



Universiteit
Leiden
The Netherlands

Policy versus Practice. Language variation and change in eighteenth- and nineteenth-century Dutch

Krogull, A.

Citation

Krogull, A. (2018, December 12). *Policy versus Practice. Language variation and change in eighteenth- and nineteenth-century Dutch*. Retrieved from <https://hdl.handle.net/1887/67132>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/67132>

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

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/67132>

Author: Krogull, A.

Title: Policy versus Practice. Language variation and change in eighteenth- and nineteenth-century Dutch

Issue Date: 2018-12-12

Policy versus Practice

Language variation and change in
eighteenth- and nineteenth-century Dutch

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
<http://www.lotschool.nl>

ISBN: 978-94-6093-306-6
NUR 616

Copyright © 2018: Andreas Krogull. All rights reserved.

Policy versus Practice

Language variation and change in
eighteenth- and nineteenth-century Dutch

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 12 december 2018
klokke 15.00 uur

door

Andreas Krogull

geboren te Gelsenkirchen
in 1985

Promotor Prof. dr. M.J. van der Wal
Co-promotor Dr. G.J. Rutten

Promotiecommissie Prof. dr. M.G. Kossmann
Prof. dr. J. Schaeken
Dr. R. Vosters (Vrije Universiteit Brussel)
Prof. dr. E. Ziegler (Universität Duisburg-Essen)

Het onderzoek voor dit proefschrift is uitgevoerd als onderdeel van het project *Going Dutch. The Construction of Dutch in Policy, Practice and Discourse (1750–1850)*, gefinancierd door een Vidi-beurs van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), toegekend aan Dr. Gijsbert Rutten.

Table of contents

Acknowledgements	IX
CHAPTER 1	
Introduction	1
1 Main research objectives	1
2 Sociolinguistic situation: Standard and non-standard Dutch	3
3 Dissertation outline	6
CHAPTER 2	
Historical background	9
1 Historical-sociolinguistic context	9
1.1 Socio-political overview	9
1.2 Enlightenment movement	10
2 Metalinguistic discourse	13
2.1 Elitist grammar (1700–1740)	14
2.2 Civil grammar (1740–1770)	15
2.3 National grammar (1770 onwards)	15
3 Language policy: The <i>schrijftaalregeling</i> of 1804/1805	16
3.1 Siegenbeek (1804): National orthography	16
3.2 Weiland (1805): National grammar	19
4 Language-in-education policy	21
CHAPTER 3	
Theoretical framework	23
1 The field of historical sociolinguistics	23
2 Principles and challenges	24
3 Language histories from different perspectives	27
3.1 Traditional language histories	27
3.2 Language histories from below	28
4 Data and corpora	30
5 Prescriptivism, language planning and policy	32
CHAPTER 4	
Corpus and methodology	37
1 Introduction	37
2 Compiling the <i>Going Dutch Corpus</i>	38
2.1 Collection and selection of data	38

2.2	Transcription procedure and conventions	43
2.3	Size and structure of the final corpus	47
3	Variational dimensions of the <i>Going Dutch Corpus</i>	49
3.1	Genre dimension	49
3.1.1	Private letters	51
3.1.2	Diaries and travelogues	55
3.1.3	Newspapers	59
3.2	Diachronic dimension	60
3.3	Spatial dimension	61
3.3.1	Regions	62
3.3.2	Centre versus periphery	65
3.4	Social dimension	66
3.4.1	Social class	66
3.4.2	Gender	67
4	Individual dimension and the <i>Martini Buys Correspondence Corpus</i>	69
5	The <i>Normative Corpus of the Northern Netherlands</i>	70
6	Procedure and methodological remarks	72
6.1	Systematic methodological procedure for linguistic analyses	72
6.2	Final remarks on statistical methods	73
CHAPTER 5		
	Orthographic variables (1): Syllable-final /xt/	75
1	Discussion in Siegenbeek (1804)	75
2	Eighteenth-century normative discussion	79
3	Previous research	85
4	Corpus analysis	86
4.1	Method	86
4.2	Results	87
5	Discussion	98
CHAPTER 6		
	Orthographic variables (2): Final /t/ in <i>d</i>-stem verbs	101
1	Discussion in Siegenbeek (1804)	101
2	Eighteenth-century normative discussion	102
3	Previous research	107
4	Corpus analysis	108
4.1	Method	108
4.2	Results	110
5	Discussion	117
CHAPTER 7		
	Orthographic variables (3): Word-medial and word-final /s/	121
1	Discussion in Siegenbeek (1804)	121

2	Eighteenth-century normative discussion	123
3	Previous research	127
4	Corpus analysis	128
4.1	Method	128
4.2	Results	128
5	Discussion	136
CHAPTER 8		
	Orthographic variables (4): Long <i>e</i>'s in open syllable	139
1	Discussion in Siegenbeek (1804)	139
2	Eighteenth-century normative discussion	142
3	Previous research	149
4	Corpus analysis	150
4.1	Method	150
4.2	Results	152
5	Discussion	161
CHAPTER 9		
	Orthographic variables (5): West Germanic *<i>ī</i>	165
1	Discussion in Siegenbeek (1804)	165
2	Eighteenth-century normative discussion	167
3	Previous research	173
4	Corpus analysis	174
4.1	Method	174
4.2	Results	175
5	Discussion	183
CHAPTER 10		
	Morphosyntactic variables (1): Neuter relative pronouns	187
1	Relativisation in Dutch	187
2	Discussion in Weiland (1805)	190
3	Eighteenth-century normative discussion	194
4	Previous research	198
5	Corpus analysis	199
5.1	Method	199
5.2	Results	203
6	Discussion	211
CHAPTER 11		
	Morphosyntactic variables (2): Masculine and feminine singular and plural relative pronouns	215
1	Corpus analysis	215
1.1	Method	215
1.2	Results	219

2	Discussion	228	
3	Relative pronouns: General conclusion	231	
CHAPTER 12			
	Morphosyntactic variables (3): Genitive case		235
1	The genitive case in Dutch	235	
2	Discussion in Weiland (1805)	236	
3	Eighteenth-century normative discussion	238	
4	Previous research	244	
5	Corpus analysis	245	
	5.1 Method	245	
	5.2 Results	250	
6	Discussion	268	
CHAPTER 13			
	Inter- and intra-individual variation and change: The Martini Buys family correspondence		273
1	Introduction and research objectives	273	
2	The Martini Buys family correspondence	276	
	2.1 Compiling the <i>Martini Buys Correspondence Corpus</i>	276	
	2.2 Size and structure of the <i>Martini Buys Correspondence Corpus</i>	276	
	2.3 Family background	278	
3	Corpus analysis	279	
	3.1 Method	279	
	3.2 Results	280	
4	Discussion	290	
CHAPTER 14			
	Conclusion		295
1	Introduction	295	
2	Measuring policy success: Orthography and morphosyntax	295	
3	Genre, space, gender and the individual: Assessing the external variables	301	
4	Policy versus practice	306	
5	Concluding remarks	308	
	References		309
	Appendix: Transcription conventions for handwritten ego-documents		327
	Dutch summary (<i>Samenvatting</i>)		331
	Curriculum Vitae		341

Acknowledgements

I would like to take the opportunity to thank a number of people who have played a vital part while writing this dissertation, directly or indirectly.

First of all, I want to thank my supervisors Marijke van der Wal and Gijsbert Rutten. I am very grateful for your time and commitment, your sharp eyes and honest, but always constructive feedback throughout the entire process – and beyond! Our regular meetings provided plenty of food for thought and motivation to work on another and yet another chapter. Four and a half years later, I am convinced that I could not have wished for better supervisors. It has been a great pleasure being part of the *Going Dutch* team and representing our project in so many different occasions and places, from Eichstätt to New York.

This dissertation would not have been the same without our wonderful research assistants Christa Bouwmans, Hielke Vriesendorp and Gillan Wijngaards. Thank you for helping me transcribing my corpus data, and spending countless hours neatly separating the *ij*'s from the *j*'s and *y*'s.

Beyond *Going Dutch*, I want to thank Tanja Simons for refreshing discussions over coffee and for being a helpful companion during my time as lecturer. Also, thanks to my hard-working colleagues in PhD office 1.03, which, to my own surprise, I enjoyed so much that I stayed until the finishing touches.

I truly believe that my PhD years would not have been such a pleasant experience without the extended HiSoN community and all the inspiring and lovely people I have met at conferences and summer schools through the years (...too many names to mention here).

Certainly not forgetting my Duisburg-Essen roots, I want to thank Sandra Jansen in particular for introducing me to the world of linguistics and being a great mentor during my formative years in academia. Also, *hartelijke dank* to the Dutch department in Essen for their support over the years.

While I had to leave a few dear people behind when moving to Leiden, I am happy that they are still with me. Many heartfelt thanks to my family and close friends for their support and understanding when I prioritised my research-related activities all too often. *Meine lieben Eltern, ich danke euch für euer Verständnis, wenn ich mich mal wieder zu lange nicht gemeldet habe.*

Last but not least: Thank you, Joost, for going on this journey with me, for your genuine interest in everything I have been working on and even sharing my fascination for Sibilla Martini's spelling behaviour. Off to new adventures!

Leiden, October 2018

CHAPTER 1

Introduction

1 Main research objectives

In this dissertation, I examine language variation and change in late eighteenth- and early nineteenth-century Dutch. More specifically, my main research objective is to investigate the effects of the national language policy introduced in the early 1800s on actual language usage in the Northern Netherlands¹, roughly corresponding to the area of the present-day Netherlands.

Socio-historically, the decades around 1800 can be considered a fundamental phase of Dutch nationalism and nation building, which also had major implications for linguistic matters. As the concepts of nation and language were closely intertwined during this period, a homogeneous standard variety of Dutch was called for in order to symbolise ‘the’ Dutch nation. The strong nationalist discourse ultimately resulted in a top-down language policy at the beginning of the nineteenth century. In the so-called *schrijftaalregeling* ‘written language regulation’ of 1804/1805, the spelling and grammar of Dutch were officially regulated on behalf of the national government. The standard language rules laid down in the official orthography (Siegenbeek 1804) and the official grammar (Weiland 1805) became the norm for the administrative and educational domains. Moreover, they were implemented in the national language-in-education policy, aiming at the spread of Standard Dutch across the population at large (cf. Chapter 2 for a detailed outline).

As the first political interference with linguistic issues and, moreover, the first official codification of Dutch, the *schrijftaalregeling* clearly marks a turning point in the long history of Dutch standardisation. More generally, these concrete language policy measures also constitute a crucial intervention in the sociolinguistic situation of the Northern Netherlands, discursively introducing a clear-cut distinction between ‘Standard Dutch’ and ‘non-standard’ varieties (Section 2).

From a historical-sociolinguistic perspective, the decades around 1800 are thus a highly interesting period. Surprisingly, the intriguing relationship between language policy and language practice in these early years of Dutch nationalism has hardly been studied so far. This research gap in Dutch historical (socio)linguistics is sought to be filled in the research programme *Going Dutch. The Construction of Dutch*

¹ Throughout this dissertation, I will use the term *Northern Netherlands* (for Dutch *Noordelijke Nederlanden*) in order to refer to the area of the Low Countries that roughly corresponds to the present-day Netherlands. In contrast, the area referred to as *Southern Netherlands* (for Dutch *Zuidelijke Nederlanden*) more or less corresponds to the Dutch-speaking part of present-day Belgium.

in *Policy, Practice and Discourse (1750–1850)*², conducted at Leiden University. Addressing three tightly interrelated topics, the research of the *Going Dutch* project is tripartite. First, it will be discussed why and how the rise of the standard language ideology was construed in public and academic discourse on linguistic diversity (e.g. Rutten 2016a, 2016b). Secondly, it will be investigated how this ideology was implemented through language and language-in-education policies (cf. Schoemaker & Rutten 2017; Schoemaker 2018). Thirdly, the effectiveness of these policies will be assessed through an analysis of their influence on actual language use. The latter is, in fact, the central research objective of my PhD sub-project.

The question at the heart of this dissertation is whether and to which extent the national language policy of the early 1800s exerted influence on patterns of language variation and change. In other words, how successful was the Dutch *schrijftaalregeling* in spreading Standard Dutch as ‘the’ Dutch language across the population? Did the citizens of the Dutch nation adopt the officialised standard norms in their actual language use, as envisaged by the national government?

The effectiveness of the top-down language policy will be assessed through a systematic analysis of authentic language usage data drawn from the newly compiled *Going Dutch Corpus*, a diachronic multi-genre corpus of late eighteenth- and early nineteenth-century Dutch (1770–1790, 1820–1840), which comprises more than 420,000 words (cf. Chapter 4).

Building on the research tradition in historical sociolinguistics at Leiden University, most notably the *Letters as Loot* research programme³ on variation and change in seventeenth- and eighteenth-century Dutch (Rutten & van der Wal 2014; cf. also Nobels 2013, Simons 2013), my dissertation, as part of the *Going Dutch* programme, seeks to fill in more *witte vlekken* ‘blank areas’ in the history of Dutch (van der Wal 2006). My research shifts the focus to variation and change in the intriguing period of political and linguistic nationalism in the eighteenth and nineteenth centuries, which has largely remained understudied from a historical-sociolinguistic point of view.

In general terms, it is also the aim of my dissertation to contribute to the international discussion on the success of language planning and policy in historical contexts. Not only for the Dutch case, but for most European languages, hardly anything is known about the impact of concrete policy measures on actual language practice in the past (e.g. Langer 2011, Rutten et al. 2014a, Moliner & Ziegler 2017). Therefore, it is important to gain new insights into how these external (top-down) endeavours to influence language usage affected patterns of variation and change.

² The research programme *Going Dutch. The Construction of Dutch in Policy, Practice and Discourse (1750–1850)* is funded by a Vidi grant of the Netherlands Organisation for Scientific Research (NWO), awarded to Gijsbert Rutten.

³ The research programme *Letters as Loot. Towards a non-standard view on the history of Dutch* (2008–2013) was initiated and directed by Marijke van der Wal and funded by her Free Competition grant by the NWO.

2 Sociolinguistic situation: Standard and non-standard Dutch

The historical event that marks the point of departure of this dissertation is the introduction of the Dutch *schrijftaalregeling* in 1804/1805. As briefly mentioned in Section 1, this ‘written language regulation’ encompasses the publication of the first officialised reference works of Standard Dutch, viz. the national orthography by Matthijs Siegenbeek (*Verhandeling over de Nederduitsche spelling*, 1804) and the national grammar by Petrus Weiland (*Nederduitsche spraakkunst*, 1805). The codification of spelling and grammar was commissioned by the national government, laying down the written standard rules to be used in the administrative and educational domains. In that sense, the official regulations of the *schrijftaalregeling* were concrete language policy measures during the early years of Dutch nationalism.

For the ongoing process of Dutch standardisation, which goes back to the sixteenth and seventeenth centuries (e.g. van der Wal 1995), the official codification of spelling and grammar on behalf of the government has to be regarded as a decisive turning point. Although there had been a vivid normative tradition before 1800, with numerous spelling guides and grammar books being published particularly throughout the eighteenth century, these earlier normative works were mainly targeted towards a fairly restricted readership of educated men, such as ministers and poets. However, when a homogeneous variety of Dutch was called for in the later decades of the eighteenth century, conceptualised as a symbol of the Dutch nation, the intended audience was gradually widened (see Rutten 2009a). The shift in normative discourse clearly mirrors the socio-political developments of the nation-building decades around 1800. Rutten (2016c: 213) points out that

[t]he crucial ideological step includes the aim to spread one form as the language among the population, to eradicate all other forms such as local dialects, to reconceptualise the preferred variety as the only ‘real’ variety of ‘the’ Dutch language, and to develop a national educational system to disseminate the one ‘neutral’ variety.

Against the background of a strong nationalist debate, in which nation and language formed one cohesive ideological framework, metalinguistic discourse thus aimed at the population at large – men and women, from all layers of society and from all regions of the Northern Netherlands. With respect to the nationalist orientation of metalinguistic discourse around 1800, the period under investigation in this dissertation is significantly different from the earlier stages of standardisation in the sixteenth and seventeenth centuries (e.g. van der Wal 1995).

In the Northern Netherlands, the standard language ideology emerged in the decades around 1800 and was almost immediately implemented in language and language-in-education policies (Rutten 2016a). The official policy measures of the early nineteenth century have to be considered a crucial intervention in the contemporary sociolinguistic situation. Metalinguistically and ideologically, the Dutch language was split into two rigid categories, viz. standard and non-standard. On the one hand, Standard Dutch, as regulated in the *schrijftaalregeling* and codified in Siegenbeek’s orthography (1804) and Weiland’s grammar (1805), was discursively

constructed as ‘the’ Dutch language of ‘the’ Dutch nation. On the other hand, all other forms and varieties deviating from the national standard, i.e. the traditional dialects and localisable variants, were categorised as and, in fact, degraded to non-standard. While elevating the national standard variety to the only ‘real’ version of Dutch, the language policy at the same time suppressed the so-called *platte taalen* ‘vulgar languages’ (van der Ploeg 1800), which were considered to constitute a threat to the (linguistic) unity of the young nation.

