **Figure 1** shows, this MDS dimension draws a clear dividing line between the Roman War and all the other books of the *Morte*, giving a strong indication that this book does not map onto a continuous scale of authorial style of the other books. With respect to the specific types of word frequencies that are represented in the first dimension of the MDS *all the other books are more or less indistinguishable* (**Figure 1**).

§20 While Books I and III–VIII all show values of between 0.2 and 0.4 on the horizontal axis (MDS1), the “Roman War” section in Book II is centred at 0.9. We applied a standard statistical test (T-test) to investigate how likely it is that Book II

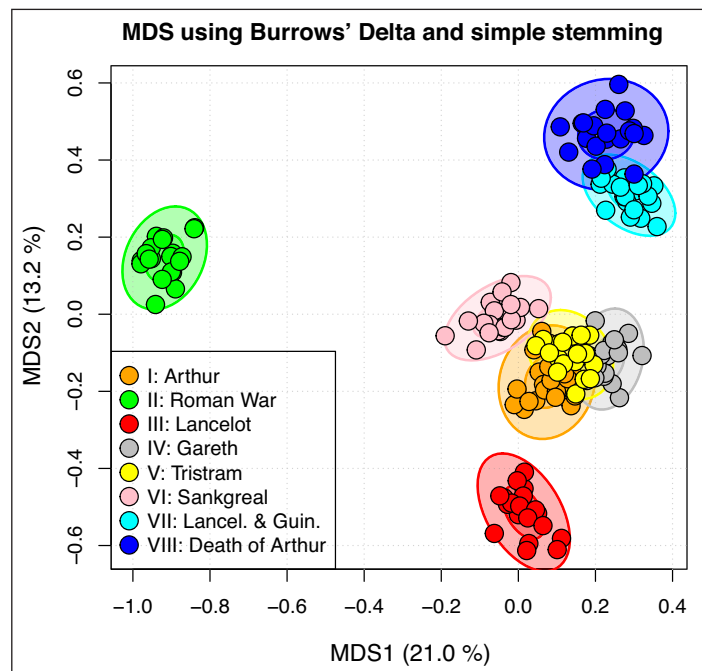


Figure 1: Multi-Dimensional Scaling (MDS) scatterplot of high frequency words in the *Morte*. Each filled circle represents a sample of 5,000 tokens (obtained by “simple stemming”, see ‘Methods’), the colour indicating which of the eight “books” (Table 1) it was taken from. The ellipses are drawn at 50 and 95% confidence intervals calculated from the 20 samples that were drawn from each “book”. The MDS was calculated from Burrows’ Delta distances between samples. The two first dimensions of the MDS (MDS1 and MDS2) are shown. The proportion of variance represented in each dimension is given in brackets.

belongs to the same population (i.e. has the same style, based on MDS1) as all the other books, and obtained a very clear-cut answer in that, in all probability, it does not belong to the same population ($p < 1e-15$). By contrast, the differences between the other books of the *Morte* are located entirely in the second dimension of the MDS shown in **Figure 1** and must therefore originate from different types of stylistic features than those that set the “Roman War” section apart from all the other books.

§21 The first two dimensions of the MDS appear to represent two almost entirely independent effects, which have a signal that is far above the noise level contained in the data: the first effect might be regarded as that of third-party involvement in the writing of the “Roman War” section and the second effect is that of whether the respective section is based on French or English source texts. Higher MDS dimensions, such as the third MDS dimension, shown as the z axis in **Figure 2**, represent somewhat less important, or statistically less pronounced, features. These higher dimensions reveal a tendency to further separate the individual books. In

