**New methods, old data: using digital technologies to explore nineteenth century letter writing practices**

Emma Moreton, Coventry University, aa0282@coventry.ac.uk

Chris Culy, Independent Researcher, culy@mac.com

1. Introduction

The sourcing, preservation and documentation of migrant letter collections is gathering pace, with the Internet providing a significant new forum for the dissemination of long-hidden archives.[[1]](#footnote-1) Important studies of, for example, English (Gerber 2006), Scottish (Erickson 1972), Irish (Miller 1985; 2003; 2008), Welsh (Conway 1961) German (Kamphoefner et al. 1988), Swedish (Barton 1990) and Norwegian (Zempel 1991) migration have demonstrated the value of using personal letters to gain a fuller and deeper understanding of both the complex social processes of migration and the conditions and daily lives of the migrants themselves.They have also enriched our understanding of how the form of the letter itself, for migrants more than any other group, “functioned primarily to reconfigure personal relationships made vulnerable by distance” (Elliott et al. 2006: 17) and to reinforce familial bonds.

Whilst the benefits of uncovering, documenting and analysing historical letter collections are clear, this growing body of research sometimes appears somewhat fragmented. Researchers often work in isolation from one another and their projects evolve independently. Details of letter collections are sometimes difficult to find, and access to resources can be restricted by copyright and intellectual property concerns. This can lead to collections being missed or overlooked and/or the reduplication of work, with letters often being transcribed several times by different projects – projects which have their own research aims and, quite often, their own transcription and markup practices.

There seems to be a good case, therefore, for developing a collaborative, cross-disciplinary approach to working with (migrant) letter collections: the digital humanities (defined here as the use of digital technologies to address humanities research questions), arguably, offers one possible solution. Once digitised in a consistent and formalised way, it is possible for different archives to interconnect and for the letter content to be analysed in a more systematic and replicable way, using, for instance, corpus and visualisation tools that allow us to notice patterns in the data (within and across collections) which we might not have seen otherwise. Additionally, digital tools allow for the easy reanalysis of the data in the future if new questions arise, thus enabling different projects to build on previous research and carry out comparative studies across collections.

Drawing on developments in the field of digital humanities, this chapter uses a combination of qualitative and computational approaches to explore letter-writing practices within the context of nineteenth-century mass migration to America. Our aim is to explore and evaluate the advantages and limitations of digital techniques and methods for the analysis of historical migrant letter collections. In so doing, we hope to show how digital technologies can shed new light on old data in ways which might allow us to better understand and interpret the migrant experience.

The chapter focuses on the letters of two sisters, Annie and Julia Lough, who migrated from Ireland to America in the late nineteenth century. First, a close qualitative reading of the letters is carried out to identify topics in the discourse. Corpus and visualisation tools are then used to examine four topics in more detail, namely: Future Letters (any reference to letters which are going to be sent, or which the author hopes/expects to receive in the future), Previous Letters (any reference to letters which have already been sent or received at the time of writing), Remittance (money orders) and Enclosure (e.g. photographs, gifts, newspaper cuttings etc.). The chapter demonstrates how these features of the letters – often overlooked because of their formulaic nature – are important to our understanding of the migrant experience, providing useful insights into how, through correspondence, vulnerable familial roles and relationships were negotiated, maintained and reinforced.[[2]](#footnote-2) The chapter concludes by considering the extent to which digital technologies offer new opportunities for the transhistorical study of migrant correspondence more generally.

2. The (female) Irish migrant

The post-famine period (circa. 1850s-1920s) was a time that saw a significant increase in female migration from Ireland to America. Economic changes in Ireland, including declining wage earning capabilities due to the deindustrialisation of the Irish countryside, as well as changes in inheritance practices from partible to impartible inheritance systems (in turn, leading to changes in marriage trends), contributed to “a massive post-famine emigration by young, unmarried women” (Miller 1985: 3). As Diner explains, “in the society created in the wake of the Famine only one son per family could inherit land”; the same son typically entered into matrimony (1983: 10). Additionally, the widespread application of the dowry system meant that only one daughter within a family could realistically expect to marry. Her unmarried sisters had, as Diner puts it, “limited options”: “They could remain single in the countryside” with very little chance of work, they could seek employment in cities such as Dublin, or further afield, or they could join the “millions of other young Irish women crossing the Atlantic to seek fortune or family in the United States” (1983: 12). By the second half of the nineteenth century Ireland had become “a nation characterized by late and reluctant marriage as well as by a massive voluntary exodus” (ibid.: 8).

According to Diner, Irish women differed from other migrant women in several ways. For one, “they were the only significant group of foreign-born women who outnumbered men” (1983: xiv). Indeed, after 1880, young women constituted the majority of the departing Irish (Miller 1985: 392). However, while female Irish migrants appeared to enjoy a certain level of autonomy and independence – they migrated alone, played an important economic role within the notional family hierarchy, and sought to advance themselves socially – this, Diner argues, “did not mean that that they thought only of themselves” (1983: xiv). Rather, their actions were largely motivated by “family loyalties” and “a commitment to Irish Catholic culture and to its way of life” (ibid.). The female Irish migrant, then, was required to balance old world traditions with New World ways of life, maintaining relationships with those back home whilst simultaneously integrating into American society. The letters they wrote home arguably provides insight into how this delicate and complex family dynamic was negotiated and performed.

Previous research which uses Irish migrant letters as a primary data source has typically focused on one of two areas: 1) the language of the letters (these studies have often involved corpus linguistic methods of analysis) and 2) the subject matter of the letters (these studies have generally involved a qualitative reading of the letters). While linguists have been especially interested in exploring the linguistic features of Irish English as well as aspects of language change and variation (see, for instance, McCafferty and Amador 2012, 2014; Amador and McCafferty 2015; Van Hattum 2014, and Montgomery 1995), historians have been more concerned with what the migrants wrote about (see, for instance, Miller (1985) and Fitzpatrick (1994)). The current study builds on research in both of these areas: linguistics and social history. Similar to Miller and Fitzpatrick, it begins by examining what the Lough sisters wrote about to see whether similar topics and themes emerge. It then uses computational and visualisation tools to examine four topics in particular: Future Letters, Previous Letters, Remittance and Enclosure to see what these often mundane and formulaic sections of the discourse might reveal about letter writing practices, the communicative function of migrant correspondence, and the ways in which family roles and relationships were constructed through language.

3. The Lough letters

The Lough (pronounced lɒk) family letters are taken from a much larger body of Irish migrant correspondence collected by Kerby Miller, Emeritus Professor, University of Missouri. Miller himself has explored this wider collection in several pioneering works on Irish migration (see, for instance, Miller (1985, 1995, 2008) and Miller et al. (2003)), and his archive of over 5,000 letters has been referred to by many scholars, including Emmons (1990), Koos (2001), Bruce (2006), Corrigan (1992), Nolan (1989) and Noonan (2011).

The six Lough sisters – Elizabeth, Alice, Annie, Julia, Mary and Maggie – came from a Roman Catholic family in Meelick, in what was then called Queen’s County (now County Laois), Ireland.[[3]](#footnote-3) The sisters were daughters of Elizabeth McDonald Lough and James Lough who lived on a small holding consisting of two fields, one of which, according to family legend, was sold to pay for the sisters’ passages to America. The Lough family was, according to Miller, not of the lowest class as both parents and daughters were able to write. Apart from Mary and Maggie, all the Lough sisters migrated to America between 1870 and 1884.

There are 99 letters in the Lough collection, the majority of which were written by Annie (39 letters / 20,405 words), the third sister to migrate in 1878, and Julia (35 letters / 12,220 words), the fourth sister to migrate in 1884. This chapter focuses on these two larger collections, which, from hereon in, will be referred to as the *ALC* (*Annie Lough Corpus*) and the *JLC* (*Julia Lough Corpus*). Obviously, there are issues to do with representativeness when working with private letter collections, as we are unable to account for letters that have been lost, misplaced, or remain undiscovered. Indeed, it is very possible that many more letters were exchanged within the Lough family and their wider social network. Nevertheless, the Lough sisters are representative of the type of female Irish migrant described by Diner; their letters, therefore, offer a good starting point for exploring and understanding the female migrant experience.