The new sociolinguistic situation from the early nineteenth century onwards, involving the categorisation of Dutch into standard and non-standard varieties, could be described as diglossic (Rutten 2016c: 215), referring to the hierarchical situation in which Standard Dutch represents the H-variety, as opposed to any kind of non-standard Dutch representing the L-variety⁴. In the sociolinguistic situation before 1800, there was no officialised split into standard and non-standard Dutch in the Northern Netherlands. Nevertheless, traditional language histories have often given the impression of a similarly rigid form of diglossia, even before the government’s interference with linguistic matters. Like for most European languages, the historiography of Dutch usually describes a sociolinguistic situation with a more or less uniform supraregional variety of the written language on the one hand, and the localisable spoken dialects on the other hand (Rutten 2016c: 200). However, it is important to note that traditional language histories are largely based on the printed language of a small upper layer of society, mainly found in literary and formal texts. The written language in these published sources indeed suggests that the standard language gained more and more ground over the centuries. Whereas there is still a fair amount of variation in seventeenth-century printed language, it is commonly assumed that linguistic uniformity must have been consolidated in the eighteenth century (Rutten & van der Wal 2014: 3).

However, from an alternative perspective on language history, which is commonly referred to as *language history from below* and studies the language in handwritten ego-documents such as letters and diaries (cf. Chapter 3), the sociolinguistic situation in the Early and Late Modern periods can hardly be narrowed down to the implied diglossia of a supraregional written language and the traditional spoken dialects. In fact, recent historical-sociolinguistic research has shown that these ego-documents contain much more localisable features and generally much more variability than contemporary published texts (Rutten 2016c: 200; cf. also Rutten & van der Wal 2014), even in the eighteenth century. Linguistic uniformity in the sense of a supraregional written ‘standard’ might have been consolidated in printed language, as suggested by traditional language histories, but certainly not in handwritten ego-documents. These sources *from below* represent a

⁴ In Ferguson’s (1959: 327) seminal paper on diglossia, he distinguished between the superposed variety as the H (‘high’) variety, and the regional dialects as the L (‘low’) varieties. According to Auer (2005: 12), the “linguistic repertoire which is not bilingual but contains a structurally related standard variety in addition to the vernacular varieties (dialects) is called diglossic”. As one of the main characteristics of diglossia, “the standard represents the H-variety and is used for writing and (if spoken at all) for formal situations, whereas the dialect as the L-variety is not (usually) written” (ibid.).

certain kind of intermediacy between standard and dialect, and can be characterised as hybrid, combining localisable traces of the spoken language with features typical of (more or less) standardised writing (cf. Martineau 2013). In the Dutch case, the hybridity of ego-documents in a supposedly diglossic situation in the Northern Netherlands makes it difficult to apply this dichotomous concept to the sociolinguistic space before 1800 (Rutten 2016c: 197). Therefore, the situation before the official language policy is best described as a diaglossic⁵ repertoire, in which “there is no clear-cut separation of standard and dialect” (Auer 2005: 27), ranging from a wide variety of spoken dialects to supralocal writing traditions.

With respect to the diaglossic continuum without a formally standardised variety of Dutch, the top-down language policy of the early nineteenth century with its *schrijftaalregeling* was undoubtedly a major intervention in the sociolinguistic space. For the first time, the conceptual categorisation into standard and non-standard varieties was officialised. Describing the change from a diaglossic to a (metalinguistically constructed) diglossic sociolinguistic situation, Rutten (2016c: 214) argues that

[h]ierarchization of forms and varieties existed well before 1800, but it is with the construction of the written variety of Dutch used by socio-economically privileged groups as the ‘neutral’ variety of Dutch that the diaglossic continuum is discursively split into standard and non-standard [...] This means that, in the northern Netherlands in the era of nationalism, a diglossic interpretation of sociolinguistic space results from a metalinguistic, ideological operation carried out in and applied to a diaglossic situation.

While the splitting of the continuum into standard and non-standard Dutch was supposed to have far-reaching consequences for the language use of the Dutch population, a historical-sociolinguistic perspective calls for an empirical investigation of whether this was actually the case. Therefore, the central question of this dissertation is to what extent the Dutch government succeeded in disseminating the national standard variety of Dutch across the population at large. Although the standard language norms codified in Siegenbeek (1804) and Weiland (1805) were meant to be used in the administrative and educational domains⁶, they were not formally mandatory, certainly not in private writing. Therefore, it remains uncertain how effective the top-down policy measures actually were in spreading the officially constructed Standard Dutch. This research gap in the history of Dutch will be addressed in my dissertation.

⁵ See Rutten (2016d) for a critical discussion of the concepts of *diaglossia* and *diglossia* applied to the histories of Dutch, German and English, in reference to Auer (2005, 2011).

⁶ Interestingly, Marynissen (2005, cf. also Marynissen & Nübling 2010) notes that the prescriptions in Siegenbeek’s (1804) orthography also served as the point of reference for the official spelling regulation of family names in the Northern Netherlands in 1811.

3 Dissertation outline

In addition to this introductory chapter, the following three chapters provide the historical, theoretical and methodological background of my dissertation. To begin with, **Chapter 2** outlines the socio-historical context of the period under investigation, i.e. the second half of the eighteenth and the first half of the nineteenth century. Apart from the main socio-political and discursive developments during this fundamental phase of Dutch nation building, this chapter also serves as an introduction to the language and language-in-education policies of the early 1800s, with a particular focus on the *schrijftaalregeling*, i.e. Siegenbeek's (1804) orthography and Weiland's (1805) grammar.

Chapter 3 presents the theoretical framework of historical sociolinguistics, in which my research is embedded. This relatively young but thriving field of study will be introduced, highlighting its principles and challenges, its alternative perspective on language histories, as well as its data and corpora. As a central topic of this dissertation, the (under-researched) role of language planning and policy and their implications for actual language use will also be addressed.

Chapter 4 then introduces the corpus and methodology of this dissertation. In particular, the newly compiled *Going Dutch Corpus* of late eighteenth- and early nineteenth-century Dutch will be presented in detail, discussing the collection and transcription of data as well as the variational dimensions integrated into the corpus design. The *Going Dutch Corpus* is a diachronic multi-genre corpus of more than 420,000 words, which comprises handwritten ego-documents (private letters, diaries and travelogues) and printed texts (newspapers) from various regions, both from the centre and the periphery of the Northern Netherlands. The ego-documents were written by men and women. The corpus serves as the basis for the empirical analyses of linguistic variables in Chapters 5–12. Furthermore, Chapter 2 also introduces the normative corpus, i.e. an exhaustive selection of normative publications from the eighteenth century until Siegenbeek (1804) and Weiland (1805). Combining quantitative with qualitative methods, these primary sources will be consulted in order to examine and assess the interplay between norms and usage.

Chapters 5–12 contain the corpus-based analyses of eight linguistic variables, comprising five orthographic and three morphosyntactic case studies. Starting with spelling, the selected variables address both consonantal and vocalic issues, all of which were commented on and regulated in the official orthography by Siegenbeek (1804). The first three case studies investigate consonantal variables: (1) the orthographic representation of syllable-final /xt/ in etymologically distinct words (as <cht>, <gt>, <ght>) in **Chapter 5**, (2) final /t/ in particular forms of *d*-stem verbs (as <dt>, <d>, <t>) in **Chapter 6**, and (3) word-medial and word-final /s/ derived from Wgm. **sk* (as <s>, <sch>) in **Chapter 7**. The next two case studies examine vocalic variables: (4) the orthographic representation of etymologically distinct long *e*'s in open syllables (as <e>, <ee>) in **Chapter 8**, and finally (5) Wgm. **i* (as <ij>, <ÿ>, <y> and alternative forms) in **Chapter 9**.

Moving on to grammatical issues, the following chapters focus on the analysis of three morphosyntactic variables, which were discussed in eighteenth-century metalinguistic discourse and regulated in the official grammar by Weiland (1805). In these case studies, I investigate (1) neuter relative pronouns in **Chapter 10**, (2) masculine and feminine singular and plural relative pronouns in **Chapter 11**, and (3) the synthetic genitive case versus the analytical *van*-construction in **Chapter 12**. All based on data from the *Going Dutch Corpus*, it is the aim of these eight orthographic and morphosyntactic case studies to systematically investigate and assess the possible effects of the Dutch nationalist language policy on actual usage patterns.

While Chapters 5–12 focus on usage in the language community at large, **Chapter 13** adds micro-level insights on the impact of language policy by zooming in on intra- and inter-individual variation and change. Based on the additional *Martini Buys Correspondence Corpus*, specifically compiled for this analysis, the use of five features will be revisited in private letter writing across three generations of family members. While building on the (macro-level) insights gained in Chapters 5–12, this chapter allows a complementary perspective on the effectiveness of the official language regulations by taking into account individual behaviour and the factors of age and lifespan change.

The findings of all previous chapters will be brought together in **Chapter 14**, in which I provide a comprehensive discussion of the results. It is the aim of this concluding chapter to give answers to the central research objectives of this dissertation. First and foremost, it will be assessed whether and to which extent the early nineteenth-century language policy exerted influence on patterns of language variation and change. In other words, were the officialised standard norms of the *schrijfaalregeling* successfully adopted in actual language practice? Bringing together the results of all case studies, I will discuss whether differences in policy success can be identified, both with regard to individual linguistic variables, and with regard to spelling and grammar in general.

Making use of all external variables integrated in the *Going Dutch Corpus*, I will also investigate whether and how we can detect variation between the genres and, on a broader level, between manuscript and print. Geographically, I will examine whether differences can be identified between the various regions of the Northern Netherlands, or between the centre and the periphery. Furthermore, the issue of gender variation will be addressed, examining whether the language policy ‘reached’ male and female writers to different extents.

To sum up, this dissertation aims to provide a sophisticated assessment of the effectiveness of early nineteenth-century language policy measures in the Northern Netherlands. Given the fact that the influence of language planning and policy on actual language practice is still a largely understudied topic in historical (socio)linguistics, I also seek to suggest a methodological approach to examine and measure the possible impact of language policy systematically and based on a corpus of authentic usage data. In this respect, the Dutch national language policy of the early 1800s offers a highly interesting and concrete case in language history.

CHAPTER 2

Historical background

1 Historical-sociolinguistic context

1.1 Socio-political overview

In the Northern Netherlands, like in many other parts of (western) Europe, the second half of the eighteenth century and the first half of the nineteenth century witnessed the emergence and rise of a strong nationalist movement. The decades around 1800⁷ in particular can be considered the fundamental years of Dutch nation building, grounded on the socio-political ideals of a homogeneous nation and inclusive citizenship⁸. Towards the end of the eighteenth century, the Northern Netherlands began to undergo a series of radical changes and transformations (e.g. van Sas 2004: ch. 1; Klok & Mijndhardt 2004), which were also “beneficial to the standardisation of Dutch” (Willemyns 2013: 107; cf. de Bonth et al. 1997: 369), resulting in a national language policy and the official codification of the Dutch spelling and grammar in the early 1800s.

In the course of the 1780s, the socio-political climate in the Dutch Republic, also known as the Republic of the Seven United Netherlands (*Republiek der Zeven Verenigde Nederlanden*), was characterised by internal political conflict, in particular by the growing tensions between the republican Patriots (*patriotten*) and the royalist Orangists (*orangisten*). The opposition to the reigning House of Orange, with Prince William V (1748–1806) as its stadtholder, had increased ever since the beginning of the Fourth Anglo-Dutch War (1780–1784). The French Revolution, which had started in 1789, further strengthened the republican and anti-Orangist sentiments of the Patriots, who demanded more freedom and power for the civilian population (e.g. van Sas 2004: ch. 8). However, without a centralised state organisation, it was difficult for the Patriots to start a revolution. Another important political conflict, which divided the Dutch Republic internally, was related to the tensions between the old but mouldering federalism and the more progressive unitarianism, inspired by the ideas of the French Revolution (cf. also Rutten 2016a: 16).

The most profound socio-political developments took place in the so-called Batavian-French era, in which the Northern Netherlands were under a strong French influence. In 1794, the troops of the revolutionary French Republic invaded the Southern Netherlands (under Austrian rule at that time), which were occupied

⁷ For a detailed account of the nation-building period around 1800, see Klok & Mijndhardt (2004).

⁸ The notion of *inclusive citizenship* refers to the socio-political ideal that all members of a given society, rather than only a selected group, are concerned.

and annexed by France in 1795. The same year, the French troops also invaded the Dutch Republic, which became a vassal state of France, forcing the last stadtholder William V to flee into exile in England. With the backing of the French, the Patriots now seized power throughout the country.

In 1795, the peace treaty with France, known as the Treaty of The Hague, led to the foundation of the Batavian Republic (*Bataafsche Republiek*), as the Northern Netherlands were called between 1795 and 1801. Following the French model, the democratic government of the Batavian Republic gradually replaced the federal structure of the old Dutch Republic by a unitarianist form. The year 1796 saw the establishment of the first national parliament, the National Assembly (*Nationale Vergadering*), which was “basically the country’s first representative body” (Kloek & Mijnhardt 2004: 25). The design of a new constitution was one of the parliament’s primary tasks. Approving a proposal that was clearly unitarianist, the first constitution of 1798 (*Staatsregeling voor het Bataafsche Volk*) laid the foundation for the Dutch nation-state. Although the initial constitution underwent many changes in subsequent years, turning the Batavian Republic into the Batavian Commonwealth (*Bataafsch Gemeenebest*) between 1801 and 1806⁹, it became apparent that the new national government could intervene in domains that had never been matters of political concern before (Kloek & Mijnhardt 2004: 26).

By the end of the eighteenth century, the government appointed so-called *agenten* ‘agents, ministers’, including the Minister of National Education (*Agent van Nationale Opvoeding*), Johan Hendrik van der Palm (1763–1840). Education, as well as language, had become issues of political interest and were considered important means to promote the national unity. Van der Palm called for the nationalisation of the school system, which also had decisive consequences for linguistic matters. In fact, it was from this period onward that “grammar and spelling have been focal points of Dutch educational policy, discursively constructing the alleged rules of written Dutch as ‘the’ rules of ‘the’ language of ‘the’ Dutch nation” (Rutten 2016c: 124). These fundamental socio-political developments in the process of Dutch nation building were embedded in the broader ideological context of Enlightenment, which will be outlined in Section 1.2 below.

1.2 Enlightenment movement

With regard to ideology, the eighteenth century was typically characterised by the Enlightenment movement. Like in many parts of western Europe, discourse in the Northern Netherlands aimed at the spread of enlightenment through the population as a whole and, ultimately, at the creation of a homogeneous nation (Rutten 2016b: 45-46). Two of the main themes during the eighteenth-century

⁹ In 1806, after the Batavian regime, Dutch independence came to an end with the creation of the Kingdom of Holland (1806–1810), ruled by Louis Napoleon, the brother of Napoleon Bonaparte. The French period ended with the fall of Napoleon Bonaparte in 1813. Two years later, the Northern and Southern Netherlands were unified into the United Kingdom of the Netherlands (cf. also Kloek & Mijnhardt 2004: 21-30).

movement of Dutch *volksverlichting*, literally ‘folk enlightenment’, were emancipation and social action (Rutten 2016b: 47). Particularly in the second half of the century, many private and semi-public initiatives were actively engaged in the Enlightenment discourse and participated in prize contests on the education of man (cf. e.g. Los 2005; Mijnhardt 1987). Learned and cultural societies were founded in many towns and cities from the 1760s onwards. Rutten & van Kalmthout (2018: 16) argue that “[b]y means of competitions, debates, treatises and readings [...] learned societies contributed in their own way to the national and international exchange of new scientific insights”. Some of these societies were also concerned with language, most notably the *Maatschappij voor Nederlandsche Letterkunde* ‘Society for Dutch Language and Literature’ (henceforth referred to as the Maatschappij) and the *Maatschappij tot Nut van ‘t Algemeen* ‘Society for the Benefit of the Common Good’ (henceforth: the Nut). The Maatschappij, founded in 1766, had its origins in the student societies of university towns like Leiden, and was prominently involved in language-oriented activities. Furthermore, the Nut, founded in 1784, was one of the most influential societies during the Enlightenment period, well known for its contribution to emancipation and social action. Its central aim was social change by spreading enlightened knowledge across all layers of society (e.g. Klok & Mijnhardt 2004). The Nut also strongly argued for a revolution in the national school system, which entailed the production of new school books (Simons & Rutten 2014: 54-55).

The role these learned societies played for the national language and language-in-education policies of the early 1800s can hardly be underestimated. The ideology of inclusive citizenship, explicitly represented by societies like the Nut, coincided with the central ideas of the government, increasingly considering education as a powerful instrument to create national unity. In 1796, the government of the Batavian Republic approached the Nut for expert advice for a nationally-organised educational policy. In the highly influential report *Algemeene Denkebeelden over het Nationaal Onderwijs* ‘General Ideas on National Education’, published in 1798, a broad range of topics and policy measures were discussed, such as a renewed school system, school inspection, teacher training and curricula. Generally, the Nut emphasised the need for mother-tongue education and grammar teaching in the national school system (Rutten 2016b: 46).