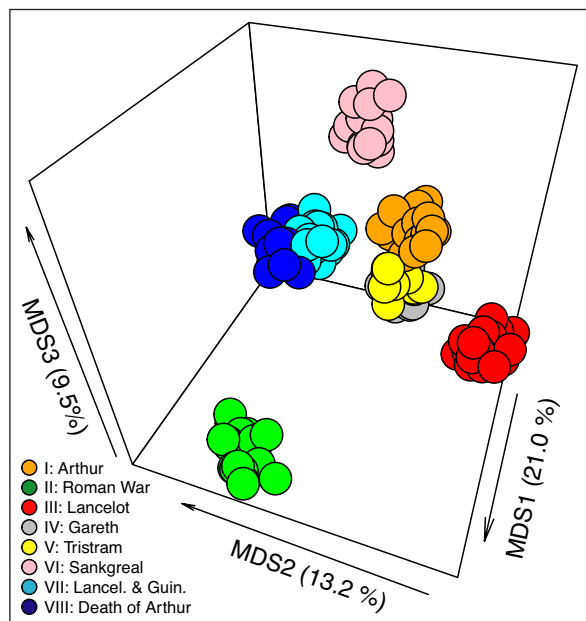


Figure 2: Three-dimensional representation of the first three dimensions of a multi-dimensional scaling of one-word tokens in the *Morte*. The MDS is the same as shown in Figure 1.

MDS3, the “Tale of the Sankgreal” (Book VI) is separated from Books I, III and IV; in even higher MDS dimensions up to MDS7 the differences between Books I, III and IV are revealed (Supplementary Figure S3). The eighth MDS dimension is the first one to show no distinguishable signature anymore.

§22 At this stage it should be noted that MDS, like PCA is an unsupervised technique, that does not take class labels (i.e. which book a sample originates from) into account. Thus, in principle, the MDS analysis is blind and even-handed with regard to our classification of the different sections of the work into books and the different source-text languages involved and only represents the distances between samples. As we have sampled with replacement, there is a possibility of the degree to which the samples taken from the same book cluster together being slightly overstated in the MDS—especially for the shorter “books”. Despite this caveat, the overall class separation revealed in the MDS remains reliable, as it is representative of actual differences between the classes with respect to the underlying data. This makes it all the more significant that the different MDS dimensions reveal a consistent and increasingly fine-grained degree of distinction between all of the different books.

§23 Moreover, we also obtained very similar results when we used two-word tokens and character 3, 4 or 5-grams instead of one-word tokens (compare **Figure 1** to **Figures 3** and **4** and Supplementary Figures S1 and S2) or applied different Distance measures (Supplementary Figures S1 and S2). This offers further evidence that the linguistic signal we have detected in the MDS is strong and reliable.

§24 To offer a better insight into which high-frequency words are distinctive for the different “books” of the *Morte*, we have used the results obtained with “dictionary lemmatization” as the basis for an MDS bi-plot showing the loadings of the 57 lemmas that have contributed most strongly to the first two dimensions of the MDS (**Figure 5**). The results suggest a greater use of romance words related to formal battle (such as “counceylle” and “bataille”) in the Roman War section, which is located on the left half of this MDS. This section also shows an increased use of third-person pronouns “his”, “them” and “they”. By contrast, the other “books” based on a major Middle English source, which are located in the right-hand upper quadrant

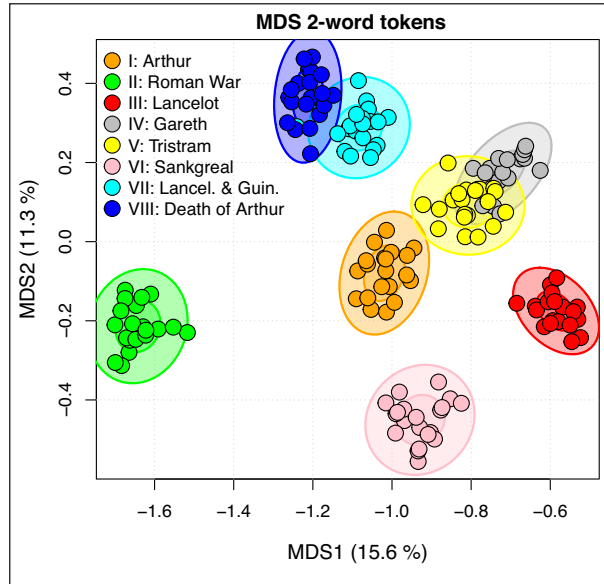


Figure 3: Multi-dimensional scaling scatterplot of two-word tokens in the Morte. Each filled circle represents a sample of 7,500 two-word tokens (obtained by “simple stemming”). The two first dimensions of the MDS are shown, the colours and symbols have the same meaning as in Figure 1.

