Towards a historical corpus of Canadian English letters and diaries

Giulia Rovelli

University of Bergamo

ABSTRACT

The paper reports on the compilation and illustrates the main features of a corpus of manuscript ego documents written in English by both adults and children in the nineteenth and early-twentieth centuries in Canada – a crucial time in the history of the country and in the development of the Canadian variety of English. The corpus is part of a larger project that intends to contribute to Canadian studies and English historical sociolinguistics by a) tracing the evolution of the specificities of Canadian English in a diachronic perspective; b) investigating the contribution of different languages or varieties of the same language to the development of Canadian English; and c) analyzing how demographic and social differences are encoded in and have influenced language use in Canada. Some preliminary findings are also presented.

Keywords: Canadian English, corpus linguistics, historical sociopragmatics, ego documents, personal letters, diaries.

1. A short history of Canadian English

The term "Canadian English" was first used in 1857 by Reverend A. Constable Geikie with regard to the "lawless and vulgar innovations" that, in his opinion, characterized the English language as spoken in Canada (DCHP-1, Avis et al. 1967; Chambers 1993). Non-derogatory uses of the expression, which may provide evidence of increasing language awareness and nationalism and, consequently, of the language having entered its endonormative stabilization phase (Schneider 2007; Dollinger 2014: 104), emerged only much later (Dollinger 2019a: 20). This is also suggested by

the relatively late publication of such metalinguistic materials as dictionaries (e.g., Avis et al.'s 1967 *Dictionary of Canadian English: The Senior Dictionary*, Avis et al.'s 1967 *Dictionary of Canadianisms on Historical Principles* and Barber's 2004 *Canadian Oxford Dictionary*) and usage guides (e.g., Fee and McAlpine's 1997 *Guide to Canadian English Usage*). However, while "Canadian English" gained currency only in the second half of the twentieth century, the English language has had a much longer history in Canada.

Although English was introduced to Newfoundland in 1497, the territory remained a separate political entity until 1949. For this reason, the language developed fairly independently from what is now referred to as Standard Canadian English (Schneider 2007: 238; Boberg 2010: 26). The origins of Canadian English (CanE) are therefore generally traced back to the late eighteenth century and to the first two of the five waves of immigration that shaped Canada's population and, ultimately, the main features of this variety (Dollinger 2015: 26). Indeed, while the Treaty of Paris (1763), which concluded the Seven Years' War (1756-1763), sanctioned the passage of France's North American colonies (i.e., Nouvelle-France) to Britain, British rule did not initially modify the demographic dominance of French (Edwards 1998: 19-22).

The first wave of immigration took place in the wake of the American Revolution, between 1776 and 1812. It consisted in two main groups of American settlers. The first, and smaller group, generally referred to as the United Empire Loyalists, was made up of anti-revolutionary Americans from Connecticut, Massachusetts, Maine and Rhode Island who chose to maintain their allegiance to the British Crown and settled mostly in Nova Scotia. The second, and more prominent group, consisted of the so-called "late Loyalists", Americans from Pennsylvania, New Jersey, New York and Vermont who, on the other hand, were more interested in new land for agricultural purposes and first settled modern-day New Brunswick and Ontario (Chambers 1998: 258-261; Boberg 2012: 163). Since the latter group of first-wave immigrants effectively represented the "founder population" (Schneider 2007: 240) of the territory that was to become the financial and industrial hub of the country (i.e., Ontario, Chambers 2004: 225), it has been described as the most significant from both a linguistic and cultural point of view (Bloomfield 1948; Mackey 1998: 22).

The second wave of immigration took place between 1812 and 1865 and consisted of mostly government-sponsored immigration from the British Isles, and specifically from Ireland, Northern England and Scotland,

which was supposed to have two main aims: on the one hand, it would "dilute" the American base in Canada, which was (wrongly) suspected of being disloyal to the British Crown, and, on the other, it could relieve Britain of paupers and demobilized soldiers (Boberg 2012: 164; Dollinger 2019a: 14). While this second wave of immigration increased the prestige of all things British, including language forms, on Canadian soil, its impact has been described as quite limited from a linguistic point of view as, with the exception of a few isolated regions (e.g., Cape Breton Island, the Ottawa valley and Peterborough County, see Chambers 2004: 228), it did not alter the basically North American character of the language (Chambers 1998: 262; Dollinger 2019a: 14).

The third wave of immigration has been traced back to the period between 1890 and 1914 and consisted of immigration from not only Scotland and Ireland, but also Continental Europe, especially from Germany, Scandinavia and Ukraine (Chambers 1998: 264). The fourth wave took place in the wake of WWII, between 1946 and 1970, and consisted a highly diverse immigrant population from Europe, Asia and Latin America (Dollinger 2019a: 12). Finally, the fifth wave began in the 1990s and has been described as characterized by an even more varied immigrant population (Dollinger 2019a: 12). While certainly fundamental for the development of the country and its identity, the latter three waves of immigration had a more limited linguistic impact, as later immigrant populations tend to assimilate very quickly and conform to dominant patterns.

Because of this, for a long time Canadian English has been described simplistically as a conservative and homogenous language, and as a variety of American English with a British orientation – a rather restrictive definition, which fails to recognize the independence and autonomy of Canadian English (Dollinger 2008: 134, 2012b, 2015), effectively relegating it to a secondary position. Many studies have recently been carried out in an attempt to rectify the situation and highlight the specificities of this variety of English, especially from the point of view of lexis and phonology (e.g., Clarke et al.'s 1995 discussion of the Canadian Shift, Berger's 2005 study of lexical variables, Chambers's 2006 investigation of Canadian Raising, Dollinger's 2008 analysis of the use of modal auxiliaries in early Ontario English, and Boberg's 2010 monograph on Canadian English, among others). Even so, Canadian English continues to be a relatively under-researched variety, especially from a historical perspective (Dollinger 2008, 2012a, 2019b; Boberg 2010, 2017).