Tables 1 and 2 show how frequently Annie and Julia wrote home and whom they wrote to. Focusing on Annie’s correspondence first of all, Table 1 shows that her earlier letters were addressed to her mother, the first of which was sent in around 1878 (although the letter itself is not dated) from Queenstown, County Cork, Ireland, just before Annie was about to set sail for America. After 1895 (around the time of her mother’s death), Annie began writing to her sister, Mary, and the correspondence continued into the late 1920s. Annie wrote to Mary regularly during this 30 to 35 year period, often sending letters at Easter and Christmas, or on the anniversary of a family member’s death. Several of Annie’s letters are not dated but their content would suggest they were written from 1920 onwards. All but the first letter were sent from Winsted, Litchfield County, Connecticut.

Table 1: Overview of the *ALC*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Day | Month | Year | From (location) | Recipient | To (location) | No. Words |
| 1 | 18 | June | - | Queenstown, Ireland | Mother | Meelick, Ireland | 356 |
| 2 | 03 | March | 1890 | Winsted, America | Mother & Sister | Meelick, Ireland | 480 |
| 3 | 29 | October | 1891 | Winsted, America | Mother | Meelick, Ireland | 1055 |
| 4 | 15 | December | 1891 | Winsted, America | Mother | Meelick, Ireland | 487 |
| 5 | 23 | March | 1892 | Winsted, America | Mother | Meelick, Ireland | 1017 |
| 6 | 30 | March | 1893 | Winsted, America | Mother | Meelick, Ireland | 971 |
| 7 | - | December | - | Winsted, America | Mother | Meelick, Ireland | 208 |
| 8 | - | - | - | Winsted, America | Mother | Meelick, Ireland | 645 |
| 9 | 17 | March | 1895 | Winsted, America | Sister | Meelick, Ireland | 612 |
| 10 | 18 | May | 1899 | Winsted, America | Sister | Meelick, Ireland | 541 |
| 11 | 16 | February | 1901 | Winsted, America | Sister | Meelick, Ireland | 394 |
| 12 | 21 | September | 1901 | Winsted, America | Sister | Meelick, Ireland | 441 |
| 13 | 10 | December | 1902 | Winsted, America | Sister | Meelick, Ireland | 365 |
| 14 | 03 | April | 1906 | Winsted, America | Sister | Meelick, Ireland | 431 |
| 15 | 20 | June | 1906 | Winsted, America | Sister | Meelick, Ireland | 332 |
| 16 | 30 | November | 1906 | Winsted, America | Sister | Meelick, Ireland | 302 |
| 17 | 12 | December | 1912 | Winsted, America | Sister | Meelick, Ireland | 632 |
| 18 | 08 | December | 1913 | Winsted, America | Sister | Meelick, Ireland | 514 |
| 19 | 11 | December | 1914 | Winsted, America | Niece | Meelick, Ireland | 398 |
| 20 | 31 | April | 1918 | Winsted, America | Sister | Meelick, Ireland | 664 |
| 21 | 06 | May | 1918 | Winsted, America | Sister | Meelick, Ireland | 884 |
| 22 | 14 | July | 1918 | Winsted, America | Sister | Meelick, Ireland | 863 |
| 23 | 14 | August | 1919 | Winsted, America | Sister | Meelick, Ireland | 857 |
| 24 | 21 | March | 1920 | Winsted, America | Niece | Ireland | 469 |
| 25 | 21 | March | 1920 | Winsted, America | Sister | Meelick, Ireland | 649 |
| 26 | 07 | December | 1919/1920 | Winsted, America | Sister | Meelick, Ireland | 396 |
| 27 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 435 |
| 28 | 31 | March | 1924 | Winsted, America | Sister | Meelick, Ireland | 538 |
| 29 | 29 | September | 1925 | Winsted, America | Sister | Meelick, Ireland | 237 |
| 30 | 28 | March | 1928 | Winsted, America | Sister | Meelick, Ireland | 206 |
| 31 | 18 | October | 1928 | Winsted, America | Sister | Meelick, Ireland | 870 |
| 32 | 04 | November | - | Winsted, America | Nephew | Ireland | 513 |
| 33 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 261 |
| 34 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 400 |
| 35 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 489 |
| 36 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 476 |
| 39 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 207 |
| 38 | - | - | - | Winsted, America | Sister | Meelick, Ireland | 307 |
| 39 | 01 | December | 1919 | Winsted, America | Sister | Meelick, Ireland | 503 |

Focusing on Julia’s correspondence, Table 2 shows that 23 of the 35 letters were addressed to Julia’s mother, while 12 were addressed to her sister, Mary. Most of Julia’s letters (33 out of 35) were sent between 1884 to 1895. Some of the letters are not dated, but their content has allowed them to be placed within an approximate timeframe. Most of the letters dated between 1884, when Julia first migrated, and 1894 were sent from Winsted, Connecticut. In around 1895 Julia relocated to Torrington, Connecticut.

Table 2: Overview of the *JLC*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Day | Month | Year | From (location) | Recipient | To (location) | No. Words |
| 1 | 27 | September | 1884 | Queenstown, Ireland | Mother | Meelick, Ireland | 40 |
| 2 | - | - | 1884 | Ireland or England | Mother | Meelick, Ireland | 98 |
| 3 | 20 | December | 1884 | Winsted, America | Sister | Meelick, Ireland | 519 |
| 4 | - | - | 1884-1894 | Winsted, America | Mother | Meelick, Ireland | 190 |
| 5 | - | December | 1888 | Winsted, America | Mother | Meelick, Ireland | 342 |
| 6 | 03 | November | 1889 | Winsted, America | Mother | Meelick, Ireland | 444 |
| 7 | 02 | December | 1889 | Winsted, America | Mother | Meelick, Ireland | 436 |
| 8 | - | - | 1889-1890 | Winsted, America | Mother | Meelick, Ireland | 487 |
| 9 | - | - | 1889-1894 | Winsted, America | Mother | Meelick, Ireland | 259 |
| 10 | 09 | March | 1890 | Winsted, America | Mother | Meelick, Ireland | 463 |
| 11 | 10 | August | 1890 | Winsted, America | Mother | Meelick, Ireland | 366 |
| 12 | - | December | 1890 | Winsted, America | Mother | Meelick, Ireland | 350 |
| 13 | 18 | January | 1891 | Winsted, America | Sister | Meelick, Ireland | 348 |
| 14 | 25 | January | 1891 | Winsted, America | Mother | Meelick, Ireland | 351 |
| 15 | 30 | March | 1891 | Winsted, America | Mother | Meelick, Ireland | 225 |
| 16 | 18 | October | 1891 | Winsted, America | Mother | Meelick, Ireland | 317 |
| 17 | 14 | December | 1891 | Winsted, America | Mother | Meelick, Ireland | 300 |
| 18 | 11 | May | pre-1892 | Winsted, America | Mother | Meelick, Ireland | 400 |
| 19 | 01 | September | 1892 | Winsted, America | Mother | Meelick, Ireland | 396 |
| 20 | - | - | 1892-1893 | Winsted, America | Mother | Meelick, Ireland | 321 |
| 21 | 21 | March | 1893 | Winsted, America | Sister | Meelick, Ireland | 423 |
| 22 | - | May | 1893 | Winsted, America | Mother | Meelick, Ireland | 305 |
| 23 | - | July | 1893 | Winsted, America | Mother | Meelick, Ireland | 340 |
| 24 | 03 | September | 1893 | Winsted, America | Mother | Meelick, Ireland | 356 |
| 25 | 10 | October | 1893 | Winsted, America | Mother | Meelick, Ireland | 334 |
| 26 | - | December | 1893 | Winsted, America | Mother | Meelick, Ireland | 451 |
| 27 | 25 | March | 1894 | Torrington, America | Sister | Meelick, Ireland | 183 |
| 28 | 24 | May | 1893-1894 | Torrington, America | Sister | Meelick, Ireland | 477 |
| 29 | - | - | 1889-1894 | Torrington, America | Sister | Meelick, Ireland | 354 |
| 30 | 04 | June | 1894 | Winsted, America | Sister | Meelick, Ireland | 736 |
| 31 | - | November | 1895 | Winsted, America | Sister | Meelick, Ireland | 469 |
| 32 | 08 | July | 1895 | Queenstown, Ireland | Sister | Meelick, Ireland | 44 |
| 33 | - | August | 1895 | Torrington, America | Sister | Meelick, Ireland | 416 |
| 34 | 17 | March | 1919-1920 | Torrington, America | Sister | Meelick, Ireland | 331 |
| 35 | 09 | November | 1927 | Torrington, America | Sister | Meelick, Ireland | 349 |