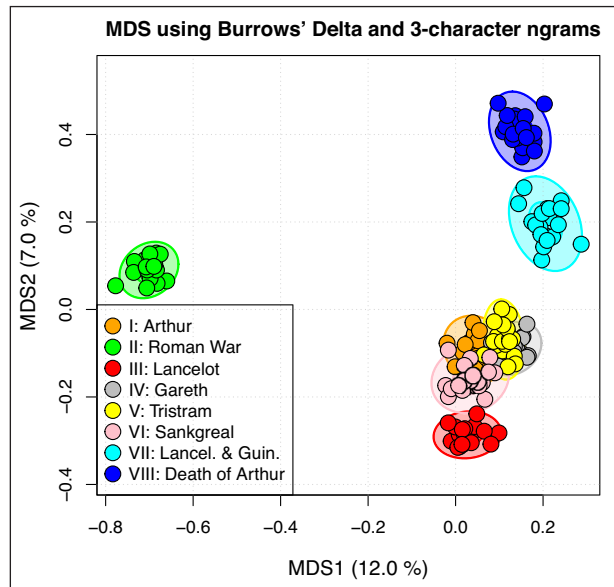


Figure 4: Multi-dimensional scaling scatterplot of character 3-grams in the Morte. Each filled circle represents a sample of 5,000 character 3-grams (three consecutive characters within a word). The two first dimensions of the MDS are shown, the colours and symbols have the same meaning as in Figure 1.

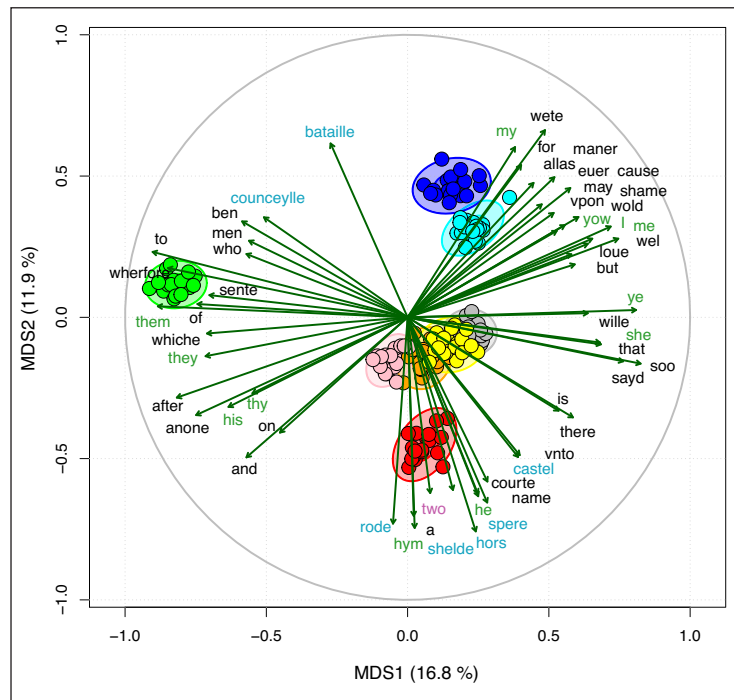


Figure 5: MDS bi-plot showing the loadings as arrows. Only variables with an arrow length of at least 0.6 in the unit circle were included. Labels show the most frequent word of a group of words that are combined into one variable (lemma). This MDS was calculated from the dictionary lemmatized corpus and the distance function was Burrows' Delta.

of the MDS, show a greater use of personal pronouns denoting direct speech, such as “I”, “me”, “my” and “yow”, as well as a greater use of “allas”, which also suggests direct speech. At the same time, those “books” based on an Old French major source text, located in the lower half of the MDS, show a greater usage of words related to chivalry and questing, such as “sheld”, “hors”, “rode”, “spere” and “castel”, and greater use of the numbers “one”, “two”, “thre” and “four” (see **Table 4**), which may be related to descriptive details, than those “books” based on major Middle English sources.

