(Extra)Ordinary letters: A view from below on seventeenth-century Dutch
Nobels, J.M.P.

Citation

Version:	Corrected Publisher's Version
License:	[Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)
Downloaded from:	https://hdl.handle.net/1887/20900

Cover Page

The handle http://hdl.handle.net/1887/20900 holds various files of this Leiden University dissertation.

Author: Nobels, Judith Maria Petrus
Title: (Extra)Ordinary letters. A view from below on seventeenth-century Dutch
Issue Date: 2013-05-21

Chapter 2. Corpus and methodology

The seventeenth-century *Letters as Loot* corpus comprises 595 letters written by 441 different writers. [15] These letters were captured against the background of the Second and Third Anglo-Dutch Wars (1665-1667 and 1672-1674 respectively) and were written between 1661 and 1675.[16] The majority of the letters stem from 1664 (47%) and from 1672 (28%). The corpus in its entirety comprises about 245,000 words and is not parsed.[17] The *Letters as Loot* corpus is split up into three sub-corpora. A first sub-corpus comprises all the autograph letters, while a second sub-corpus is made up of letters that are non-autographs. The third sub-corpus contains letters of which I have not been able to establish whether they are autographs or not. These last two corpora will often be combined in the research.

These sub-corpora were created because the status of a document (autograph or not) determines the use that can be made of it in historical sociolinguistic research. This matter will be dealt with in further detail in chapter 3. In the present chapter I will focus on the practicalities of

---

[15] It is important to note that a *writer* is not the same as an *individual*. A *writer* is a person of a certain age, with a specific regional background, belonging to a particular social class. During the lifetime of people these characteristics change (e.g. everyone ages, some people rise or fall on the social ladder) so that the same *individual* can represent different *writers* at different stages of his/her life. This will be illustrated in §2.3.1.

[16] The years in which the letters were written do not correspond exactly to the period in which England and the Dutch Republic were officially at war. There are several reasons for this. Firstly, ships had sometimes been under way for a while before they were confiscated and could thus have been carrying letters written before the start of a war. It also happened that people aboard a ship had a personal archive of letters they had once received. This archive could contain letters written several months or even years before the capture of the ship. Furthermore, privateering did not seem to be completely restricted to official times of war. It seems to have taken place during the build-ups and the aftermaths of wars as well.

[17] The term 'word' should not be interpreted literally here. We counted as 'words' elements separated from each other by spaces. Not all the 'words' in the corpus can thus be viewed as proper words. Some are syllables (e.g. when the term *vereenicht* 'reunited' is spelled as *ver_eenicht* 're_united'), some are more random parts of words (e.g. when *je* 'you' is spelled as *j_e* 'yo_u'), and some are a combination of words or of a word and a part of another word (e.g. when dealing with clitics). In spite of the fact that the term 'word' cannot be interpreted literally in the context of the dimensions of the corpus, I will use this term throughout the dissertation, given that the optional alternative term 'token' can be mistaken for an occurrence of a specific 'type'. The exact number of words in the totality of this corpus is 244,637.

compiling the corpus in §2.1. In §2.2 I will discuss the different independent variables that are of importance for the case studies that will follow. I will briefly describe why they are relevant for my investigations and how they were put into practice. In §2.3 I will describe the contents and structure of the corpus. Some methodological issues will be discussed in §2.4 and the conclusion of this chapter is presented in §2.5.

2.1. Developing the corpus

The reliability and quality of my historical sociolinguistic investigations depend to a large extent on the reliability and size of the corpus used. The *Letters as Loot* corpus was therefore compiled with the utmost care and was rendered as large as possible, a process which eventually took two years to complete. In what follows I will describe the procedures that were followed in compiling the corpus. More detailed information about the creation of the sub-corpora will be provided in chapter 3.

2.1.1.   Preparation
Different steps needed to be taken in order to get from a collection of about 38,000 Dutch letters in the London National Archives to a workable corpus fitted for sociolinguistic research. The letters in the National Archives needed to be selected, photographed, transcribed, provided with metadata and organised in such a way that sociolinguistic research of the letters would become feasible.

Selection procedure and photographs
Marijke van der Wal visited the National Archives in Kew (London) in 2007 and in 2008 to explore the wealth of letters preserved in the High Court of Admiralty's archives. During these visits she selected a fair amount of letters and photographed them. After these explorations, other members of the *Letters as Loot* team, including the present author, visited the National Archives twice a year in 2009 and in 2010 to pursue this work. For the seventeenth-century *Letters as Loot* corpus the focus was on the boxes dating from the Second (1665-1667) and Third Anglo-Dutch War (1672-1674).

       The letters were selected for photographing based on a number of features: language, text type, and condition of the paper and/or ink. Only letters written in Dutch were selected and priority was given to private letters, although a small number of business letters was included as well. Most of the documents in the *Prize papers* have been preserved remarkably well, although some letters have become difficult to decipher due to tears in the

paper, faded ink or ink eating into the paper. Partly or wholly illegible letters were not selected for photographing. The photographed content of a few High Court of Admiralty boxes was provided to us by the *Koninklijke Bibliotheek* (The Royal Library, KB) who are participating in a project called *Metamorfoze* together with the *Nationaal Archief* (the Dutch National Archives).[18]

Transcriptions

Back in the Netherlands, these digital pictures were sent to members of the Wikiscripta Neerlandica Project. This project was set up by Marijke van der Wal in 2007 and involved a team of volunteers who provided diplomatic transcriptions of letters from the HCA archives. The transcription protocol and an example of a transcription can be found in appendices A and B.

During various correction phases as many transcription and interpretation problems as possible were solved. The volunteers sent their transcriptions back to the *Letters as Loot* research assistant who carried out a first check. The transcription was compared to the photographs one letter at a time. The transcriptions of seventeenth-century letters were always double-checked meticulously by the present writer and a last correction, aimed at filtering out any remaining problems and illegible fragments, was carried out by Marijke van der Wal. Each letter in the seventeenth-century *Letters as Loot* corpus has thus gone through three phases of correction. The final transcriptions resulting from this project can therefore be considered as accurate and reliable.

The text files

The final transcriptions were converted into Text files in order to create a corpus that is searchable with the computer program WordSmith, a popular corpus linguistics tool.[19] Deletions, problematic readings, words written in full that were originally abbreviated, and best guesses and suggestions for missing words were all tagged.[20]

---

[18] The *Metamorfoze* project is a national programme for the preservation of the Dutch paper heritage. The programme was initiated by the Dutch Ministry of Education, Cultural Affairs and Science and is carried out by a joint venture of the *Koninklijke Bibliotheek* and the *Nationaal Archief*. For more information: <www.metamorfoze.nl> [08/11/2012]. The boxes were HCA 30 226-1, HCA30-227-1, HCA 30-227-2, and HCA 30-223.

[19] WordSmith was developed by Mike Scott from the University of Liverpool. For general information on WordSmith, see <www.lexically.net/wordsmith/index.html> [08/11/2012]

[20] See appendix C for the protocol used to convert transcriptions into Text files.

The database
When creating a corpus one needs to be able to store contextual information about the corpus texts. It is also very useful when these texts can be organised in different ways depending on which element the researcher is interested in. Therefore the metadata of the letters need to be searchable. The *Letters as Loot* database provides these facilities. This database was developed by Marijke van der Wal and Coen Zimmerman in 2008 and was adapted slightly throughout the first couple of years in which it was used.

