



Universiteit  
Leiden

The Netherlands

## **Linking Words to Writers: Building a Reliable Corpus for Historical Sociolinguistic Research**

Nobels, J.M.P.; Wal, M.J. van der; Langer, N; Davies, S;  
Vandenbussche, W

### **Citation**

Nobels, J. M. P., & Wal, M. J. van der. (2012). Linking Words to Writers: Building a Reliable Corpus for Historical Sociolinguistic Research. In N. Langer, S. Davies, & W. Vandenbussche (Eds.), *Language and History, Linguistics and Historiography. Interdisciplinary Approaches*. (pp. 343-362). Bern: Peter Lang. Retrieved from <https://hdl.handle.net/1887/32241>

Version: Not Applicable (or Unknown)

License:

Downloaded from: <https://hdl.handle.net/1887/32241>

**Note:** To cite this publication please use the final published version (if applicable).

# Linking Words to Writers: Building a Reliable Corpus for Historical Sociolinguistic Research<sup>1</sup>

## Abstract

Building a corpus of seventeenth-century private letters suitable for historical sociolinguistic research, we face the problem caused by the contemporary rate of literacy. As part of the seventeenth-century population was either illiterate or partly literate, we have to establish whether letters are written by the senders themselves or not, before matching specific language use with the social rank of the sender of a letter. Therefore we have developed a procedure, based on both form and content analysis, in order to identify letters as autographical or non-autographical. The analyses in this identification procedure would not be possible without interdisciplinary research and collaboration with historians, archivists and artificial intelligence researchers, which shows that crossing the borders of linguistics can be vital for achieving a reliable corpus. A spin-off of our identification procedure is a sub-corpus of non-autographical letters, which will allow us to examine the practice of so-called encoding in Dutch private letters for the first time.

## 1 A treasure for historical linguists

Examining the linguistic past from the perspective of the language history from below, historical linguists consider the language of proximity, found in ego documents such as private letters and diaries, as an indispensable source

<sup>1</sup> A preliminary, shorter version of this article, titled *Tackling the Writer-Sender Problem: the newly developed Leiden Identification Procedure (LIP)*, was published in the journal *Historical Sociolinguistics and Sociohistorical Linguistics* 9 (2009).

for reliable data.<sup>2</sup> Until recently, for the history of the Dutch language of the seventeenth and eighteenth centuries linguists had to rely mainly on ego documents written by men from higher ranks in society. Ego documents from women in general and from both men and women of lower and middle classes were available only in small numbers, scattered over various provincial, municipal and personal archives in the Netherlands.<sup>3</sup> This situation has changed considerably as historians rediscovered the Dutch documents in the High Court of Admiralty's archives, kept in the National Archives (Kew, UK). Apart from a wide range of other material including treatises on seamanship, plantation accounts, textile samples, ships' journals, poems and lists of slaves, this collection of so-called *sailing letters* comprises about 38,000 Dutch letters, both commercial and private, from the second half of the seventeenth to the early nineteenth centuries. These sailing letters were confiscated during the wars fought between the Netherlands and England. What makes this huge collection of letters so interesting for linguists are the 15,000 private letters, written by men, women and even children of all social ranks, including the lower and middle classes. They offer an unprecedented opportunity to gain access to the everyday and colloquial language of the past and will consequently enable linguists to get a view on the Dutch language history from below.

As part of the research project *Letters as loot. Towards a non-standard view on the history of Dutch*, we explore this highly valuable source for both the seventeenth and the eighteenth century.<sup>4</sup> In this chapter we will focus on the challenges of the seventeenth-century material and discuss our fruitful collaboration with researchers from other disciplines.

- 2 For a discussion of the concept of 'language history from below', see for instance Elspaß (2007a: 3–9; 2007b: 155).
- 3 In this article we use the term *social class* as a synonym of *social rank* and not in its nineteenth-century meaning.
- 4 The research project *Letters as loot. Towards a non-standard view on the history of Dutch* was initiated by the programme leader Marijke van der Wal (Leiden) and funded by The Netherlands Organisation for Scientific Research.; cf. also <<http://www.brievenalsluit.nl>> (Dutch and English version).