§25 Finally, we have included a table of the loadings of the second MDS dimension in particular, where the differences between the “books” based on Old French sources in the lower half of the bi-plot and the “books” based on Middle English sources in the upper half of the bi-plot are particularly visible (**Table 4**). This table shows that the “books” based on Middle English sources make a comparatively

Table 4: Loadings in the second MDS dimension of personal pronouns, numerals and synonymous words. A negative value corresponds to an arrow component pointing south in Fig. 5, and a positive value corresponds to an arrow component pointing north.

| Pronoun | MDS2 loading | Pronoun | MDS2 loading | Synonym | MDS2 loading | Numeral | MDS2 loading |
|---------|--------------|---------|--------------|----------|--------------|---------|--------------|
| hym | -0.70 | they | -0.14 | knowe | -0.38 | two | -0.62 |
| he | -0.62 | yow | 0.26 | countrey | -0.27 | four | -0.53 |
| thou | -0.49 | me | 0.28 | land | 0.49 | one | -0.52 |
| his | -0.32 | I | 0.32 | wete | 0.66 | thre | -0.51 |
| thy | -0.27 | my | 0.60 | | | | |

greater use of pronouns associated with direct speech (such as “my”, “me” and “I”), while the “books” based on Old French sources make a greater use of “hym”, “he”, “his” and “they”. Interestingly, where the latter “books” make frequent use of “thou” and “thy”, the former “books” make frequent use of “yow” and “your”. To give two examples of synonym usage, the “books” based on Middle English source texts make frequent use of the words “wete” and “land”, while the “books” based on Old French source texts make frequent use of the words “knowe” and “countrey”.

4. Discussion

§26 We will now relate our findings to the four questions raised in our introduction.

1. What type of written source (if any) did Malory use for his Book IV, “The Tale of Sir Gareth”?

§27 Failing the discovery of a lost romance text containing a clear source or analogue story to the “Tale of Sir Gareth”, this question cannot be fully resolved and must remain to some degree a point of speculation. However, it is nonetheless possible to draw conclusions about the nature of the source of Gareth, based on how this section of Malory’s work compares to the other sections of his work, for which his sources are known. If we accept the premise discussed above, that the linguistic differences between the “books” of Malory’s work are the result of his use of different

sources for different sections of his *Morte*, then we can also infer that those of his books with a greater linguistic resemblance to each other are based on sources that linguistically resemble each other. As we have seen, this is confirmed in our MDS graphs, which show the samples taken from sections that have French sources (Books I, III, V, and VI, marked in orange, red, yellow and pink respectively) clustering together in the lower right-hand quadrant of the graph in **Figure 1**. The samples taken from the “Tale of Sir Gareth” (Book IV, marked in grey) emerge as part of this cluster, an affinity that is further confirmed by the MDS generated for two-word tokens (**Figure 3**), in which the high-usage vocabulary of the “Tale of Sir Gareth” is represented as most resembling that of the tales of “Sir Tristram” and “Sir Lancelot” both tales based primarily on Old French source texts. Judging solely by the usage of high-frequency words analyzed here, our results therefore show that “The Tale of Sir Gareth” is linguistically significantly more similar to those “books” of the *Morte* that have a French written source. We can therefore conclude that Malory is most likely to have been working with a now-lost written source in French when writing his “Tale of Sir Gareth”. This supports Norris’ hypothesis of a lost French source (Norris 2008, 159) and is, from a purely pragmatic point of view, all the more convincing in that it suggests that Malory’s approach to writing this tale was consistent with his approach to writing the other seven sections of his work, in that he worked with a written source and that, as in five of the seven other sections of his work, that source was French.

2. How unified are the different sections of the *Morte*?

§28 Since literary unity remains a shifting goalpost that must always be interpreted in relation to other comparable collections, this is a question that cannot be resolved by our digital analysis. However, our analysis does show that the different “books” of the *Morte* are linguistically distinct from each other. Given that the different “books” are adaptations of different major source texts, the fact that they are linguistically heterogeneous may suggest that the different linguistic features of the various source texts have not been fully homogenized into a unified text corpus, but instead remain

traceable in the linguistic differences between the different "books". The fact that the group of "books" based on Middle English source texts are more similar to each other than the group of "books" based on Old French source texts confirms that suggestion to some degree.