The *Letters as Loot* database contains information about the letters' finding place at the National Archives and the correction process each letter has been through. Furthermore it assembles information about the letter (text type, quality of the handwriting, date of writing, number of words), about the sender and the addressee of the letter (name, whereabouts, occupation, social class, age, religious background, place of birth, relationship with addressee or sender) and about the contents (which people, places and events are mentioned). The *Letters as Loot* database also has a very useful comprehensive search function which allows researchers to look for specific letters or see which fields have not been completed yet. It is of course in the researcher's best interest to gather as much information as possible in the database so that a large and balanced corpus can be created: the more information is known about a letter and its writer, the larger the chances are that they can be categorised succesfully according to the different independent variables of importance and that the language in the letter can be used in as many investigations of the influence of various variables as possible. In §2.2 I will show how letters and their writers were categorised. For screen shots of the database, see appendix D.

Some data which were needed to complete the database could be found in the letter itself, but for other information more research was required. At a first stage, the internet was used to find relevant information about the letter's sender and addressee: a number of Dutch archives offer the possibility to do limited research online, many genealogists publish their findings on the web, and there are public databases which contain information about ships and their crew.[21] If neither the letter nor the internet

---

[21] Online research is possible with e.g. the online register of baptism of Amsterdam, the notarial archive and the digital registers of marriage, baptism and death of Rotterdam, the online registers of marriage, baptism and death of Vlissingen, and the 'Zeeuwen gezocht' website (<www.zeeuwengezocht.nl> [08/11/2012]) offering all sorts of genealogical information about people in Zeeland. Information about ships and their crew can be found in a database of VOC ships (<http://www.vocsite.nl/schepen/lijst.html> [08/11/2012]), a database of people aboard VOC ships (<http://vocopvarenden.nationaalarchief.nl/> [08/11/2012]) and

procured sufficient information, Dutch archives offered a final possibility of finding the details needed. This archival research for the seventeenth century was coupled to archival visits necessary to determine the status (autograph, non-autograph or letter of uncertain authorship) of the letters. More information about the archival research will thus be provided in the chapter which is dedicated to the autograph problem and the Leiden Identification Procedure, chapter 3.

2.1.2. Determining the letters' status using the Leiden Identification Procedure

Information about the sender's gender, social class and region was not enough to create a reliable corpus for the seventeenth century. The letters also needed to be assigned to one of the three different sub-corpora on the basis of their status (autograph, non-autograph, letter of uncertain authorship). Autograph letters are letters that have been written by the senders themselves. Non-autograph letters are letters that have been written for the sender of the letter by someone else. Letters of uncertain authorship are letters for which it is unclear whether they should be classified as autographs or as non-autographs. In order to be able to distinguish between these three different types of letters, the Leiden Identification Procedure was developed. In chapter 3 this procedure will be discussed in detail. For now it suffices to note that each letter was assigned a status (autograph, non-autograph, letter of uncertain authorship) and wherever the true writer of a non-autograph or a letter of uncertain authorship was not identified, a unique code was given to the writer in question.[22]

2.2. The independent variables

One of the elements that makes this dissertation unique in the field of Dutch historical linguistics is the fact that the focus in the case studies presented here is on social variation, and on variation related to social class and gender in particular. However, there are other external factors as well that are taken into account in the case studies of language variation and change in the seventeenth-century Dutch Republic in this dissertation, namely text type,

---

in the Slave Voyages database (<http://www.slavevoyages.org/tast/database/ search.faces> [08/11/2012]).
[22] The codes for writers are constructed as 'X:x'. The capital 'X' stands for a code that indicates the region in which the writer was active (e.g. CAR for the Caribbean islands) and the lower case 'x' is a number or a letter indicating one particular writer for that region.

region, and age of the writer. Finally, in some case studies, language-internal factors are examined as well. In what follows, I will discuss each of the language-external factors that will be of importance for this dissertation and describe how they are operationalised in my research.

In §2.2.1 I will present the variable text type. The independent variable region will be presented in §2.2.2. The most important independent variables, gender and social class, are discussed in §2.2.3 and §2.2.4 respectively. The importance and the operationalisation of the variable age of the writer will be discussed in §2.2.5. Finally, in §2.2.6, I will discuss the factors of education and writing experience. These last two factors will not function as independent variables in my investigations, but they can and will be examined indirectly.

2.2.1. Text type

The first independent variable that will be introduced is text type. This variable has two variants: private and business. In chapter 1, it has been shown that text type is related to the extent to which language use in the text can be described as 'language of immediacy'. The private letters of the seventeenth-century *Letters as Loot* corpus are expected to contain language use more closely related to spoken Dutch than the business letters. It is very important to note here that the case studies in this dissertation are mainly focused on the language use in private letters, given the fact that the objective of this dissertation is to describe several aspects of the everyday Dutch of the seventeenth century. The *Letters as Loot* corpus therefore mainly consists of private correspondence. Some aspects of the language use in the small sub-set of business letters will be brought to attention in chapters 4 (on forms of address), 5 (on reflexivity and reciprocity) and 7 (on apocope of the final schwa).

For seventeenth-century letters, a straightforward decision on the text type is not always possible, for the business and personal lives of seventeenth-century people were more interwoven than it is nowadays (Kooijmans 1997: passim). It is not unusual, for instance, that letters between business partners contain references to the health of friends and family members or that the term *vriend* 'friend' is used as a form of address. Also, when merchants were overseas, family members at home were sometimes relied on for help in the family business. It thus sometimes occurs that a letter from a merchant to his wife mainly consists of businesslike requests: wives were asked to pay this merchant or that friend, to collect money here or there, or to take care of goods that had been sent to the Netherlands. A consequence of this intertwing of private and business life is the existence of hybrid letters. An example of such a letter is the letter Jan Fransz Doens wrote from Surinam to his wife Neeltje Schuijen in

Vlissingen.[23] After an apology for not having written earlier, Jan turns to business and does not change the subject anymore. The following passage is just a fragment of the business part of the letter:

> *Liefste Ick ben met freer Jacob Soetelijnck aen*
> *Monsr Jacobes vaader sendende de somme van 10903 lb netto*
> *Suijcker en daer noch bij de somme van 2619 lb letterhoudt*
> *Daer noch bij ben Ick aen ul sendende de somme van 3583 lb*
> *Suijckerbruijt te weeten met de tarra van de vaeten daer*
> *ul moet van aftrecken: Comt dan netto suijcker 3222 lb en*
> *De tarra is netto 361 lb De suijcker moet ul in stillighheijt door*
> *Abraham den elt laeten verkoopen en ul moet hier seer*
> *sekreet in In sijn Jae ul moet het teegen u eijgen susters niet*
> *seggen of teegen u Eijgen broers want de suijcker comt*
> *op Een ander man sijn Risikoe over: Soo drae als Monsr van*
> *Der beke de suijcker verkocht heeft: soe moet ul de suijcker*
> *van ons verkoopen of voor hem soot ul beliet te doen*
> *En laet Monsr vander beke ul dan de gerechte derde part*
> *van alles wat Ick hem gesonden hebben geven en eijst*
> *de Rekeninge wat het goet verkocht is*

'My dearest, together with brother Jacob Soetelijnck
I am sending to Mister Jacob's father the sum of 10,903 lb. net
in sugar and with it the sum of 2,619 lb. of letterwood.
On top of that I am sending you the sum of 3,583 lb.
of sugar cones from which you must deduct the tare of
the barrels: that gives you 3,222 lb. of sugar net and
the tare is 361 lb. net. You must have the sugar sold in
secret by Abraham den Elt and you have to conceal it well.
Why, you cannot even tell your own sisters or
your own brothers, because the sugar is transported
at another man's risk. As soon as Mister van
Der Beke has sold the sugar, you must sell our
sugar or sell it just before he does, if you wish.
And then have Mister van Der Beke give you the
third share – which you are entitled to – of everything I sent him
and demand the bill of the goods that have been sold.'