## 2 Confiscated letters

England and the Netherlands were rivals and enemies for centuries: no fewer than four Anglo-Dutch Wars were fought and in various other wars of the eighteenth and the beginning of the nineteenth century England and the Netherlands stood at opposite sides. Warfare implied *privateering* (in Dutch *kaapvaart*): private ships (privateers) authorized by a country's government attacked and seized cargo from ships owned by the enemy. Privateering was a longstanding legitimate activity, practiced by all seafaring European countries and regulated by strict rules. The conquered ship and all its cargo, called a prize, were considered as loot for the privateer, if the rules had been followed by the book. In England it was the High Court of Admiralty (HCA) that had to establish whether the current procedures were properly followed. In order to be able to decide whether the ship was a lawful prize, all the papers on board, both commercial and private, were confiscated and claimed by the High Court of Admiralty. After the legal procedure, the confiscated letters stayed in the High Court of Admiralty's Archives. This is how a huge number of Dutch letters from Dutch ships taken by privateers ended up stocked in hundreds of boxes in the British National Archives. To fully appreciate the huge number of letters it is important to note that in very many cases the ships' cargo contained a lot more mail than the crew's own correspondences. Ships sailing to the Caribbean (West India) and to East India often took mailbags on board and thus functioned as mail carriers between the Netherlands and those remote regions, and vice versa (Van Vliet 2007: 47–55; Van Gelder 2006: 10–15).

Gathering dust in the HCA archives for centuries, only a very small part of the collection of confiscated papers has been examined for specific historical research in the last decade of the twentieth century. The actual size of the collection came to light in 2005 when the historian Roelof van Gelder made an indispensable, but still rough inventory of the Dutch HCA material.<sup>5</sup>

5 Cf. van Gelder's report (Van Gelder 2006). The inventory is available now at the website of the Royal Library in The Hague (<<http://www.kb.nl/sl>> accessed 14 August 2009).

Samples of private letters prove that the sailing letters represent what we would like to call the language of proximity of people from all social ranks. This can be illustrated by the sample of letters written to the crew of admiral De Ruyter's fleet in 1664 when that fleet was sailing to the African coast: 63 per cent of the letters were written by the spouses of the addressees and 22 per cent by their parents (Van der Wal 2006: 10; Van Vliet 2007: 57). Among the addressees of the letters we find a great variety of professions: from rear admiral, captain and navigating officer to carpenter, sailor and assistant barber and a lot of professions in between. Assuming that close relatives belong to the same social rank as indicated by the professions of the addressees, we may conclude that the letters represent the language of all social ranks, including middle and lower classes.

### 3 Daily practice: Selecting letters and building an electronic corpus

Before being able to analyze the linguistic data of the letters from a socio-historical perspective, we have had to address the complicated tasks to make a selection from the 15,000 private letters and to build an electronic corpus. We have chosen to make two cross-sections in the huge material, one in the seventeenth century (1664–1674) and the second one in the eighteenth century (1776–1784): the periods of the second and third Anglo-Dutch Wars (1665–1667 and 1672–1674) and of the fourth Anglo-Dutch War (1780–1784) and the American War of Independence (1775–1783), respectively.<sup>6</sup> From these two periods letters were selected and digital photographs

6 The cross-sections correspond with two subprojects of our programme: *Everyday Dutch of the lower and middle classes. Private letters in times of war (1665–1674)*, the PhD-project carried out by Judith Nobels, and *A perspective from below. Private letters versus printed uniformity (1776–1784)*, Tanja Simons' PhD-project. In the third subproject *Rewriting the language history of Dutch*, Gijsbert Rutten and Marijke van der Wal compare the two periods and evaluate the ultimate results of examining the sailing letters.

taken, which photographs are being transcribed to form an electronic corpus.<sup>7</sup> As a working tool a database, specially developed for our project, is available to store and retrieve all information on the letters. A careful selection from the letters available for the two periods has to guarantee an appropriate representation of both male and female writers, of different relative age and social classes. The regional origin of the writers will be taken into consideration too. The letters only provide us with the sender's and addressee's name and address, but more details can be found in the registers of marriage or baptism, kept in local archives. Those, however, are often incomplete – if they have been kept at all – and it is not surprising therefore that we often encounter difficulties in determining a sender's age or social background.<sup>8</sup> It is the huge number of letters that enables us to compile such representative corpora in spite of these frequent setbacks.<sup>9</sup>