3. Was Malory as a reviser responsible for the differences between the Caxton and Winchester versions of his *Morte*?

§29 One of the weaknesses of our study is that in its current form it only draws on the Caxton version of the *Morte*. This is an issue that we plan to remedy in a follow-up study that will make a comparative analysis of high-frequency words and two-word phrases in both the Winchester and the Caxton version of the *Morte*. Until this comparative study has been undertaken, our conclusions regarding the "Roman War" section of the *Morte* must remain partial and preliminary. Nonetheless, our present results provide strong confirmation of the discrepancies between the "Roman War" section contained in the Caxton edition and the remaining sections of the *Morte* contained in the Caxton. Thus, as the first dimension of the MDS graphs illustrate most clearly, the "Roman War" samples analyzed by our MDS of high frequency words stand clearly apart from the samples drawn from all the other sections, by clustering in the extreme left hand side of the graph. Nor does our analysis show the "Roman War" section overlapping with the two other tales based on major English language sources. Thus the samples drawn from Books VII and VIII, which are based on the Stanzaic *Morte Arthur* and, to a lesser degree, the Alliterative *Morte Arthure*, cluster together in the upper right hand section of the MDS graph, at quite some distance from the samples drawn from Book II (the "Roman War" section). Taking a simplistic view of the longstanding controversy surrounding the authorship of the Caxton edition of the "Roman War" section, the most obvious explanation for the strong linguistic discrepancy between the "Roman War" section of the Caxton *Morte* and the other sections would be that this section was not written by Malory alone or possibly was not written by Malory at the same time as the other sections. This becomes even more likely in light of the fact that the "Roman War" section of the *Morte* is also the

only one of Malory's "books" that does not exhibit marked linguistic similarities to the other sections that were based on source texts in the same language (Middle English and Old French respectively).

4. What was Malory's role in creating the *Morte*, and where should he be placed on the continuum between "faithful translator" and "original author"?

§30 The process of translating, selecting, arranging and formulating a narrative based on another text is a complex one, whose nuances cannot be fully grasped by quantitative methods. In this regard, quantitative assessment of linguistic differences between the tales remains a blunt instrument for understanding the role of an adaptor. It follows that our results are likely to be most illuminating when supplemented by more in-depth qualitative analysis that goes beyond the scope of this study.

§32 Nonetheless, our results do provide a concrete basis for considering the ways in which Malory's work may be reproducing not only thematic, but also linguistic aspects of his source texts. Thus the fact that the high-frequency words used in those sections of Malory's work that are based on Old French sources resemble each other more closely than the high-frequency words used in those sections of the *Morte* that are based on Middle English sources suggests that Malory was not only using the plots and storylines of his sources, but that his language, too, has been formed by his interaction with his source texts. This implies a transitory imprint of linguistic influences from his source texts on Malory's authorial style, casting a new and different light on his authorial process. While Malory's style of adaptation is perhaps most notable for his condensation and abbreviation of wide-ranging material (Kennedy 1981, 28), this imprint reveals the tangible effects of Malory the reader engaging with the language of his source texts in the course of rewriting them. Indeed, such an imprint implies that the process of rewriting texts cannot be neatly divided into the faithful translation of a text on the one hand, and the "original" reworking of a storyline on the other. Rather, the adaptor's process of

“borrowing” a plotline from a source text cannot be detached from his or her active engagement with the language of the source text in question and the subsequent momentary or lasting impression that engagement might leave on the adaptor’s authorial voice and identity.

§33 In light of this, the term “adaptor” seems particularly apt for Malory, in that it emphasizes the reshaping of material for a new purpose or context but gives due weight to the extent to which the “material itself” in both its abstract narrative content and the language in which that content is expressed often remains recognizably the same. It follows that in the case of Malory’s adaptation of his sources we cannot speak of “old wine in new skins”, as such images of adaptation create a dichotomy between the plot as the “substance” of a source text and the language or formal structure of the source text as the “exchangeable receptacle” in which that substance is contained. Instead, our study suggests that the language and content of his sources work together to shape Malory’s *Arthuriad*.