Julia’s pattern of letter writing differs quite noticeably from her sister Annie’s. Annie’s letters are fairly evenly distributed and there are no major gaps in her correspondence. Most of Annie’s letters (27 out of 39) were addressed to younger sister Mary and were sent after their mother’s death. In contrast, most of Julia’s letters were addressed to their mother. After their mother’s death, Julia wrote somewhat sporadically to Mary and there was a 24-year gap in Julia’s writing between letter 33 (sent in 1895) and letter 34 (sent in 1919-1920). Overall, Annie tended to write longer letters than Julia with an average word count of 523 (versus 349 for Julia). Additionally, Annie’s letters were more varied in length than Julia’s.

4. Topic identification

To prepare the letters for analysis they were first transcribed and saved in Plain Text format. While automated semantic taggers provide a useful overview of the content of a corpus, they can sometimes miss more subtle topics and themes in the discourse. A trial study using *Wmatrix* (Rayson 2009) as a starting point, for example, showed that it was sometimes difficult to ascertain those sections of the discourse that were describing, for instance, homesickness, which is often expressed indirectly or through the use of metaphor. Unsurprisingly, there is no specific semantic domain for homesickness in the UCREL semantic tagset. This meant that it was necessary to look for semantic domains that might indicate expressions of homesickness in the discourse. Obvious domains might be *E1+ Emotional* or *E4.1- Sad*, or perhaps *N2.2- Distance: Far*; however, descriptions of remembering, forgetting and dreaming are also typically used by the Lough sisters in the context of describing feelings of homesickness – something which could only be identified through a contextualised reading of the correspondence. As such, a close reading of the letters was necessary to identify sequences in the discourse that appeared to be lexically related to see what topics emerged.

The topics were then annotated following XML conventions. The section below, for example, is about education and is marked with the opening tag <education> to show where this section begins and the closing tag </education> (with forward slash) to show where it ends, as follows:

**<education>**I hope you keep them to school all you can when they grow bigger you can not send them very well I suppose there are not many of the nuns alive now that was there when I went to school we call them Sisters here I think we have 15 or more of them here they have a very nice convent they go to Mass with the school children and go see the sick we have two other schools besides publick schools that is for any one wishes to go there and there is a High school also**</education>** (*ALC*, 10 December 1902).

In cases where the discourse could be interpreted in more than one way, two or more tags were assigned. This meant that a section could be said to be about just one topic, or it could be said to be about a number of topics. In the example above, where the text is annotated with the tags <education></education>, an alternative interpretation might be the topic ‘Ireland and America’ (any reference to life in Ireland and/or America). In this situation, the annotation would be as follows:

**<education><IrelandAmerica>**I hope you keep them to school all you can when they grow bigger you can not send them very well I suppose there are not many of the nuns alive now that was there when I went to school we call them Sisters here I think we have 15 or more of them here they have a very nice convent they go to Mass with the school children and go see the sick we have two other schools besides publick schools that is for any one wishes to go there and there is a High school also**</IrelandAmerica></education>** (*ALC*, 10 December 1902).

Additionally, it is possible for one or more topics to be embedded within another main topic. In the example below, for instance, the main theme is work; however, within this section Annie makes a comment regarding attitudes towards work in Ireland versus America, so the tag <IrelandAmerica> has been embedded within <work>, as follows:

**<work>**is Maggie home yet or is she going to stay home all winter idle spending her money and weering out her nice clothes I think she is very foolish for her self she had ought to stay in a good place when she had one she wont get one like it for a while again I wonder she comes home to stay there now **<IrelandAmerica>**she ought to be working for herself and considering the wey things are at home now what she would do if she was here no one ever thinks of staying a week away from work unless they were sick or out of work**</IrelandAmerica></work>** (*ALC*, 15 December 1891).

Tagging for topics allowed us to get a sense of what the sisters wrote about – their preoccupations, perceptions and experiences. It also allowed us to extract all instances of a particular topic to explore how language is used within that specific context.

Table 3 lists the topics that were identified in the *ALC* and *JLC*, organised alphabetically, while Figure 1 shows the same information represented visually. While the topic categories were developed from scratch (i.e. we did not approach the data with any preexisting ideas about what we expected to find), most of the topics identified can be mapped onto Fitzpatrick’s thematic index (1994: 643-649).[[4]](#footnote-4) There are, however, a few differences, namely: a distinction is made between the topics Memories and Recollections and there is a separate category for Homesickness. Additionally, there are separate categories for Future Letters and Previous Letters and a distinction is made between remittances and other types of enclosure. Finally, the topics Postal System, World War I, Reassurance and Identity are evident in the Lough letters, but these categories do not appear in Fitzpatrick’s index. As with most qualitative approaches to topic identification, the process is largely a subjective and intuitive one. However, through tagging the topics (thereby making transparent our reading of the letters) it is possible for other researchers to build on and/or challenge our interpretation of the discourse. Indeed, one of the benefits of using XML markup to annotate the letters is that it becomes possible for other researchers to comment on this annotation and to add their own layers of interpretative markup. The markup process can thus become a collaborative effort with each iteration being fully explained and documented within the markup itself.

The ‘Freq.’ column shows how often the topics occurred across all letters and the ‘Av. p/letter’ column gives the average frequency of a topic per letter. Of course, it is possible that a topic may be mentioned several times in the same letter, or it may not be mentioned at all. Counting the number of times a topic occurs thus offers only one way into the letters. Counting the number of words attributed to each occurrence of a particular topic, on the other hand, arguably provides a more accurate reflection of the content of a letter, or a letter collection (i.e. it tells us what percentage of the letter or letter collection was spent discussing health, education or work, for instance). The ‘Word’ column, therefore, provides the raw word count for each topic while the ‘%’ column shows what percentage of the corpus was attributed to that particular topic.

Tagging the letters for topics, placing the letters in a digital frame, arguably allows the researcher to look more objectively at the spread of topics within and across collections, thereby encouraging a more neutral analytic approach. What is perhaps striking about Table 3, for example, is how rarely topics like Migration, Work and Education seem to crop up, even though these topics are precisely the ones many researchers on migrant letters have focused on (e.g. Koos 2001; Miller 1985; Miller *et al*. 1995). Instead, the far greater focus of both Annie’s and Julia’s letters is on Family and Friends (82 occurrences (39.94% of the overall content) in the *ALC* and 58 occurrences (34.81%) in the *JLC*), as well as the subtle national comparisons identified under Ireland and America (59 occurrences (22.14%) in the *ALC* and 66 occurrences (18.57%) in the *JLC*). It is perhaps unsurprising that these are the two most frequent topics in the Lough sisters’ letters given the context in which they wrote; however it would be interesting to compare different letter collections, taking into consideration sociobiographic variables, to see whether these topics are typical of migrant correspondence more generally.

Other topics, meanwhile, feature heavily because they provide a recurring structure to the letters, reflecting the rituals and demands of letter-writing itself. For instance, the topics Future Letters, Previous Letters and Enclosure are a significant part of Annie’s and Julia’s correspondence (and, arguably, correspondence more generally – see, for instance, Fitzmaurice (2012, 2015)). Table 3 shows that although these topics have very high frequencies (i.e. they may be mentioned several times in each letter), they only take up a small percentage of the overall discourse (7.74%, 8.24% and 0.24% in the *ALC* and 5.25%, 9.71% and 5.72% in the *JLC*). In other words, these tend to be short, often formulaic sections, which occur within most letters.