2. Aims

The present paper reports on the compilation of the new *Corpus of Canadian English Letters and Diaries* (CCanDL), a historical corpus of manuscript ego documents written in Canada between the late-eighteenth and early-twentieth centuries, which is being developed at the University of Bergamo, Italy, in collaboration with the *Rural Diary Archive* and the *McLaughlin Library* of the University of Guelph, Ontario (CA).[1] The project intends to contribute to Canadian English studies, which, as lamented by some scholars, still lack a diachronic dimension in full detail (Dollinger 2008, 2012a, 2019b; Boberg 2010, 2017). In addition, it aims to advance studies in English historical sociolinguistics by a) tracing the origins and diachronic evolution of Canadian English and its specificities; b) investigating the contribution of different languages (most importantly French and some of Canada's indigenous languages) and of different varieties of English to the development of Canadian English; and c) analyzing how demographic and social differences are encoded in and have influenced language use in Canada.

3. Corpus design

As the main goal of the project is to investigate the origins and diachronic evolution of the English language as used in Canada, CCanDL has been designed to include private letters and diaries written in the several provinces of Canada between the late-eighteenth and early-twentieth centuries. Personal letters and diaries have been chosen precisely because of their private nature: since they were written for the authors themselves (in the case of diaries) or for their immediate family (in the case of letters, but sometimes applicable to diaries as well), these genres generally provide evidence of less monitored linguistic choices, which might more closely reflect vernacular language – that is, how language is used in everyday situations (Tieken-Boon van Ostade 2005; Dossena 2012; Allen 2014; Dollinger 2014). Indeed, letters and diaries represent prototypical texts of immediacy, which are often characterized by orality, and for this reason they have been described as the closest approximation to spoken language (Elspaß 2012; Van der Wal – Rutten 2013). Although focusing on written

---

[1]   Gratitude is expressed to both institutions for their generous cooperation. Once completed, the corpus will be made freely available in order to facilitate research into the historical development of Canadian English.

texts has its limitations, in that it restricts the investigation to the language use of literate informants who had the means of putting their thoughts on paper, this approach represents the only viable option – a necessary evil if one wants to analyze how language was used before the invention and widespread availability of the means to record and reproduce actual speech. Indeed, since the earliest recordings date back to the late nineteenth century,[2] much of the period under examination here would be excluded. Moreover, as stated by Hickey (2017), early recordings are normally very short, thus limiting the range of data; they generally consist of readings of set pieces, making free speech rarely available; and they tend not to present social stratification (Hickey 2017: 1-4).

The timeframe for the collection of such documents has been established as between 1776 and 1950. Following Dollinger (2008), the start date for the corpus has been set to 1776, the year which gave birth to the United States and, as a consequence, to British North America, namely, Canada (Chambers 2004: 225). The end date, on the other hand, has been set to 1950, as it was in the wake of WWI and WWII that linguistic nationalism started to peak in Canada, thus resulting in linguistic awareness and autonomy, which have been described as the direct consequences of the "coming of age" of Canada, and in the production of Canadian metalinguistic materials (Dollinger 2012b: 452).

The texts to be included in the corpus are being retrieved from the *Rural Diary Archive*, an online archive that collects rural diaries written in Ontario (which are being transcribed by volunteers all over the world),[3] and from the *McLaughlin Library* manuscript archives. As their being based in Ontario would most likely bias the corpus towards that particular region and most certainly provide further evidence for the Loyalist Base Theory (cf. Bloomfield 1948), other archives will also be searched so as to provide as

---

2   While the first attempts to record and reproduce sound waves date back to the phonautograph, a mechanical sound-recording device invented in 1857 by Édouard-Léon Scott de Martinville, the first device that could actually record and play back sounds was Thomas Alva Edison's phonograph, invented in 1877 (cf. https://www.britannica.com/science/acoustics/Amplifying-recording-and-reproducing#ref527605).

3   Although using manuscript materials that have being transcribed by volunteers is not entirely unproblematic, as it might result in unsystematic biases and errors, this approach allows the compilation of a much bigger corpus than would be possible if the author alone were to transcribe the material. Moreover, the *Rural Diary Archive* offers prospective volunteer transcribers instructions on how to transcribe texts and deal with most issues. Finally, since the Archive collects both the digitized and the transcribed version of each file, the original text can be referred to and checked in case of doubtful transcriptions.

balanced an account as possible. In order to be included in the corpus, the authors of the manuscript materials in question must have either been born in Canada or emigrated there permanently. On the other hand, the corpus does not include the following text types:

a) official or semi-official letters, whose authorship is more difficult to trace, as the sender might not have been the encoder of the letter, and, precisely because of their official or formal purpose, tend to be characterized by controlled, rather than spontaneous, linguistic choices; [4]

b) published material, since it is impossible to differentiate the author's language from the editor's or publisher's on account of editorial interventions;

c) memoirs and other types of autobiographical narratives that entail a certain degree of identity construction and reconstruction, which, because of their partly fictional nature, would be further removed from actual vernacular language use.

Each document included in the corpus will be accompanied by a detailed fact file (see Fig. 1) that reports the author's biographical and demographic information – data which are essential for sociopragmatic studies. These include the writers' name and gender; their place of birth and ethnicity, which might offer an indication of the contribution that different groups of speakers and their original dialects had on the development of Canadian English; their date of birth and age at the time of writing, which may help shed light on the diachronic evolution of the language; and location, which could provide some clues to its geographical differentiation. The fact files will also include information on the writers' religion and occupation, which could give an indication of their level of education; and close relatives, which, on the other hand, may be used to trace communities of practice. Such information will be retrieved directly from the archive and, when available, enriched by census data. The most important pieces of information will also be reported in the file name, which will include the diarist's name, gender, year of writing, any outstanding detail (e.g., ethnicity, religious affiliation, age, etc.), the standardized date of the entry, and the date as it appears in the diary itself (e.g. F. 1887 (German) – Middagh, V. – 1887-01-01 (1 Saturday); M.c. 1866 – Clarke, C.K. – 1866-02-21 (Wedensday Feb 21st 1866)). This will

---

[4]   Although more insecure writers might have relied on their more experienced relatives and friends, or even scribes, to help them write personal letters as well, this is less likely to be the case than for official or semi-official correspondence.

allow the creation of *ad hoc* sub-corpora that could be set up to investigate the contribution or idiosyncrasies of a specific writer or socio-demographic group.