Additional Files

The additional files for this article can be found as follows:

- **Supplementary Figure S1.** The first two MDS dimensions of three different types of tokenization (“simple stemming”, “Porter-like stemming” and “dictionary lemmatization”) that were combined with three different types of distance functions. Colours and symbols have the same meaning as in Figure 1. DOI: <https://doi.org/10.16995/dm.86.s1>
- **Supplementary Figure S2.** The first two MDS dimensions of three different types of tokenization (character n-grams of length 3, 4, and 5) that were combined with three different types of distance functions. Colours and symbols have the same meaning as in Figure 1. DOI: <https://doi.org/10.16995/dm.86.s2>
- **Supplementary Figure S3.** The first eight dimensions of a multi-dimensional scaling of one-word tokens in the *Morte* (obtained by “simple stemming” and Burrows’ Delta distance). The MDS is the same as shown in

Figure 1, only here every dimension, from the first to the eighth, is shown as an independent row. The solid lines that are overlaid onto the circles representing the samples are drawn at the average value of each “book” in that dimension. DOI: <https://doi.org/10.16995/dm.86.s3>

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Both authors contributed equally. The corresponding author is Miriam Edlich-Muth.

Conceptualization: mem, cem

Methodology: mem, cem

Software: Cem

Validation: cem

Formal Analysis: cem

Investigation: mem, cem

Resources: mem, cem

Data Curation: mem, cem

Writing – Original Draft Preparation: mem

Writing – Review & Editing: mem, cem

Visualization: cem

Supervision: n.a.

Project Administration: mem

Funding Acquisition: n.a.

Editorial Contributions

Recommending editor: Mike Kestemont, University of Antwerp, Belgium

Recommending referees: François Laramée, University of Ottawa, Canada; Simone Rebola, University of Verona, Italy

Section/Copy editor: Nathir Haimoun, University of Lethbridge (Canada) Journal Incubator

Layout editor: Mahsa Miri, University of Lethbridge (Canada) Journal Incubator

References

- Archibald, Elizabeth.** 2013. "Malory and Late Medieval Arthurian Cycles." In *Traditions and Innovations. The Study of Medieval English Literature: The Influence of Derek Brewer*, edited by Charlotte Brewer and Barry Windeatt, 173–187. Cambridge: Boydell and Brewer.
- Batt, Catherine.** 1989. "Malory's Questing Beast and the Implications of Author as Translator." In *The Medieval Translator: The Theory and Practice of Translation in the Middle Ages*, edited by Roger Ellis, 143–166. Cambridge: D.S. Brewer.
- Batt, Catherine.** 2002. *Malory's Morte D'Arthur: Remaking Arthurian Tradition*. Basingstoke: Palgrave Macmillan.
- Benedict, Kimberley.** 2004. *Empowering Collaborations: Writing Partnerships between Religious Women and Scribes in the Middle Ages (Studies in Medieval History and Culture, 27)*, 1st Edition. New York; London: Routledge. DOI: <https://doi.org/10.4324/9780203491577>
- Bennett, J. A. W., and Walter Oakeshott,** eds. 1963. *Essays on Malory*. Oxford: Clarendon Press.
- Binongo, José Nilo G.** 2003. "Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution." *CHANCE* 16(2): 9–17. DOI: <https://doi.org/10.1080/09332480.2003.10554843>
- Burrows, J. F.** 2002. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17(3): 267–287. DOI: <https://doi.org/10.1093/lc/17.3.267>
- Clark, David Eugene.** 2014. "Hearing and Reading Narrative Divisions in the 'Morte Darthur.'" *Arthuriana*, 24(2): 92–125. DOI: <https://doi.org/10.1353/art.2014.0027>
- Davidson, Roberta.** 2008. "The 'Freynshe booke' and the English Translator: Malory's 'Originality' Revisited." *Translation and Literature* 17(2): 133–149. DOI: <https://doi.org/10.3366/E0968136108000198>