The following sections use digitial humanities techniques to examine these topics (as well as the topic Remittance, a specific type of enclosure) in more detail. Whilst these topics may not be very rich in terms of their content, they may provide useful insights into how transatlantic relationships were structured, reinforced and maintained through correspondence. First, we annotate the topics for information regarding the participants involved. We then analyse the topics across the letters using a combination of numerical and visual presentations. The visualizations allow us to identify patterns and trends across the letters, which would be more difficult to notice through reading alone. The analyses were done using a combination of manual and automatic methods using a variety of tools. Most of the counting was done automatically using programs we developed in Python 3, though the counting of the interactants was done manually. Calculations of percentages were done in spreadsheets, while all of the visualizations were created with the R programming language using the ggplot2 package (Wickham 2016). Finally, we evaluate the use of digital humanities tools for exploring topics and themes in the discourse.

Table 3: Topics in the *ALC* and *JLC*, listed alphabetically[[5]](#footnote-5)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ALC topics | Freq. | Av. p/letter | Word | % | JLC topics | Freq. | Av. p/letter | Word | % |
| Childbirth | 1 | 0.03 | 92 | 0.45 | Daily life | 8 | 0.23 | 320 | 2.62 |
| Death | 18 | 0.46 | 1712 | 8.39 | Death | 9 | 0.26 | 691 | 5.65 |
| Education | 6 | 0.15 | 255 | 1.25 | Education | 2 | 0.06 | 53 | 0.43 |
| Enclosure | 49 | 1.26 | 1354 | 6.64 | Enclosure | 17 | 0.49 | 699 | 5.72 |
| Family and friends | 82 | 2.10 | 8149 | 39.94 | Family and friends | 58 | 1.66 | 4255 | 34.82 |
| Future letters | 86 | 2.21 | 1580 | 7.74 | Future letters | 41 | 1.17 | 642 | 5.25 |
| Health and illness | 25 | 0.64 | 1197 | 5.87 | Health and illness | 24 | 0.69 | 563 | 4.61 |
| Homesickness & separation | 7 | 0.18 | 230 | 1.13 | Homesickness & separation | 28 | 0.8 | 737 | 6.03 |
| Identity | 1 | 0.03 | 9 | 0.04 | Identity | 6 | 0.17 | 234 | 1.91 |
| Ireland and America | 59 | 1.51 | 4517 | 22.14 | Ireland and America | 66 | 1.89 | 2269 | 18.57 |
| Marriage | 6 | 0.15 | 274 | 1.34 | Migration | 1 | 0.03 | 136 | 1.11 |
| Migration | 7 | 0.18 | 453 | 2.22 | News event | 10 | 0.29 | 476 | 3.90 |
| News event | 12 | 0.31 | 800 | 3.92 | Previous letters | 49 | 1.4 | 1186 | 9.71 |
| Postal system | 4 | 0.10 | 129 | 0.63 | Recollections | 31 | 0.89 | 978 | 8.00 |
| Previous letters | 71 | 1.82 | 1681 | 8.24 | Religion | 48 | 1.37 | 1854 | 15.17 |
| Reassurance | 5 | 0.13 | 231 | 1.13 | Remittance | 16 | 0.46 | 539 | 4.41 |
| Recollection | 39 | 1 | 1274 | 6.24 | Reunion | 10 | 0.29 | 213 | 1.74 |
| Religion | 17 | 0.44 | 961 | 4.71 | Transportation | 1 | 0.03 | 60 | 0.49 |
| Remittance | 15 | 0.38 | 438 | 2.15 | Weather and seasons | 31 | 0.89 | 774 | 6.33 |
| Reunion | 6 | 0.15 | 152 | 0.74 | Work | 23 | 0.66 | 885 | 7.24 |
| Transportation | 1 | 0.03 | 69 | 0.34 |  |  |  |  |  |
| Weather and seasons | 15 | 0.38 | 616 | 3.02 |  |  |  |  |  |
| Work | 27 | 0.69 | 823 | 4.03 |  |  |  |  |  |
| World War I | 4 | 0.10 | 396 | 1.94 |  |  |  |  |  |

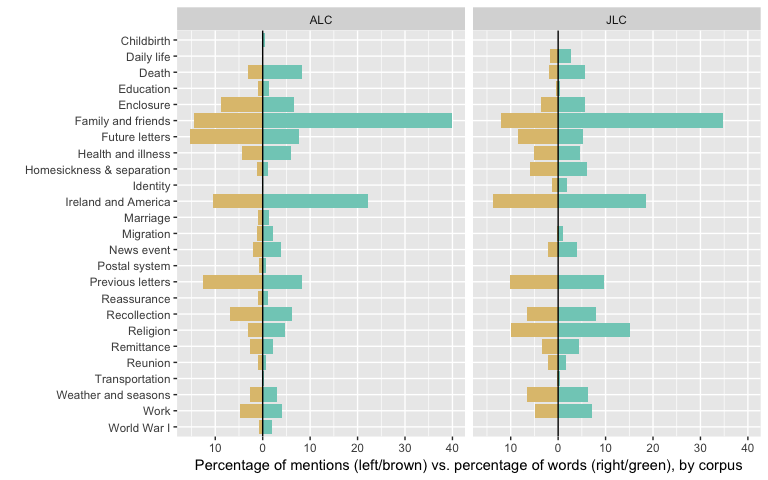


Figure 1: Topics as proportion of mentions vs. proportion of words, by corpus

There are two interesting aspects of the topics that we can see from the visualization in Figure 1. The first aspect is that a few topics show a large difference between their percentage of mentions as a topic and the percentage of words they occupy. For example, occurrences of “Family and Friends” are 10-15 % of the topic instances, but they take about 35-40% of the total words in the letters. In other words, a disproportionate amount of the content of the words is taken up by the topic Family and Friends. Ireland and America shows a similar, if less striking pattern.

The second interesting aspect of the topics is seeing the similarities and differences between the two corpora. While overall the distributions of the topics are similar in the two corpora, there are striking instances of topics that occur (primarily) in one corpus or the other. For example, ALC contains Education, Marriage, and Reassurance, while JLC does not but does contains Daily Life and Identity.

5. Using XML markup to capture information about participants

Having identified and annotated the different topics in the *ALC* and *JLC*, the next stage was to capture information about the participants involved. In other words, who was expecting a letter from whom? Who had received a letter from whom? And who was sending/receiving enclosures and remittances? To do this, each instance of the topics Future Letters, Previous Letters, Remittance and Enclosure was annotated for person information, following eXtensible Markup Language (XML) conventions.

There are various applications of XML.[[6]](#footnote-6) For this project, we referred to the *Text Encoding Initiative (TEI) Guidelines* when carrying out the encoding process. (The TEI provides a set of guidelines as well as a markup language for encoding texts in the humanities and is considered the de facto standard.) As an application of XML, TEI markup language must do certain things for it to be XML compliant – that is, it must obey certain rules for it to be considered well-formed. Those rules (as explained in the *TEI Guidelines*) are summarised below:

1. A tag (in the form of angle brackets <>) must explicitly mark the start and end of each *element*. Elements represent the structural components of a text, such as the body, paragraphs and line breaks. Elements of one type (say, a paragraph) may be embedded within elements of another type (the body, for instance).
2. There must be a single element enclosing the whole document: this is known as the root element.
3. Each element apart from the root element must be completely contained by another element; elements cannot partially overlap one another.
4. In addition to elements, there are *attributes*. Each element can possess one or more attributes. The TEI describes attributes as ‘information that is in some sense descriptive of a specific element occurrence but not regarded as part of its content’ (TEI Consortium 2008, p. xlii).
5. Other key characteristics of well-formed XML are the case-sensitivity of tag names and the fact that special reserved characters (‘<’ and ‘&’) must be escaped with entity references.[[7]](#footnote-7)

In the example below, for instance, a letter by Julia Lough to her Mother, the opening tag <futureLett> and closing tag (with forward slash) </futureLett> tell us that this section of the discourse is concerned with the topic Future Letters. The information contained within the <interaction></interaction> tags provides more details regarding who is sending (role type: from) and who will receive (role type: to) this future letter. The unique identifiers LoughPers\_003 and LoughPers\_006 used in the @key attribute tell us that in this occurrence of Future Letters Julia is referring to a letter that Annie is going to write to their Mother (Elizabeth McDonald Lough).