Name: George Hill

Surname: Detlor

Gender: Male

Year of Birth: 1794

Year of Death: 1883

Age at Time of Writing: 28-55

Place of Birth: Ontario

Ethnicity: Irish
- Father's place of birth: US
- Mother's place of birth: US

Residence: Ontario
- County: Lennox & Addington
- Township: Fredericksburg

Religion: Methodist

Occupation: Farmer, Merchant, Mill-builder, Politician

Relationships:
- Father: Col. John Detlor (United Empire Loyalist)
- Mother: Jerusha (United Empire Loyalist)
- Wife: Maria Moore Roblin
- Sons: John St George, William Valentine, Titus Simon, Samuel McLean, Egerton Ryerson
- Daughters: Amelia Ann, Mary Matilda, Elizabeth Sophia, Elizabeth Lockwood, Helen, Jane Jerusha

Materials in corpus:

3 diaries (1822-1849)

Figure 1. Example of a fact file. Photo courtesy of *Rural Diary Archive*

As such, the corpus allows investigations in historical sociopragmatics at any level – phonological, morphological, syntactic, lexical, and pragmatic. Since CCanDL will include letters and diaries written at different times in

different parts of Canada and by different types of writers, diachronic (i.e., time-related), diatopic (i.e., regional) and diastratic (i.e., related to social class) studies may also be envisaged. The following section presents a brief case study which was carried out by retrieving data from the CCanDL corpus sampler that is available at this stage.

4. Preliminary findings

4.1 The CCanDL sampler

At the time of writing (July 2023), the CCanDL sampler includes a selection of 78 diaries written in present-day Ontario which were collected from transcribed material already available in the *Rural Diary Archive*, for a total of 1,221,244 tokens. Ontario was chosen as the starting point, firstly, for chronological reasons; secondly, for its importance in the history of Canada from a political, economic and linguistic point of view; and, thirdly, for the availability of material (Chambers 2004: 225; Baskerville 2005; Whitcomb 2007). The composition of the CanDL Sampler is shown in Table 1 below.

Table 1. Composition of the CCanDL Sampler

Gender	Males	26
	Females	18
Age	Adults (26-54yo)	42
	Young adults (20-25yo)	15
	Older people (60-84yo)	11
	Adolescents (17-19yo)	8
	Children (8-9yo)	2
Ethnicity	Canadians	28
	3rd-generation immigrants	13
	English ethnicity	4
	German ethnicity	2
	Irish ethnicity	5
	Scottish ethnicity	2
	2nd-generation immigrants	6
	English ethnicity	4
	Irish ethnicity	2

	no information	9
	African ethnicity	1
	English ethnicity	2
	German ethnicity	1
	Irish ethnicity	2
	Scottish ethnicity	3
Ethnicity	Immigrants	16
	American ethnicity	1
	Dutch ethnicity ?	1
	English ethnicity	4
	English ethnicity ?	2
	Irish ethnicity	1
	Scottish ethnicity	4
	Scottish ethnicity?	2
	Welsh	1
	Presbyterians	11
	Methodists	10
	Anglicans	9
	Baptists	5
Religion	Quakers	3
	Disciples	2
	Universalists	1
	Congregationalists	1
	Roman Catholics	1
	Farmers	18
	& newspaper men	1
	& minister	1
	& merchants	1
	& politician	1
	& students	1
	& millers and distillers	1
Occupation	& newspaper men	1
	Farm women	15
	& students	1
	& teachers	1

	Family of merchants	2
	Merchants	2
	Apprentice surveyor	1
Occupation	Carpenter	1
	Family of farmers	1
	Marble cutters	1
	Teachers	1

The diaries were penned by a total of 44 diarists, 26 of which were male and 18 female, who were born between 1794 and 1894 and died between 1870 and 1984. While the greatest share of diaries (42, corresponding to 56%) was written by adults (a category that includes people aged between 26 and 54), the corpus sampler also contains 15 diaries (19%) written by young adults (namely people aged between 20 and 25), 11 diaries (14%) written by an older person (which includes people aged between 60 and 84), 8 diaries (10%) written by adolescents (in this case, 17-19-year olds), and 2 diaries (3%) written by a child (an 8-9-year old, cf. Table 1).

The diaries so far included in the corpus were written between 1822 and 1919, thus offering a glimpse into Canadian English over much of the nineteenth century. That being said, however, the sub-sections are not entirely comparable to each other; the 1830s and 1840s are not represented in the corpus yet; while the 1820s are only scarcely present – see Fig. 2, which shows the proportion of the chronological sub-sections in the corpus sampler.
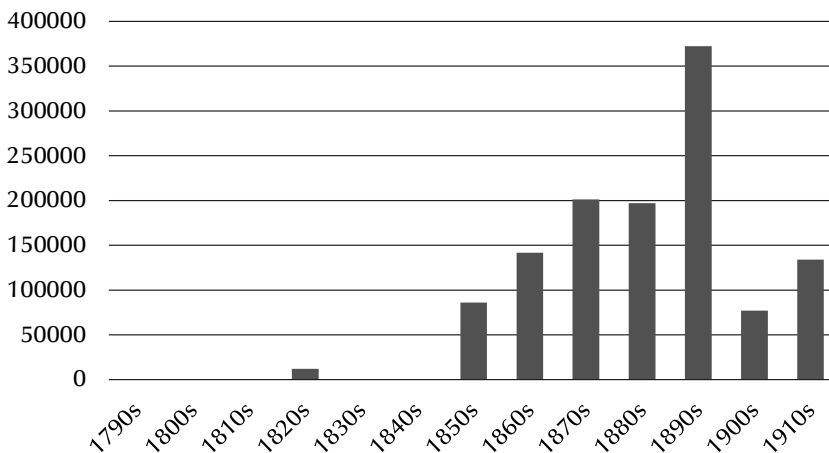

Figure 2. Proportion of the chronological sub-sections of the CCanDL corpus sampler

Although the corpus is not as balanced as seems possible given the scope of the material, this will soon be amended as new material is being made available and prepared.