In addition to primarily educational aspects, Enlightenment discourse also addressed linguistic matters. Possibly the most explicit arguments with regard to the *moedertaal* ‘mother tongue’ can be found in the well-known prize essay *Het belang der waare volksverlichting* ‘The importance of the true enlightenment of the people’, written by Hidde Wibius van der Ploeg (1769–1853) and anonymously published by the Nut in 1800. Highlighting the mother tongue as “what binds the individual members of the population together, irrespective of their social position or gender” (Rutten 2016b: 49), van der Ploeg (1800: 35) argued that a true enlightenment of the Dutch population would not be possible without a ‘general and rule-based knowledge’ of the nation’s language:

Doch ik [...] beweer, dat waare volksverlichting niet denkbaar is, ‘zonder eene algemeene en naar regels geleerde kennis van de Moedertaal des Lands;’ ja, mijns oordeels, is het nodig, dat het volk, zal het verlicht kunnen heeten, of kunnen worden, in alle Departementen van den Staat, **niet alleen gelijkluidend spreekte, maar ook zodaanig, als men in de volksschriften gewoon is te schrijven en in de openlijke aanspraaken zich uit te drukken;** op dat het volk in staat zij, bij het leezen van de eerste, en het aanhooren van de laatste, alles te verstaan, en zich, in het gemeene leeven, weder overal verstaanbaar te maaken. [emphasis mine]

‘But I argue that the true enlightenment of the people is not imaginable without a general and rule-based knowledge of the mother tongue of the country. Yes, in my opinion, it is necessary that the people, will they ever be or become enlightened, in all departments of the State, **not only speak identically, but also in the way that is commonly written in popular publications and in public speeches,** so that the people are able to understand everything when reading the former and hearing the latter, and, in the common life, can make themselves understood everywhere again.’

According to Rutten (2016b: 50), van der Ploeg’s position was “a clear call for nationwide homogenization both in the spoken and in the written language”. What is more, van der Ploeg (1800: 35; 129) pled for the eradication of so-called *platte taalen* ‘vulgar languages’, i.e. regional dialects, and thus explicitly rejected the use of non-standard Dutch to the benefit of the (yet-to-be-codified) national standard variety:

‘Er zou dus een groot stuk der verlichting gewonnen zijn, indien in een land gene, zo genoemde, *Platte Taalen* gevonden wierden, die zeer hinderlijk zijn in het onderwijs der jeugd, en dus ook in de algemeene verlichting, welke laatste toch alleen, bij het gros des volks, door het eerste kan verkregen worden (van der Ploeg 1800: 35)

‘A great part of enlightenment would be gained, if so-called *vulgar languages* were not found in the country, which are true hindrances in the education of the young, and therefore in the general enlightenment, too, which can only be reached through education among the majority of the population’ (translation by Rutten 2016b: 50)

In other words, van der Ploeg argued that the use of dialects, or any variety other than the national standard variety, was not only an impediment to the education of the youth, but also to the enlightenment of people in general. Strikingly, he even extended the use of the standard language to the private domain, demanding from school teachers to ensure that children always use the *zuivere moedertaal* ‘pure mother tongue’ not just at school, but also among each other and at home with their parents (Rutten 2016a: 22).

It seems evident that this section can only give a glimpse at the Enlightenment discourse in the Northern Netherlands¹⁰. However, it is important

¹⁰ See, for instance, Rutten (2016b) for a more comprehensive outline of the ideological framework, especially in the context of standardisation and the national language policy.

to note that the emergence of language and language-in-education policies in the early 1800s (cf. Sections 3 and 4, respectively) should not only be regarded as the outcome of the socio-political changes with regard to nationalism and nation building, but also against the background of the bigger ideological framework of Enlightenment. The general debates on *volksverlichting* and the participating learned societies, aiming at the spread of enlightenment through the population at large, actively contributed to the nationalisation of language and education. In fact, together with the parallel developments in eighteenth-century metalinguistic discourse (cf. Section 2), they paved the way for the official *schrijftaalregeling* ‘written language regulation’, as investigated in this dissertation.

2 Metalinguistic discourse

As outlined in Section 1, the socio-political and ideological developments in the Northern Netherlands are reflected in eighteenth- and early nineteenth-century metalinguistic discourse. Particularly in the crucial period around 1800, the political ideology of nationalism was also closely tied to linguistic nationalism (cf. Blommaert & Verschueren 1998). From around 1750 onwards, language became “a socializing force” and “a means to establish a community, a nation, and to improve civil society” (Noordegraaf 2004: 218). In other words, the mother tongue was instrumentalised to create a Dutch national identity and, ultimately, a homogeneous Dutch nation. The conception of language as a national symbol increasingly called for a standardised variety of (written) Dutch, which has its roots in eighteenth-century metalinguistic discourse. According to Rutten (2016b: 41),

[t]he major change in the history of standardization [...] occurs in the eighteenth century, when important concepts such as the mother tongue, hierarchization and polishing are radicalized, brought together into one coherent language ideology, and combined with social and political ideas about the nation and about social action and emancipation.

It is also in this period that the so-called standard language ideology (cf. Chapter 3) emerged, which is “intrinsically connected to the nation-building processes of that time” (Rutten 2016b: 41).

There had been a vivid normative tradition in the Northern Netherlands throughout the eighteenth century with the publication of numerous spelling books and grammars ever since the earliest decades. In the course of the century, however, metalinguistic discourse “underwent a social turn” (Rutten 2009a), in which the question of target audience gained in importance. According to Noordegraaf (2004) and van de Bilt (2009: 60–68), the social orientation is one of the defining characteristics of eighteenth-century metalinguistic discourse. From rather diverse language planning activities in the early 1700s, the debates developed into a national language policy in the period around 1800 (Rutten 2016a: 14). Rutten (2012a) distinguishes three different periods in the normative tradition, in which the intended audience was gradually widened: from the period of *elitist*

grammar (1700–1740; cf. Section 2.1) to the period of *civil grammar* (1740–1770; cf. Section 2.2) and, finally, the period of *national grammar* (from 1770 onward; cf. Section 2.3). These three stages paved the way for the national language policy, and more specifically the *schrijftaalregeling* in 1804/1805, with the codification of the first official orthography (Siegenbeek 1804) and grammar (Weiland 1805) (cf. Section 3).

2.1 Elitist grammar (1700–1740)

In the early decades of the eighteenth century, language planning activities were chiefly individual efforts rather than concerned with the construction of a national language (Rutten 2016a: 20). Many of the normative grammars from the first half of the eighteenth century still followed the *vondelianist* tradition, which was explicitly founded on the language of seventeenth-century literary authors, mainly the Dutch poet Joost van den Vondel (1587–1679). With regard to the intended audience, the grammars of the early eighteenth century were primarily targeted towards an elite readership of educated men such as ministers and poets, hence the term *elitist grammar*. As grammars also signalled intellectualism and scholarship in this period, it was almost impossible to study Dutch grammar without knowledge of Latin and/or Greek. In this respect, linguistic education and knowledge of grammar clearly distinguished the intellectual elite from the middle and lower classes (Rutten 2009a: 57). The focus of elitist grammars was on specific written registers, mainly poetry, literary prose and sermons. Rutten (2016b: 39) sums up that this type of grammar was “based on the written language and targeted towards the written language – it is first and foremost a textual discipline, in which references to the spoken language hardly occur”.

Despite the restricted focus on specific registers and audiences, two of the most influential normative grammars of the eighteenth century were published in this period: Arnold Moonen’s *Nederduitsche spraakkunst* ‘Dutch grammar’ of 1706, and Willem Sewel’s *Nederduitsche spraakkunst* ‘Dutch grammar’ (1708, 1712) (cf. Rutten 2009a: 58). Ten Kate’s *Aenleiding tot de kennisse van het verbevene deel der Nederduitsche sprake* ‘Introduction to the knowledge of the sublime part of the Dutch language’ (1723) had a particularly strong historical focus and comparative perspective (cf. also van der Wal 2002a). Other publications from the fairly diverse period of elitist grammar include Petrus Francius’ introduction to his Dutch translation of Gregorius Nazianzenus’ *Van de mededeelzaamheid* (1699), David van Hoogstraten’s *Aenmerkingen over de geslachten der zelfstandige naemwoorden* ‘Remarks on the gender of nouns’ (1700), Jakobus Nylöe’s short grammar *Aenleiding tot de Nederduitsche taal* ‘Introduction to the Dutch language’ (1703), Adriaen Verwer’s Latin grammar of Dutch called *Linguae Belgicae idea grammatical, poetica, rhetorica* (1707) and Balthazar Huydecoper’s *Proeve van taal- en dichtkunde* ‘Essay on linguistics and poetics’ (1730).

2.2 Civil grammar (1740–1770)

In the second half of the eighteenth century, metalinguistic discourse saw the first stage of widening to the extent that the rules for spelling and grammar, as laid down in normative publications, should become common knowledge to all adult citizens of the Northern Netherlands. The intended audience encompassed both men and women as well as the youth of the upper and middle classes, but still excluded the lower classes. Rutten (2009a: 58) argues that the education of Dutch *burghers* ‘citizens’ and the evolution of *civil grammar* were “the linguistic counterparts of the democratic revolutions of the later 18th century”. Compared to elitist grammars, normative works published in this period no longer required any knowledge of Latin and/or Greek. Generally, they were characterised by a comparatively simplified approach, rephrasing complex grammatical issues in a less classical but more comprehensive vocabulary, and also employing educational strategies (Rutten 2009a: 58).

For the period of civil grammar between 1740 and 1770, Kornelis Elzevier’s *Proef van een nieuwe Nederduitsche spraekkonst* ‘Outline of a new Dutch grammar’ (1761), Frans de Haes’ *Nederduitsche spraekkunst* ‘Dutch grammar’ (1764), Jan van Belle’s *Korte wegwijzer, ter spel- spraak- en dichtkonden* ‘Short introduction to orthography, grammar and poetry’ (1748) and *Korte schets der Nederduitsche spraekkonst* ‘Short sketch of the Dutch grammar’ (1755), and Kornelis van der Palm’s *Nederduitsche spraekkunst, voor de jeugdt* ‘Dutch grammar, for the youth’ (1769) are usually regarded as the most important grammars (Rutten 2009a: 56).

2.3 National grammar (1770 onwards)

The third and final period of eighteenth-century metalinguistic discourse is characterised by the further widening of the target audience, which mirrored the inclusive idea that all members of the Dutch nation should be trained in the grammar of their national language (Rutten 2016b: 42). The *national* grammars from the late eighteenth century onward, in fact, addressed all inhabitants of the nation, and often specifically children. From an activity of certain social groups in the early decades of the eighteenth century, grammar had turned into a matter of national concern, aiming at the society as a whole. Instead of dividing the Dutch society, knowledge of grammar was now used to separate the Dutch nation from other nations (Rutten 2009a: 58–59).

Among the important normative works published in the period of national grammar are Ernst Zeydelaar’s *Nederduitsche spelkonst* ‘Dutch orthography’ (1774), Klaas Stijl & Lambertus van Bolhuis’ *Beknopte aanleiding tot de kennis der spelling, spraakdeelen, en zinteecken van de Nederduitsche taal* ‘Concise introduction to the knowledge of the spelling, parts of speech and punctuation of the Dutch language’ (1776), Adriaan Kluit’s *Vertoog over de tegenwoordige spelling der Nederduitsche taal* ‘Treatise on the present spelling of the Dutch language’ (1777), Lambertus van Bolhuis’ *Beknopte Nederduitsche spraekkunst* ‘Concise Dutch grammar’ (1793) and the

anonymous *Rudimenta of gronden der Nederduitsche spraake* ‘Fundamentals of the Dutch language’, written by Gerrit van Varik (and published by the Maatschappij tot Nut van ‘t Algemeen) in 1799.

It was particularly in this period of national grammar that the one language—one nation ideology came into existence (Rutten 2016b: 45), explicitly linking the mother tongue, and the knowledge thereof, with the concepts of citizenship and nationhood. These ideas paved the way for the *schrijftaalregeling* with official regulations for the Dutch orthography and grammar (Section 3).

3 Language policy: The *schrijftaalregeling* of 1804/1805

The early nineteenth century saw the first official codification of a standardised variety of (written) Dutch, laid down in a national orthography and a national grammar. For the first time in the long standardisation history of Dutch, the government was concerned with language and actively involved in the regulation of spelling and grammar. The codification itself was initiated by the Minister of National Education, Johan Hendrik van der Palm (who was the son of the aforementioned Kornelis van der Palm, cf. Section 2.2). In fact, it was one of the minister’s tasks to ‘take all possible measures to purify and cultivate the Dutch language and to regulate its spelling’, as mentioned in the *Instructie voor den Agent der Nationale Opvoeding* ‘Instruction for the Minister of National Education’ (1798: 6):

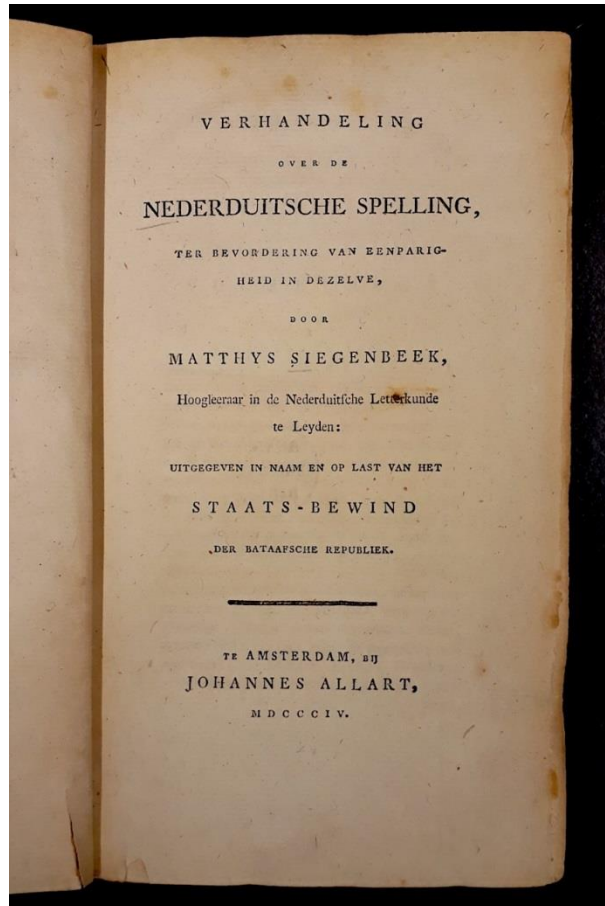
Hy zal alle mogelyke middelen beramen, om de Nederduitsche taal te zuiveren, te beschaven, en derzelver spelling op eenen gelyken voet interigten.

Carried out on behalf of the national government, the *schrijftaalregeling* of 1804/1805 comprised two complementary works, which, according to Noordegraaf (2018: 146), “can be seen as inaugurating the final phase of the codification of the Dutch standard written language”. First, the national orthography of 1804 was codified by Matthijs Siegenbeek (Section 3.1). Secondly, the national grammar, codified by Petrus Weiland, was published one year later in 1805 (Section 3.2). These official regulations for both spelling and grammar were intended to be used in the administrative and educational domains.

3.1 Siegenbeek (1804): National orthography

In the context of the normative tradition in the Northern Netherlands, Matthijs Siegenbeek (1774–1854) is first and foremost known as the codifier of the national orthography of Dutch, published in 1804 as the *Verhandeling over de Nederduitsche spelling ter bevordering van eenparigheid in dezelve* ‘Treatise on the Dutch spelling for the promotion of uniformity therein’ (cf. Figure 1).

Figure 1. Title page of Siegenbeek's orthography (1804, Leiden University Libraries).



Traditionally, however, Siegenbeek, who was born in Amsterdam and trained to be a clergyman, has often been remembered as the first professor of Dutch – even though this claim was repeatedly adjusted in recent years. When Siegenbeek was inaugurated as extraordinary professor of Dutch rhetoric at Leiden University in 1797, several academic professors had been involved in teaching Dutch linguistics, literature and rhetoric at universities before him. Nevertheless, Siegenbeek was the first professor to hold a chair solely devoted to Dutch (Vis 2004: 10; cf. also Rutten 2018: 26). In 1799, his extraordinary chair was changed into a regular chair and widened to Dutch language and literature¹¹.

Apart from his university duties, Siegenbeek was actively involved as a long-term board member of one of the first and most important learned societies,

¹¹ For a comprehensive outline of Matthijs Siegenbeek's activities in the field of Dutch studies and his linguistic publications in particular, see Rutten (2018).

the aforementioned *Maatschappij der Nederlandsche Letterkunde* (cf. Section 1.2). Between 1803 and 1822, he was the secretary of this Society. Siegenbeek also published numerous works in the fields of literary history and rhetoric but also linguistics. Although he discussed a fairly wide range of linguistic topics, Siegenbeek is mainly associated with the *schrijftaalregeling* and his national orthography. As outlined before, it was the Minister of National Education, Johan Hendrik van der Palm, who commissioned Siegenbeek to write the official orthography of Dutch. His *Verhandeling over de Nederduitsche spelling* was published in 1804 *in naam en op last van het Staats-bevind der Bataafsche Republiek* ‘in the name of and by order of the government of the Batavian Republic’, as mentioned on the title page.

Siegenbeek’s orthography was based on a set of three main principles (e.g. van der Wal & van Bree 2008: 318-322; van de Bilt 2009: 206-207), the most fundamental of which was the phonological principle. Commonly known as *Schrijf, zoo als gij spreekt* ‘Write as you speak’ (Siegenbeek 1804: 13), spelling had to be in accordance with the pronunciation: “de spraak dat gene is, ‘t welk door het schrift moet worden uitgedrukt, zoo behoort natuurlijk de eerste ten rigtsnoer te strekken voor het laatste” ‘speech is what needs to be expressed by writing, which is why the former must, of course, serve as a guideline for the latter’ (1804: 14). Rutten (2018: 34) assumes that Siegenbeek was “probably well aware of many regionally and socially conditioned phonetic differences”, when he argued that ‘the most pure and most polite pronunciation’ (*de zuiverste en meest beschaafde uitspraak*, 1804: 18, 26), as heard in his native region of Holland (1804: 20), should serve as a guide for the spelling. In addition to the phonological principle, which Siegenbeek admitted to be insufficient for a fully-fledged orthography, he proposed two more principles. Apart from pronunciation, one must also take into account the *Afleiding der Woorden* ‘derivation of the words’ as well as *het algemeen erkend en aangenomen gebruik* ‘the generally acknowledged and accepted usage’. With respect to *Afleiding*, Siegenbeek not only referred to the etymological principle in the strictest sense, but also encompassed what is commonly known as the principle of *gelijkvormigheid*, literally ‘uniformity’, which implies morpheme consistency. He argued that ‘it is impossible to know the actual power and meaning of the words, without the necessary knowledge of their origin and derivation’ – “dat het onmogelijk is, de eigenlijke kracht en betekenis der woorden wel te kennen, zonder de noodige kennis van derzelver oorsprong en afleidinge” (Siegenbeek 1804: 29).