- Eder, Maciej.** 2015. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Digital Scholarship in the Humanities* 30(2): 167–182. DOI: <https://doi.org/10.1093/llc/fqt066>
- Eder, Maciej, Jan Rybicki, and Mike Kestemont.** 2016. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8(1): 107–121. DOI: <https://doi.org/10.32614/RJ-2016-007>
- Field, P. J. C.** 1995. "The Source of Malory's 'Tale of Gareth.'" In *Malory: Texts and Sources*, edited by P. J. C. Field, 246–260. Cambridge: D.S. Brewer.
- Fox, John, and Sanford Weisberg.** 2011. *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks, CA: SAGE.
- Gordon, E. V., and Eugene Vinaver.** 1937. "New Light on the Text of the Alliterative 'Morte Arthure.'" *Medium Ævum* 6(2): 81–98. DOI: <https://doi.org/10.2307/43626034>
- Grzybek, Peter.** 2014. "The Emergence of Stylometry: Prolegomena to the History of Term and Concept." In *Text within Text – Culture within Culture*, edited by Katalin Kroó and Peeter Torop, 58–75. Budapest: L'Harmattan.
- Hanks, D. Thomas, Jr.** 2003. "The Rhetoric of the Folk Fairy Tale in Sir Thomas Malory's 'Tale of Sir Gareth.'" *Arthuriana* 13(3): 52–67. DOI: <https://doi.org/10.1353/art.2003.0051>
- Hoover, D.** 2004a. "Delta Prime?" *Literary and Linguistic Computing* 19(4): 477–495. DOI: <https://doi.org/10.1093/llc/19.4.477>
- Hoover, D.** 2004b. "Testing Burrows's Delta." *Literary and Linguistic Computing* 19(4): 453–475. DOI: <https://doi.org/10.1093/llc/19.4.453>
- Jannidis, Fotis, Steffen Pielström, Christof Schöch, and Thorsten Vitt.** 2015. "Improving Burrows' Delta: An Empirical Evaluation of Text Distance Measures." In *Digital Humanities Conference 2015*.
- Juola, Patrick.** 2008. "Authorship Attribution." *Foundations and Trends in Information Retrieval*, 1(3): 233–334. DOI: <https://doi.org/10.1561/1500000005>
- Kennedy, Edward.** 1981. "Malory and His English Sources." In *Aspects of Malory*, edited by Toshiyuki Takamiya and Derek Brewer, 27–55. Cambridge: D. S.

Brewer; Woodbridge, Suffolk: Boydell & Brewer; Totowa, N. J.: Rowman & Littlefield.

- Kestemont, Mike.** 2012. "Stylometry for Medieval Authorship Studies: An Application to Rhyme Words." *Digital Philology: A Journal of Medieval Cultures* 1(1): 42–72. DOI: <https://doi.org/10.1353/dph.2012.0002>
- Kestemont, Mike.** 2015. "A Computational Analysis of the Scribal Profiles in Two of the Oldest Manuscripts of Hadewijch's Letters." *Scriptorium* 69: 159–177.
- Kestemont, Mike, Sara Moens, and Jeroen Deploige.** 2015. "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux." *Digital Scholarship in the Humanities* 30(2): 199–224. DOI: <https://doi.org/10.1093/llc/fqt063>
- Kestemont, Mike, Walter Daelemans, and Guy De Pauw.** 2010. "Weigh your Words—Memory-Based Lemmatization for Middle Dutch." *Literary and Linguistic Computing* 25(3): 287–301. DOI: <https://doi.org/10.1093/llc/fqq011>
- Koppel, Moshe, Jonathan Schler, and Shlomo Argamon.** 2009. "Computational Methods in Authorship Attribution." *Journal of the American Society for Information Science and Technology* 60(1): 9–26. DOI: <https://doi.org/10.1002/asi.20961>
- Lewis, C. S.** 1963. "The English Prose Morte." In *Essays on Malory*, edited by J. A. W. Bennett and Walter Oakeshott, 7–28. Oxford: Clarendon Press.
- Love, Harold.** 2002. *Attributing Authorship: An Introduction*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511483165>
- Lumiansky, Robert.** 1964. *Malory's Originality: A Critical Study of Le Morte Darthur*. Baltimore: John Hopkins Press.
- Lynch, Andrew.** 2006. "A Tale of 'Simple' Malory and the Critics." *Arthuriana* 16(2), 10–5. DOI: <https://doi.org/10.1353/art.2006.0065>
- Mairey, Aude.** 2006. "These Trewe Conclusions in Englissh: Langues, Cultures Et Autorités Dans L'Angleterre Du XIVE Siècle." *Revue Historique* 1(637): 37–57. DOI: <https://doi.org/10.3917/rhis.061.0037>
- Matthews, William.** 1997. "A Question of Texts." *Arthuriana* 7(1): 93–133. DOI: <https://doi.org/10.1353/art.1997.0017>