In the selection process we take into account the contemporary circumstances of literacy and illiteracy. Although the rate of literacy in the Netherlands was high compared to other European countries at the time, part of the population could neither read nor write (Frijhoff & Spies 1999: 237). Those who were able to read may not have had any writing skills, as reading and writing were taught in succession, not simultaneously (Blaak 2004: 13; Kuijpers 1997: 501; Van der Wal 2002: 9–13). Besides, those who learnt to write may have had little writing practice. So when Van Doorninck and Kuijpers (1993:14) calculate that in 1670 in Amsterdam 70 per cent of the men and 44 per cent of the women could write their own name, we must realize that some of these signers were probably not capable of pro-

7 Anticipating that transcribing hundreds of letters from scratch would be too time-consuming for the researchers, Marijke van der Wal started the Wikiscripta Neerlandica project in 2007. Participants in this volunteer project provide first transcriptions, which are checked three times by different members of the Letters as Loot team before they are accepted as final transcriptions for the electronic corpus.

8 Sometimes the letter itself does not even provide us with the sender's and addressee's full name and address. And to make matters even more complicated, seventeenth-century surnames were more often than not patronymic and some first names – like Jan, Cornelis, Claes, Pieter and Jacob for men, and Trijn, Mary, Neel, Guurt, Griet and Anna for women – were very frequent (Van Deursen 2006: 31–33).

9 Important criteria for sociolinguistic research will be amply met with at least hundred (probably more) carefully selected letter writers in each corpus.

ducing anything more than just their signature (Frijhoff & Spies 1999:237). These fully or partly illiterates had to ask professional scribes (such as the ship's writer or a public writer) or acquaintances in possession of writing skills (whom we will designate as social scribes) to write letters for them, if communication with loved ones elsewhere was called for.

Surprisingly little is known about this so-called *encoding practice* in the Netherlands of the seventeenth century. We can only guess for instance who the public writers were and where they worked on the basis of research on French public writers in Paris at the time of the Ancien Régime (Métayer 2000). One of the questions that have remained unanswered until now is how the actual encoding of letters by public or social scribes took place. Did the senders dictate the entire letter word for word or did they just indicate which tidings needed to be conveyed by the letter? Did the *encoders* make use of the same formulae for every letter or did they allow for variation?<sup>10</sup> And in which respects did social and professional scribes differ in their encoding practices?

For our linguistic research it is crucial to establish whether a letter is autographical or not. When the sender and writer are not one and the same person, the language use in the letter cannot be immediately related to the identified sender and it is usually not possible to identify the anonymous writer either. Even if the encoder can be identified, the language use in the letter cannot simply be ascribed to him or her, since we do not know how much influence the sender had on the encoder's writing process. In the selection stage therefore a first analysis of the letters needs to establish whether or not sender and writer are identical in order to avoid the risk of

10 Note our use of the terms *sender*, *writer*, *scribe*, and *encoder*. The *sender* of the letter is the person in whose name the letter is written, the person whose thoughts are conveyed in the letter. The *scribe* or *writer* of the letter is the person who performed the mechanical act of writing the letter. Sometimes the scribe of a letter is not its sender, for instance when the sender of the letter is illiterate and has paid a professional writer to produce the letter. In these cases, we call the writer of the letter an *encoder*. An *encoder* is a person who has written a letter for someone else. It is important to note that our use of the term *encoder* differs from the use in Dossena (2008) and Dossena & Tiekens-Boon van Ostade (2008).

linking specific language use to the social rank, age or gender of someone who did not write the letter at all. Without such an analysis we would be unable to guarantee the validity of our results.

To solve this major problem of distinguishing between self-writing senders and encoders we developed a procedure which combines script and content analysis. For part of the script analysis we benefited greatly from the expertise of artificial intelligence researchers, as will become clear below.

## 4 Tackling the writer-sender problem: Various clues

### 4.1 Reference to the writing process

In developing what we will hereby introduce as the Leiden Identification Procedure (LIP), we identified a number of form, content and sender characteristics, clues that can reveal whether the letter is an autograph or not. Explicit reference to the writing process is a first and an obvious content clue, one which does not, however, occur very often. A convincing example is a letter written by a sailor to his father in which he explains his bad handwriting by referring to lying in his bunk due to illness:<sup>11</sup>

ende al ist wat qualick geschreven ick en kon dat nijet helpen want ick nogh ijn de  
koije lagh dat ick nogh niet eel ghenesen en was  
(and if it is written a bit badly, I couldn't help it, because I was still in the bunk since  
I was not yet entirely recovered)

While this letter irrefutably shows that it is autographical, other letters prove the opposite. An example of such a letter is one written on behalf of Elisabeth Bernaers.<sup>12</sup> The letter to her husband is written in the first person

11 Letter 0938–0939 in our corpus.

12 Letter 3–1–2008, 129–130 in our corpus.

singular and signed with the name of Elisabeth Bernaers, but below this signature we find the line that identifies the scribe: 'door mij gescreven maey ken pieters ul dochter' (written by me, Maaïke Pieters, your daughter) (see Figure 10).