Census information revealed that most diarists included in the corpus sampler (28) were born in Ontario. Of these, 13 may be described as at least third-generation immigrants, since their parents were born in Ontario as well; 6 may, on the other hand, be identified as second-generation immigrants (i.e., the sons and daughters of immigrants); while no information is available for the rest. Although they were born in Canada, 10 were originally from England, 9 from Ireland, 5 from Scotland, 3 from Germany and 1 was of African ethnicity – most likely a former American refugee.[5] The corpus sampler, however, also includes a number of first-generation immigrants (11) who had emigrated to Canada from either Europe (4 each from England and Scotland, and 1 each from Ireland and Wales) or the US (1 diarist). No census information is available for the remaining 5 diarists, although the archive describes them as of English (2), Scottish (2) and Dutch (1) ethnicity (cf. Table 1). While the contribution of United Empire Loyalists and late-Loyalists (see above) is more difficult to trace as there was no such thing as an American ethnicity, emigrants who settled in Ontario later on might provide a clearer indication of the import of the second (and third) wave of immigration on Canadian English.

As shown in Fig. 3, which plots the distribution of the diaries in space, most documents cover much of present-day rural Ontario. However, there are some notable peaks: 12 diaries were written in Wellington County, a predominantly rural county originally formed in 1837 and which is part of the Greater Golden Horseshoe, while 4 diaries each were written in Oxford County, which was established in 1798, and Leeds and Grenville County, originally two separate counties which were created by Governor John Graves Simcoe in 1792 in an area that had already been settled by United Empire Loyalists.[6]

Religion-wise, the overwhelming majority of the diarists may, quite unsurprisingly, be generally described as Protestants, an umbrella term that covers a very rich array of denominations, the most significant of which are Presbyterians (11 diarists), Methodists (10 diarists), Anglicans (9 diarists) and

---

5   Canada was one of the "promised lands" of the Underground Railroad, a complex, clandestine network of people and safe houses that helped people enslaved in Southern plantations to reach freedom in North America, cf. https://www.thecanadianencyclopedia.ca/en/article/underground-railroad.

6   Cf. https://sites.rootsweb.com/~onleedsg/.

Baptists (5 diarists), but which also include Quakers (3 diarists), Disciples (2), Universalists (1) and Congregationalists (1). The corpus sampler also includes 1 Roman Catholic of Scottish origin (cf. Table 1).
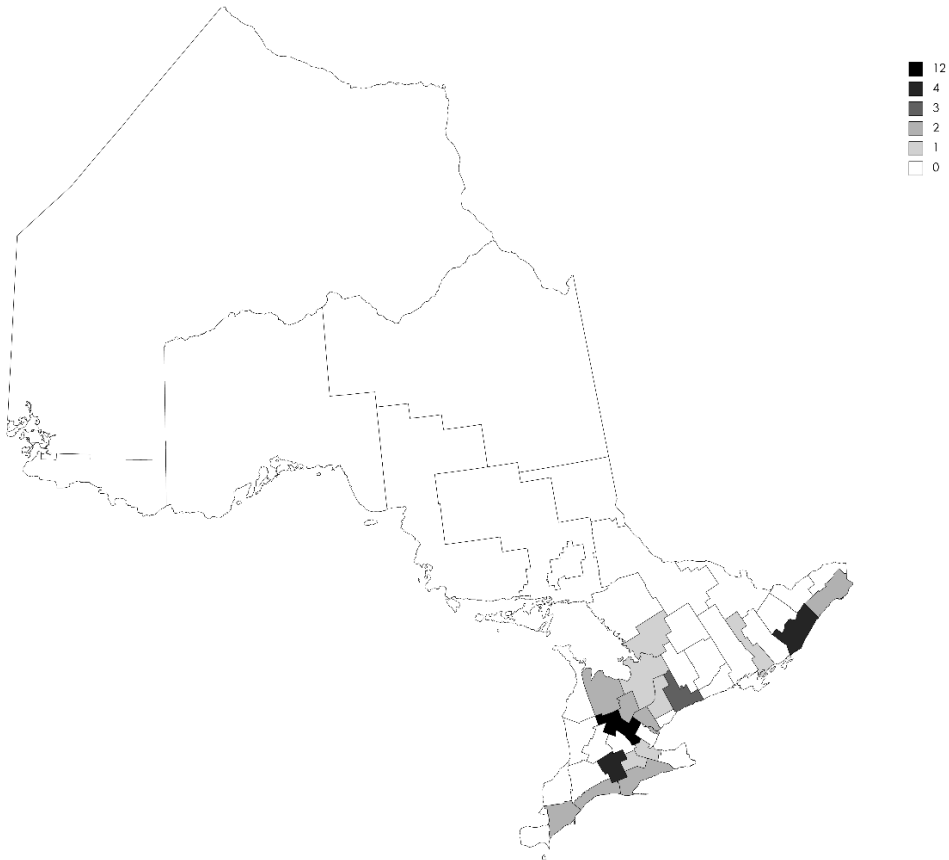

Figure 3. Distribution of the diaries in the different counties of Ontario

Although there are no explicit data on the level of education of the diarists, who were all at least literate enough to put pen to paper, some indications may be retrieved from their occupations. Quite predictably, the majority may be described as farmers (18) and farm women (15). However, while the farm was their primary occupation, some of them also covered other positions. The diarists thus also include 3 merchants, 2 newspaper men, 2 teachers, 1 apprentice surveyor, 1 carpenter, 1 marble cutter, 1 miller and distiller, 1 minister, and 1 politician, in addition to some members of their families (cf. Table 1). The corpus sampler therefore includes, at least in

theory, the diaries of both educated and, most importantly for the purposes of the project, minimally-schooled writers, whose language might more clearly reflect vernacular usage, thus allowing for diastratic investigations as well.