In order to decide on the text type of letter, the following rule of thumb was used: if the sender and addressee of the letter were closely related to each

---

[23] Letter 17-06-2009 086-087 in the corpus (HCA 30-223).

other (e.g. husband and wife, father and son, cousin and cousin, nephew and uncle) the letter was classified as private, even if it contained information about business. If the sender and intended receiver of the letter were not closely related and if the letter did not contain any private messages other than greetings for the addressee's family and wishes for the addressee's good health, the letter was classified as a business letter.

2.2.2. Region
Region is an important factor of influence on language use and language change. First of all, different dialects and regiolects are used in different regions. These dialects and regiolects are not necessarily limited to spoken Dutch, but can influence the (spontaneously) written Dutch as well. Secondly, different regions may have a different socio-economical status. Supra-regional variants and standard languages are usually established in the socio-economical and political centre of a language area and as a consequence often contain relatively many elements of the dialects spoken in this centre. These elements can therefore start to spread to other regions as well. It is thus important to include region as a factor. But how should region be put into practice in the analyses of the *Letters as Loot* corpus?

For practical reasons, letters that were written in the Netherlands were grouped geographically at the level of the current Dutch provinces (see figure 2.1). Admittedly, there is still a large amount of dialectal variation within a province, but seeing the size of the corpus and the number of other factors that will also be taken into account (gender, age and social class), it is more practical to work with a few broad categories rather than with a large number of small categories. The regions that provided us with the bulk of letters are Zeeland, South Holland and North Holland. A few letters can be linked to other provinces of the Netherlands (such as Gelderland and Friesland), but considering their small numbers they were classified under the left-over category of 'Other'. This category also contains letters linked to other present-day countries, such as Norway, Germany, and Belgium. The category 'Unknown' comprises the letters that cannot be linked to a region.

Figure 2.1: The present-day provinces of the Netherlands

The region of North Holland is a special case. It is the province that is best represented in the seventeenth-century *Letters as Loot* corpus: almost half of the letters in the entire corpus were written by writers stemming from this province (286 letters out of 595 letters in the entire corpus). The letters linked to North Holland are not distributed evenly across the entire province. More than half of the letters linked to North Holland (182) originate from the province's largest city: Amsterdam. The letters linked to Amsterdam were separated from the letters linked to other towns or cities in the province for several reasons. Firstly, such a large number of letters are related to the city of Amsterdam that this city simply deserves its own category. Secondly, the city of Amsterdam was a very dynamic city: it was an important seaport and it had a large number of inhabitants, among whom many immigrants (Hart 1976: 135-181; Frijhoff & Spies 1999: 154-155, 160-161, 165-168; Sogner & Van Lottum 2007). The language use in this city might therefore differ substantially from the language use in the smaller cities and villages in the

rest of the province due to the contact between different languages and dialects, as has been argued (for Amsterdam and other urban centres in the Netherlands) by Boyce & Howell 1996, Boyce-Hendriks 1998, Boyce-Hendriks & Howell 2000, Goss 2002, Howell 2006, and Goss & Howell 2006. Thirdly, the city of Amsterdam is located in the south of the province of North Holland. The dialects spoken in this area are known to resemble South Holland dialects more closely than West-Frisian dialects, which occur in the villages to which the majority of the rest of the North Holland letters in the corpus are linked. On these last two grounds, one can expect the linguistic data for Amsterdam to differ from the data for the rest of the province of North Holland. If Amsterdam is not treated as a separate category, these potential differences cannot be examined and the data for North Holland can become distorted.

The regions that are distinguished in the case studies of this dissertation are thus Zeeland, South Holland, North Holland (Amsterdam), North Holland (rest of the province), 'Other', and 'Unknown'. How a letter was assigned to one of these regional categories depended on whether the letter was an autograph or not and on whether the letter was a private letter or a business letter. In the remainder of this section, I will discuss the different paths that were followed to identify the regions to which the letters could be linked.

Autograph letters
For autographs it was attempted to discover the sender's current or last place of residence in the Netherlands, assuming that this was a place in which the sender had his/her roots or to which the sender was strongly linked in any case. For letters sent from the Netherlands, the place from which the letter was sent is usually mentioned in the header of the letter. The writer of an autograph letter written in Middelburg, was thus assumed to have been living in Middelburg at the time of writing. If no counter indications were found, letter writers writing from Middelburg were also assumed to originate from Middelburg. The language use in the letter was then linked to the province of Zeeland.

For letters sent from overseas to the Netherlands, the region was decided upon using the address of the letter as an indication. A sender writing to his family in a Dutch city, had probably been living in this city as well until he or she left. For instance, the above-mentioned Jan Fransz Doens had written his letter himself. He wrote from Surinam, which does not give us much of a clue about his previous place of residence in the Netherlands. However, Jan wrote to his wife, who lived in Vlissingen in the province of Zeeland. He had therefore probably also lived in Vlissingen until he left for

Surinam. Jan's letter was therefore classified as a letter linked to the region of Zeeland.

For *business* letters sent from abroad, the method of using the address to identify the sender's last place of residence in the Netherlands seems somewhat less reliable. Married couples generally lived together, but business partners did not necessarily have to reside in the same city. However, if the contents of business letters sent from abroad did not provide us with any other indications, the address was used as a point of departure for a search online or in Dutch archives. More often than not, the address of a business letter sent from abroad indeed gave away the sender's regional background.

If the letters themselves could not provide the answer, the location of the letters in the archives of the High Court of Admiralty was used as a last resource. If a letter was discovered in a box which only contains letters written from North Holland to Batavia, there is a good chance that the letter is a North Holland one. However, given that the content of some boxes in the archive can be linked to different regions at the same time and given that the content of some of the boxes is jumbled up, extreme caution was asked for. This piece of evidence was therefore only used in order to get a first lead. All of the classifications were always verified when looking for more information about the sender online or in Dutch archives.

Non-autograph letters and letters of uncertain authorship
For non-autograph letters written in the Netherlands, there are several factors influencing the methods that can be used to link the letter to a region. Whenever the writer of the letter (i.e. not the person who sent the letter and whose message is conveyed, but the person who did the actual writing) was known, his or her place of residence was traced (starting from the place name mentioned in the header of the letter) and this place determined the region to which the letter was linked.[24] If the writer's name was not known

---

[24] It only happens occasionally that the name of the writer of a non-autograph letter is known, for instance when this writer is mentioned explicitly in the letter itself. An example of such a letter is given in §3.2.1 in the discussion of content clues. It can also happen that the corpus contains a number of letters that have been written in the same handwriting but that have been sent by different people. Archival research can then show that the sender of one of these letters is also the writer of all of the letters. For instance, archival research carried out by Juliette Sandberg has shown that Elsje Wijbrants, sender of letter vliet-7 in the corpus, was able to write and indeed did write her letter herself. But letter vliet-20 in the corpus, a letter sent by Marte Reijnders, is written in the same handwriting as Elsje's letter. Therefore we know for certain Marte Reijnders has not written her letter herself, but that Elsje Wijbrants is the actual writer of this letter.

or if the known writer could not be linked to a certain region, everything depended on the place where the letter was written.

If the non-autograph letter was written by an unknown writer in the Netherlands, the place name mentioned in the header of the letter was used to determine the region to which the writer of the letter was probably most closely linked. Whenever a place name was not mentioned, information about the sender's place of residence was traced. When found, this information was extended to the writer of the letter for it is plausible that the actual writer of a non-autograph letter lived in the same region as its sender.