#### 4.2 *Identical handwriting and the GIWIS-programme*

The second clue applies if we find two letters written in the same hand, but sent by different people, in which case at least one of the letters is non-autographical. Illustrative examples are two letters written on 10 December 1664 in Saint-Kitts, in the roadstead of Basseterre.<sup>13</sup> Although the first letter in Figure 11 is signed with the name of Claeijs Pietersen and the second in Figure 12 is signed by Jan Lievensens, the handwriting and layout of the letters are so similar that both of them must have been written by one and the same person. The similarity is particularly striking in the header and in the closing formula of the letters.

Since the content of the letters does not indicate that one of the senders is better educated or of higher rank than the other, and since the letters have been written aboard a ship, in a very neat and professional handwriting, we are tempted to assume that a third person (maybe the ship's writer, the clergyman or one of the petty officers) wrote the letters for both Claeijs and Jan. Although in this case we cannot know for certain who the writer was, it is clear enough that we should not mark these letters as autographs. We have to bear in mind that this 'same hand clue' can only be applied if enough letters are available to compare the handwritings of letters written around the same time in the same area. Of course, this clue cannot give a decisive answer about the status of letters that cannot be linked to other epistles in this way. We have to allow for the possibility that letters from different senders and written by the same scribe have not survived or have not been discovered yet.

13 Letters 3b-1-2008, 187-188 and 3b-1-2008, 203-204 in our corpus.

Since comparing the handwriting of different letters takes up a considerable amount of time, we are fortunate to benefit from the expertise of a team of artificial intelligence specialists at the University of Groningen. This team, under the direction of Lambert Schomaker, has developed a computer programme that is able to compare a sample of handwriting to a large set of handwritings and identify similar samples. This Groningen Automatic Writer Identification System (GRAWIS) was originally meant for forensic purposes, but with a few modifications it can also be applied to historical texts (Bulacu 2007, Bulacu & Schomaker 2007a, Bulacu & Schomaker 2007b). A modified version of this programme, called GIWIS (Groningen Intelligent Writer Identification System), was developed for us by Axel Brink (Groningen).

GIWIS allows us to compare the handwriting of one specific letter to an entire set of letters. After the necessary preparatory work (which involves uploading pictures into the programme and selecting sections of the pictures that are suitable for processing), GIWIS lists the ten samples that resemble the handwriting under investigation the most with just one click of the mouse. The programme can compare handwritings using different features, such as the slant of the script and the thickness of the quill strokes. At this stage, the powers of perception of the researcher come in, for the programme always lists samples that are supposed to show a similar handwriting, even if the overlap between samples is very small to almost non-existent. It is therefore the researchers' responsibility to check whether one of the listed 'matches' is a real match. Although human beings are still undoubtedly better at recognizing matching handwritings, computers are quicker at scanning large sets of examples. Using the GIWIS programme wisely will save us a lot of time and it need not affect our conclusions for the worse.

Not only the handwriting between letters can be compared, but also the handwriting within one and the same letter can be scrutinized in order to establish whether the letter is an autograph or not. If the sender's name or signature at the bottom of the letter differs noticeably from the body of the letter, the sender may not have written the letter him/herself. This is certainly the case if the handwriting in the letter itself is rather neat and steady while the signature shows an inexperienced hand. It is very likely

– because of the educational circumstances sketched above – that there were people whose writing experience was just sufficient to sign their name, but who were not able to produce an entire letter (Frijhoff & Spies 1999: 237). Apparently, some senders wanted to sign the letters that had been written for them, maybe from a point of honour, as a proof of authenticity, or as a more personal sign of life.