<futureLett>  
  <interaction>  
     <role type="from" key="LoughPers\_003" name="Annie Lough"/>  
     <role type="to" key="LoughPers\_006" name="Elizabeth McDonald Lough"/>  
  </interaction>  
I think she is going to write home before Christmas  
</futureLett>

(*JLC, 2 December 1889, Julia to her Mother*)

The unique identifier is used to create an association between the <persName> element and the @key attribute in the markup and the corresponding <person> element and @xml:id attribute located in a separate personography file, containing all known information about that particular participant (their maiden and married names and nicknames, their date of birth, occupation, educational background, places of residence etc., as well as their relationship to other participants in the corpus). An example Personography file can be found in Appendix A. See (1)-(4), below, for further examples of how the topics were annotated for participants.

1. <futureLett>  
     <interaction>  
        <role type="from" key="LoughPers\_004" name="Julia Lough"/>  
        <role type="to" key="LoughPers\_005" name="Mary Lough"/>  
     </interaction>  
   I wonder why Mary never writes to me for herself I have been expecting a letter from time to time but I suppose she has too much to take up her time now  
   </futureLett>  
    (*JLC, 1 Sept 1892, Julia to her Mother*)   
    [i.e. Julia is expecting a letter from Mary]
2. <previousLett>  
     <interaction>  
        <role type="from" key="LoughPers\_004" name="Julia Lough"/>  
        <role type="to" key="LoughPers\_006" name="Elizabeth McDonald Lough"/>  
     </interaction>  
   I received all your letters. I was so surprised to get a letter in your own dear hand writing. I think you done just splendid it was a very nice letter and I am very thankful to you I shall always treasure it as the only letter I ever received from my dear mother and will often read it. I assure you I did not cry so much in a long time as when I read your letter but I do think you ought to often write and not say that is your last  
   </previousLett>  
    (*JLC, 9 March 1890, Julia to her Mother*)   
    [i.e. Julia received a letter from her Mother]
3. <enclosure>  
     <interaction>  
        <role type="from" key="LoughPers\_005" name="Mary Lough"/>  
        <role type="to" key="LoughPers\_004" name="Julia Lough"/>  
     </interaction>  
   so you do not know how much I felt when I looked upon your face again if only in a picture it is a very poor picture and I always considered you good looking and if it was done decent it would be a good picture Lizzie and Nan said they would never know you but I should I noticed in particular how nice your hands looked and you had flowers under the side of you hat. I am very thankful to you I am sure you had trouble to get it  
   </enclosure>  
    (*JLC, 4 June 1894, Julia to Mary*)   
    [i.e. Mary sent an enclosure to Julia]
4. <remittance>  
     <interaction>  
        <role type="from" key="LoughPers\_004" name="Julia Lough"/>  
        <role type="to" key="LoughPers\_006" name="Elizabeth McDonald Lough"/>  
     </interaction>  
   I am sending you one pound and five shillings to get two masses said for fathers and Mothers souls I suppose you can include both. The one pound is your xmas present and let me know will the five shillings pay for the two masses  
   </remittance>  
    (*JLC, n.d. November 1894, Julia to Mary*)   
    [i.e. Julia sent a remittance to Mary]

Once annotated in this way it is possible to extract topic and participant information so as to analyse differences in letter writing practices between the two sisters.

6. Analysing differences in letter writing practices

Table 4 provides an overview of the number of occurrences of each topic, Future Letters, Previous Letters, Remittance and Enclosure, across all letters in the *ALC* and *JLC*.

Table 4: Frequency Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ALC | Av p/letter | JLC | Av p/letter |
| futureLett | 87 | 2.23 | 42 | 1.20 |
| previousLett | 71 | 1.82 | 49 | 1.40 |
| enclosure | 49 | 1.26 | 16 | 0.46 |
| remittance | 14 | 0.36 | 14 | 0.40 |

Some of the letter writing practices/patterns as they emerge over the course of letters are highlighted more clearly in the following charts. In Figure 2, we have the distributions of the topics in each of the two corpora. One of the striking things we can observe is how much more Annie talks about enclosures than Julia does. There are also strong differences between the two sisters with respect to remittances, though for both of them the topic is much less common than the other three. On the other hand, their patterns for previous and future letters are broadly similar, with more mentions of future letters early in the sequences.

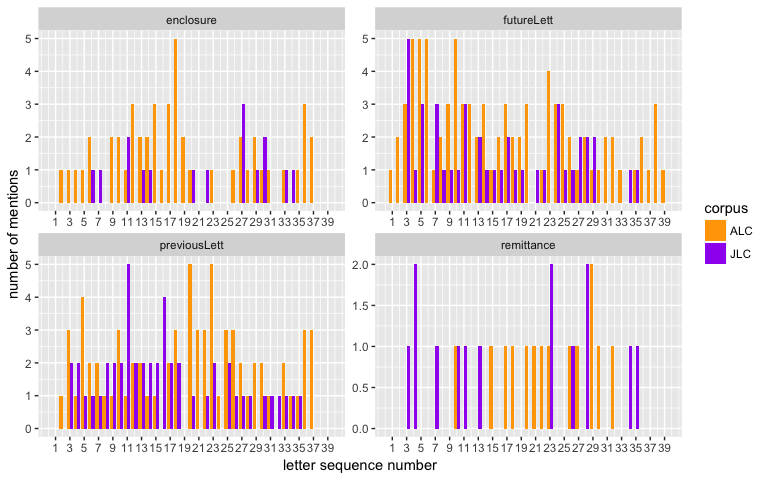


Figure 2: Absolute frequency of selected themes in ALC and JLC, grouped by theme

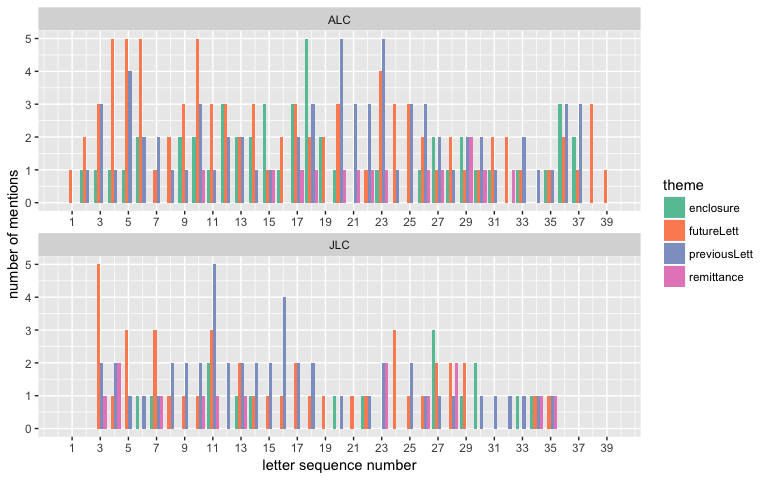


Figure 3: Absolute frequency of selected themes in ALC and JLC, grouped by author

In Figure 3, we have the same information as in Figure 2, but viewed from the perspective of the authors. Here we can see that Annie’s discussion of these four topics is much denser than Julia’s. Annie almost always mentions several topics in a letter, and very often touches on all four topics. Julia, on the other hand, typically mentions just one or two of the topics. While this could be connected to Annie’s longer more varied writing in general, there is no a priori reason why these particular topics should be so much denser in Annie’s letters.