When working on his *Verhandeling*, Siegenbeek was undoubtedly influenced by his eighteenth-century predecessor Adriaan Kluit. In fact, many of Kluit’s (1763, 1777) spelling choices were followed – and officialised – by Siegenbeek. Van de Bilt (2009: 203-212) even claims that Kluit himself played a crucial role in the codification of Dutch. Interestingly, Kluit was also one of the language experts who had been consulted by J. H. van der Palm, before he officially approved Siegenbeek’s orthography.

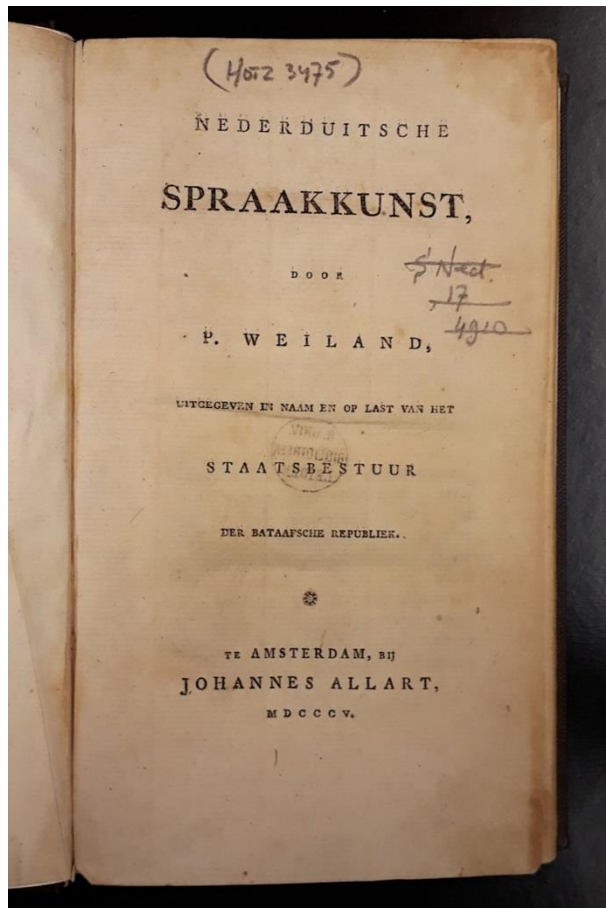
Assessing Siegenbeek’s role in and contribution to the history of Dutch linguistics, Rutten argues that “[t]hroughout his career, Siegenbeek was in defence of Dutch, where *Dutch* should be interpreted as a cultivated, normalised, and

uniform variety modelled after the written language of well-known authors, symbolically representing the Dutch nation” (2018: 27), concluding that “asking Siegenbeek to design the national orthography was clearly the right choice” (2018: 43).

3.2 Weiland (1805): National grammar

Pieter, later Latinised as Petrus, Weiland (1754–1844) made an impact on the history of Dutch as a prolific lexicographer and grammarian, most notably as the codifier of the national grammar of Dutch, his *Nederduitsche Spraakkunst* ‘Dutch grammar’ of 1805 (cf. Figure 2).

Figure 2. Title page of Weiland’s grammar (1805, Leiden University Libraries).



Before that, Weiland, who was born in Amsterdam, studied theology in Leiden and became a Remonstrant minister in Rotterdam. In 1789, Weiland was elected as a member of the *Maatschappij der Nederlandsche Letterkunde*. Confronted with several failed attempts of the Society to compile a comprehensive, explanatory dictionary of Dutch, Weiland suggested that he would be willing to work on a dictionary on his own. This proposal ultimately resulted in eleven volumes of his *Nederduitsch Taalkundig Woordenboek* ‘Dutch Linguistic Dictionary’ (1799–1811)¹².

In addition to his work on lexicography, Weiland also published on orthography and grammar. In 1801, when the Batavian government was developing concrete plans for a national language policy, Weiland was requested by the Minister of National Education to codify the Dutch grammar, which he accepted (Noordegraaf 2018: 148). His *Nederduitsche Spraakkunst* was published in 1805, i.e. one year after Siegenbeek’s orthography, and again *in naam en op last van het Staatsbestuur der Bataafsche Republiek* ‘in the name of and by order of the government of the Batavian Republic’.

In fact, it was the first – and also the last – authorised grammar to be prescribed by a Dutch government (Noordegraaf 2018: 145). Furthermore, it is interesting to note that Weiland’s *Spraakkunst* had been peer-reviewed by an advisory board of Leiden scholars, including Matthijs Siegenbeek and Adriaan Kluit, who approved his proposal.

Weiland’s national grammar was evidently modelled upon the works of the influential German grammarian Johann Christoph Adelung (1732–1806), especially the *Umfständliches Lehrgebäude der deutschen Sprache* of 1782. According to van Driel (1992: 226), “Weiland attempted to become the Dutch Adelung”, and he was even accused of plagiarism (Noordegraaf 2018: 155). Apart from Adelung, whom Weiland called *Duitschlands grooten Taalleraar* ‘Germany’s great language teacher’ (1805: XVI), the *Spraakkunst* was also heavily influenced by Lambert ten Kate’s *Aenleiding tot de kennisse van het verbevene deel der Nederduitsche sprake* (1723). Weiland divided his grammar into two parts, referred to as *Spelling* and *Woordvoeging*, respectively. In the comprehensive first part on phonetics and morphology, he discussed, for instance, the sounds of Dutch and its parts of speech. This part is actually an adaptation of the almost 200-page introduction in the first volume of his dictionary (1799). In the second part, Weiland focused on syntactic matters.

Noordegraaf (2018: 154) explains that “Weiland first and foremost sought to provide a practical grammar: an authoritative and solid resource for forming an opinion about the correctness of contemporary Dutch language use”. In this respect Weiland was fairly successful, as his authorised grammar was reprinted many times well into the second half of the nineteenth century and, moreover, adapted for use in schools (e.g. *Nederduitsche spraakkunst ten dienste der scholen* in 1806).

¹² For a comprehensive outline of Petrus Weiland’s linguistic activities and particularly his 1805 grammar, see Noordegraaf (2018).

4 Language-in-education policy

Parallel to the national language policy, which resulted in the official *schrijftaahregling* in 1804/1805, as discussed in Section 3, the period of Dutch nation building in the years around 1800 also saw fundamental changes in the educational system¹³. Language and education were closely intertwined during the Batavian-French period, when the national debates led to “proposals to make grammar and spelling obligatory subjects in primary school, and to concrete language-in-education laws aimed at the top-down dissemination of grammatical knowledge in the school system” (Rutten 2016: 124-125).

Generally, in Late Modern European nationalism, the implementation and dissemination of the standard language variety were first and foremost educational issues (Rutten, Krogull & Schoemaker accepted). In the case of the Northern Netherlands, the appointed Minister of National Education, J. H. van der Palm, was given the task to nationalise the educational system by bringing it under government control. He was handed an instruction, the *Instructie voor den Agent van Nationale Oproeding* (1798), with a detailed list of his tasks and responsibilities, which included, for instance, a proposal for the reform of the primary school system (article 3), the compilation of a list of prescribed school books (article 4) and the establishment of a school inspection system (article 8). The latter would enable the government to exert influence on practices in the classroom (cf. also Schoemaker & Rutten 2017). As pointed out in Section 3, the regulation of the language also fell under the responsibilities of the Minister (article 15), which led to the official codification of the Dutch orthography and grammar by Siegenbeek (1804) and Weiland (1805), respectively.

From 1801 onwards, a series of educational laws were passed. These acts were largely inspired by the 1798 report *Algemeene Denkebeelden* proposed by the Nut society (cf. Section 1.2). The first two education acts were issued in 1801 and 1803, but were barely enforced due to the many constitutional changes during the unsteady Batavian-French period. The third and final education act of 1806, however, was put into practice and remained in effect until 1857 (when a new school act was passed). In the 1806 act, knowledge of the Dutch language was explicitly mentioned as one of the fundamental elements in the national school curriculum, alongside reading, writing and arithmetic (Boekholt & de Booy 1987: 99). For the first time, orthographic and grammatical issues of Dutch became a central part of primary education.

As noted before, J. H. van der Palm initiated the regulation of the Dutch spelling and grammar, which were meant to be adopted in the administrative domain, comprising all printed documents by the government. Furthermore, the prescriptions laid down in Siegenbeek (1804) and Weiland (1805) became the valid

¹³ The language-in-education policies in the second half of the eighteenth and the first half of the nineteenth century are examined in depth in Bob Schoemaker’s PhD sub-project, which is also part of the Leiden-based research programme *Going Dutch. The Construction of Dutch in Policy, Practice and Discourse (1750–1850)* (Schoemaker 2018; cf. also Schoemaker & Rutten 2017)

norms for the educational domain and were thus highly recommended to schoolteachers throughout the country (cf. *Notulen* 1804):

Dat dezelve Spelling zal worden gevolgd en in acht genomen in alle Onderwijsboeken, welke van 's Lands wege, ten dienste der Scholen, zullen worden uitgegeven, met aanschrijving aan alle Schoolopzieners, om hunne beste pogingen aan te wenden, ten einde deze Spelling alom in de Scholen worde geadopteerd. (Siegenbeek 1804: xvi-xvii).

'That this spelling will be followed and regarded in all school books which will be published on behalf of the country for the use in schools, with an instruction for all school inspectors to make their best efforts in order to adopt this spelling everywhere in the schools.'

Dat de regels en gronden van taalkunde, bij deze Nederduitsche Spraakkunst vastgesteld, zullen worden gevolgd in alle onderwijsboeken, welke van 's Lands wege ten dienste der Scholen zullen worden uitgegeven, met aanschrijving aan alle Schoolopzieners, om hunne beste pogingen aan te wenden, ten einde dezelve regels en gronden alom in de Scholen worden geadopteerd. (Weiland 1805: xiv)

'That the rules and principles of language, as laid down in this Dutch grammar, will be followed in all school books which will be published on behalf of the country for the use in schools, with an instruction for all school inspectors to make their best efforts in order to adopt these rules and principles everywhere in the schools.'

In the interest of the nation's unity, the language-in-education policy of the early 1800s aimed at the widespread dissemination of knowledge of the national standard variety. Language and education formed a coherent framework in order to ensure that all future citizens of the nation would acquire the official language norms during their formative school years (Boekholt & de Booy 1987: 96-97). The question whether and to what extent the government succeeded in spreading the national language through the Dutch population at large will be at the heart of this dissertation. In Chapters 5-12, the effects of the early nineteenth-century language and language-in-education policy measures on actual language use are analysed systematically on the basis of authentic usage data in the *Going Dutch Corpus*. The theoretical framework for these analyses will be introduced in Chapter 3.

Theoretical framework

1 The field of historical sociolinguistics

The research presented in this dissertation is embedded in the theoretical framework of historical sociolinguistics. Over the past three decades, this field of study has developed into a mature and well-established linguistic (sub)discipline, which focuses on the reconstruction and study of language in its historical and social context. Historical sociolinguistics is described as “a hybrid subfield subsisting on the interdisciplinary character of sociolinguistic methodology” (Conde-Silvestre & Hernández-Campoy 2012: 1), drawing on findings and principles of (modern-day) sociolinguistics and historical linguistics, as well as social history. Nevalainen & Raumolin-Brunberg (2003: 10) point out that

[r]econstructing how language changes diffuse socially is one of the major tasks, if not *the* major task, of historical sociolinguists. This is an area where the ‘historical’ in ‘historical sociolinguistics’ is connected not only with historical linguistics but also with social history.

From an advanced cross-disciplinary perspective, historical sociolinguistics is also connected with even more associated fields of study, such as corpus linguistics, philology and dialectology (Nevalainen & Raumolin-Brunberg 2012: 27).

Suzanne Romaine’s (1982) *Socio-historical Linguistics: Its status and methodology*, a systematic study of relative markers in Middle Scots, is usually considered as the initiation of historical sociolinguistics as a specialised field of study (Auer et al. 2015: 2). Terminologically, her book was the first to include *sociohistorical linguistics* in its title. The same term was also used by Tiekens-Boon van Ostade (1987) in her study of auxiliary *do* in eighteenth-century English. The alternative term *historical sociolinguistics*, which has been established as the most common term internationally, first appeared in Milroy’s (1992) *Linguistic Variation and Change: on the Historical Sociolinguistics of English*. It also occurred in the title of a theoretical chapter on historical sociolinguistics by Nevalainen (1996). These pioneering works were largely inspired by the empirical research methods in modern-day sociolinguistics and addressed the need to apply such an approach to language history (Rutten et al. 2014b: 1).

It should be noted that there had been many publications on what would nowadays be considered as historical-sociolinguistic research before Romaine’s (1982) study, without using either of the terms *sociohistorical linguistics* or *historical sociolinguistics* (Auer et al. 2015: 2). In a handbook chapter on historical sociolinguistics, Romaine (2005: 1696) explicitly refers to the influential paper by Weinreich et al. (1968), which already emphasised the need to incorporate external

factors into language change theory in the late 1960s and thus “laid the foundations for an approach to language that was inherently historical and social” (Auer et al. 2015: 2). While the above-mentioned seminal works in historical sociolinguistics suggest a strong focus of the field on the language history of English, the 1980s and 1990s also saw the parallel development of a German tradition (e.g. Mattheier 1988; Mihm 1998).

The first decade of the twenty-first century witnessed the gradual development and expansion of historical sociolinguistics. Nevalainen & Raumolin-Brunberg’s (2003) book on language change in Tudor and Stuart England can be regarded as an introduction to the field of historical sociolinguistics, both explaining sociolinguistic concepts and demonstrating methods to apply these concepts to language in the past. Furthermore, two handbook chapters (Romaine 2005; Roberge 2006) as well as the paper by Willemyns & Vandenbussche (2006) manifested historical sociolinguistics as a thriving field of study. The same period also saw the establishment of the *Historical Sociolinguistic Network* (HiSoN), which has organised numerous international conferences, workshops and annual summer schools ever since.

A landmark for the field was undoubtedly the publication of *The Handbook of Historical Sociolinguistics* (Hernández-Campoy & Conde-Silvestre 2012), which testified to the multi-faceted research on different topics, language areas and time periods. The institutionalisation of historical sociolinguistics as an independent discipline within linguistics was further advanced by the launch of two academic book series, viz. *Advances in Historical Sociolinguistics* (John Benjamins), edited by Marijke van der Wal and Terttu Nevalainen, and *Historical Sociolinguistics. Studies on Language and Society in the Past* (Peter Lang), edited by Nils Langer, Stephan Elspaß, Joseph Salmons and Wim Vandenbussche. In 2015, publisher De Gruyter launched the *Journal of Historical Sociolinguistics*, a new academic journal specifically devoted to historical-sociolinguistic research, which is edited by Gijsbert Rutten, Anita Auer, José del Valle, Rik Vosters and Simon Pickl.

In the following, Section 2 briefly outlines the basic principles and challenges in historical sociolinguistics. The different approaches to language histories, both from a traditional perspective and from below, will be discussed in Section 3. Section 4 provides an overview of historical-sociolinguistic data collections and corpora. The interrelated topics of prescriptivism, language planning and policy (and their possible effects on actual language usage) will be addressed in Section 5.

2 Principles and challenges

As mentioned in Section 1, historical sociolinguistics largely draws on insights and concepts from present-day sociolinguistics, assuming that the fundamental principles and mechanisms of language variation and change do not change over time. This assumption is based on the so-called *uniformitarian principle*, which, according to Labov (1972: 275), implies that “the forces operating to produce

linguistic change today are of the same kind and order to magnitude as those which operated five or ten thousand years ago”. Although Bergs (2012: 96) agrees on the fact “that language must always have been variable, that different social groups and genders had different ways of speaking, and that people have always been aware of these differences”, he also warns of what he refers to as *ideational anachronism*, i.e. the danger of hastily transposing modern concepts like social class or gender to historical settings. Therefore, Auer et al. (2015: 5) argue that

it is the task of historical sociolinguists to reconstruct a broad picture of the social context in which the language varieties under investigation were used, drawing on the inductive method to identify the social conditions of language variation and change, ensuring empirical, social and historical validity.

While modern sociolinguistics is primarily based on spoken language data, language use in earlier periods – at least before the advent of audio recordings – can only be accessed by written sources. Addressing the limitations of historical (written) data, which are naturally fragmented records of the past and often preserved by chance, Labov (1994: 11) remarks that historical linguistics can be described as “the art of making the best use of bad data”. In that sense, the well-known *bad data problem* might be even more challenging for historical sociolinguists, as background information on the informants’ age, gender, regional origin and mobility, education, and so on, are often hard to find and reconstruct. Nevertheless, the dependence on supposedly ‘bad’ data¹⁴ should not be overstressed and can, at least to a certain extent, be compensated by “systematicity in data collection, extensive background reading and good philological work”, as Nevalainen & Raumolin-Brunberg (2003: 26) point out. They further admit that these tasks are “demanding and time-consuming, but by no means unrealizable” (ibid.).