Although this third signature clue can sometimes offer convincing proof of the non-autographical nature of a letter, it is to be handled with caution, for it seems to be the case that experienced writers sometimes used a larger handwriting for their name or signature as part of their stylistic habit (see Figure 13 as an example).<sup>14</sup> A seemingly different handwriting in the sender's signature therefore does not always point to a different identity for sender and writer. The letter might have been written by an experienced writer who wanted to emphasize his/her signature by using a larger or somewhat different hand. Only if the signature seems to have been written by a less experienced writer than the person who wrote the body of the letter we can be certain that we are dealing with a non-autograph (see for instance Figure 14 as an example of this).

#### 4.3 *Occupation and social rank*

The fourth clue is related to the occupation and social status of the writer. If the letter's contents reveal enough about the life of the sender for us to determine his/her occupation, we can estimate how likely it is that the sender of the letter was an experienced writer. Captains, helmsmen, salesmen, doctors, lawyers, book keepers, clergymen and ship's writers had to master writing in order to study or carry out their profession (Van Doorninck & Kuijpers 1993: 46–50; 58–61). These occupations are typical for men. For women it is more difficult to determine whether they needed to be able to write. They rarely mention anything in the letters about the jobs

14 Cf. the letter model written by the seventeenth-century writing-master Hendrik Meurs (Croiset van Uchelen 2005: 37).

they might have had in order to secure an additional income, but we may assume that a lot of these jobs involved some kind of domestic work or retail trade, which did not require writing skills (De Wit 2008: 148–149). Wives of captains and skippers may be an exception: when their husbands were at sea, many of them were responsible for these seamen's businesses and part of their administration and it would probably have been more difficult for these women to handle all the paper work if they themselves could not write (De Wit 2008: 161–162, Bruijn 1998: 67).<sup>15</sup> People of a higher social rank were also very likely to be experienced writers, because it is plausible that their parents were wealthy enough to offer them an education that included the costly writing instruction (Frijhoff & Spies 1999: 238). A man's social rank can thus be determined through his occupation or his title; for a woman's social rank, we rely on the social rank of her husband or family.<sup>16</sup> Although the fact that these people could write does not necessarily mean that they actually wrote their private letters themselves, we nevertheless assume that they did if there is no proof otherwise. Just like all the other clues, this fourth clue has to be used carefully too.

The fifth clue is very closely connected to the previous one; it is in fact an elaboration upon the fourth one. If we can find out a sender's social status, we can compare the level of experience of the handwriting and the number and nature of corrections in the letter to the expected level of education.<sup>17</sup> Neatly written letters of low ranked senders are of particular interest: they may well be non-autographical. Clearly, however, this clue is problematic in two ways. Is it always possible to tell the difference

15 Evidence for this is also to be found in some sailing letters. Cf. the letters of Katelynen Haexwant to her husband Leendert Ariensen Haexwant, rear admiral, in which she informs him about financial matters (Van Vliet 2007: 314–333).

16 Social rank or status is certainly determined by more elements than just occupation as Frijhoff and Spies rightly note (1999: 188), but information about these other elements for our letter writers is often more difficult to retrieve.

17 'The level of experience' of a handwriting is a subjective criterion to some extent, but we believe that it is possible to distinguish different levels of experience based on various features such as whether the letters have been drawn graph by graph or not, the regularity of the handwriting in form and size, and whether the lines slope or not.

between an experienced, but sloppy hand and an inexperienced one? Secondly, we might be at risk of falling into the trap of circularity. It is of vital importance to avoid creating our corpus based on expectations about the language use of people of a certain rank, because that is exactly what we want to research. This clue needs to be restricted to our expectations about *the handwriting* i.e. about the way in which people wrote, not about *what* they wrote (the *language use*).

The last and most objective way to determine whether a letter's writer and sender are the same person, is to compare the handwriting and/or signature used in the letter with other samples of the sender's handwriting that are known to be authentic. It is not always easy to find these samples, but it is possible. For certain cities with accessible and searchable archives, we can retrieve a surprising number of such samples with the help and advice of archivists.