In Figures 4 and 5, we have the same information as in Figures 2 and 3, but showing the topics in relation to the length (in words) of the letters. There is not much difference here, except that a couple of longer letters by Julia diminish the prominence of certain topics. For example, looking at Figure 5, the five instances of futureLett in letter three are similar in proportion to the three instances of futureLett in letter five, simply because letter three is longer than letter five (519 words vs. 342 words). While, these differences are small and the exception rather than the rule (and seemingly more common for the *JLC* corpus), they are worth noting in passing.

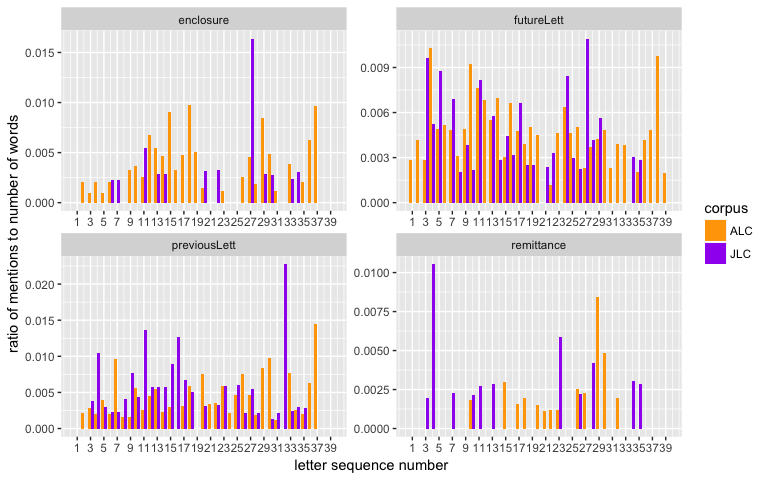


Figure 4: Relative frequency of selected themes in ALC and JLC, grouped by theme

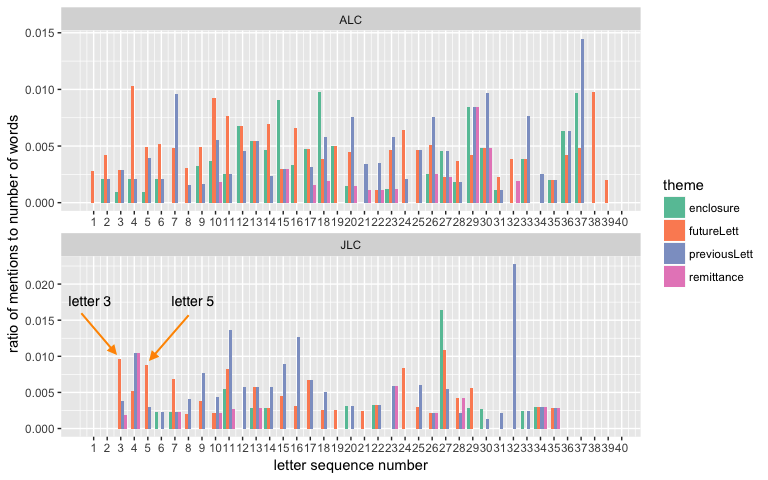


Figure 5: Relative frequency of selected themes in ALC and JLC, grouped by author

Using the markup described in section 5, we were then able to look more closely at the interactants, i.e. the people being talked about in Annie’s and Julia’s letters. Figure 6 gives an overview of the patterns of interactants. Each subchart shows the patterns for the “to” interactant labelled at the top of the subchart, with the corresponding “from” interactants in the rows. The numbers of mentions in Annie’s letters are indicated by the size of the circles, while triangles indicate the number of mentions in Julia’s letters.

Most obviously, we see that the combinations of interactants is restricted (e.g. only Annie refers to Friend; only Annie, Mother and Julia refers to NN=nieces and nephews, etc.). However, there are two more subtle observations that we can make. First, both corpora have the same combinations of interactants (i.e. at every line intersection in the chart we have both a circle for *ALC* and a triangle for *JLC*). This means that not only are Annie and Julia talking about the same interactants, but the interactants have the same to and from roles in both corpora. The second subtle observation is that Annie and Julia talk about the interactants in roughly the same proportions, with the natural exception that each talks about themselves more than about others.



Figure 5: Interactants and themes in the corpora (circles are for ALC, triangles for JLC)

8. A closer look at the topics Future Letters and Previous Letters

Digital humanities techniques (annotating the Lough letters for topics and personography information as well as using computational and visualisation tools to extract metadata and present it in different ways and from different viewpoints) has allowed us to identify potentially interesting features of the letters.

Overall Annie’s discussion of the four topics (Future Letters, Previous Letters, Remittance and Enclosure) is a lot denser than Julia’s. Additionally, whereas Annie seems to cover all four topics in each letter, Julia may cover just one or two.

While Annie sent significantly more enclosures (typically newspaper cuttings) than Julia, both sisters sent remittances. There are, however, some differences in the spread of these remittances over the course of their correspondence. Julia, for instance, consistently sent remittances to her mother earlier on in her correspondence. The remittances then stopped for some time (possibly because these were Julia’s childrearing years). In her later letters (after her mother’s death), the remittances continue, this time to her younger sister, Mary. Annie, in contrast, primarily sent remittances to her younger sister, Mary, regularly and without any significant gaps.

Both Annie and Julia frequently talk about future and previous letters in their correspondence, with both sisters tending to write more about future letters (letters which they plan/hope to send/receive) in their earlier correspondence. In her later correspondence, Annie (unlike Julia) talks in some depth about previous letters she received.

A closer look at the interactants involved in the four topics shows that both sisters corresponded with, and wrote about, the same individuals – typically family members in Ireland and America. However, in the context of writing about previous and future letters, Annie tended to refer to a wider network of people. In other words, while Julia tended to focus on letters sent between her and her mother (*I received your letter*, *I will write soon*), Annie was more likely to write about letters that her Mother and Sister received or can expect to receive from other family members. Given that Annie mainly wrote to her younger sister Mary, while Julia mainly wrote to her mother, our initial findings could suggest that different letter writing practices reflect different familial roles. Indeed a closer look at the language of the topic Future Letters (using the corpus tool *Sketch Engine* (Kilgarriff and Kosem 2012)) showed that Annie tended to use more direct (write soon) and indirect (I hope you will write soon) when referring to future correspondence. In contrast, Julia (mainly writing to her mother) was less likely to instruct the recipient to write a letter. Rather the pattern *I will write soon* is most typical in Julia’s letters.

Table 5 shows the most frequent lemmas within the topic Future Letters in both the *ALC* and *JLC*. A closer look at the lemma WRITE (which occurs more than twice as often in the *ALC* when compared with the *JLC*) revealed that it is typically used in the following structures in the *ALC*:

1. *when you write + infinitive* (as in, *when you write* + *let me know how are they*, *remember me to her*, *tell Lizzie she must write to me*, *let me know if old Mrs Cassen is deed yet)* (19 occurrences)
2. *I hope you will write* (as in, *I hope you will write her often*) (15 occurrences)
3. *and write* + *adverb/noun phrase* (as in, *and write soon to me*, *and write a longer letter next time*) (7 occurrences)
4. *I will write* (as in, *I will write again by Christmas* or *I will write to you soon again*) (7 occurrences)

Julia, in contrast, typically uses the lemma WRITE in the structure *I/she + will + write* (9 occurrences) and *I/she + is going to write* (6 occurrences), with 5 occurrences referring to the process of learning to write (as in, *she is not able to read or write yet*). Only 5 of the 37 instances of the lemma WRITE in the *JLC* instruct the recipient of the letter (*you*) to write. In short, while both sisters refer to future letters in their correspondence, Annie is much more likely to be instructing or directing the recipient to write, whereas Julia is more likely to be confirming and reassuring the recipient that she will write again soon.