- Moorman, Charles.** 1995. "Desperately Defending Winchester: Arguments from the Edge." *Arthuriana* 5(2): 24–30. DOI: <https://doi.org/10.1353/art.1995.0022>
- Noguchi, Shunichi.** 1995. "The Winchester Malory." *Arthuriana* 5(2): 15–23. DOI: <https://doi.org/10.1353/art.1995.0028>
- Norris, Ralph.** 2006. "Minor Sources in Caxton's Roman War." *Studies in Philology* 103(1): 68–87. DOI: <https://doi.org/10.1353/sip.2006.0004>
- Norris, Ralph.** 2008. *Malory's Library*. Woodbridge: Boydell & Brewer.
- Piotrowski, Michael.** 2012. *Natural Language Processing for Historical Texts*. California: Morgan & Claypool Publishers.
- R-Core Team.** 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.r-project.org>.
- Reynolds, Noel, Schaalje, G. Bruce, and John Hilton.** 2012. "Who Wrote Bacon? Assessing the Respective Roles of Francis Bacon and his Secretaries in the Production of his English Works." *Literary and Linguistic Computing* 27(4): 409–425. DOI: <https://doi.org/10.1093/lc/fqs020>
- Sanders, Arnold.** 2006. "Sir Gareth and the 'Unfair Unknown': Malory's Use of the Gawain Romances." *Arthuriana* 16(1): 34–46. DOI: <https://doi.org/10.1353/art.2006.0054>
- Shepherd, Stephen, and Thomas Malory.** 2003. "Le Morte Darthur". New York: Norton.
- Simko, Jan.** 1957. *Word-Order in the Winchester Manuscript and in William Caxton's Edition of Thomas Malory's Morte Darthur (1485): A Comparison*. Halle: Niemeyer.
- Smith, Jeremy J.** 1996. "Language and Style in Malory." In *A Companion to Malory*, edited by S. G. Edwards and E. Archibald, 97–113. Suffolk: Boydell and Brewer.
- Soetaert, Karline.** 2017. plot3D: Plotting Multi-Dimensional Data. *R package version 1.1.1*. Accessed September 4, 2019. <https://cran.r-project.org/web/packages/plot3D/plot3D.pdf>
- Souvay, Gilles, and Jean-Marie Pierrel.** 2009. "LGeRM: Lemmatization de mots en moyen français." *Traitement Automatique des Langues* 50(2): 21.
- Stamatatos, Efstathios.** 2006. "Ensemble-based Author Identification Using Character N-Grams." In *Proceedings of the 3rd International Workshop on Text-*

based Information Retrieval, 41–46. Accessed September 30, 2019. <https://pdfs.semanticscholar.org/349e/921f56d29b71fc1442ed44724862948fde5c.pdf>

- Takamiya, Toshiyuki.** 1993. "Editor/Compositor at Work: The Case of Caxton's Malory." In *Arthurian and Other Studies Presented to Shunichi Noguchi*, edited by Suzuki and Mukai. Suffolk, UK: D.S. Brewer, 143–151.
- Thompson, John J.** 2016. "Print, Miscellaneity and Impact of Oral Performance: Shaping the Understanding of Late Medieval Readers." In *Readings on Audience and Textual Materiality*, edited by G. Allen, Carrie Griffin and Mary O'Connell. London: New York: Routledge, 9–22.
- Tieken-Boon van Ostade, I. M.** 1995. *The Two Versions of Malory's Morte Darthur: Multiple Negation and the Editing of the Text*. Cambridge: Brewer.
- Van Dalen-Oskam, K., and J. van Zundert.** 2007. "Delta for Middle Dutch: Author and Copyist Distinction in Walewein." *Literary and Linguistic Computing* 22 (3): 345–362. DOI: <https://doi.org/10.1093/lc/fqm012>
- Vinaver, Eugène,** ed. 1947. *The Works of Sir Thomas Malory*. Oxford: Clarendon Press.

How to cite this article: Edlich-Muth, Christian, and Miriam Edlich-Muth. 2019. "A Computational Approach to Source Adaptation in Thomas Malory's *Morte Darthur*." *Digital Medievalist* 12(1): 5, pp. 1–26. DOI: <https://doi.org/10.16995/dm.86>

Published: 30 December 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Digital Medievalist is a peer-reviewed open access journal
published by Open Library of Humanities.

OPEN ACCESS The Open Access logo, a stylized 'O' with a person icon inside.