In addition to their fragmented nature, written sources from the past share the fundamental bias that there are no direct records of the spoken language in historical writing. However, it is often claimed that speech is primary and thus essential for the understanding of language variation and change processes (Auer et al. 2015: 7; cf. also Schneider 2013: 57). This means that for a thorough study of variation and change that has to rely on written documents only, the relationship between speech and writing has to be critically (re)assessed. Historical sociolinguists in particular have to tackle the questions “how oral registers can be reconstructed from written sources, and thus [...] how written and oral language are interrelated” (Auer et al. 2015: 7).

¹⁴ Nevalainen & Raumolin-Brunberg (2003: 27) further remark that “[t]he comments on bad data easily lead to an impression that historical texts would be inferior to present-day material in every respect”. They argue that this is “not necessarily the case”, listing a number of advantages of these texts. In contrast to modern sociolinguistics, historical data have the major advantage of real-time analyses of language change. Moreover, these texts represent genuine communication from the past, unaffected by the participation of the researcher.

In search of suitable, i.e. ‘oral’-like written data for investigating language variation and change in the past, the field of historical sociolinguistics heavily relies on the influential frameworks by Biber (1988) and Koch & Oesterreicher (1985), who criticise the rigid distinction between speech and writing. The traditional dichotomy of spoken and written language, as Elspaß (2012: 157) remarks, is too “simplistic and even misleading”. Suggesting an alternative and more sophisticated model, Koch & Oesterreicher (1985) therefore differentiate between the *medium* and the *conception* of language. On the one hand, the linguistic medium is a strict dichotomy between the phonic and the graphic code. On the other hand, the conception of language distinguishes linguistic utterances on the basis of the communicative strategies they utilise. In contrast to the medial dichotomy, the conception has to be regarded as a continuum, on which all types of texts and genres can be located according to their specific situational and communicative parameters. The two poles on this conceptual continuum are referred to as the ‘language of immediacy’ (*Sprache der Nähe*) and the ‘language of distance’ (*Sprache der Distanz*). The pole of immediacy represents texts that are prototypically ‘oral’ (dialogic, informal, unplanned, and so on, like an intimate conversation), whereas texts close to the pole of distance are prototypically ‘literate’ (monologic, formal, planned, and so on, like a legal contract) (Elspaß 2012: 157). In fact, some (medially) spoken genres, such as sermons, mainly utilise literate strategies, whereas some (medially) written genres, such as private letters, are relatively close to the ‘oral’ pole of the continuum (Koch & Oesterreicher 1985: 23). Elspaß (2012: 157) highlights the importance of the linguistic conception as opposed to the linguistic medium:

The distinction between the conceptual poles of the continuum is primary to an understanding of the ‘orality’ or ‘writtenness’ of language. The medium is secondary, as a written text can be read out and a recording of a spoken text can be transcribed at any time.

For historical sociolinguists, who seek to find written data that reflect the spoken language, or rather orality, as much as possible, genres close to the ‘language of immediacy’ are thus highly valuable sources. Private letters are generally characterised as one of the most oral written genres (cf. Biber 1988), which make them “first-class primary data” for historical-sociolinguistic research (Nevalainen & Raumolin-Brunberg 2003: 29). More generally, so-called ego-documents, a cover term for texts from the private sphere “in which an author writes about his or her own acts, thoughts and feelings” (Dekker 2002: 14), for instance letters, diaries and travelogues, have proven to be useful sources as they are “usually close to speech and relatively unaffected by conventions of writing” (Auer et al. 2015: 7; cf. also Chapter 4). In historical sociolinguistics, these comparatively ‘oral’-like ego-documents in particular are at the heart of the so-called *language history from below*, an alternative approach to traditional language histories, which will be outlined in Section 3.

3 Language histories from different perspectives

3.1 Traditional language histories

For most European languages, including Dutch, traditional language historiography was first and foremost concerned with the unification and standardisation process of a given language. Until the late twentieth century, language histories were strongly dominated by the teleological view on language change as “the inexorable march towards a uniform standard” (Elspaß 2007: 3), presenting a more or less linear development from medieval variation to present-day uniformity, largely restricted to the evolution of an ideal and invariant standard variety.

Language histories are, in fact, greatly influenced by an underlying standard language ideology, defined by Lippi-Green (2012: 67) as “a bias toward an abstract, idealized homogeneous language, which is imposed and maintained by dominant bloc institutions and which names as its model the written language”. The origins of this ideology go back to the eighteenth century as a side effect of the period of nation building (cf. Chapter 2; cf. also Rutten 2016b). Referring to the standard language ideology, Milroy (2012: 579) emphasises that its influence “on traditional historical descriptions of major modern languages cannot be overstated”. The consequences of such a single-stranded approach are particularly noticeable in introductory textbooks on the subject, which easily give the impression that a language’s post-medieval history is equivalent to the history of its standard variety. Similarly, Watts (2012: 585) argues that traditional language historiography is grounded on “an implied teleology [...] that standard languages are the only valid objects of study for a language history”. He metaphorically calls this approach a tunnel vision view, “which projects from a source domain of the restriction, unidirectionality, and perhaps even darkness of passage through a tunnel onto the target domain of the abstract concept of language history” (Watts 2012: 585; cf. also Watts & Trudgill 2002). By narrowing historical descriptions of languages down to their standardisation process, that is, ignoring or rejecting the existence of non-standard(ised) forms and varieties (also referred to as erasure), incomplete language histories are written (Watts 2012: 579).

In historical sociolinguistics, the teleological view on language history has often been criticised for being “one-sided, partial, biased, largely based on a limited collection of text sources” (Rutten et al. 2014b: 2). Indeed, language historiography primarily described the evolution of printed languages (Elspaß 2012: 156). Partly motivated by the underlying standard language ideology, partly due to the wider availability and easier accessibility of (printed) data, language historiographers mainly focused on literary and formal texts from the higher registers (van der Wal 2006). However, as Schneider (2013: 59) remarks, these texts “normally display categorical, invariant usage and fail to reflect natural speech behaviour and associated processes”. Furthermore, literary and formal writings have a strong social and gender bias, as they were almost exclusively produced by well-educated men from the upper ranks of society. According to Elspaß (2007: 4-5), this restricted group of elite writers probably constituted no more than five per cent of

the population, at least in the case of nineteenth-century Germany. In other words, a substantial part of language use and language users of the past is not represented in these conventional language histories.

Examples of this traditional perspective on the history of Dutch, with a strong focus on standardisation and the increasing uniformity in literary and formal texts, can be found in van der Sijs (2004) and van der Wal & van Bree (2008), among many others. The specific case of negation in Dutch shows that these linguistic histories often portrayed a clear development in which single negation, as opposed to bipartite negation, had been selected as the norm by the mid-seventeenth century and then became standardised (e.g. van der Sijs 2004: 534–537; van der Wal & van Bree 2008: 217–218). This might indeed be the case in more formal registers and texts produced by literary authors. However, new historical-sociolinguistic insights gained from seventeenth-century private letters give evidence that “[i]n the spoken language, or perhaps more generally in informal registers, both spoken and written, bipartite negation remained in use” (Rutten 2016c: 204; cf. also Rutten & van der Wal 2014: ch. 10). This concrete counterexample illustrates that the study of a limited collection of texts can draw an incomplete picture of the linguistic past. As Elspaß (2007: 4) argues, the “neglect of texts ‘below’ the surface of printed language [...] has led to a language historiography in which a major part of both the language community (i.e. those writers with no access to printing) and their written language production is simply not represented”.

3.2 Language histories from below

In order to provide an alternative approach to language history, historical sociolinguists have suggested the *language history from below* (e.g. Elspaß et al. 2007). To begin with, it should be emphasised that the term *from below* does not correspond with the Labovian concepts indicating linguistic phenomena from above or below the level of awareness (Labov 1994: 78). Furthermore, *from below* as a specific perspective on language history does not refer to linguistic changes starting in the lower social classes (cf. also Rutten & van der Wal 2014: 5). In line with previous historical-sociolinguistic studies, and following Elspaß’s (2007: 5) definition, I will use the term *from below* in reference to “the social ranks below the highest social class and to texts representing everyday language that could thus be considered as below formal registers such as the language of literature”.

The central aim of the *language history from below* approach is to fill in the substantial gaps – or *witte vlekken* ‘blank areas’, as van der Wal (2006) puts it – in traditional language historiography, and thus to contribute to a more complete reconstruction of languages in the past. Auer et al. (2015: 6) point out that “[o]ne of the core concerns of historical sociolinguistics [...] is the effort to overcome the social bias connected to class, education and literacy inherent in written sources that has afflicted historiography”.

With respect to textual sources, the change of perspective involves a shift from printed, often literary and formal texts to handwritten, more ‘oral’-like documents. Although we do not have direct records of historical spoken language, as noted in Section 2 before, there are certain genres which reflect the ‘language of immediacy’ as much as possible and can, therefore, be considered as valuable sources for historical-sociolinguistic research. The approach *from below* usually takes advantage of these ego-documents, comprising various types of texts that have been “important in people’s private lives and personal experiences” (Rutten et al. 2014b: 1). It has repeatedly been argued in historical-sociolinguistic research (e.g. Nevalainen & Raumolin-Brunberg 2003, Elspaß 2005, Rutten & van der Wal 2014) that in order to gain access to authentic language use in historical contexts, we need to find those textual sources that are “as close to actual speech as possible, only in written form” (Sevic 1999: 34). In Koch & Oesterreicher’s (1985) terms, ego-documents like private letters, diaries and travelogues are considered to be conceptually more oral and thus closer to the ‘language of immediacy’ (Section 2) than the conceptually much more ‘literate’ writings, on which conventional language histories were grounded.

Moreover, ego-documents also have the advantage that they were produced by much wider parts of the language community than just the elite. Private letters and other types of ego-documents were written by both men and women from different social classes (including the lower and middle classes), from different regions, and so on. Forming the basis of the *from below* approach, ego-documents thus offer a remarkable opportunity of filling in the gaps left by language historians. Elspaß (2012: 161) argues that this approach is “a necessary counterweight” to the teleological historiography dominated by the standardisation perspective.

Nonetheless, it has also been remarked more recently by Rutten et al. (2014b: 2; cf. also Fairman 2007) that

purposely and explicitly leaving aside the more ‘standard’-like textual sources found in print, in literature, in elite documents, and setting aside the possible influence of supraregional writing conventions, language norms and prescriptions, may run the risk of presenting another one-sided view of language history.

Thus, in order to avoid a description of language history that becomes just as one-sided and biased as the traditional approach, historical sociolinguists have called for an integrated approach to language history, combining the two perspectives from above (i.e. in the traditional sense) *and* below (Rutten et al. 2014b: 2):

Based on the considerable research tradition in historical sociolinguistics that has come into existence over the past few decades, the time has now come to integrate both perspectives, and to reassess the importance of language norms, standardization and prescription on the basis of sound empirical studies of large corpora of texts.

In this dissertation, such an integrated perspective is taken in order to investigate the effectiveness of the national language policy and officially standardised language norms on actual usage in the Northern Netherlands. In other words, central issues typically known from traditional language histories, such as standardisation, language norms and prescriptions, will be reassessed on the basis of new insights gained from the perspective from below. In concrete terms, the possible ‘success’ of the Dutch language policy in the early 1800s, with its officialised regulations for spelling and grammar, will be examined on the basis of a newly compiled corpus of eighteenth- and nineteenth-century Dutch that makes use of handwritten ego-documents from the private sphere (in line with recent historical-sociolinguistic research), but also integrates more ‘standard’-like printed texts (cf. also Elspaß & Niehaus 2014). In this respect, I consider the different perspectives on language history discussed in this section as complementary rather than contradictory. The corpus and methodology applied in this dissertation will be presented in more detail in Chapter 4.

4 Data and corpora

Most studies in the field of historical sociolinguistics are based on empirical data, usually combining quantitative and qualitative findings. In fact, the collection of suitable data has played a central role in historical-sociolinguistic research ever since the advent of corpus linguistics in the 1980s. A number of corpora have been compiled for various languages and time periods.

In the early days, the focus was on diachronic multi-genre corpora, which allowed a systematic comparison of different registers and genres over time. Without any doubt, genre was considered as “the key external variable in sociolinguistically informed studies of language change” (Nevalainen & Raumolin-Brunberg 2012: 23). An example of such a multi-genre corpus is the *Helsinki Corpus of English Texts*, commonly regarded as the first historical corpus of English (cf. <<http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus>>). The corpus counts more than 1.5 million words in total, comprising text samples from the Old, Middle and Early Modern English periods, which are enriched by metadata on the selected texts and their authors.

Another diachronic multi-genre corpus of English is *ARCHER: A Representative Corpus of Historical English Registers*, initiated by Douglas Biber and Edgar Finegan in the 1990s. In its most recent version (*ARCHER* 3.2, completed in 2013, cf. <<http://www.projects.alc.manchester.ac.uk/archer>>), the corpus comprises 3.3 million words of British and American English texts from the seventeenth to the twentieth centuries, covering a range of both prototypically written texts, for instance medical and scientific prose, but also more speech-like genres like letters and diaries.

Following the model of historical corpora of English, especially *ARCHER*, Martin Durrell’s *GerManC* project (Durrell et al. 2012) aims to provide a representative sample of German in the period between 1650–1800. The *GerManC*

corpus comprises 800,000 words, covering texts from eight genres, including more orally oriented genres like personal letters, sermons and drama, as well as more print-oriented genres like scientific and legal texts. Adding a spatial dimension to the corpus, the selected texts derive from five different regions in the language area (cf. <<http://www.llc.manchester.ac.uk/research/projects/germanc>>).

Particularly from the 2000s onwards, historical sociolinguistics has witnessed a shift from textually balanced multi-genre corpora towards single-genre corpora (Nevalainen & Raumolin-Brunberg 2003: 27). With respect to the remarkable value of ego-documents for historical sociolinguistics¹⁵, as discussed in Section 3, it is “no coincidence that many of the corpora explored for this research consist of private letters” (Rutten & van der Wal 2014: 4; cf. also Elspaß 2005, 2012: 47–48). One of the pioneering historical corpora of letter data is the *Corpus of Early English Correspondence* (CEEC) (Nevalainen & Raumolin-Brunberg 2003), compiled at the University of Helsinki under the direction of Terttu Nevalainen. The central aim of the project was to test how the methods used in present-day sociolinguistics could be applied to historical data. The original 1998 version of the CEEC contains c. 2.6 million words, comprising around 6,000 letters by 778 informants, written in the period 1410–1681. More recently, the corpus was extended to a time span of four centuries (1400–1800), and to 11,819 letters by 1,066 writers (Palander-Collin et al. 2009: 14). For the field of historical sociolinguistics, the CEEC can certainly be considered ground-breaking as it was the first corpus which was tagged with fairly elaborate sociolinguistic metadata on the informants (cf. <<http://www.helsinki.fi/varieng/CoRD/corpora/CEEC>>).

For the history of German, Stephan Elspaß’s (2005) corpus of nineteenth-century emigrant letters, written by men and women from the lower and lower-middle ranks of society, testifies to the relevance of the approach from below. The corpus of emigrant letters (*Auswandererbriefe*) comprises around 700,000 words, 648 letters and 273 writers (Elspaß 2012: 47). Also integrating a geographical dimension, the corpus represents data from different regions of the investigated language area.

In order to fill in the major gaps in the language history of Dutch, the *Letters as Loot* corpus (Rutten & van der Wal 2014: 16; cf. also Nobels 2013, Simons 2013) was compiled at Leiden University under the direction of Marijke van der Wal. It contains private letters from the late seventeenth and late eighteenth centuries (i.e. two diachronic cross-sections: 1660s–1670s, 1770s–1780s), which were confiscated as so-called ‘Prize Papers’ during the Anglo-Dutch Wars and have been kept in the British National Archives. The *Letters as Loot* corpus counts 424,000 words and a selection of 933 private letters produced by 716 writers. These letters are a unique collection, as they were produced by both men and women from all ranks of society and different regional backgrounds. Furthermore, the electronic *Letter as Loot* corpus is lemmatised, tagged for parts of speech, and includes detailed metadata (cf. <<http://brievenalsbuit.inl.nl>>).

¹⁵ See Elspaß (2012: 162–163) for an overview of various corpora of ego-documents.

When we look beyond the Germanic languages, France Martineau's *Corpus de français familier ancien* is certainly worth mentioning, as it is "a unique sociohistorical French megacorporus of 20,000 non-literary texts, such as family letters, diaries and account books" (Martineau 2013: 132). The corpus is based on original documents found in archives in France and North America, dating from the seventeenth to the early twentieth century. For Portuguese, Rita Marquilha's *CARDS – Unknown Letters* project compiled a historical digital archive of Portuguese letters, spanning four centuries from the sixteenth to the nineteenth centuries. The succeeding project *P.S. Post Scriptum* focuses on private letters from different social backgrounds, written in Portugal and Spain during the Early Modern period and covering writers from different social background (cf. Marquilha 2012; <<http://ps.clul.ul.pt>>).