Finally, as shown in Table 2, the corpus includes both what may be described as personal diaries, which contain long entries focusing on the writer's personal thoughts and experiences, and other types of diaries, which may take the form of family books, account books and private chronicles, which generally record events of family and village life, weather reports and news, and which were also read and used by other members of the family (cf. Elspaß 2012: 163).

Table 2. Examples of diary entries included in the corpus

Examples of personal diary	Examples of family books / account books / private chronicles
Mary came home from Nass.{Nassaga-wago} with her new teeth, they look just splendid better that ever expected, I picked ten qts of strawberries this afternoon, and put them down, I had no idea the little patch would do so well, sat up last night with Mrs McGregors Babies, twins, I think one of the little things will hardly live, John Hannah is real sick I am afraid Mrs. Hannah be worn out waiting on him, I am afraid the neighbours are not doing their duty. Mary had letter from Brittain. I think By the tone of it that he thinks I persuaded Anne not to go with him, I feel real hurt about it for I am sure I never said one word against him, but perhaps it will all come right sometime and if it never does I need not mind for God knows all things even the very inner most thoughts of our hearts – I went for a drive in the evening it was quite chilly. I don't think I was very good company, I felt so dull and my thoughts seemed so scattered [F. 1884-1885 – Hill, M. – 1885-07-10]	Cold East rain all day, don't think the rain ceased once Strong wind with it, finished the dress, also helped look over & can some gooseberries [F. 1888-1891 – Mott, P. – 1888-06-28]
	To swamp cutting wood. Went to Etta Ellerton's funeral. Bought a buggy pole from J. Hammond. [M. 1893-1896 (German) – Main, S. – 1894-01-22]
	Warm to day sowed our barley this afternoon William Morros stable catched fire but was put out [M. 1866-1867 – Beattie, W. – 1866-05-10]
	A lovely day, Sarah, Mary Ann, Carroll, and myself went out to Marcia's and spent the day. Bill was here for a few minutes, Andrew hoed the mangolds, Harold took Fannie and went out in the boat. [F. 1878 – Jones, F.A. – 1878-06-20]

4.2 Corpus analysis

As spelling idiosyncrasies can reveal a lot about a language and its speakers, in what follows, a brief case study on orthography will be presented, in order to illustrate the different types of information that can be retrieved from the corpus, which in this case concern phonology, dialectology, pragmatics, and sociolinguistics.

Since Canadian English is generally described as a variety of American English, in order to retrieve idiosyncratic or marked spellings as compared to present-day norms, the corpus keywords were calculated using AmE06, a 1-million-word corpus of general American English from the Brown Family, as reference corpus. The keywords were then filtered according to their frequency in ascending order in the reference corpus, thus retrieving spellings that are only present in the target corpus and which, for this reason, may be considered typical of early vernacular Ontario English, the variety of English represented in the corpus sampler. The keywords list thus arranged is shown in Table 3.

As shown in Table 3, the majority of these (in normal type) may be described as simply erratic spellings (e.g. *nise*, *recieved*, *takeing*, *clowdy*, *kiled*), which are probably more indicative of the writers' level of education and literacy, rather than of their accent or orientation. Indeed, while certainly irregular by today's standards, these orthographies do not seem to reflect a different pronunciation of a specific word, but rather an attempt on the part of less educated and more insecure diarists to write on the basis of what we could describe as phonetic spelling (see Fairman 2000 and 2007).

On the other hand, some of the idiosyncratic orthographies included in Table 2 (highlighted in bold) appear in this list simply because they follow the British, rather than American, standard. While these do not necessarily give an indication of how the words were pronounced, such spellings are nonetheless extremely interesting, since they may be considered as indicative of the writers' background, schooling (cf. Gold 2004) and, possibly, of their orientation. Although connections between orthographical choices and ideology have been proven difficult to make (cf. Grue 2013), orientation represents a particularly significant concept with regard to Canadian English, which has frequently been described as alternatively attracted by either of the two poles: the US, where most Canadians actually came from and an influential neighbor in many respects; and Britain, the colonial administrating power (Heffernan et al. 2010: 3; Dollinger 2012b: 451; Boberg 2016: 112; Dollinger 2019b: 56). Indeed, while most scholars agree

Table 3. Idiosyncratic or marked spellings as compared to present-day norms in CCanDL

WORD	FREQ (CCanDL)	FREQ (AmE06)	WORD	FREQ (CCanDL)	FREQ (AmE06)
ploug*	974	0	buisy	38	0
staid	595	0	makeing	37	0
nise	416	0	**favour**	37	0
geting	351	0	**saviour**	34	0
cuting	231	0	showry	33	0
waggon	204	0	cemetry	32	0
metin	174	0	*meatin*	31	0
recieved	93	0	*thrashin*	31	0
thrash	91	0	**neighbour**	30	0
threatning	89	0	*seting*	29	0
blustry	88	0	peices	27	0
choars	85	0	**labour**	26	0
harowed	77	0	**parlour**	26	0
puting	77	0	afternon	25	0
beaitful	76	0	calld	24	0
centre	72	0	dollers	24	0
takeing	71	0	choped	24	0
freinds	66	0	clening	22	0
grey	66	0	choping	22	0
spliting	65	0	borowed	22	0
tok	64	0	finishd	22	0
stoped	62	0	burried	22	0
favourable	53	0	kiled	22	0
dissagreeable	49	0	togather	22	0
stopt	48	0	**cheque**	21	0
comeing	47	0	pullin	20	0
comenced	45	0	refference	19	0
comming	42	0	loged	19	0
proffitable	42	0	privelege	19	0
peice	40	0	especialy	18	0
clowdy	39	0	sowd	18	0