Table 5: Most frequent lemmas in the topic Future Letters in the *ALC* and *JLC*

|  |  |  |  |
| --- | --- | --- | --- |
| **ALC** | | **JLC** | |
| **Lemma** | **Freq.** | **Lemma** | **Freq.** |
| you | 134 | I | 74 |
| I | 85 | be | 66 |
| and | 74 | you | 66 |
| write | 72 | to | 57 |
| to | 63 | write | 37 |
| will | 59 | and | 32 |
| me | 50 | will | 29 |
| be | 46 | she | 25 |
| hope | 42 | have | 23 |
| soon | 39 | letter | 21 |
| all | 35 | a | 21 |
| know | 28 | the | 21 |
| send | 28 | all | 19 |
| the | 27 | not | 17 |
| let | 27 | hope | 16 |
| when | 26 | me | 16 |
| a | 25 | well | 16 |
| from | 23 | from | 14 |
| get | 21 | her | 14 |
| how | 20 | when | 13 |
| time | 19 | it | 12 |

Finally, a keyword search of the topics Future Letters and Previous Letters in both the *ALC* and *JLC* revealed a noticeable amount of emotive language. Both sisters express anxiety over not receiving a letter (*I am very anxious to hear from Mary I hope to get good newes from her* (*ALC*)), they write about crying at the sight of a letter from home (*I assure you I did not cry so much in a long time as when I read your letter* (*JLC*)), and they write about feelings of happiness upon hearing news about loved ones (*I am very happy to hear from you and to hear you are in the enjoyment of good health* (*JLC*)). Whilst these topics may appear on the surface to be rather formulaic in nature (*I hope you will write* is the most common n-gram in the topic Future Letters in both corpora, for instance) their function is, arguably, much deeper than that: these topics reassure the recipient that they are missed and thought about and as such perform a key role in maintaining familial relationships.

9. Conclusion

This chapter set out to explore the use of digital humanities techniques for examining letter writing practices in the context of nineteenth century mass migration to America. Specifically, our aim was to explore how digital/visualisation tools (i.e. new methods) can help us to notice patterns and trends in old data. Whilst our findings may not be entirely surprising and/or revealing, through repeating the process we have described here it may be possible to see whether our observations are typical (or not) of other letter collections.

There are two aspects of digital methods used here, XML markup and visualisation, each of which provides certain benefits. The XML markup is a way of providing annotations which has several advantages over non-digital markup. One advantage is that it is easy to have detailed, overlapping annotations, such as a topic within a topic. Another advantage is that it is easy to check for consistency of the annotations. A third advantage of digital markup is that it is easy to tally the various annotations for further analysis, including input to the visualisations.

The visualisations are made possible not only by having the digital data, but by the existence of high level tools that allow for the rapid creation of complex visualisations, which in turn facilitates exploration of the data in various ways. Visualisations enable us to discover patterns that we might not have otherwise noticed, or even thought to look for, such as the similarity of the roles of the interactants in *ALC* and *JLC* shown in Figure 5. Visualisations also allow to us communicate patterns we have discovered, and allow others to discover new patterns. Of course, any visualisation embodies assumptions and perspectives, and the visualisations presented here are no different: they were chosen from a variety of alternatives to highlight certain aspects of the data, to communicate certain ideas.

In short, both methods (markup and visualisation) allow us to analyse the Lough letters in ways that would have been extremely difficult to do in a non-digital fashion. The use of digital methods allows for quick, detailed exploration of the letters. For example, the charts in Figures 2-5 took minutes to prepare digitally from the calculated counts, not hours. It is also very easy to change perspectives, such as from the counts in Figures 2 and 3 to the ratios in Figure 4 and 5. In addition, using digital annotations allows for different future analyses without having to re-annotate the letters. Of course, since this corpus is relatively small, all of the work here *could* have been done manually, at a much greater cost in time and effort. However, for larger corpora, it simply would not be feasible to these kinds of studies without digital methods.

Digital methods are not a panacea, of course. For this study, the annotation of themes had to be done manually, since there are no digital tools for the task. Digital analysis tools typically depend on the existence of large annotated corpora similar to the one to be analysed, and there are no corpora of immigrant letters annotated for these themes, or even similar themes. At this point, we can only speculate whether automatic theme analysis will ever be possible for corpora such as the one studied here. Our view is that certain aspects will be possible to annotate automatically, aspects which occur in other, larger corpora (e.g. death, religion, weather, etc). However, more narrow topics (e.g. future letters, Ireland and America) may be more difficult.

The results presented in this chapter are provided as illustrative guides to the kinds of insights and findings made available by these methods. The next step will be to carry out similar studies across other collections to see the extent to which the Lough letters reflect letter writing practices of other migrant families and groups. For example, do male and female migrants tend to write as frequently and about similar topics? Do they send similar numbers of remittances and enclosures? Are the topics Future Letters and Previous Letters prevalent across other migrant letter collections, and to what extent do letter writing practices (such as those described in this chapter) reflect different familial roles? Whilst identifying topics and themes in the discourse is a largely subjective process, the use of digital markup makes that process transparent, allowing others to build on, contribute to and challenge the various categories. The process of topic identification (and the resultant analysis of those topics) thus becomes a collaborative process and opens up the possibility for cross disciplinary discussion.

References

Amador Moreno, C. P. and McCafferty, K. (2015) ‘“Sure this is a great country for drink and rowing at elections”: Discourse markers in the *Corpus of Irish English Correspondence*, 1750–1940’, in Amador Moreno, C. P, McCafferty, K. and Vaughan, E. (eds.) *Pragmatic markers in Irish English*. Amsterdam: John Benjamins, pp. 270–291.

Barton, H. A. (1990) *Letters from the Promised Land: Swedes in America, 1840-1914*. Minneapolis: University of Minnesota Press.

Bruce, S. U. (2006) *The Harp and the Eagle: Irish-American Volunteers and the Union Army, 1861-1865*. New York: New York University Press

Conway, A. (1961) *The Welsh in America: Letters from the Immigrants*. Minneapolis: University of Minnesota Press.

Corrigan, K. P. (1992) I gcuntas De muin Bearla do na leanbhain’: eisimirce agus an Ghaeilge sa naou aois deag. In P. O’Sullivan (ed.), *The Irish World Wide*. Leicester: Leicester University Press. pp. 143–161.

Diner, H. R. (1983) *Erin’s daughters in America*. Baltimore, MD: The Johns Hopkins University Press.

Elliott, B. S., Gerber, D. A. and Sinke, S. (eds.) (2006) *Letters Across Borders : Epistolary Practices of International Migrants*. London: Palgrave Macmillan.

Emmons, D. M. (1990) *The Butte Irish: Class and Ethnicity in an American Mining Town, 1875-1925*. Urbana: University of Illinois Press.

Erickson, C. (1972) *Invisible Immigrants: The Adaptation of English and Scottish Immigrants in Nineteenth-Centrury America*. London: Weidenfeld & Nicolson.

Fitzmaurice, S. (2015) Aristocratic Letters. In A. Auer, Schreier, D. & Watt, D (eds.) *Letter Writing and Language Change*. Cambridge: Cambridge University Press.

Fitzmaurice, S. (2012) Sociability: Conversation and the Performance of Friendship in early eighteenth century letters. In Ulrich Busse & Axel Hübler (eds.) *The Meta-communicative Lexicon of English Now and Then: A Historical Pragmatics Approach.* Amsterdam: John Benjamins. pp. 21-43.

Fitzpatrick, D. (1994) *Oceans of Consolation: Personal Accounts of Irish Migration to Australia*. Cork: Cork University Press.

Gerber, D. A. (2006) *Authors of Their Lives: The Personal Correspondence of British Immigrants to North America in the Nineteenth Century*. New York: New York University Press.

Kamphoefner, W. D., Helbich, W. and Sommer, U. (1988) *News from the Land of Freedom: German Immigrants Write Home*. Ithaca: Cornell University Press.

Kilgarriff, A. and Kosem, I. (2012) Corpus Tools for Lexicographers. In S. Granger and M. Paquot (eds.), *Electronic Lexicography*. New York: Oxford University Press. Pp. 31- 56. Available from: http://www.sketchengine.co.uk.

Koos, G. (2001) The Irish Hedge Schoolmaster in the American Backcountry. *New Hibernia Review* 5: 9-26.