Currently, one can also notice a trend (back) towards multi-genre corpora, combining handwritten ego-documents with printed data such as newspapers. Elspaß & Niehaus (2014: 51-52), for instance, suggest a corpus design for German which considers emigrant letters as historical data *from below* on the one hand, and regional newspapers as historical data *from above* on the other. Furthermore, an intriguing corpus project on nineteenth-century Icelandic is currently being compiled at the University of Iceland, comprising private letters, school assignments and newspapers (cf. <<http://www.arnastofnun.is/page/LCLV19>>; cf. also van der Feest Viðarsson 2017).

Specifically for the research presented in this dissertation, a new multi-genre corpus of late eighteenth- and early nineteenth-century Dutch was compiled. The *Going Dutch Corpus* contains (handwritten) ego-documents, viz. private letters, diaries and travelogues, as well as (printed) newspapers. Two diachronic cross-sections, viz. 1770–1790 and 1820–1840, allow a systematic analysis of the effectiveness of the official language policy introduced in the Northern Netherlands in 1804/1805. The texts in the corpus cover seven regions, both from the centre and the periphery, and in the case of ego-documents, represent both male and female writers. The design and compilation of the *Going Dutch Corpus* will be presented at length in Chapter 4.

5 Prescriptivism, language planning and policy

Among the variety of issues that have been addressed in historical sociolinguistics are the interrelated topics of linguistic prescriptivism, language planning and policy, all of which are "examples of endeavors to influence the way that languages are used within a speech community" (Auer et al. 2015: 8). Often with a strong focus on prescriptivism, a great deal of attention in historical-(socio)linguistic research has been paid to the metalinguistic comments made by grammarians and norm authorities (e.g. Tieken-Boon van Ostade 2008; cf. also Auer & Gonzáles-Díaz 2005: 31). More recently, however, there has been growing interest in the interplay of prescriptivism and normativity on the one hand, and actual language practices on the other (e.g. Rutten et al. 2014a; Anderwald 2016; cf. also Ziegler 2007).

While the study of the relationship between language norms and language usage is not new as such, the advances in (corpus) linguistics in recent decades, notably the field of historical sociolinguistics, have made it possible to reassess the interplay of norms and usage through the systematic study of large corpora. The edited volume by Rutten et al. (2014a), in fact, discusses language norms and language use in the histories of Dutch, English, French and German from the seventeenth to the nineteenth centuries, in most cases on the basis of corpus studies. What has widely been neglected, though, is the potential influence of language planning and language policy in historical contexts. Up to today, it is very much unclear whether and to what extent these external factors have affected linguistic patterns of variation and change.

Language planning is commonly divided into three major parts: status planning, corpus planning, and acquisition planning¹⁶. Wright (2012) embeds these three types of language planning in the context of nation-building and nationalism. According to her definition, status planning is “the process whereby state elites identify a language variety as the national language to be used in all the formal functions of state business” (Wright 2012: 65). The conscious activity of corpus planning, which includes, for instance, the codification of linguistic norms in orthographies and grammars, typically evolves from the “growing desire to achieve and maintain linguistic cohesion”. Therefore, it also plays an essential role in the process of (linguistic) unification within the nation (Wright 2012: 68). Finally, acquisition planning aims at the promotion of the national language through the educational system. Wright (2012: 71) points out that school was “the institution where the ideology of one people, one territory and one language could be translated into reality”, leading to a generation of children who become literate in the national standard language.

Terminologically differentiated from language planning, language policy, according to Ager (2001: 5-6), is “official planning, carried out by those in political authority, and has clear similarities with any other form of public policy”. In that sense, language policy “represents the exercise of political power, and like any other policy, may be successful or not in achieving its aims” (*ibid.*). In the Dutch case, as examined in this dissertation, various language planning activities led to a concrete national language policy in the early nineteenth century on behalf of the Batavian government (cf. Chapter 2). The policy measures comprised the publication of an officialised orthography (Siegenbeek 1804) and grammar (Weiland 1805) for the

¹⁶ Ager (2001: 6) provides broad definitions of the three types: “Status planning modifies the status, and hence the prestige, of languages or language varieties within society, often by modifying the way the language codes groups or individuals use are perceived (‘deliberate efforts to influence the allocation of functions among a community’s languages’). Corpus planning in the traditional sense is what communities do to the forms of the language (‘graphization, standardisation, modernization and renovation’), but is sometimes also subdivided into codification of the existing language together with its elaboration and modernization by adding new terms or styles and controlling neologisms. Acquisition planning affects the ‘acquisition, reacquisition or maintenance of first, second or foreign languages’” (cf. also Cooper 1989).

educational and administrative domains. Investigating the effectiveness of this *schrijftaalregeling* ‘written language regulation’, I therefore use the term language policy (rather than planning) in order to refer to the top-down endeavour of the government in the early 1800s to exert influence on actual language usage.

The ‘success’ of concrete language planning and/or policy measures, but also of prescriptivism in a more general sense, has hardly been examined in historical-(socio)linguistic research. Empirical studies based on large-scale corpora are needed for a systematic assessment of how planning and policy measures affect usage patterns. There have been only a few efforts to test their effectiveness quantitatively. The study by Auer & González-Díaz (2005), for example, aims to reassess the effects of eighteenth-century prescriptivism on actual language use in English by closely comparing two corpora: a so-called precept corpus (i.e. a collection of metalinguistic comments) and a usage corpus based on data from the *Helsinki Corpus of English Texts* and *ARCHER* (cf. Section 4). This method of assessing the influence of prescriptive grammars on actual usage had been introduced in studies on German standardisation (e.g. Takada 1998; Langer 2001).

A more recent empirical study investigating the success of prescriptivism is presented in Anderwald (2016). Focusing on verbs and verb categories in nineteenth-century English, Anderwald’s study correlates the linguistic norms laid down in British and American English grammars with a corpus-based analysis of actual language change. For the French case, the article by Poplack et al. (2015) contributes to the discussion on prescriptivism and usage by studying the evolution of morphosyntactic features (such as the future temporal reference, cf. also Poplack & Dion 2009) in the grammatical tradition and a corpus of usage data.

Addressing the possible influence of prescriptive norms on language usage in the history of Dutch, Nobels & Rutten’s (2014) paper on seventeenth-century Dutch studies two features, viz. negation and the genitive case. For the eighteenth century, Simons & Rutten (2014) examine the representation of final *n* and, again, the genitive case. Both studies, which are based on data from the *Letters as Loot* corpus (cf. Section 4), show that there is only limited evidence that language users adhered to the prescriptive norms codified in normative publications. The seventeenth-century data, for instance, show that bipartite negation, against its stigmatisation in metalinguistic discourse, remained a common and fairly frequent variant across large parts of the population. In contrast, the eighteenth-century case study on the deletion of final *n* reveal a clear influence of norms (in favour of <en>) on usage. Nevertheless, it should be emphasised that both periods under investigation in Nobels & Rutten (2014) and Simons & Rutten (2014), i.e. the late seventeenth and late eighteenth centuries, respectively, were before the introduction of the first Dutch language policy in the early 1800s – and thus before the standardisation of Dutch became a national concern and was regulated on behalf of the government. In fact, hardly anything is known about the influence of these official language regulations and their possible normative influence on actual usage patterns.

The aim of my dissertation is to fill this research gap by investigating the effectiveness of the concrete language policy measures in the Northern

Netherlands. As a top-down policy, the introduction of the Dutch *schrijftaalregeling* and its implementation into the national school system have to be regarded as a decisive intervention in the sociolinguistic situation of the early nineteenth century (cf. Chapter 1). The extent to which these changes affected language practices will be examined based on a newly compiled multi-genre corpus of late eighteenth- and early nineteenth century Dutch (cf. Chapter 4 for a detailed outline of the *Going Dutch Corpus*). More generally, my dissertation also aims to contribute to the broader historical-sociolinguistic discussion on the impact of language planning and policy in the past (and how to measure it quantitatively). Not only in the history of Dutch, but in most European language histories, this has been an astonishingly understudied topic of research.

Corpus and methodology

1 Introduction

For the specific purpose of measuring and assessing the normative influence of the early nineteenth-century language policy on actual language usage, a new diachronic multi-genre corpus of more than 420,000 words was compiled. The *Going Dutch Corpus*, named after the research programme for which it was built, represents late eighteenth- and early nineteenth-century Dutch in the Northern Netherlands. Based on the assumption that linguistic changes affect different genres to different degrees, the corpus comprises data from three different types of authentic text sources, viz. (1) private letters (approx. 210,000 words), (2) diaries and travelogues (approx. 140,000 words), and (3) newspapers (approx. 70,000 words). In line with historical-sociolinguistic research and the *language history from below* approach (cf. Chapter 3), this corpus design takes into account handwritten and conceptually more ‘oral’-like ego-documents, i.e. private letters, diaries and travelogues. On the other hand, the corpus also incorporates printed and published texts, in this case newspapers, which are commonly regarded as fairly standardised writing. The three genres of the *Going Dutch Corpus* and their use in historical-sociolinguistic research will be introduced in more detail in Section 3.1.

Considering the historical event of the official *schrijftaalregeling* ‘written language regulation’ in 1804 and 1805 as the main point of departure, two diachronic cross-sections of twenty years each were chosen. The nineteenth-century period of 1820–1840 represents the generation of language users *after* the introduction of Siegenbeek’s (1804) official orthography and Weiland’s (1805) official grammar, i.e. those writers who had (probably) received the national education with its corresponding language norms. Symmetrically, the eighteenth-century period of 1770–1790 represents the generation of language users *before* the officialised language norms were introduced. The diachronic dimension of the corpus will be discussed in Section 3.2.

With regard to the considerable degree of regional variation in the investigated language area, the *Going Dutch Corpus* covers seven regions of the Northern Netherlands, which are based on present-day provincial boundaries: Friesland, Groningen, North Brabant, North Holland, South Holland, Utrecht and Zeeland. This selection of regions comprises both the urbanised centre (i.e. North and South Holland, Utrecht) and more peripheral parts of the language area (i.e. Friesland, Groningen, North Brabant). The spatial dimension of the corpus, addressing variation between individual regions as well as between the centre and the periphery, will be discussed in Section 3.3.

Integrating a social dimension into the corpus design, the texts in the two sub-corpora of ego-documents were written by both men and women, mainly from

the (upper) middle to the upper classes. The variables of social class and gender will be introduced in Section 3.4.

In more general terms, the present chapter aims to give detailed insights into the compilation process of the *Going Dutch Corpus* and the methodology applied in this dissertation. Section 2 first outlines the collection and selection of corpus data (2.1), then describes the transcription procedure and conventions (2.2), and finally presents an overview of the size and structure of the final corpus (2.3). Section 3 introduces four variational dimensions and their independent variables investigated in all corpus-based case studies, viz. the genre dimension (3.1), the diachronic dimension (3.2), the spatial dimension (3.3), and the social dimension (3.4). The additional dimension of inter- and intra-individual variation and change will be discussed in Section 4, briefly presenting the specifically compiled *Martini Buys Correspondence Corpus*. Taking into account the developments in metalinguistic discourse, Section 5 introduces the *Normative Corpus of the Northern Netherlands*, a collection of eighteenth-century normative publications. Finally, Section 6 outlines the systematic methodological approach to the linguistic analyses in Chapters 5–12, followed by some final remarks on statistical methods.

2 Compiling the *Going Dutch Corpus*

Compiling a multi-genre corpus requires awareness of the specific characteristics and methodological challenges of its textual sources. In order to collect a representative and well-balanced sample of a certain genre on the one hand, and to meet the need of comparability with other genres on the other hand, customised approaches and selection criteria were developed for the *Going Dutch Corpus*. Section 2.1 discusses the collection and selection of corpus data. The transcription procedure and conventions will be presented in Section 2.2.

2.1 Collection and selection of data

The most crucial difference between the three genres included in the *Going Dutch Corpus* concerns the medium of texts, i.e. handwriting and print (cf. Rutkowska & Rössler 2012: 219). In fact, two of the three sub-corpora (i.e. private letters, diaries and travelogues) represent largely unpublished and handwritten ego-documents, whereas the third sub-corpus (i.e. newspapers) contains published and printed texts. Therefore, this section addresses the collection and selection of ego-documents and newspapers separately.

Ego-documents

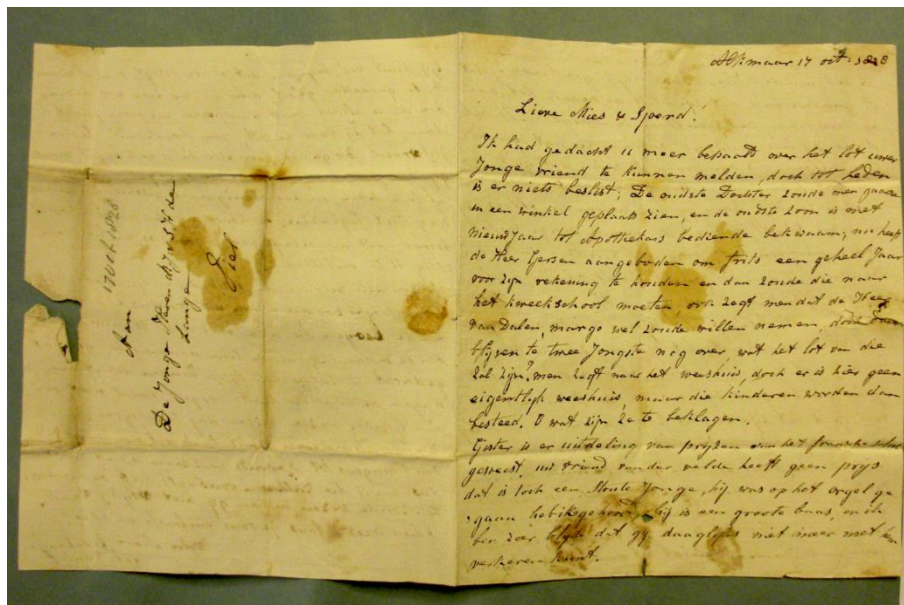
The compilation of the two sub-corpora of handwritten ego-documents, i.e. private letters as well as diaries and travelogues, started with a preparatory phase of thorough research and planning, in order to detect what kind of material was

actually available for the two investigated periods of 1770–1790 and 1820–1840, and where to find these sources. The ego-documents selected for the *Going Dutch Corpus* were collected from numerous municipal, regional and provincial archives spread all over the Netherlands, as listed below (categorised by province):

- Friesland: *Historisch Centrum Leeuwarden, Tresoor* (both Leeuwarden),
- Groningen: *Groninger Archieven* (Groningen),
- North Brabant: *Stadsarchief Breda* (Breda), *Brabants Historisch Informatie Centrum* ('s-Hertogenbosch), *Regionaal Archief Tilburg* (Tilburg),
- North Holland: *Regionaal Archief Alkmaar* (Alkmaar), *Stadsarchief Amsterdam* (Amsterdam), *Noord-Hollands Archief* (Haarlem), *Westfries Archief* (Hoorn),
- South Holland: *Archief Delft* (Delft), *Nationaal Archief* (The Hague), *Regionaal Archief Dordrecht* (Dordrecht), *Erfgoed Leiden en Omstreken* (Leiden), *Stadsarchief Rotterdam* (Rotterdam),
- Utrecht: *Het Utrechts Archief* (Utrecht),
- Zeeland: *Zeeuws Archief* (Middelburg).

As a rule, I visited the above-mentioned archives in order to request original manuscript sources and to take digital photographs (cf. Figure 1 for an example).

Figure 1. Private letter by Anna de Lange (17 October 1828, de Lange family archive, *Regionaal Archief Alkmaar*).



Only in a few exceptional cases, scans of the original documents were provided by staff members of the archives.

The digital images were inventoried according to a standardised format, for instance *Alkmaar_DeLange_79011_520_let12*. These file names include information about the place of the archives (i.e. *Regionaal Archief Alkmaar* in the town of Alkmaar), the name of the family archive (i.e. de Lange family). Furthermore, the file name contains the exact accession (79011) and inventory numbers (520), which makes it easy to trace back the origin of the documents. The abbreviation *let12* refers to the genre (i.e. private letter) and the individual code assigned to the document within a given inventory number (usually containing more than one document).

For the eighteenth-century data of private letters, the *Going Dutch Corpus* takes advantage of the extended *Letters as Loot* corpus, which had previously been compiled as part of the research programme *Letters as Loot. Towards a non-standard view on the history of Dutch* at Leiden University (2008–2013), directed by Marijke van der Wal (<<http://brievenalsbuit.inl.nl>>). The project investigated variation and change in the so-called sailing letters, confiscated during the wars fought between the Netherlands and England, and kept in the National Archives in Kew (London). This unique and linguistically highly valuable collection of Dutch private letters from the seventeenth and eighteenth centuries comprises texts from all ranks of the society, written by both men and women (Rutten & van der Wal 2014).

In order to achieve comparability with the letters specifically collected for the *Going Dutch Corpus*, a set of criteria had to be introduced. The selected texts from the extended *Letters as Loot* corpus comprise 104 private autograph letters (59,496 words) from the eighteenth-century period of 1776–1784. These letters were written by men and women from the upper middle class (UMC) and upper class (UC), which correspond with the social ranks predominantly represented in the *Going Dutch Corpus* (cf. Section 3.4.1). Geographically, the so-called ‘regions of residence’, i.e. the regions where letter writers were born and raised, largely match the regional categories of the *Going Dutch Corpus* (cf. Section 3.3.1), except for the region of North Holland. In this case, the *Letters as Loot* corpus distinguishes between ‘North Holland (Amsterdam)’ and ‘North Holland (rest of the province)’ as two separate categories (Rutten & van der Wal 2014: 11–12):

Amsterdam is considered separately for geographical as well as demographic reasons. Geographically, the city of Amsterdam is located in the south of North Holland, separated from the northern parts of North Holland by water. Demographically, Amsterdam was a highly urbanized metropolis, attracting many immigrants from the rural areas of Holland and from other provinces of the Netherlands, as well as from abroad, mainly from German-speaking regions.