on the American base of Canadian English, mostly as a result of the first wave of immigration into Canada (Dollinger 2008: 134, 2019b: 55), until the 1970s it was the British variety that carried prestige in official life (Boberg 2010: 34; Tagliamonte 2014: 201; Dollinger 2019b: 60), a phenomenon which is generally referred to as "Canadian Dainty" (cf. Chambers 2004). This seems to be confirmed by an analysis of the diarists' origins. Indeed, of the 39 diarists who displayed a British orientation, 24 were born in Canada and only 4 in England (the place of birth of 3 diarists is unknown, 5 were born in Scotland, and 1 each in Ireland, the US and Wales). Moreover, 31 of them also made use of forms that by today's standards are indexed as American, as shown in Examples 1 to 4, and, while showing preference for a particular spelling, at times they also spelled the same word following the other standard, as shown in Examples 5 to 8:

(1)   "I drove to Oshawa for a load of bran with the **wagon**" [M. 1899 – Geddes, M.D. – 1899-01-03 (Jan 3rd Tuesday)].

(2)   "Mr. Hudson rode down from Toronto yesterday with a **grey** mare he bought for us" [M. 1899 – Geddes, M.D. – 1899-03-21 (March 21)].

(3)   "Mr Ludy and Johnny worked at **parlor** and sitting room on Dec 8th" [F. 1901-1902 – Watson, J.].

(4)   "[…] placing them 16 in apart from **centre** to **centre**" [F. 1901-1902 – Watson, J.].

(5)   "A day or two after Mr Ingram, who has been our near **neighbor** ever since we came here, sent a lovely cup and saucer and plate" [F. 1877-1907 – Simpson, E. – 1893-01-30 (30)].

(6)   "Berrie and I went to call on our new **neighbour** Mr Carroll and had a very pleasand [sic.] time" [F. 1877-1907 – Simpson, E. – 1881-04-04 (4)].

(7)   "I finished **plowing** the south front field this forenoon and started to plow in the pea field" [M. 1896 – Sunter, W. – 1896-09-09 (SEPTEMBER WEDNESDAY 9 1896)].

(8)   "I have been **ploughing** today in the summer Fallow and got very well along" [M. 1857 – Sunter, W. – 1857-06-18 (THURSDAY, June 18th, 1857)].

While American spellings (odd-numbered examples) generally prevail, most diarists alternate between the two conventions, not only in spelling different words, which may simply be the result of custom, rather than orientation, but also in spelling the same word, which, on the other hand, might be read as evidence of minimal schooling and of linguistic insecurity.

In addition to these first two categories, there are also a number of phonetic spellings in Table 2 (e.g. *blustry*, *clening*, *meatin*, *seting*) which are of particular interest from a dialectological point of view, since they seem to reproduce in writing the diarists' own pronunciation of these words. Of these, 9 involve vowel sounds (underlined in the table), while 8 involve consonants (in italics in the table). Although some of these phonetic spellings are found only in the diaries of one specific writer (e.g. "tok" in the diary of William Rea and "geting" in the diaries of Courtland Olds) and may thus be described as belonging to their idiolect, others are found in the works of several diarists, suggesting possible dialect and accent features that characterized this particular variety of English.

Among the most frequent phonetic spellings that involve vowels, we find the <thrash> spelling for "thresh" (and its inflections),[7] which occurs 690 times in the corpus (98 instances of "thrash", 373 of "thrashing" and 219 of "thrashed"), against 482 occurrences of the <thresh> spelling (128 instances of "thresh", 294 of "threshing" and 60 of "threshed"). This might, again, be treated as a symptom of minimal schooling and literacy, as also evidenced by the fact that some diarists seem to be undecided about the two forms. However, this form appears in the diaries of 25 different writers (17 males and 8 females), most of whom had been born in Canada, even though they were of different ethnicities (7 each of English and Scottish, 5 of Irish, 2 of German, 1 each of Welsh and African origin, besides 2 whose ethnicity in unknown). Moreover, as shown in Table 4, it is found, albeit in different proportions, in all chronological subsections of the corpus, except for the pre-1850s one, which is also particularly unpopulated (only 12,066 tokens) and contains the diaries of only one writer. For these reasons, the <thrash> spelling might indicate that the word was pronounced with a lower /æ/ vowel, instead of the standard mid /ɛ/ vowel. While there are a number of possible alternative explanations, and further research is needed to either confirm or disprove this hypothesis, it may be read as evidence for

---

7   The <thrash> spelling is recorded in the *OED* as a now chiefly regional variant of "thresh" (first attested in 1364, https://www.oed.com/dictionary/thrash_v?tab = meaning_and_ use#1344545110).

the beginning of the phenomenon commonly referred to as Canadian Shift (Dollinger 2008; Labov et al. 2006; Boberg 2019).

Table 4. Normalized frequency (10,000 words) of the two variants of "thresh" and its inflections in the corpus

	pre-1850s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	WHOLE CORPUS
thrash*	0	32.4	6.5	2.1	3.8	3.6	7.1	1.2	5.7
thresh*	47.2	0.5	3.5	3.3	3.4	5.4	1.4	2.1	4

Another quite frequent and widespread pronunciation feature emerging from the corpus, evidenced by the <-ry> instead of the regular <-ery> spelling in such words as "showry" (showery), "blustry" (blustery) and "cemetry" (cemetery), appears to be schwa syncope, that is, the deletion of post-tonic /ə/, a phenomenon that is quite common in informal fast speech (Polgárdi 2015). Since these particular spellings emerge from the writings of 17 different diarists (11 males and 6 females) and, as shown in Table 5, are found in all decades covered by the corpus, with the exception of the very limited pre-1850s section, the phenomenon may be regarded not as part of the idiolect of one specific person, but as generally more widespread.