This is easy to show: if people could not write, they could ask friends or family to write the letter for them or they could go to a professional writer. There is no reason why these people would have their letter written by someone far away from home. The actual writer was usually someone from their direct environment. The seventeenth-century *Letters as Loot* corpus contains numerous examples which illustrate this. There is for instance Maartje Jaspers who wrote a letter for her sister-in-law Annetje Barens. Both women lived in Rotterdam. There is Antheunis Verbrugge who wrote letters for his mother Maaike Andries in Vlissingen. Or Marretie Flipse who wrote letters for her sister Elisabeth Flipse Amelingh in the city of Amsterdam. Of course it is possible that the writer of a non-autograph letter originally came from a different region than the region in which he or she was writing this letter, but it is assumed that such cases are in the minority. Besides, even if some of such writers originated from a different region, they were clearly living in another region at the moment of writing and thus stood in (close) contact with people and the language from this last region.

If a non-autograph letter written by an unknown writer was written abroad, there was no chance to link this letter to a specific region with any certainty. The place in which it was written does not necessarily say anything about the Dutch region it could be linked to. Secondly, it is dangerous to assume that the region to which the sender of the letter is linked is also the region to which the actual writer of the letter is linked: while it is true that on many ships a large part of the crew originated from one and the same city and people writing letters for each other abroad may have been neighbours at home, it is also true that members of a ship's crew could have very diverse regional backgrounds.[25] The same goes for Dutch

---

[25] An example of this can be found in letter KB 227-2 010-011 in the corpus (HCA 30 227-2). In this letter, Jan Eghberts, originating from Amsterdam, informs his mother who is living in the same city that he has sent a small keg of oil to the wife of his assistant. This woman lives in Vlissingen. Jan Eghberts and his assistant were workmates – and maybe even friends – working on the same ship. But they were linked to different regions.

people living in the colonies abroad. They could easily befriend other people living in those colonies with very different regional backgrounds. Non-autograph letters written by unknown writers abroad were therefore classified as letters for which the region of the writer is unknown.[26]

Letters of uncertain authorship were handled in the same way as non-autograph letters written by an unknown writer. If they were written in the Netherlands, the place where the letter was written was decisive for the region. If they were written abroad, the letters were treated as letters for which the region is unknown.

Foreign writers
It sometimes happened that a letter was written in Dutch while I suspected or knew (from elements in the language use or from references in archives) that its writer had a foreign background (e.g. Scandinavian or German). This is not surprising due to the fact that the Dutch Republic – and the large cities in the Dutch Republic in particular – counted a large number of German and Scandinavian immigrants in the seventeenth century (Hart 1976 126-127, 162-171; Kuijpers 1997: 510; Kuijpers 2005: 336, 379; Sogner & Lottum 2007: 155). Some of these immigrants married Dutch people and settled in the Netherlands for good, which explains why these immigrants sometimes wrote letters in Dutch. A fragment from such a Dutch letter written by an immigrant is presented below. [27] It is a fragment from a letter written by Annetie Harms who was born in Bentheim (Germany) to her husband Harmen Gerritsen, a Dutchman who was born in Kampen (in the Dutch province of Overijssel). The couple lived in Amsterdam at the time of writing (November 1664). Annetie's letter is written in Dutch, but her language use differs in some respects from the language use typically found in letters written by people who were born in Amsterdam. A very striking feature of her Dutch is for instance the spelling of the preposition *te* 'to', which she spells as <to> or <tho>:

---

[26] Letter 3-1-2008 283-285 in the corpus (HCA 30-228).

[27] Originally, the region for these letters was marked as *neutral* in order to keep these non-autograph letters apart from autograph letters of which the writers could not be linked to a region successfully. However, since these *neutral* letters are treated in the same way as letters for which the region is unknown as far as investigations into regional distributions of certain features is concerned and since the autograph and the non-autograph letters are clearly distinguished from each other in the *Letters as Loot* corpus anyway, I will not use this category in this dissertation in order to avoid confusion.

> *en ul schreijft wan daer **tho** komen het welck ijn dese tijet van oerlog*
> *nijt nijet gheraet saem en ijs en ock met en vremt **to** ghan en en mens*
> *ijs sterveijcllijck eijn ijck ul nijt en vonde waer sou ijck met meijn*
> *leve keijnt dan hen en onse leijven her kon meij ock komen **tho***
> *haelen so most daet onnosele keijnt swerven van de en plaes ohp de*
> *andere*

> 'And you write me to go there, which is not wise in this time of war.
> And travelling with a stranger. And a human is mortal. And if I
> could not find you there, where would I go with my sweet child?
> And our sweet Lord could come and get me too and then that
> innocent child would have to wander from one place to another.'

For letters such as this one the same procedure as described above was used
to determine the region to which the letter was most closely linked.
Regardless of the foreign background of the writer, such letters could thus be
assigned to one of the Dutch regions, although the letters were marked in the
database as letters with foreign influence. Annetie's letter, for instance, was
categorised as a letter linked to North Holland (Amsterdam). I included
letters from these foreign writers in my corpus in this manner, because I
want to treat speakers of Dutch with another native tongue as full members
of the Dutch language community in the seventeenth-century Republic. To
exclude them from this study would be in direct conflict with what this
dissertation is trying to achieve: to fill the gaps in the sociolinguistic history
of Dutch and to present a more complete picture of the variation that was
present in the Dutch everyday language of the seventeenth century.

2.2.3. Gender
Gender is the first of the three variables that will only be used in research on
letters from the sub-corpus of autograph letters. This social variable has
repeatedly proved to be a strong variable in sociolinguistic research
(Nevalainen & Raumolin-Brunberg 2003: 110). Although I categorise
writers as male or female solely on their biological sex, I prefer to use the
term 'gender' rather than 'sex'. The former term emphasises the importance
of the specific social roles and practices that come with the two sexes and
stresses that "no biological determinism is intended" (Nevalainen &
Raumolin-Brunberg 2003: 110; Cheshire 2002: 423-424).
        Men and women in the seventeenth-century Dutch Republic held
different positions in society. They had different economical, socio-cultural
and legal roles (De Wit 2005: 61, 2008: 138). For instance, married women
were legally not allowed to handle their own affairs; they always needed a

male guardian (De Wit 2005: 61, 2008: 138).[28] Furthermore, although there were jobs that could be done by men as well as by women, there were occupations typical of the particular sexes. Seagoing occupations were typical of men, for instance, while care giving occupations were typical of women (Van Deursen 1988: 7-8; De Wit 2005: 71, 2008: 138). Some guilds even excluded women from membership (De Wit 2005: 71, 2008: 138). According to Van Deursen (1988: 11), a typical women's occupation was an occupation that did not require a large capital or much schooling. This suggests that women typically received less schooling than men, which is confirmed by Kuijpers (1997: 513). This has consequences for women's literacy of course: Van Doorninck and Kuijpers (1993: 14) calculated that about 70% of the men must have been able to write in Amsterdam in 1670, compared to only 44% of the women.

These differences between seventeenth-century men and women could be reflected in aspects of their language use. Therefore, gender was taken up as an important independent variable in the case studies of this dissertation. Luckily, it was easy to determine the gender of the writers of autograph letters based on the sender's name. For the few cases in which the sender's name was missing, it was possible to decide on the gender based on the relationship between the letter writer and the intended receiver or based on information about the sender's activities in the letter. If the sender was writing the letter to 'my beloved husband', for instance, the sender was obviously female. Take the letter written to Adriaen Nousters.[29] The sender never mentions his/her name, but near the end of the letter there is a closing formula that says *bij mijn ul moeder* ('written by me your mother'). The sender was thus obviously female.