## 5 Searching the archives for authentic signatures or handwriting samples

The seventeenth-century sources that can offer authentic signatures or handwriting samples differ for various regions. In Amsterdam, for instance, newlyweds were requested to sign the register of marriage; in Rotterdam, the names of the newlyweds were recorded by the clergyman in charge of the ceremony. The latter practice makes the Rotterdam registers of marriage unsuitable for our search for authentic signatures. Luckily, the Rotterdam municipal archives keep about half a million notarial acts dating from the sixteenth until the nineteenth century which offer us a good chance of retrieving authentic signatures.<sup>18</sup> In the cities of Middelburg and Vlissingen

18 Juliette Sandberg successfully applied these notarial acts as a tool in her MA-thesis *Vergeet min niet te schrijven al gij kent. Een zoektocht naar Hollandse levens en taalnormen in zeventiende-eeuwse brieven*. (Leiden University, 2009).

gen in the province of Zeeland, the situation is more complicated. Many of the seventeenth-century Zeeland archives have been lost, among which the registers of the city of Middelburg. In Vlissingen, the few remaining registers of marriage do not contain any signatures, but the information in the archive of the Audit Office of Zeeland is more promising.<sup>19</sup>

The Audit Office of Zeeland kept a detailed track of the expenses of the admiralty of Zeeland. Salesmen who had delivered goods or people who had worked for the Zeeland admiralty, had to send a request to the Audit Office in which they described what they had done and how much they expected to be paid. If the request was approved by the Audit Office, the creditor could collect his money and had to sign for receipt on the same document that had originally been sent in as a request. All these documents have been kept in the enormous archive of the Audit Office of Zeeland. It contains a wealth of signatures and samples of handwriting not only of sailors, soldiers, merchants and labourers, but also of their relatives who collected their wages. Sadly enough, a detailed inventory of the Audit Office's archive is not available.<sup>20</sup> Although this makes a targeted search almost impossible, we succeeded in retrieving a couple of authentic signatures from senders of our letters. One of them is Jacob van de Velde who sent a letter to his wife and added a signature with distinctive flourishes (see Figure 15). In the Audit Office's archive we found his request for a delivery of wine and other goods, written in a similar handwriting and signed with the unmistakably identical signature (see Figure 16). In this case we can convincingly identify Van de Velde's letter as autographical.

- 19 We are indebted to Albert Meijer of the Zeeland Archive (het Zeeuws Archief) for the information he provided about the Audit Office of Zeeland and his kindness and perseverance in helping us in our search for signatures and handwriting samples. More information about the archive of the audit office (Rekenkamer van Zeeland) is to be found on the website of the Zeeland Archive: <<http://www.zeeuwsarchief.nl>> accessed 14 August 2009.
- 20 Currently, within the Metamorfoze-project, an inventory is made for a small part of the archive, linked to business in Surinam, but at present, July 2009, it is not clear whether a further inventory will follow. For the Metamorfoze-project cf. <<http://www.metamorfoze.nl/>> accessed 14 August 2009.

## 6 Combining the clues

Some of the above mentioned clues are very telling, while some clues only become important if a number of other indications cannot provide conclusive results. To visualize this we transformed the list of clues into a flow chart (Figure 17) that takes different priorities into consideration and that allows us to examine every letter thoroughly, as well as efficiently in the approach.

The flow chart starts with the content of the letter. If it mentions explicitly that the letter is autographical or non-autographical (box 1 and 2), we need not look for further evidence and can go straight to the relevant conclusion (A or B). If the content does not offer any information about the writing of the letter, we should check our corpus for letters in the same handwriting, but sent by someone else (box 3). If we find such letters, we can check whether they are all sent by people of low status, but written in an experienced hand (box 4). If this is the case, chances are high that we are dealing with letters written by a professional writer (C). If they are not all neatly written letters sent by people of low status, we can only learn more about the potential writer if we look for signatures or handwriting samples of the senders concerned (D).

If the letter is the only letter in our corpus which shows a certain hand, we can scrutinize the signature (box 5). If it is not written in the same hand as the body of the letter and if it seems to be written in a less experienced hand, we are probably dealing with a non-autographical letter (B). If the hand in the signature does not seem to be different from that in the rest of the letter, it is time to take into account the occupation and social status of the sender (box 6). As we have seen before, if the writer is a salesman, a captain, a helmsman, a lawyer, a doctor, a clergyman, a ship's writer, someone of high social status, or his wife or child, it is quite probable that the letter is autographical (A). If the sender of the letter falls into neither category, the only option left is to compare the sender's handwriting with what we would expect of someone with his/her status (box 4). If the handwriting is very neat, while the sender is of low status, it might be possible that a

professional writer or a friend who was an experienced writer interfered (C).<sup>21</sup> If the handwriting does not seem to be very deviant from what we would expect, the letter might be self-written. But because the writer could have been a non-professional writer as well, the only way to find out for certain is to look at authentic samples of handwriting or signatures (D).