Table 5. Normalized frequency (10,000 words) of schwa syncope and retention in the corpus

	pre-1850s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	WHOLE CORPUS
*ry	0	0.3	0.5	0.6	3.3	0.2	3.5	3	1.3
*ery	0	4.9	1.3	2	1.4	2.3	2.7	6.2	2.6

Although schwa syncope might in some cases also be indicative of Canadian Dainty (Chambers 2004) and may thus be described as characteristic of Canadian English, it is more probably related to the genre and register under investigation here, which, though written, display many features that are typical of texts of immediacy (Elspaß 2012: 157). As such, this finding may be more significant as evidence of the spontaneity and of the underlying oral nature of these specific texts, which seem to be particularly well-suited for the purposes of future investigations.

Phonetic spellings involving consonants, on the other hand, cluster in two groups. The first, evidenced by such spellings as "metin", "meatin"

and "thrashin", may be traced back to the very widespread phenomenon commonly (though improperly) referred to as g-dropping – that is, the pronunciation of final /ŋ/ as /n/, which in Late Modern times came to be associated with the lower social classes and lack of education (Beal 2004: 160-161). These spellings, however, are only found in the diaries of David Rea and his brother William, and may thus be considered characteristic of their own idiolects, rather than of Canadian English as a whole.

The second group consists of such spellings as "geting", "cuting", "spliting" and "seting", which are used quite consistently by 4 different diarists. These may, once again, be considered as symptomatic of the limited level of education of the diarists – all of whom were farmers born in the first half of the nineteenth century. However, such spellings may also point to tapping or flapping – that is, the pronunciation of intervocalic /t/ as /ɾ/, thus suggesting an American orientation and the attempt, at least on the part of some diarists, to reproduce in writing what they probably perceived as a weaker sound.

5. Concluding remarks

Although confined to just one element – orthography – and carried out on a corpus sampler, the analysis has highlighted some features which may provide a better insight into the characteristics of early Canadian English, including further evidence of the so-called "dual-standard", which has long been described as typical of this variety, and of possible early instances of the Canadian Shift. Moreover, as shown by the preliminary investigation, since the corpus focuses on prototypical texts of immediacy (Elspaß 2012: 157), where spontaneous and less controlled linguistic choices may have been more frequent, it may also help to shed light on the history of vernacular usage more in general. Furthermore, the corpus was also designed to complement and integrate itself with other similar corpora, including the *Corpus of Early Ontario English* (CONTE, Dollinger 2006), which includes letters, diaries and newspaper articles written in Ontario between 1776 and 1899, and the *Petworth Emigration to Canada Corpus* (PECC, Dollinger 2019c), which includes letters written by Southern English emigrants to Ontario between 1832 and 1842. Once completed, the *Corpus of Canadian English Letters and Diaries* may therefore represent a promising tool, as it is expected to make a significant contribution to our understanding of the characteristic features and diachronic (i.e., time-related), diatopic (i.e., regional) and diastratic (i.e., related to social class) developments of one of the major, albeit slightly under-researched, regional varieties of English.

REFERENCES

Primary sources

Corpus of Canadian English Letters and Diaries (CCanDL). Compiled by G. Rovelli
in collaboration with the *Rural Diary Archive* and the McLaughlin
Library, University of Guelph, Ontario (CA).
Oxford English Dictionary Online (OED online)
http://www.oed.com/, accessed July 2023.

Secondary sources

Allen, A.
2014   "A non-standard standard? Exploring the evidence from 19th century
vernacular letters and diaries". In: A. Auer – D. Schreier – R. J. Watts
(eds.) *Letter Writing and Language Change*. Cambridge: Cambridge
University Press, 202-222.
Avis, W. S. – P. D. Drysdale – R. J. Gregg – M. H. Scargill (eds.)
1967   *Dictionary of Canadian English: The Senior Dictionary.* Toronto: Gage.
Avis, W. S. – C. Crate – P. Drysdale – D. Leechman – M. H. Scargill – C. L. Lovell
(eds)
1967   *A Dictionary of Canadianisms on Historical Principles (DCHP-1).* Toronto:
Gage.
Barber, K. (ed.)
2004   *Canadian Oxford Dictionary* (2nd edn.). Toronto: Oxford University
Press.
Baskerville, P. A.
2005   *Sites of Power. A Concise History of Ontario*. Oxford: Oxford University
Press.
Beal, J.
2004   *English in Modern Times*. London: Hodder Education.
Berger, C. M.
2005   *The Dialect Topography of Canada: Method, Coverage, Interface and
Analyses*. Dissertation (Master's). Vienna: University of Vienna.
Bloomfield, M. W.
1948   "Canadian English and its relation to eighteenth century American
speech", *The Journal of English and Germanic Philology* 47 (1), 459-67.
Boberg, C.
2010   *The English Language in Canada. Status, History and Comparative
Analysis*. Cambridge: Cambridge University Press.
2012   "Standard Canadian English". In: R. Hickey (ed.) *Standards of English.
Codified Varieties around the World*. Cambridge: Cambridge University
Press, 159-178.

2016   "Newspaper dialectology: Harnessing the power of the mass media to study Canadian English", *American Speech* 91 (2), 109-138.

2017   "Archival data on earlier Canadian English". In: R. Hickey (ed.) *Listening to the Past: Audio Records of Accents of English.* Cambridge: Cambridge University Press, 375-394.

2019   "A closer look at the short front vowel shift in Canada", *Journal of English Linguistics* 47 (2), 91-119.

Chambers, J. K.

1993   "'Lawless and vulgar innovations': Victorian views of Canadian English". In: S. Clarke (ed.) *Focus on Canada*. Amsterdam: John Benjamins, 1-26.

1998   "English: Canadian varieties". In: J. Edwards (ed.) *Language in Canada*. Cambridge: Cambridge University Press, 252-272.