In contrast, the category of ‘North Holland’ in the *Going Dutch Corpus* comprises both Amsterdam and the rest of the province. Previous research, including a historical-sociolinguistic study of negation in seventeenth- and eighteenth-century Dutch, has shown that the metropolis of Amsterdam is “not exceptionally progressive compared to the other regions”, and that “it perfectly fits

into the overall north-to-south pattern: it is less progressive than North Holland, and more progressive than South Holland” (Rutten & van der Wal 2013: 118). It did not seem fully justified to split North Holland into two separate categories again, which is why it is treated as one single regional category in the *Going Dutch Corpus*. The selected *Letters as Loot* texts, generally labelled as ‘North Holland’ here, thus actually comprise data from both Amsterdam and the rest of the province.

Table 1 provides an overview of the selected eighteenth-century private letters taken from the extended *Letters as Loot* corpus, distributed across the seven regions of the *Going Dutch Corpus* (i.e. FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland) and across genders.

Table 1. Selection of eighteenth-century private letters taken from the extended *Letters as Loot* corpus.

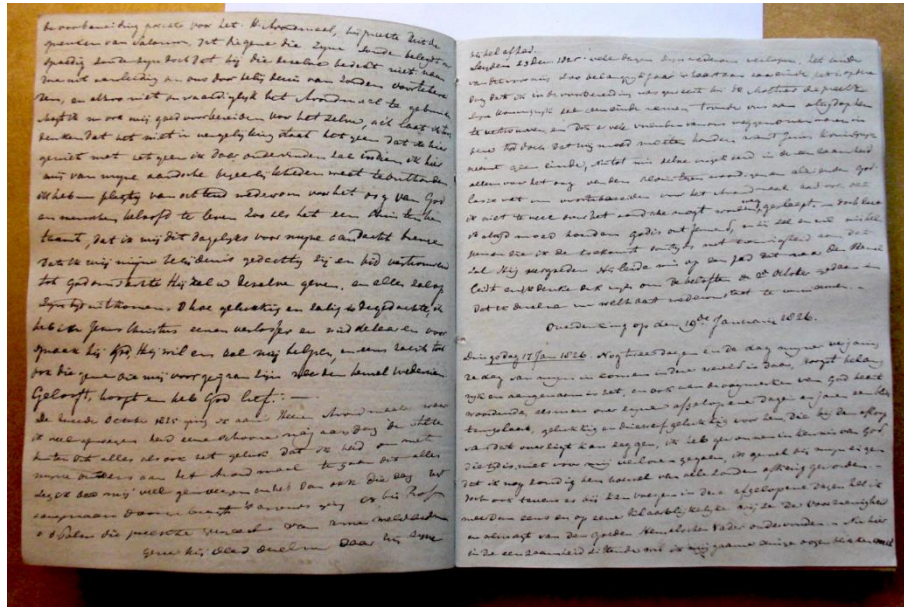
	FR	GR	NB	NH	SH	UT	ZE	Total
Male	5	4	2	18	13	8	6	56
Female	8	–	3	16	11	2	8	48
Total	13	4	5	34	24	10	14	104

It shows that the regions are not equally represented in quantitative terms. Most texts, in fact, stem from the western coastal regions of North Holland, South Holland and, to a lesser extent, Zeeland. Although the texts from *Letters as Loot* form the basis of the eighteenth-century data (104 out of 200 private letters), additional material had to be collected from various Dutch archives in order to fill the gaps in the under-represented regions, especially Groningen, North Brabant and Utrecht, but also Friesland and Zeeland.

Similar to the collection of private letters, the sub-corpus of diaries and travelogues also started out with exploratory research on the availability of suitable material. The fundamental works by Lindeman et al. (1993; 1994), providing a comprehensive inventory of ego-documents in the Northern Netherlands, as well as the corresponding and highly valuable website of the *Center for the Study of Egodocuments and History* <<http://www.egodocument.net>> by Arianne Baggerman and Rudolf Dekker, served as a starting point for the compilation of this sub-corpus. However, whereas private letters were relatively numerous and easy to find in Dutch archives, it was considerably more challenging to collect an appropriate amount of diaries and travelogues. Not surprisingly, there were far fewer eighteenth- and nineteenth-century diarists than letter writers, which affected the availability of suitable texts for the periods under investigation.

The actual procedure to collect and inventory diaries and travelogues was similar to the collection of private letters. Again, I visited various archives in order to take digital photographs of original manuscript documents (cf. Figure 2 for an example).

Figure 2. Diary by Pieter Glaudius Hubrecht (1825/1826, Hubrecht family archive, *Erfgoed Leiden en Omstreken*).



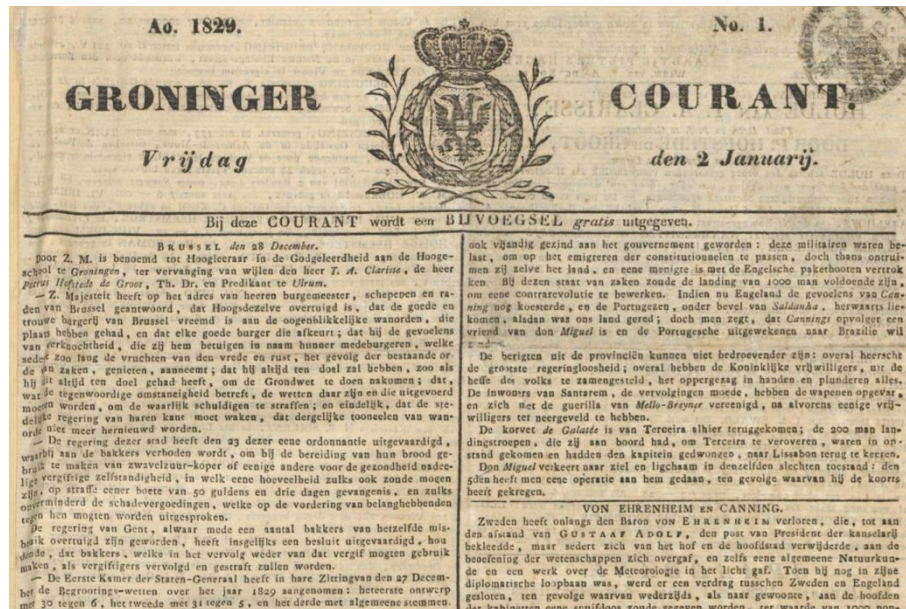
The selected archival sources were inventoried according to the same standardised format, e.g. *Leiden_Hubrecht_529_457_dia01*, with the abbreviation *dia* referring to the genre of diaries and travelogues.

Newspapers

Unlike the sub-corpora of handwritten ego-documents, which are based on original archival sources, the sub-corpus of newspapers was compiled on the basis of digital scans taken from the *Delpher* website (<<http://www.delpher.nl>>) (cf. Figure 3 for an example).

Delpher is an online service, which gives free access to a vast amount of Dutch newspapers, magazines and books from the seventeenth to the twentieth century. For the *Going Dutch Corpus*, digitised newspapers were selected from the eighteenth-century period of 1770–1790, and from the nineteenth-century period of 1820–1840.

Taking into account regional variation, a representative newspaper was selected for each of the seven regions, preferably published in both periods and accessible on the *Delpher* website. Only in the case of North Brabant, data had to be taken from two different newspapers, both of which were published in the city of ‘s-Hertogenbosch, viz. the *‘s Hertogenbossche courant* (1770–1790) and the *Noord Brabander* (1820–1840).

Figure 3. Title page of the *Groninger courant* (2 January 1829, *Delpher.nl*).

The final sub-corpus contains text samples from the following newspapers (categorised by province):

- Friesland: *Leeuwarder courant*,
- Groningen: *Groninger courant*,
- North Brabant: *'s Hertogenbossche courant*, *Noord Brabander*,
- North Holland: *Oprechte Haarlemsche courant*,
- South Holland: *Leydse courant*,
- Utrecht: *Utrechtsche courant*,
- Zeeland: *Middelburgsche courant*.

2.2 Transcription procedure and conventions

In order to make the selected material machine-readable and analysable for corpus-linguistic software tools such as *WordSmith*, all texts were manually transcribed and saved as electronic text files. As the genres included in the *Going Dutch Corpus* represent two fundamentally different types of data, i.e. handwritten and printed texts, the transcription procedures and conventions will be discussed separately for ego-documents and newspapers.

Ego-documents

Based on digital images of archival sources, the selected ego-documents were diplomatically transcribed, which means that the original spelling, punctuation and word boundaries were retained and not normalised according to contemporary standards. With regard to the handwritten nature of private letters, diaries and travelogues, detailed guidelines were essential in order to guarantee a consistent transcription process (cf. Appendix I for the full transcription conventions).

Aspects that were taken into consideration include ambiguities (<ambig>word</ambig>¹⁷), illegibilities (<illeg/>), deletions (word), insertions (<ins>word</ins>), underlining (<u>word</u>), hyphenation (<reg orig="wo|rd">word</reg>), capitalisation and intra-word spacing, as well as line and page breaks. Illustrating the use of these tags, the example below provides the extract of a transcribed private letter taken from the *Going Dutch Corpus*:

Waarde Dogter! Breda den 16 augustus <u>1825</u>
 UEd brief met couvert er om, heb ik wel ontrangen
 daar ik hem vrijdag s'avonds heb ontrangen konde
 ik s'zaterdag niets meer van UEd goed laten <reg orig="was|sen">wassen</reg>
 omdat ik <ins>er</ins> twee leegen<ins>dagen</ins> op volgen
 en ik zend
 UEd nu maar zo veel als ik in den Groote <reg orig="Lesse|naar">Lessenaar</reg>
 kan bergen en waar van het zjysje dat
 ik UEd zend in UEd roode jas gespeld is, de jas
 is wel vuil maar om dat er nog plaats in was
 heb ik er hem in gedaan UEd kund hem dan maar
 met u ander goed laten wassen [...]

'Dear daughter! Breda the 16th of August 1825
 I have well received your letter with [the] envelope around it.
 As I received it on Friday evening, I could not
 have your laundry washed on Saturday
 because two vacant days follow and now I send
 you as much as I can store in the big lectern
 and of which the siskin that
 I send you is tacked on your red jacket. The jacket
 is dirty but because there was still space in it
 I put it in there. You could have it
 washed with your other laundry then [...]

The representation of *ij/y*, i.e. one of the central orthographic variables under investigation, needed special attention before and during the transcription process. In order to be able to investigate the use of this variable in detail, four main variants were distinguished (cf. also Chapter 9):

¹⁷ The tag <ambig> was used to indicate unclear or ambiguous spellings and words. The suggested transcriptions are too uncertain to be taken into consideration for the corpus-based analysis of orthographic features in particular, but may still be useful for the study of morphosyntactic features and further context-related matters.

- (1) <ij>, i.e. double-dotted <ij> (*lange ij*) with <i> and <j> written as two separate characters,
- (2) <ÿ>, i.e. double-dotted <y>,
- (3) <y>, i.e. (undotted) <y> (*Griekse y*),
- (4) <°y>, i.e. other variants, e.g. single-dotted <y>, <y> with accent marks or other diacritics (positions of dots and accents are irrelevant here).

The introduction of clearly defined conventions, ensuring that the transcriptions are as consistent as possible, is crucial especially with regard to the number of people involved in the transcription process. For the most part, the transcriptions of texts were carried out by the project's research assistants Christa Bouwmans and Hielke Vriesendorp, as well as by myself¹⁸. After the first transcription phase, each document was thoroughly double-checked either by one of the assistants or myself, comparing the first transcription to the corresponding digital images in order to detect and fix possible transcription errors. Even though a few remaining inconsistencies cannot be excluded, the final transcriptions can be considered as accurate and reliable.

Within the sub-corpus of private letters, the transcription of eighteenth-century data marked a special case. As mentioned in Section 2.1.1, a substantial number of private letters was taken from the *Letters as Loot* corpus. The applied conventions, however, were slightly different from the transcription guidelines followed here (cf. Nobels 2013 and Simons 2013 for a detailed description). Consequently, even minor deviations would have resulted in an inconsistent use of tags and, in some cases, to different transcriptions of specific variants¹⁹. For the sake of consistency and comparability, all transcriptions from the *Letters as Loot* corpus were modified according to the conventions of this dissertation. New text files were created for each of these external transcriptions, applying the same tags as for all other private letters in the *Going Dutch Corpus*.

Enriching the transcriptions with a basic set of metadata, headers were added to each text file, both for the newly collected letters and for those taken from the *Letters as Loot* corpus. These headers contain information on the provenance of the original archival documents, the genre (either 'letter' or 'diary'), as well as the date and place of writing. Furthermore, all headers provide information on the transcriber, the number of words, and (optionally) notes about the transcriptions. A header example is given below:

¹⁸ I also want to thank our MA students Brenda Assendelft, Anne Rose Haverkamp and Marlies Reitsma for contributing some initial transcriptions as part of their Master's theses.

¹⁹ In the *Letters as Loot* corpus, the highly variable representations of the *ij/y* variable were transcribed in a less complex manner, only taking into account the two variants <ij> and <y>. All occurrences were manually revised in order to consider the four variants distinguished in the *Going Dutch Corpus* (i.e. <ij>, <ÿ>, <y>, <°y>).

```
<header>
DOCUMENT: Alkmaar_DeLange_79011_520_let12
ARCHIVE: Regionaal Archief Alkmaar
GENRE: letter
DATE: 1828-10-17
PLACE: Alkmaar
TRANSCRIPTION: HV
NOTES:
WORD COUNT: 592
</header>
```

For the sake of a convenient corpus analysis, each transcription selected for the *Going Dutch Corpus* was given a file name, for example *LET-2-NH-F-Alkmaar_DeLange_79011_520_let12*. In addition to the name of the source document (*Alkmaar_DeLange_79011_520_let12*), these file names contain the standardised codes for genre²⁰ (LET), period²¹ (2), region²² (NH) and gender²³ (F).

One final remark from a more practical point of view concerns the condition of archival documents and the legibility of handwriting. Both factors, at least in some cases, could influence the selection of data. In fact, (parts of) ego-documents in very poor condition and/or with hardly legible handwritings had to be neglected. As Van Bergen & Denison (2007: 4) rightly note, “deciphering the letters could be at least as time-consuming as the actual transcription”.

Newspapers

While the transcription of handwritten ego-documents was indeed a time-consuming and, depending on the legibility of handwriting, challenging procedure, the transcription of newspapers turned out to a comparatively straightforward task. First of all, the texts selected for this sub-corpus were manually transcribed²⁴ based on digitised newspapers on the *Delpher* website. All transcriptions were provided by research assistant Hielke Vriesendorp and double-checked by myself. Again, all texts were transcribed diplomatically, intended to be as close to the original as possible. The legibility of the newspaper texts was generally unproblematic, although ambiguous and even illegible readings did occur, mostly due to scan quality or folds in the paper. Those instances were explicitly marked in the transcriptions with the corresponding tags for ambiguities (`<ambig>word</ambig>`) or illegible words (`<illeg/>`).

²⁰ The codes for genres are: LET (= private letters), DIA (= diaries and travelogues), NEW (= newspapers).

²¹ The codes for periods are: 1 (= period 1, 1770–1790), 2 (= period 2, 1820–1840).

²² The codes for regions are: FR (= Friesland), GR (= Groningen), NB (= North Brabant), NH (= North Holland), SH (= South Holland), UT (= Utrecht), ZE (= Zeeland).

²³ The codes for genders are: F (= female), M (= male).

²⁴ Initial attempts to use OCR software (optical character recognition) resulted in transcriptions which were too unreliable and would have required a fair amount of manual correction.

In terms of typographic variation, capitalisation and other types of emphasis, i.e. usually words or passages in italics (<emph>*word*</emph>), were also transcribed in their original form. Furthermore, line and page breaks were taken into account as two fundamental aspects of layout²⁵. This was mainly done for practical reasons in order to find back specific passages in the original scans more conveniently.

Similar to the transcriptions of ego-documents, all files in the sub-corpus of newspapers contain a header with a basic set of metadata, including the name and source of the document, the genre (i.e. ‘newspaper’), the date(s) and place of publication, and the word count (i.e. a standardised amount of 5,000 words). An example is given below:

```
<header>
DOCUMENT: Groninger Courant
ARCHIVE: Delpher
GENRE: newspaper
DATE: 1828-01-01, 1829-01-02, 1830-01-01, 1831-01-04
PLACE: Groningen
TRANSCRIPTION: HV
NOTES:
WORD COUNT: 5064
</header>
```

Each text file in the final *Going Dutch Corpus* was assigned a code, for instance *NEW-2-GR.txt*, comprising information on the relevant independent variables, i.e. genre (NEW), period (2) and region (GR).

2.3 Size and structure of the final corpus

The aim was to compile a diachronic multi-genre corpus, which comprises 420,000 words, and can be divided into three sub-corpora, representing three different genres, viz. (1) private letters (approx. 210,000 words), (2) diaries and travelogues (approx. 140,000 words) and (3) newspapers (approx. 70,000 words). Table 2 provides an overview of the intended corpus size and structure, serving as a starting point for the actual collection and selection of data for the *Going Dutch Corpus*.