The letters that fall into category A are identified as autographical, those of the categories B and C as non-autographical. The letters in category D might prove to be either autographical or non-autographical or letters with an uncertain status, depending on the authentic handwriting samples or signatures that can be traced.

We tested this working method on thirty letters that were written in 1664 on the island of Saint-Kitts, which was then known as Sint-Christoffel to the Dutch. Twenty of the letters were positively identified as autographical (among which the letter written by Jacob van de Velde), two letters were identified as non-autographical (the letters written by Claeijs Pietersen and Jan Lievensens), and we were left with eight letters that have an uncertain status because we could not find any authentic signatures or samples to back up our assumptions.

Two more remarks about our procedure have to be made. Firstly, if particular striking clues occur in a letter at first sight, there is no harm in skipping steps in the flowchart. The chart's chief purpose is to help us analyze letters that do not immediately signal whether they are autographical or not. Secondly, we note that it is not always possible to be one hundred percent certain about the status of a letter without the evidence of authentic handwriting or signature samples.

21 Other clues that suggest an experienced or even a professional writer are: names in the text or in the signature written in a slightly larger hand, embellishments and flourishes in the margins, and a cursive hand.

## 7 Concluding comments and research perspectives

Our newly developed Leiden Identification Procedure (LIP) allows us to distinguish three categories of letters in our corpus: autographical, non-autographical and letters of uncertain status. This achievement has major consequences for our research. It enables us to build a reliable corpus of autographical letters that is fit for our historical sociolinguistic research, a corpus that will allow us to relate linguistic characteristics to social variables such as age, gender and social status. The non-autographical and uncertain letters should be kept well apart from the autographical letters to avoid working with linguistic material that is unsuitable for our research aims. Evidently, if some senders did not write their letter themselves, their social characteristics cannot be related to the language use in the letter. Moreover, since it is impossible to distinguish between the input of the sender and that of the encoder, it is dangerous to link the professional or social scribe's social variables (if known at all) to the language use of the letter.

The non-autographical letters, unsuitable for our sociohistorical research, can nevertheless serve a different purpose: they allow us to examine encoding by Dutch professional and social scribes, a widespread practice which until now has never been examined systematically. Keeping professional scribes apart from social scribes (friends or family of the sender who were not professional writers), we may discover whether these groups use different encoding strategies. Again it is not an easy task to distinguish professional from social writers, but the level of experience of the handwriting and the number of letters written by the same scribe might be good indicators. If in the end this line of research delivers clear-cut differences between non-autographical and autographical letters, these results could be useful to reduce the category of uncertain letters.

The letters with an uncertain status should be kept apart from the autographical and the non-autographical subcorpora, as they can be used neither for sociolinguistic research nor for research into encoding practices. There is no urgent reason why we should try to include these letters into our research project. Even after discarding the uncertain letters, we will still have plenty of letters at our disposal to conduct reliable sociolinguistic research and research into encoding practices.

The possibility that a considerable number of seventeenth-century private letters in our corpus could be non-autographical need not be an impediment for our socio-historical research. In this article we have shown that it is feasible to distinguish autographical letters from non-autographical letters in many cases. Determining whether letters are autographical or not is obviously of vital importance to an analysis of their language. It needs to take place before the actual linguistic research can start, because without the knowledge of the autographical status of the letters we could easily draw false conclusions about the historical sociolinguistic situation. It is important to note that in applying the procedure outlined here we benefited greatly from interdisciplinary research and extensive collaboration with archivists, historians and artificial intelligence specialists. In this case crossing the boundaries of our linguistic discipline has proven to be essential to lay a solid base for a reliable historical letter corpus that will give us access to the everyday and colloquial language of the past and broaden our view on the language history from below.

## References

- Blaak, Jeroen. 2004. *Geletterde levens. Dagelijks lezen en schrijven in de vroegmoderne tijd in Nederland 1642–1770*. Hilversum: Verloren.
- Bruijn, Jaap. 1998. *Varend verleden. De Nederlandse oorlogsvloot in de zeventiende en achttiende eeuw*. Amsterdam: Balans.
- Bulacu, Marius. 2007. *Statistical pattern recognition for automatic writer identification and verification*. Unpublished doctoral thesis. Groningen: Artificial Intelligence Institute, Universiteit Groningen.
- Bulacu, Marius & Lambert Schomaker. 2007b. 'Automatic handwriting identification on medieval documents'. In: Werner, Bob et. al. (eds), *Proceedings Of the 14th International Conference on Image Analysis and Processing (ICIAP 2007)*, 279–284.
- Bulacu, Marius, & Lambert Schomaker. 2007b. 'Text-independent writer identification and verification using textural and allographic features'. In: *IEEE Trans on Pattern Analysis and Machine Intelligence (PAMI), Special Issue – Biometrics: Progress and Directions* 29/7, 701–717.