2004   "'Canadian dainty': The rise and decline of Briticisms in Canada". In: R. Hickey (ed.) *Legacies of Colonial English. Studies in Transported Dialects*. Cambridge: Cambridge University Press, 224-241.

2006   "Canadian raising. Retrospect and prospect", *Canadian Journal of Linguistics* 51 (2-3), 105-118.

Clarke, S. – F. Elms – A. Youssef

1995   "The third dialect of English: Some Canadian evidence", *Language Variation and Change* 7, 209-228.

Dollinger, S.

2006   "Oh Canada! Towards the corpus of early Ontario English". In: A. Renouf – A. Kehoe (eds.) *The Changing Face of Corpus Linguistics*. Amsterdam and New York: Rodopi, 7-25.

2008   *New-Dialect Formation in Canada. Evidence from the English Modal Auxiliaries*. Amsterdam: John Benjamins.

2012a   "Canadian English in a real-time perspective". In: A. Bergs – L. Brinton (eds.) *English Historical Linguistics. An International Handbook*. Berlin: De Gruyter Mouton, 1858-1879.

2012b   "On the autonomy and homogeneity of Canadian English", *World Englishes* 31 (4), 449-466.

2014   "Emerging standards in the colonies: Variation and the Canadian letter writer". In: A. Auer – D. Schreier – R. J. Watts (eds.) *Letter Writing and Language Change*. Cambridge: Cambridge University Press, 101-113.

2015   "Canadian English: A conservative variety?", *Zeitschrift für Kanada-Studien* 35, 25-44.

2019a   *Creating Canadian English. The Professor, the Mountaineer, and a National Variety of English*. Cambridge: Cambridge University Press.

2019b   "English in Canada". In: C. L. Nelson – Z. G. Proshina – D. R. Davies (eds.) *The Handbook of World Englishes* (2nd edn.). Hoboken: Wiley Blackwell, 52-69.

2019c   "'I hope you will excuse my bad writing'. Shall vs. will in the 1830s
         Petworth Emigration to Canada Corpus (PECC)". In: R. Hickey (ed.)
         Keeping in Touch. Emigrant Letters Across the English-Speaking World.
         Amsterdam: John Benjamins, 43-66.

Dossena, M.
    2012   "The Study of correspondence: Theoretical and methodological
            issues". In: M. Dossena – G. Del Lungo Camiciotti (eds.) *Letter Writing
            in Late Modern Europe*. Amsterdam and Philadelphia: John Benjamins,
            13-30.

Edwards, J.
    1998.   *Language in Canada*. Cambridge: Cambridge University Press.

Elspaß, S.
    2012   "The use of private letters and diaries in linguistic investigation".
            In: J. M. Nandez-Campoy – J. C. Conde-Silvestre (eds.) *The Handbook
            of Historical Sociolinguistics*. Oxford: Wiley Blackwell, 156-169.

Fairman, T.
    2000   "English pauper letters 1800-34, and the English language".
            In: D. Barton – N. Hall (eds.) *Letter Writing as a Social Practice*.
            Amsterdam: John Benjamins, 63-82.
    2007   "'Lower-order' letters, schooling and the English language, 1795
            to 1834". In: S. Elspaß et al. (eds.) *Germanic Language Histories 'from
            Below' (1700-2000)*. Berlin: De Gruyter, 31-43.

Fee, M. – J. McAlpine
    1997   *Guide to Canadian English Usage*. Toronto: Oxford University Press.

Gold, E.
    2004   "Teachers, texts and early Canadian English 1791-1841". In: S. Burelle
            – S. Somesfalean (eds.) *Proceedings of the 2003 Annual Conference of
            the Canadian Linguistic Association.* https://cla-acl.ca/actes/actes-2003-
            proceedings.html, accessed December 2023.

Grue, D.
    2013   "Testing Canada's 'honour': Does orthography index ideology?",
            Strathy Student Working Papers on Canadian English, 1-18.
            https://qspace.library.queensu.ca/handle/1974/8649, accessed
            December 2023.

Heffernan, K. et al.
    2010   "Preserving Canada's 'honour'. Ideology and diachronic change in
            Canadian spelling variants", *Written Language and Literacy* 13 (1), 1-23.

Hickey, R.
    2017   "Analysing early audio recordings". In: R. Hickey (ed.) *Listening to
            the Past: Audio Records of Accents of English*. Cambridge: Cambridge
            University Press, 1-12.

Labov, W. – S. Ash – C. Boberg
    2006   *The Atlas of North American English. Phonetics, Phonology and Sound
            Change*. Berlin: De Gruyter.

Mackey, W. F.
    1998   "The foundations". In: J. Edwards (ed.) *Language in Canada*.
               Cambridge: Cambridge University Press, 13-35.

Polgárdi, K.
    2015   "Syncope, syllabic consonant formation, and the distribution of
               stressed vowels in English", *Journal of Linguistics* 51 (2), 383-423.

Schneider, E. W.
    2007   *Postcolonial Englishes. Varieties Around the World*. Cambridge:
               Cambridge University Press.

Tagliamonte, S. A.
    2014   "System and society in the evolution of change: The view from
               Canada". In: E. Green – C. F. Meyer (eds.) *The Variability of Current
               World Englishes*. Berlin: De Gruyter, 199-233.

Tieken-Boon van Ostade, I.
    2005   "Eighteenth-century English letters: In search of the vernacular",
               Linguistica e Filologia 21, 113-146.

Van der Wal, M. J. – G. Rutten
    2013   *Touching the Past. Studies in the Historical Sociolinguistics of
               Ego-Documents*. Amsterdam and Philadelphia: John Benjamins.

Whitcomb, E.
    2007   *A Short History of Ontario*. Ottawa: From Sea to Sea Enterprises.

Address: Giulia Rovelli, Dipartimento di Lingue, Letterature e Culture Straniere, Università degli Studi di Bergamo. Piazza Rosate 2 – 24129 Bergamo, Italy.
ORCID code: https://orcid.org/0000-0001-6282-4097.