2.2.4. Social class

The second social variable that will only be used in research on autograph letters is the variable social class, "a central concept in sociolinguistic research" (Ash 2004: 402, Nevalainen 1996: 57). There are many different definitions of the concept of social class, but the most well-known to sociolinguists is probably the definition used in a study by Labov (1966): "an individual's life chances stated in terms of his relation to the production and acquisition of goods and services" (Ash 2004: 402). Given that the variable social class has been shown time and again to be strongly linked to language use, there is no need to explain in detail why it is deemed to be

---

[28] However, for wives of men at sea exceptions could be made. Since their husbands were often absent, these women were regarded as 'occupational widows' and were often capable of contracting all the same (De Wit 2005: 61-62, 71-7; De Wit 2008: 138).

[29] Letter 05-01-2010 080-081 in the corpus (HCA 30-225).

important for this dissertation as well. What is more interesting with respect to the implementation of this variable in the following case studies, is how this variable was operationalised.

In modern sociolinguistic research several characteristics are usually combined to determine a person's social class, such as education, occupation, income, occupation of the parents, and living area (Ash 2004: passim). For this historical corpus, however, it is impossible to determine all of these characteristics for each letter writer. Even if all these data had been kept in some archives, it would be a Sisyphean task to trace them. Therefore a simpler method was used: the social class of letter writers was determined on the basis of their occupation, "the single indicator that accounts for by far the greatest portion of the variance" (Ash 2004: 419). Only if more data were readily available, other elements were taken into account, such as the occupation of the writer's father. For female writers, whose occupation is often unknown, the social class of their (late) husbands (if the women were/had been married) or fathers (if the women were not married) was copied. This is in line with the spirit of the age, since the pre-eminence of men in the public sphere was more often than not taken for granted in the early-modern period (Frijhoff & Spies 1999: 190). A woman's social status can thus be expected to have been heavily dependent on the status of her male guardian.

The classification of the different occupations into social classes needed to be historically relevant; therefore historians' views upon the social structure in the Dutch Republic of the seventeenth century were taken into account. My classification of the different social classes was based on a framework which is commonly used among Dutch historians (Looijesteijn 2012: 221): it is used by Frijhoff & Spies (1999: 189-190), Van Leeuwen (2000: 41-42 in Looijesteijn 2012: 221), Knevel (2002: 219-220 in Looijesteijn 2012: 221), and Bruijn (2008: 16) among others. While some scholars prefer to merge particular categories, the basis of the classification remains the same throughout the publications on the Early-Modern Dutch history (Looijesteijn 2012: 221). I will describe this classification as it is presented in Frijhoff & Spies (1999: 189-190) who identify six different layers in seventeenth-century society. The first group is the patriciate: the nobility and the regents' families. The second layer consists of rich merchants, ship owners, entrepreneurs, large landowners, academics, high ranked officials and officers in the army and in the navy. The third group has amongst its members: small entrepreneurs, well-off farmers, prosperous shop owners, good craftsmen, captains, lower officers, officials, teachers, village chaplains, notaries public and clerks. The fourth group is for small farmers, low officials, small shop owners, craftsmen and modest skippers. The one but lowest group is the group of the labourers in employment,

carriers, seamen, soldiers, servants and the manual labourers. People from this group can easily sink downwards to the lowest group of the beggars, have-nots, vagrants, deserted seamen, deserted soldiers, and day labourers.

For practical purposes the number of social classes in the corpus was kept at four, a number of social divisions that is said to be ideal for sociolinguistic research and is used often in other studies (Labov 2001: 31; Nevalainen & Raumolin-Brunberg 2003: 136-137). The patriciate and the nobility are not represented in the corpus; therefore this layer was left out. The two lowest social categories mentioned by Frijhoff and Spies were merged into one category in my corpus: the lower social class. The table below gives an overview of the social categories I will use in my analyses.

Category	*Description*
Lower class **LC**	labourers in employment, carriers, seamen of low rank, soldiers, servants, manual labourers, beggars, have-nots, vagrants, deserted seamen and soldiers, day labourers
Lower-middle class **LMC**	small farmers, low officials, small shop owners, craftsmen, skippers
Upper-middle class **UMC**	small entrepreneurs, well-off farmers, prosperous shop owners and craftsmen, captains, lower officers, officials, teachers, village chaplains, notaries public, clerks
Upper class **UC**	rich merchants, ship owners, entrepreneurs, large landowners, academics, high ranked officials, officers in the army and in the navy

Table 2.1: The four social categories used in my research

I will illustrate how the social class of writers in the seventeenth-century *Letters as Loot* corpus was determined with two examples. First, the female letter writer Maertie Nanninghs. Maertie wrote several letters to her husband, Pieter Pauelsz., but none of those letters clearly states what Maertie did for a living. [30] However, it is known that Maertie's husband, Pieter, was a carpenter on a ship, an occupation that is linked to the lower-middle class. Therefore, Maertie was assigned to the group of lower-middle class writers.

---

[30] Letters 3-1-2008 091-092, 3-1-2008 093-094, 3-1-2008 097-098, 3-1-2008 099-100, 02-07-2010 206-207, and 16-06-2009 155-157 in the *Letters as Loot* corpus (HCA 30-647).

The letter writer Doede Ennes Star also presents an interesting case.[31] In a letter to his parents, Doede reveals that he had run away from the family he was staying with in Spain and that he has spent the last couple of years partly as a prisoner and partly as a mercenary in the army. At the moment of writing, Doede is working aboard a ship to pay for his crossing back to the Netherlands. On the basis of Doede's recent occupations and adventures, one could be tempted to assign him to the lower class. However, one must take into account that Doede is the son of Enno Doedes Star, a well-known Dutch admiral. Since his father was most likely a respected member of the upper class, Doede was also assigned to the upper class.

Writers whose social class could not be traced were placed in the 'unknown' group. The majority of these writers of unknown social status probably belonged to either the lower or the lower-middle class (as far as I can tell on the basis on their handwriting and the contents of their letters), but could not be placed into one category with certainty on the basis of external information. This is no surprise, for the lives of people from these classes are usually less well documented than the lives of people from the higher classes.

When dealing with these social categories, one must keep in mind that these groups were not completely separated from each other. Family ties often crossed the borders of adjoining categories and there was some social mobility in the seventeenth-century Dutch Republic (Frijhoff & Spies 1999: 190). Social climbers, people who climbed up the social ladder during their lifetime, were marked in the database. There are only five of them, which is probably far from all of the social aspirers quietly present in the corpus. It is difficult to find them, since one requires knowledge about many years of a person's life to be able to classify them as social climbers. One would need information about the occupation of the person's father (as it would indicate the social class in which the person 'starts' life) and the career of the person him/herself. It was certainly not feasible to unearth this information for every writer in the corpus, if this information was available at all.

Take for instance Arnoud Adriaensen as a typical example of a writer in the database. Arnoud wrote a letter to his wife Jacomijntje Louwers in Vlissingen.[32] Arnoud's occupation is not mentioned in his letter and I can only guess that he is not the captain of the ship he is sailing on and that he has a low or middle rank on board. In a database in the archive of Zeeland I found Arnoud as a petty officer responsible for the supplies of the ship 'The rising sun' in 1668.[33] It is unclear whether this was already a higher rank

---

[31] Letter 05-01-2009 025-026 in the *Letters as Loot* corpus (HCA 30-643).

[32] Letter 06-01-2010 160-161 in the *Letters as Loot* corpus (HCA 30-644).