- Croiset van Uchelen, Ton. 2005. *Vive la Plume. Schrijfmeesters en Pennekunst in de Republiek*. Amsterdam: De Buitenkant/ Universiteitsbibliotheek.
- Deursen, Arie van. 2006. *Een dorp in de polder. Graft in de zeventiende eeuw*. Amsterdam: Bert Bakker.
- Doorninck, Marieke van, & Erika Kuijpers. 1993. *De geschoolde stad. Onderwijs in Amsterdam in de Gouden Eeuw*. Amsterdam: Historisch Seminarium van de Universiteit van Amsterdam.
- Dossena, Marina. 2008. 'Imitatio literae. Scottish emigrants' letters and long-distance interaction in partly-schooled writing of the 19th century'. In: Kermas, Susan & Maurizio Gotti (eds), *Socially-conditioned Language Change: Diachronic and Synchronic Insights*. Lecce: Edizioni del Grifo, 79–96.
- Dossena, Marina & Ingrid Tiekens-Boon van Ostade. 2008. 'Introduction'. In: Dossena, Marina & Ingrid Tiekens-Boon van Ostade (eds), *Studies in Late Modern English Correspondence: Methodology and Data*. Bern: Peter Lang, 7–16.
- Elsaß, Stephan. 2007a. 'A twofold view 'from below': New perspectives on language histories - and language historiographies'. In: Stephan Elsaß et al. (eds), *Germanic Language Histories 'from Below' (1700–2000)*, 3–9.
- Elsaß, Stephan. 2007b. "'Everyday language' in emigrant letters and its implications for German historiography – the German case". *Multilingua* 26, 151–165.
- Elsaß, Stephan, Nils Langer, Joachim Scharloth & Wim Vandenbussche (eds). 2007. *Germanic Language Histories from Below (1700–2000)*. Berlin/New York: De Gruyter.
- Frijhoff, Willem & Marijke Spies. 1999. *1650. Bevochten eendracht*. Den Haag: Sdu Uitgevers.
- Gelder, Roelof van. 2006. *Sailing Letters. Verslag van een inventariserend onderzoek naar Nederlandse brieven in het archief van het High Court of Admiralty in The National Archives in Kew, Groot-Brittannië*. Den Haag: Koninklijke Bibliotheek.
- Kuijpers, Erika. 1997. 'Lezen en schrijven. Onderzoek naar het alfabetiseringsniveau in zeventiende-eeuws Amsterdam'. *Tijdschrift voor Sociale Geschiedenis* 23/ 4, 490–522.
- Métayer, Christine. 2000. *Au tombeau des secrets. Les écrivains publics du Paris populaire, Cimetière des Saints-Innocents, XVIe-XVIIIe siècle*. Paris: Albin Michel.
- Vliet, Adri van. 2007. 'Een vriendelijke groetenisse'. *Brieven van het thuisfront aan de vloot van De Ruyter (1664–1665)*. Franeker: Van Wijnen.
- Wal, Marijke van der. 2002. 'De mens als talig wezen: taal, taalnormering en taalonderwijs in de vroegmoderne tijd'. *De zeventiende eeuw* 18, 3–16.
- Wal, Marijke van der. 2006. *Onvoltooid verleden tijd. Witte vlekken in de taalgeschiedenis*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.
- Wit, Annette de. 2008. *Leven, werken en geloven in zeevarende gemeenschappen. Schiedam, Maassluis en Ter Heijde in de zeventiende eeuw*. Amsterdam: Aksant.

## Websites:

*Brieven als Buit / Letters as Loot:* <<http://www.brievenalsbuit.nl>>

*Koninklijke Bibliotheek / The Royal Library, Sailing Letters:* <<http://www.kb.nl/sl>>

*Metamorfoze:* <<http://www.metamorfoze.nl>>

*Zeeuws Archief / The Archive of Zeeland:* <<http://www.zeeuwsarchief.nl>>