McCafferty, K. and Amador Moreno, C. P. (2012) ‘“I will be expecting a letter from you before this reaches you”: A corpus-based study of *shall/will* variation in Irish English correspondence’, in Dossena, M. (ed.) *Letter writing in late modern Europe*. Amsterdam: John Benjamins, pp. 179–204.

McCafferty, K. and Amador Moreno, C. P. (2014) ‘“If you write soon I shall get it & will reply at once” The spread of first-person future *will* in Irish English’, *English Language and Linguistics*, 18, pp. 407–429.

Miller, K. A. (2008) *Ireland and Irish America: Culture, Class, and Transatlantic Migration*. Dublin: Field Day.

Miller, K. A. (1985) *Emigrants and Exiles: Ireland and the Irish Exodus to North America*. New York: Oxford University Press.

Miller, K. A., Doyle, D.N. and Kelleher, P. (1995) For Love and Liberty: Irish Women, Migration and Domesticity in Ireland and America, 1815–1920. In P. O’Sullivan (ed.), *The Irish World Wide*. Leicester: Leicester University Press. pp. 54–61.

Miller, K. A., Schrier, A., Boling, B. D. and Doyle, D. N. (2003) *Irish Immigrants in the Land of Canaan: Letters and Memoirs from Colonial and Revolutionary America, 1675-1815*. New York: Oxford University Press.

Montgomery, M. B. (1995) The linguistic value of Ulster emigrant letters. *Ulster folklife* 41: 26-41.

Nolan, J. A. (1989) *Ourselves Alone: Women’s Emigration from Ireland, 1885-1920*. Lexington: University Press of Kentucky.

Noonan, A. J. M. (2011) ‘Oh those long months without a word from home’, Migrant letters from mining frontiers. Cork: The Boolean, University College Cork. pp. 129–135. [Available from: <http://publish.ucc.ie/boolean/>].

Plummer, K. (2001) *Documents of Life 2: An Invitation to a Critical Humanism*. London: SAGE Publications.

Rayson, P. (2009) *Wmatrix*. Lancaster: Lancaster University. Available at: <http://ucrel.lancs.ac.uk/wmatrix/>.

Van Hattum, M. (2014) ‘“Queensland for rver & Augus un ballybug go braugh”: The expression of identity in nineteenth-century Irish emigrant letters’, in Evans, D. (ed.) *Language and identity: Discourse in the world*. London: Bloomsbury, pp. 105–122.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. available at http:ggplot2.org, accessed 11 March 2019.

Zempel, S. (1991) *In Their Own Words: Letters from Norwegian Immigrants*. Minneapolis: University of Minnesota Press.

Appendix A

Markup template for ‘Personography’ using sociobiographic information relating to Elizabeth Lough

<listPerson xmlns="http://www.tei-c.org/ns/1.0">

<person xml:id="LOUGHPers\_0001">

<persName>

<forename type="first">Elizabeth</forename>

<surname type="married">Walsh</surname>

<surname type="married" subtype="altSpelling">Welch</surname>

<surname type="maiden">Lough</surname>

<surname type="maiden" subtype="altSpelling">Locke</surname>

<surname type="maiden" subtype="altSpelling">Lowe</surname>

<addName type="nick">Liz</addName>

<addName type="nick">Lizzie</addName>

</persName>

<birth notBefore="1845-01-01" notAfter="1849-31-12">Unknown</birth>

<death notBefore="1910-01-01" notAfter="1914-31-12">approx. 1912</death>

<occupation key="HISC#79510" ref="http://historyofwork.iisg.nl/detail\_hiswi.php?know\_id=20311&lang">seamstress</occupation>

<occupation key="HISC#-1" ref=" http://historyofwork.iisg.nl/detail\_hiswi.php?know\_id=19930&lang">householder</occupation>

<socecStatus key="9" ref="http://historyofwork.iisg.nl/docs/hisco\_hisclass12\_book@\_numerical.inc">lower-skilled</socecStatus>

<education key="MS"/>

<sex value="2"/>

<faith key="RC"/>

<residence notBefore="1870-01-01" notAfter="1874-31-12">Unknown between 1870 and 1875</residence>

<residence key="LOUGHPlace\_0001" notBefore="1875-01-01" notAfter="1879-31-12">1876-1878</residence>

<residence notBefore="1880-01-01" notAfter="1914-31-12">Unknown between 1879 and 1912</residence>

<listEvent>

<event type="emigration" notBefore="1850" notAfter="1870" whereFrom="LOUGHPlace\_0006" whereTo="LOUGHPlace\_0001"><p>before 1870</p>

<desc whereFrom="LOUGHPlace\_0006" whereTo="LOUGHPlace\_0004" <name type="transportation">Unknown</name></desc>

<desc whereFrom="LOUGHPlace\_0004" whereTo="LOUGHPlace\_0014" <name type="transportation">Ship</name></desc>

<desc whereFrom="LOUGHPlace\_0014" whereTo="LOUGHPlace\_0001" <name type="transportation">Unknown</name></desc>

</event>

<event type="marriage" notBefore="1870-01-01" notAfter="1874-31-12">before 1876</event>

</listEvent>

</person>

<relationGrp>

<relation name="spouse" mutual="LOUGHPers\_0001 LOUGHPers\_0002"/>

<relation name="childOf" active="LOUGHPers\_0001" passive="LOUGHPers\_0007 LOUGHPers\_0008"/>

</relationGrp>

</listPerson>

1. See, for example: The Mellon Centre for Migration Studies, Ulster American Folk Park Museum (2012-present) *The Irish Emigration Database (IED)*. Available from: <http://www.dippam.ac.uk/ied/> [Accessed 1 October 2018].

   Immigration History Research Centre, University of Minnesota (2010-present) *Digitizing Immigrant Letters*. Available from: <http://ihrc.umn.edu/research/dil/> [Accessed on 1 October 2018]. [↑](#footnote-ref-1)
2. See, for instance, Plummer’s discussion of what he describes as ‘dross rate’: ‘Letters are not generally focused enough to be of analytic interest – they contain far too much material that strays from the researcher’s concern’ (2001: 55). As well as David Fitzpatrick’s observation that oftentimes scholars working with migrant letter collections have ‘…shared the widespread impatience of editors with material deemed “tedious for the non-specialist,” including “ritualized pious reflections” and “endless lists of persons to whom the letter-writer wishes to send his or her best regards.”’ Consequently, what has been viewed as ‘uninteresting’ or ‘irrelevant’ material within these letters has, quite often, simply been omitted (1994: 21). [↑](#footnote-ref-2)
3. The authors are indebted to Kerby Miller for the information he generously provided relating to the Lough family. [↑](#footnote-ref-3)
4. Fitzpatrick, analysing 111 letters from/to Irish migrants in Australia, identified around 140 main themes and almost 250 sub-themes within his letters. [↑](#footnote-ref-4)
5. Freq. = shows the number of times this topic occurs within the corpus / Av. p/letter = shows the average number of times this topic occurs per letter / Word = shows the number of words attributed to this topic / % = shows what percentage of the corpus is attributed to this topic. [↑](#footnote-ref-5)
6. See, for instance, the MEI (Music Encoding Initiative) Available from: <http://en.wikipedia.org/wiki/Music_Encoding_Initiative> [Accessed 1 October 2018]; and the MathML (Mathematical Markup Language) Available from: <http://en.wikipedia.org/wiki/MathML> [Accessed 1 October 2018]. [↑](#footnote-ref-6)
7. Certain characters cannot be used within XML because they have special meanings. These include the ampersand (&), “double quotes”, ‘single quotes’ and <angle brackets>. Many of the Lough letters contain, for example, the ampersand in the main body of the text, so these characters need to be ‘escaped’ with a predefined entity reference as follows: &amp;. For more on entity references see: *TEI by Example* (2004) Available from: <http://www.teibyexample.org/modules/TBED00v00.htm?target=xmlgroundrules> [Accessed 1 October 2018]. [↑](#footnote-ref-7)