[33] The database in which the information was found is the Poortvliet database (see §3.3.2).

than he had in 1664. Arnoud is not mentioned in this database again; hence I cannot say whether he was appointed to higher positions on his next journeys. Since the Old Notarial Archive of Vlissingen has been lost, no notarial deeds or wills of Arnoud or his family can be traced which may contain information about his occupation in a later state of his life. Who his father was and what he did for a living will remain a mystery as well, since no act of baptism can be found. No further information about the man seems to be available. Therefore, there is no way to tell whether he was a social climber or not. The same goes for the majority of the writers in the corpus. Only the lives of a minority of them are documented well enough to decide whether they are social climbers or not. Therefore, the influence of social mobility on language use in the seventeenth-century *Letters as Loot* corpus will not be examined.[34]

2.2.5. Age

The social variable age – the third independent variable that will only be used in research on autograph letters – can be linked to two types of linguistic change. "Age stratification of linguistic variables can reflect change in the speech of the community as it moves through time (*historical change*), and change in the speech of the individual as he or she moves through life (*age grading*)" (Eckert 2001: 151). The variable age will be examined with apparent-time research in the case studies of this dissertation. This is because the *Letters as Loot* corpus does not consist of two or more comparable sub-corpora for two or more moments in time, which would be needed for real-time research (Eckert 2001: 153). Admittedly, some letters in the corpus stem from around 1664 and other letters were written around 1672. However, the letters will be treated as letters from the same period, given that the time span between these two moments in time is short (8 years). I will thus treat the letters written around 1664 and the letters written around 1672 as letters stemming from the same period.

For practical reasons, the number of age groups was limited to three plus a group for the writers whose age could not be determined. The three age groups are: younger than 30 years of age, in between 30 and 50 years of age, and older than 50. Some people could be classified based on information in records of baptism or their birth date in genealogical overviews. For people whose year of birth was unknown, their age was estimated based on information about their family situation and their activities.

---

[34] This is unfortunate, since earlier research has proven it to be an important factor (e.g. Nevalainen 1996: 73; Nevalainen & Raumolin-Brunberg 2003: 135; Labov 1972: 286 in Nevalainen & Raumolin-Brunberg 2003: 152).

Men and women with small children or new-born babies or men and women who did not mention any children and whose parents were still alive at the time of writing were added to the youngest age group. This was done based on the impression gained from registers of marriage and baptism that people usually married (for the first time) when they were in their twenties and did not wait long to start a family. Working with this assumption, I can allot to the middle group men and women with a couple of children or with at least one child that seems to be older than five or six years old. People with grandchildren or children old enough to procure them with grandchildren and people who had retired from work or who complained about their old age were allotted to the oldest group.

It is beyond dispute that this method did not offer watertight guarantees. There will always be exceptions: people who marry at a very late or at a very young age, or couples that have their first baby only after ten or more years of marriage. However, these exceptions are not expected to influence the results greatly, given the rather large number of different writers whose language use will be examined.

2.2.6. Education and writing experience?

Other important independent variables for research on the *Letters as Loot* corpus could be the level of (writing) education and the level of writing experience of the writers, two factors that are closely linked (Elspaß 2005: 46). After all, as was described in §1.2.1, earlier research has proved (writing) education to be strongly linked to certain variables in written language use (Vandenbussche 2006: 440, 453-454, Elspaß 2005: 40-51; 67-71). Furthermore, it is not difficult to imagine that letter writers who wrote and read (letters) frequently wrote differently than letter writers who generally did not need to read or write for their livelihood and only put pen to paper in exceptional cases. This has been shown for nineteenth-century German (Mihm 1998 in Vandenbussche 2006: 453-454) and ninetheenth-century Dutch in the city of Bruges (Vandenbussche 2007). Distinguishing between 'labour-oriented' and 'writing-oriented individuals', as Vandenbussche (2006: 454) describes them, may thus be very useful for the analysis of the language use in the *Letters as Loot* corpus.

Regrettably, there is little to no information to be found in Dutch archives on the education of the seventeenth-century writers in the *Letters as Loot* corpus. The relationship between education and language use can thus not be examined directly in the private letters of the corpus. Nor is it possible to determine letter writers' exact level of writing experience, since it would require detailed knowledge about their daily lives. However, the level of education and writing experience can be taken into account indirectly through the variables social class and gender. It is the case that men and

members of the upper classes in general received a better (writing) education and had more writing experience in the seventeenth-century Dutch Republic than women and members of the lower social classes did, which is due to the different roles of men and women and the different social classes in society and the cost of writing instruction (Van Doorninck & Kuijpers 1993: 14; Kuijpers 1997: 501, 504, 513; Frijhoff & Spies 1999: 237-238). Although education and writing experience will not function as independent variables in the case studies of this dissertation, their influence on language use in the letters can and will be examined indirectly through the variables gender and social class.

2.3. The seventeenth-century *Letters as Loot* corpus

Above I have described the independent variables that are of importance for my historical sociolinguistic investigations of seventeenth-century Dutch in letters. In what follows, I will describe each of the three sub-corpora of the *Letters as Loot* corpus in general and go into details regarding the distribution of the writers and the letters across the above-mentioned variables. It should not be a surprise that the distribution of the writers across all the different categories in the corpus is not completely balanced. Since this is a historical corpus that will be studied linguistically for the first time and that should therefore contain as many letters from as many different writers as possible, groups of writers that were overrepresented were not reduced to obtain complete balance. On the other hand, there are up to 6 different independent variables that will be taken into account and, as will become clear below, there are more slots to fill than there are different writers in the collection of letters used to build the *Letters as Loot* corpus. The corpus structure is thus bound to show some gaps.

2.3.1. The sub-corpus of autographs
In this sub-corpus of autographs, most letters are private: the sub-corpus includes 260 private letters written by 202 different writers which comprises almost 118,000 words. Apart from these private letters, a small number of business letters is included: 47 business letters written by 41 different writers and comprising about 19,000 words. The complete sub-corpus of autographs contains 307 letters written by 232 different writers. The sum of the writers of private and business letters (202 + 41 = 243) exceeds the number of different writers in the entire sub-corpus of autographs (232). This is explained by the fact that 11 writers occur as writers of both private and business letters.

It is also interesting to note that we are dealing with 232 writers, but with 230 individuals. Two individuals wrote letters at different ages so that each of them actually represents two different writers (cf footnote 15). The corpus contains two letters written by the merchant Jan Jacobsen Tinnegieter. One letter was sent in 1664, but the second letter was sent eight years later, in 1672. I believe that Jan was between 20 and 30 years old when he wrote his first letter, but that he was over thirty when writing his second letter. This one individual should therefore be represented as two different writers in the corpus: as *a man younger than thirty* from Zeeland belonging to the upper-middle class on the one hand, and as *man between thirty and fifty years of age* from Zeeland belonging to the upper-middle class on the other hand. The same applies to Lieven de Wever: the corpus contains a letter written by him in 1665 and one written in 1672.

Independent variables of importance for the sub-corpus of autographs are: the sender's gender, class, age, and the region to which the sender is most closely linked. How these variables are represented in the sub-corpus of autographs will be discussed for the private and business letters separately.

Private letters
The distribution of the writers across the different regions is presented in table 2.2. With sixty-one writers from Amsterdam (NH-ams) and 43 writers that can be linked to another part of North Holland (NH), this province is best represented. The second region in line is Zeeland (Zee), with 59 writers. Twenty-two writers were linked to South Holland (SH) and 14 writers come from other regions: Flanders, Friesland and Germany. Three writers were left that could not be linked to a region with reasonable certainty.

	Zee	*SH*	*NH-ams*	*NH*	*Other*	*Unknown*	**Tot**
Writers	59	22	61	43	14	3	**202**

Table 2.2: The distribution of the writers of private autographs across the different regions

The writers of the private autographs stem from different social groups. Table 2.3 shows the distribution. Although the largest group of writers is linked to the upper-middle class, the corpus also comprises writers from the lower classes. Ten writers are members of the lower class and 36 were assigned to the lower-middle class. The upper class is represented by 17 writers. A large part of the writers categorised as 'Unknown' probably belong to one of the lower classes as well (see §2.2.4).

	LC	*LMC*	*UMC*	*UC*	*Unknown*	**Tot**
Writers	10	36	105	17	34	**202**

Table 2.3: The distribution of the writers of private autographs across the different social classes

About a quarter of the private autographs have been written by women. The language use of a total of 59 different women (in 71 letters) is therefore available for research. The rest of the letters (189 letters) have been written by 143 different male writers. The ratio between men and women may not be an ideal 1:1, but this number of seventeenth-century female writers stemming from all sorts of social layers and different regions is already unique in the history of Dutch historical sociolinguistics. The difference between the number of male and female writers in the corpus of autographs is caused by two factors. A first factor is that seventeenth-century women were on the whole less literate than men (Van Doorninck & Kuijpers 1993: 14; Frijhoff & Spies 1999: 237-238). A second factor is that it is often harder to determine whether a woman was able to write than it is the case of men, because it is often more difficult to find information about seventeenth-century women and their occupation (Frijhoff & Spies 1999: 190-191). This causes a larger share of the letters written by women to end up in the sub-corpus of letters of uncertain authorship. This difference between men and women will also be discussed in §3.2.1.

Table 2.4 shows the distribution of the writers of the private autograph letters across the different age groups. The majority of the writers are under 50 years of age: only thirteen writers are older than 50. Ninety-three writers are younger than 30, 80 writers are between 30 and 50 years old. Sixteen writers could not be assigned to one of these age groups.

	<30	*30-50*	*50+*	*Unknown*	**Tot**
Writers	93	80	13	16	**202**

Table 2.4: The distribution of the writers of private autographs across the different age groups.

To conclude this section, I include table 2.5 and 2.6 which show the distribution of the male and female writers in this sub-corpus across region, class, and age.

Men		Region						Tot Age	Tot Class
Class	Age	Zee	SH	NH ams	NH	Other	Unknown		
LC	**<30**			1	1			2	5
	30-50		3					3	
	50+								
	Unknown								
LMC	**<30**	4			7			11	20
	30-50	1	1		2	1		5	
	50+		1	1				2	
	Unknown	1			1			2	
UMC	**<30**	23	1	10	6		2	42	86
	30-50	18	2	8	5			33	
	50+			2	2			4	
	Unknown			1	4	2		7	
UC	**<30**			4				4	11
	30-50	1		1	3			5	
	50+				1	1		2	
	Unknown								
Unknown	**<30**	5				1		6	21
	30-50	2	1	4	1	1		9	
	50+								
	Unknown					6		6	
Total Region		55	9	32	33	12	2	**TOT**	**143**

Table 2.5: The distribution of the male writers of private autographs across class, region, and age.

Women		Region							
Class	Age	Zee	SH	NH ams	NH	Other	Unknown	Tot Age	Tot Class
LC	**<30**		1	1	1			3	5
	30-50			1	1			2	
	50+								
	Unknown								
LMC	**<30**	1	4	2	3			10	16
	30-50		1	3	2			6	
	50+								
	Unknown								
UMC	**<30**		1	4			1	6	19
	30-50	2	1	9	1			13	
	50+								
	Unknown								
UC	**<30**		1		1			2	6
	30-50								
	50+		1	1	1	1		4	
	Unknown								
Unknown	**<30**		3	4				7	13
	30-50	1		3				4	
	50+			1				1	
	Unknown				1			1	
Total Region		4	13	29	10	2	1	**TOT**	**59**

Table 2.6: The distribution of the female writers of private autographs across class, region, and age.

Since I want to take into account four variables (gender, age, class, and region) which each have a number of variants, there is a considerable number of slots within the sub-corpus of private autographs into which writers can be fitted (240 in total). Since this sub-corpus of the *Letters as Loot* corpus only comprises letters written by 202 different writers, it is only logical that some slots remain empty.

Business letters

A small sub-corpus of business letters was compiled in order to enable me to compare the language use of private letters with that in business letters. Since this is no more than a sideline in my research, the sub-corpus of business letters was kept small: it contains only 50 letters. Of these 50 letters, 47 letters are autographs and have been written by 41 different writers. Again the variables gender, class, age and region are of importance for these 47 autograph business letters. However, since this sub-corpus of business autographs does not contain any letters written by women or members of the lowest social class, the variable gender will not be dealt with in the following discussion and the lower class will be left out when dealing with the variable social class.

The distribution of the writers across the social classes can be gathered from table 2.7. It is undeniable that the upper-middle class is strongly represented in the business letters. This social class is even more dominant than in the sub-corpus of private letters: 35 of the 41 different writers of business letters belong to this social group.

	LMC	*UMC*	*UC*	*Unknown*	**Tot**
Writers	1	35	2	3	**41**

Table 2.7: The distribution of the writers of business autographs across the different social classes.

When it comes to region, the province of North Holland is best represented, with sixteen writers in total. Almost all of these writers (15) are linked to the Republic's largest city: Amsterdam. Eleven writers are hard to link to a particular region. The region of Zeeland follows closely with ten writers in total. Only one writer is linked to South Holland. That leaves us with three writers related to other regions: two writers who originate from Flanders and one writer, Heinrich Rode, whose name and language use reveal that he must be linked to Germany or a German speaking region.

	Zee	*SH*	*NH-ams*	*NH*	*Other*	*Unknown*	**Tot**
Writers	10	1	15	1	3	11	**41**

Table 2.8: The distribution of the writers of business autographs across the different regions.

The distribution of the writers of business letters across the different age groups is shown in Table 2.9. The age of a large group of writers could not be determined. Seventeen men had to be assigned to the 'unknown' group. This can be explained by the fact that business letters do not contain

elaborate references to the writer's private life, while it is exactly this kind of references that reveals the most about a sender's age. Out of the 24 remaining writers, nine were attributed to the youngest age group and fifteen to the middle-aged group.

	<30	*30-50*	*50+*	*Unknown*	**Tot**
Writers	9	15	0	17	**41**

Table 2.9: The distribution of the writers of the business autographs across the different age-groups.

Finally, table 2.10 shows the distribution of all the writers of business letters across age, social class and region. Again, not every slot of the table could be filled, but in this case the overrepresentation of the upper-middle class is particularly striking. When dealing with this sub-corpus of business autographs one needs to be well aware of the fact that it could almost be considered as a sub-corpus of upper-middle-class letters at the same time.

| | Business | Region | | | | | | | |
| | | | | | | | | Tot Age | Tot Class |
Class	Age	Zee	SH	NH ams	NH	Other	Unknown		
LMC	<30								1
	30-50								
	50+								
	Unknown			1				1	
UMC	<30	3		3			2	8	35
	30-50	5		8			1	14	
	50+								
	Unknown	1	1	2		2	7	13	
UC	<30			1				1	2
	30-50	1						1	
	50+								
	Unknown								
Unknown	<30								3
	30-50								
	50+								
	Unknown				1	1	1	3	
Total Region		10	1	15	1	3	11	**TOT**	**41**

Table 2.10: The distribution of the writers of business autographs across class, region, and age.

2.3.2 The sub-corpora of non-autographs and letters of uncertain authorship

There are 117 non-autograph letters in total that were written by 77 different writers. The number of words in these non-autograph letters amounts to about 45,600. This leaves 171 letters of uncertain authorship. These letters have been written by 149 different writers and comprise about 62,300 words. Since the sub-corpora of non-autographs and letters of uncertain authorship will both be used in the same way in the various case studies of this

dissertation, they can be combined into a larger corpus of 288 letters. The majority of these letters, 285 of them to be precise, are private letters. There is only one non-autograph business letter, which could not be linked to a region. Only two of the letters of uncertain authorship are business letters: they have been written by two different writers – one linked to Amsterdam and the other linked to the province of North Holland. Given these low numbers of business letters, I will not distinguish between private and business letters in the further description of this combined sub-corpus below. Although, of course, I will maintain the distinction throughout the various investigations presented in this dissertation.

The sub-corpora of non-autographs and letters of uncertain authorship combined contain 288 letters written by 222 different writers. The words add up to about 107,900. Just as in the corpus of autographs, the number of writers in the combined corpus of non-autographs and letters of uncertain authorship (222) is smaller than the sum of the writers of the separate sub-corpora (77 + 149 = 226). This is the case because of four writers who appear both in the sub-corpus of non-autographs and in he sub-corpus of letters of uncertain authorship. Take for instance the writer A:Z. This writer is responsible for 4 letters written in Amsterdam: two letters from Liesbeth Ariaans, one letter from Elisabeth Rijnhout-Goskes, and one letter from Annete Klaas. I am certain that Liesbeth Ariaans and Elisabeth Rijnhout-Goskes have not written their letters themselves, so their letters were incorporated in the sub-corpus of non-autographs. But there is still doubt about whether Annete could write. Her letter is therefore incorporated in the sub-corpus of letters of uncertain authorship. The writer A:Z thus features in both sub-corpora.

The sum of the writers of the sub-corpus of autographs and the sub-corpus of non-autographs and letters of uncertain authorship (232 + 222 = 454) does not equal the total number of different writers in the corpus given at the start of this section (441). This discrepancy is accounted for by 13 writers who are to be found both as writers of autographs and as writers of non-autograph letters. Take for instance Marretie Flipse. She sent a letter to her brother in law of which we know for certain that it is an autograph since we could retrieve her signature in the Archive of Amsterdam. But Marretie also wrote letters for her sister, Elisabeth Flipse Ameling. Marretie is thus found as a writer in the corpus of autographs as well as in the corpus of non-autographs.

Table 2.11 represents the distribution of the writers of non-autograph letters and the letters of uncertain authorship across the different regions. As always, the region of North Holland has a large number of writers, 104 in total: 64 writers are linked to Amsterdam, while 40 writers are linked to another town or city in North Holland. The regions of South

Holland and Zeeland are almost equally well represented, with 30 and 25 writers respectively. Seven writers were linked to other regions: Germany, Norway, Friesland, Flanders and Gelderland.

	Zee	*SH*	*NH-ams*	*NH*	*Other*	*Unknown*	**Tot**
Writers	25	30	64	40	7	56	**222**

Table 2.11: The distribution of the writers of the non-autograph letters and the letters of uncertain authorship across the different regions.

What seems to be surprising is that the second largest group of writers is the 'unknown' group of 56 writers. This is very different from the sub-corpus of autographs, in which the number of writers that could not be linked to a specific region was only four. There is a simple explanation for this anomaly which has to do with the fact that the writers of non-autograph letters are often unknown and the writers of letters of uncertain authorship are unknown by definition. When letters pertaining to these sub-corpora were written from abroad, this causes major problems in identifying the region to which the writer could be linked. A more detailed explanation was already given in §2.2.2.

2.3.3. Restriction on number of words per writer
The description above has shown that the seventeenth-century *Letters as Loot* corpus does not contain only one letter per writer. Of several letter writers represented in the corpus I had two or more letters at my disposal before the construction of the corpus, but I stress here that I did not use all available letters in the final version of the seventeenth-century *Letters as Loot* corpus. In order to avoid overrepresentation of linguistic data of certain writers, the number of words per writer in the corpus was restricted to a maximum of about 2000 words. This limit was chosen on the basis of the longest letter in the entire collection of seventeenth-century letters at my disposal which was provided by a writer of whom we only have one letter: the letter of Trijntje Batens to her husband, which counts 1841 words. For writers of whom there is more than one letter available to us, no letters were left out of the corpus if the sum of words of all these letters was lower than 2000. If the sum of words in the different letters exceeded this number significantly, one or more letters were not taken up in the final *Letters as Loot* corpus. Because of this limit, prolific writers do not have a (much) larger share in words than writers of whom the corpus contains only one letter.

2.3.4. Summary
Since the structure of the *Letters as Loot* corpus for the seventeenth century is rather complicated, I present a simple overview of the corpus in table 2.12. This overview lists the number of letters, writers and words comprised by each sub-corpus.

Sub-corpus	*# letters*	*# writers*	*# words*
Autographs (private)	260	202	118,000
Autographs (business)	47	41	19,000
Autographs Total	307	232	137,000
Non-autographs (private)	116	76	45,370
Non-Autographs (business)	1	1	230
Non-autographs Total	117	77	45,600
Letters of uncertain authorship (private)	169	147	62,040
Letters of uncertain authorship (business)	2	2	260
Letters of uncertain authorship Total	171	149	62,300
Non-autographs combined with letters of uncertain authorship (private)	285	219	107,410
Non-autographs combined with letters of uncertain authorship (business)	3	3	490
Non-autographs combined with letters of uncertain authorship Total	288	222	107,900
Entire corpus (private)	545	408	225,410
Entire corpus (business)	50	44	19,490
Entire corpus	595	441	244,900

Table 2.12: An overview of the structure of the seventeenth-century *Letters as Loot* corpus

2.4. A methodological remark

As far as the analysis of the data in the case studies is concerned, I have chosen to use descriptive statistics only. As explained in great detail by Vosters (2011: 218-222), there is much disagreement within (historical) sociolinguistics and the field of language variation and change about the employability of different types of tests. There are some frequently used methods within these fields, namely variable rule analysis (Tagliamonte 2006) and logistic regression (usually performed in SPSS), but the use of these two standard statistical methods in (historical) sociolinguistic research has recently been criticised as well. This is due to the fact that neither of

these methods take into account the variation between different language users (Johnson 2009, Tagliamonte & Baayen 2011). This is important for the case studies of the *Letters as Loot* corpus. In these studies, the social variables are linked to the writers of the letters. It often happens that one letter writer provides several tokens, and in these cases the letter writer "becomes a source of variation that should be brought into the statistical model" (Tagliamonte & Baayen 2011: 143). However, neither the variable rule analysis nor logistic regression treat the writer as a variable, and using these tests would thus produce unreliable results: the tests would show significance too easily.

Given this criticism and the fact that the new statistical methods suggested in Johnson (2009) and Tagliamonte & Baayen (2011) are not feasible yet, I followed Vosters (2011) in using descriptive statistics in the case studies of this dissertation. I will analyse the distributional differences of the linguistic variants under examination with the help of cross tabulation.

2.5. Conclusions

In this chapter I have shown how the *Letters as Loot* corpus was built and how the independent variables were operationalised. The next chapter will show why and how the status of the letters, the final independent variable, was determined. Both these chapters combined tell the complete story of the compilation of the corpus. The entire process of transcribing and correcting transcriptions, getting to know the social history of the seventeenth century and tracing the desired information about writers was very time-consuming, but yielded rich rewards in the form of a corpus unparalleled in the history of sociolinguistic research on seventeenth-century Dutch: a corpus of more than 240,000 words in nearly 600 different (and mainly private) letters, written by 441 writers – men *and* women – of all sorts of social backgrounds.