Table 2 also illustrates that the tripartite division of the *Going Dutch Corpus* into private letters, diaries and travelogues, and newspapers involves different sizes of sub-corpora. The underlying consideration behind the definition of (sub-)corpus sizes was the degree of (expected) uniformity and linguistic ‘standardness’, or, to put it the other way round, the degree of linguistic variation that is expected to be found in these texts.

²⁵ Catch words (so-called *custoden* in Dutch), i.e. words which are inserted at the end of a page and repeated on the following page, were only transcribed once.

Table 2. General corpus design and structure of the *Going Dutch Corpus*.

Private letters (approx. 210,000 words)				Diaries and travelogues (approx. 140,000 words)				Newspapers (approx. 70,000 words)	
1770–1790		1820–1840		1770–1790		1820–1840		1770–1790	1820–1840
Seven regions		Seven regions		Seven regions		Seven regions		Seven regions	Seven regions
♂	♀	♂	♀	♂	♀	♂	♀		

Firstly, the sub-corpus of private letters comprises approximately 210,000 words in total, with 105,000 words per period, 15,000 words per region per period, and so on. It has repeatedly been demonstrated in historical-sociolinguistic research that private letters are a particularly useful genre of ego-documents to investigate usage patterns, given the expectedly high degree of linguistic variation. Moreover, the wide availability of such texts in Dutch archives was a decisive factor in defining the intended size of the sub-corpus of private letters.

Second, the sub-corpus of diaries and travelogues comprises approximately 140,000 words. Although these texts, like private letters, belong to the group of ego-documents, they tend to be written in more ‘standard’-like language (Schneider 2013: 66). Therefore, in comparison with private letters, less linguistic variation has to be expected here, which is why this sub-corpus contains a lower number of words than the linguistically more heterogeneous letter sub-corpus. The limited availability of suitable diaries and travelogues in Dutch archives was another, more practical reason to reduce the total number of words in this sub-corpus.

Third, as the only printed sources in the *Going Dutch Corpus*, newspapers are expected to display the highest degree of linguistic uniformity. Rutten & van der Wal (2014: 3) point out that the printed language from the eighteenth-century (onwards) can be characterised as considerably uniform. Therefore, even a comparatively limited amount of data, i.e. in this case approximately 70,000 words, was considered to be sufficient for a representative sample of contemporary newspaper writing, in particular with regard to the focus on pervasive orthographic and morphosyntactic features.

Table 3 gives an overview of the actual sizes of the three sub-corpora in the final version of the *Going Dutch Corpus*, distributed across the two diachronic cross-sections and across the entire corpus. As shown, the initially intended corpus size of approximately 420,000 words was reached. In fact, the intended sizes were practically reached for all three sub-corpora and for both periods.

Table 3. Sizes of the sub-corpora in the *Going Dutch Corpus* (in absolute numbers and percentage of the total corpus).

Period	Private letters	Diaries and travelogues	Newspapers	Total
	N words (%)	N words (%)	N words (%)	N words (%)
1770–1790	105,427 (25.0)	71,157 (16.9)	35,323 (8.4)	211,907 (50.2)
1820–1840	105,299 (25.0)	69,350 (16.4)	35,322 (8.4)	209,971 (49.8)
Total	210,726 (50.0)	140,507 (33.3)	70,645 (16.8)	421,878 (100)

3 Variational dimensions of the *Going Dutch Corpus*

This section introduces the variational dimensions integrated in the *Going Dutch Corpus*. In Chapters 5–12, four dimensions will be considered in the corpus analyses of orthographic and morphosyntactic variables, viz. (1) the genre dimension (Section 3.1) with its sub-corpora of private letters (3.1.1), diaries and travelogues (3.1.2) and newspapers (3.1.3), (2) the diachronic dimension (Section 3.2) with its two twenty-year periods before and after the *schrijftaalregeling*, (3) the spatial dimension (Section 3.3), which considers variation across regions (3.3.1) as well as centre and periphery (3.3.2), and (4) the social dimension (Section 3.4) with its variables of social class (3.4.1) and gender (3.4.2). The individual dimension of inter- and intra-writer variation will be addressed separately in Section 4.

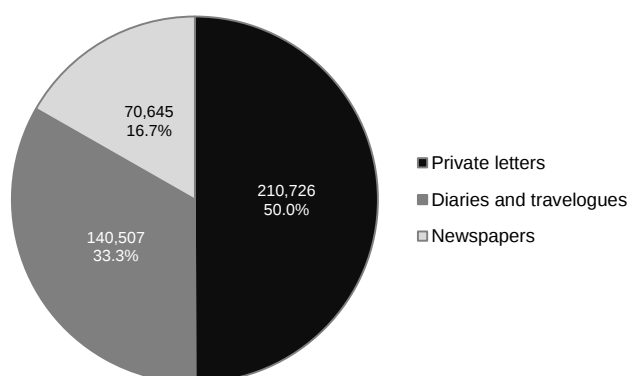
3.1 Genre dimension

Based on the assumption that linguistic changes affect different genres to different extents, the *Going Dutch Corpus* was designed as a diachronic multi-genre corpus, representing the following three genres: (1) private letters, (2) diaries and travelogues, and (3) newspapers. Figure 4 shows the relative distribution of these three genres in the final corpus.

As can be seen, the sub-corpus of private letters (50.0%) makes up one half of the entire *Going Dutch Corpus*, whereas the second half of the corpus comprises the sub-corpora of diaries and travelogues as well as newspapers with one-third (33.3%) and one-sixth (16.7%) of the data, respectively.

In terms of corpus design and, more specifically, the selection of genres, this dissertation aims to take an integrated multi-genre perspective on Dutch language history. On the one hand, it follows the historical-sociolinguistic approach *from below* by utilising handwritten ego-documents (i.e. private letters, diaries and travelogues). On the other hand, the corpus also incorporates more standard-like printed and published texts (i.e. newspapers).

Figure 4. Genre distribution in the *Going Dutch Corpus* (absolute number of words and percentage).



As outlined in Chapter 3 (cf. Section 3 in particular), traditional language histories often had a strong focus on the standardisation process and were first and foremost based on printed language, mainly literary and formal texts from the higher registers. These sources represent a fairly standardised form of writing, which often fails to fully reflect the variation found in authentic language usage. Suggesting an alternative approach to language history, historical sociolinguists therefore introduced the *language history from below* (e.g. Elspaß et al. 2007). This change of perspective involves a shift from relatively uniform printed texts to more informal handwritten sources from the private sphere, such as letters, diaries and travelogues. These ego-documents are conceptually more ‘oral’ and closer to the ‘language of immediacy’ than the sources traditionally studied in language historiography (cf. also Section 2 of Chapter 3 for Koch & Oesterreicher’s (1985) conceptual continuum).

While many historical-sociolinguistic studies have criticised the teleological view on (primarily printed) language histories for being “one-sided, partial, biased” accounts of the linguistic past (Rutten et al. 2014b: 1-2), it has also been argued that the alternative approach *from below* “may run the risk of presenting another one-sided view of language history” (ibid.). Therefore, the selection of genres in the *Going Dutch Corpus* offers an integrated perspective, considering the study of handwritten ego-documents from the ‘language of immediacy’ and printed texts from the ‘language of distance’ as complementary rather than contradictory (cf. also Elspaß & Niehaus 2014). In fact, this multi-genre corpus design allows for a direct comparison of two conceptually more ‘oral’ genres of ego-documents (i.e. private letters, diaries and travelogues) in relation to a conceptually more ‘literate’ and standardised printed genre (i.e. newspapers). At the same time, this design enables to compare manuscript to print sources, investigating possible differences on the level of the medium.

Introducing the three genres of the *Going Dutch Corpus* individually, Sections 3.1.1 and 3.1.2 focus on the two sub-corpora of ego-documents, i.e. private letters as well as diaries and travelogues, respectively. Section 3.1.3 provides an outline of the sub-corpus of newspapers.

3.1.1 Private letters

In the field of historical sociolinguistics, ego-documents have been attested a special and particularly valuable role in gaining access to authentic language use in the past (cf. Chapter 3). According to Elspaß (2012: 156), they are “as close to speech as non-fictional historical texts can possibly be and therefore cast light on the history of natural language”. Among the group of ego-documents, letters, and private letters in particular, are “the best possible data for studying everyday men and women in society, their linguistic knowledge and behaviour, as well as their social inscription” (Marquilha 2012: 31). Similarly, Martineau (2013: 133) argues that “private family letters are the best documents for historical sociolinguistics because they are the closest written documents to language of immediacy”.

As part of the *Going Dutch Corpus*, a sub-corpus of private letters of approximately 210,000 words was compiled. During the careful selection phase of Dutch letters from the late eighteenth- and early nineteenth centuries, the following criteria were considered and applied:

- (1) The selected private letters should primarily include personal matters (i.e. exclusion of business letters).
- (2) The selected private letters should be written in a symmetrical communicative setting (i.e. correspondence between family members and friends).
- (3) The selected private letters should represent ‘everyday communication’ (i.e. exclusion of letters of condolence, thanks and congratulations).

First of all, all letters had to be classified as ‘personal letters’ or ‘private letters’, implying that their primarily contain personal matters, written in the private sphere. This also means that business letters and other types of non-private letters were excluded. It should be noted, though, that the dividing line between private and business letters is often very fuzzy, especially in historical letter writing (e.g. Van Bergen & Denison 2007: 4; Włodarczyk 2013: 205). Therefore the classification of letters as private letters applied in this dissertation mainly follows the rule of thumb suggested by Nobels (2013: 27-28):

if the sender and addressee of the letter were closely related to each 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 message other than greetings for the addressee's family and wishes for the addressee's good health, the letter was classified as a business letter.

In fact, the vast majority of letters in the *Going Dutch Corpus* represents correspondence between family members. The extraordinary value of this specific type of correspondence in historical-sociolinguistic investigation is also emphasised by Martineau (2013: 132), who considers private family letters as the best possible way to gain access to authentic language use in historical contexts:

Private correspondence, especially letters to family members, are a valuable primary source of information for reconstituting the nature of exchange, and the language used in former times. Despite the use of writing as a medium, family letters reflect a fairly close relationship between the writer and the recipient in a manner similar to exchanges between friends, not always found in such oral materials as folktales or plays featuring popular characters, or even some modern sociolinguistic interviews.

Whenever family archives provided a substantial amount of suitable texts, preference was given to the more intimate relationships such as spouses, parent-child and siblings(-in-law) rather than, for instance, uncle/aunt-nephew/niece or cousin-cousin.

The second criterion is closely related to the personal content of letters, taking into account the symmetry in communication. With regard to the relationship between senders and addressees, all private letters selected for the *Going Dutch Corpus* should be written in a symmetrical communicative setting. Elspaß (2005: 55) discusses the unsuitability of sources in institutional and thus asymmetrical contexts:

Wenig tauglich sind Quellen, die in einem institutsbezogenen Zusammenhang stehen, also Bitt-, Petitions-, Beschwerdebriefe oder andere Schreiben an Behörden. Erstens kann die Autorschaft solcher Briefe sowie der Einfluss vorgefertigter Briefmuster nicht eindeutig geklärt werden, und zweitens repräsentieren sie asymmetrische Kommunikation, d. h. dass durch die geforderte Anpassung an spezialisierte und routinierte Kommunikationsformen die ‚natürliche‘ Ausdrucksweise der Alltagssprache in hohem Maße verfremdet erscheint.

He argues that in composing “letters in asymmetrical communicative settings (letters of appeal), [...] writers usually draw on discourse traditions with highly formalized discourse” (Elspaß 2012: 158), making them less suitable for the study of authentic language use. On the other hand, in symmetrical communication, “grammatical correctness, spelling or particular sets of formulae were not crucial to a successful communicational act, so that even barely literate people would take up

pen or pencil to write down texts of private interest” (ibid.). This makes them far more authentic in terms of ‘historical orality’²⁶ than, for example, letters of request.

The third criterion considered during the selection phase refers to what Elspaß (2005) calls *geschriebene Alltagssprache* ‘written everyday language’. In order to meet the criterion of ‘everyday communication’ as much as possible, more formal and formalised types of communication such as letters of condolence, letters of thanks and letters of congratulations, were preferably avoided. Nevertheless, it is important to relativise the ‘everydayness’ of these texts. In fact, Elspaß (2005: 66–67) remarks a striking contradiction with regard to ‘everyday language’ in private letters:

Ein gewisser Widerspruch zwischen der Bezeichnung “Alltagsbriefe” und der Zuordnung der Briefe zur alltagssprachlichen Kommunikation scheint darin zu bestehen, dass diese Briefe eben nicht alltäglich geschrieben wurden. [...] Obwohl durch verloren gegangene Briefe sicherlich Lücken in der Chronologie der überlieferten Briefserien bestehen, kann man doch feststellen, dass zwischen den Schreiben eines Briefwechsels oft Monate, sogar Jahre liegen. Es fällt auf, dass viele Briefe an Sonn- und Feiertagen [...], geschrieben wurden, also gerade nicht im Alltag im Sinne von ‘Wochen- und Arbeitstag’. [...] Entscheidend für die Bestimmung von Textsorten wie den Privatbriefen als Alltagstextsorte ist nicht ihre Frequenz im alltäglichen Leben, sondern die Tatsache, dass sie überhaupt nur den Lebensbereichen und Gebrauchsdimensionen des Alltags zugeordnet werden können.

Ultimately, the fact that the letters were written within the sphere of everyday life was considered as more crucial than the actual frequency and moment of writing.

To sum up, the selected letters for the *Going Dutch Corpus* contain primarily personal content, are written in a symmetrical communicative setting as found in family correspondence, and, ideally, represent everyday language. Certain writers contributed more than one letter to the corpus, which met the defined selection criteria. However, in order to avoid overrepresentation of prolific writers (cf. Nobels 2013: 51), as well as to guarantee the comparability of texts (cf. Wegera 2013: 63), the number of words per writer was restricted to a maximum of approximately 2,000 words. This limit was based on the longest letter selected for this sub-corpus, which contains 2,078 words²⁷. In practice, the data of individual letter writers may thus comprise either one long letter or a number of shorter letters.

As summarised in Table 4, the sub-corpus of private letters consists of 210,726 words in total, equally distributed across the two diachronic cross-sections (i.e. 105,427 words for 1770–1790; 105,299 words for 1820–1840) and more or less equally distributed across all seven regions (i.e. ideally 15,000 words per region per

²⁶ For a critical discussion on the notion of ‘historical orality’, see Zeman (2013).

²⁷ Similar limits were defined during the compilation of the *Letters as Loot* corpus, restricting the number of words per individual writer to a maximum of 2,000 words in the seventeenth-century cross-section (Nobels 2013: 50) and to a maximum of 2,500 words in the eighteenth-century cross-section (Simons 2013: 86).

period). The entire sub-corpus comprises 400 texts²⁸ (200 in each period), which were written by 298 different letter writers²⁹.

Table 4. General distribution of data in the sub-corpus of private letters.

Period	N texts (%)	N words (%)	N writers (%)
1770–1790	200 (50.0)	105,427 (50.0)	154 (51.7)
1820–1840	200 (50.0)	105,299 (50.0)	144 (48.3)
Total	400 (100)	210,726 (100)	298 (100)

Aiming at a well-balanced gender representation, the sub-corpus of private letters comprises data from 181 male and 117 female writers. The actual number of words, however, gives a more accurate overview of the achieved gender balance: 54.5% of the letter data was written by men, 45.5% written by women. Although male letters writers are thus slightly more prevalent in the corpus, this can be considered as a well-balanced gender representation, especially for a historical corpus.

In fact, filling the grid cells of the intended corpus design presented in Section 2.3 largely depended on the availability of archival material. This turned out to be the case for less urbanised provinces like Friesland and North Brabant. It was not possible to find an equal amount of texts written by men and women for each region and period. Therefore, the selection criteria were slightly loosened in order to reach the intended corpus size and, at the same time, not to neglect valuable data. It was decided to compensate for the gaps in some gender grid cells by adding more data from the other gender. To give an example, the eighteenth-century data from Friesland only comprise 4,645 words written by women (out of the intended 7,500 words). However, additional male data from the same period was available for this region, which was ultimately used to reach the intended number of 15,000 words. This does not mean, though, that this compensation strategy resulted in an overly male-dominated letter corpus. In the case of nineteenth-century North Brabant, for instance, the gaps in the male grid cell (only 4,333 words) were compensated by additional female data. This modification was considered for the benefit of a larger dataset and should not skew the corpus analyses considerably.

Tables 5a and 5b provide a detailed overview of the sub-corpus of private letters³⁰.

²⁸ Some letters were written by more than one hand. The transcriptions of each hand were saved as separate Text files (indicated by the codes *hand1*, *hand2* etc.) and treated as different texts in this overview, even though they were originally taken from the same archival document.

²⁹ In a few exceptional cases, letter writers contributed data for both periods. Since the two periods represent two distinct generations of language users, these writers are counted as two different individuals in this overview.

³⁰ The actual numbers of words per grid cell generally deviate from the (exact) intended numbers of 7,500/15,000 words. This is mainly due to the decision to include complete

Table 5a. Distribution of data in the sub-corpus of private letters across region and gender (P1 = 1770–1790).

P1	Male			Female			Total		
	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>
FR	23	10,889	19	15	4,645	9	38	15,534	28
GR	25	10,282	15	5	3,398	4	30	13,680	19
NB	21	10,286	13	11	5,187	7	32	15,473	20
NH	18	7,517	17	16	7,579	13	34	15,096	30
SH	13	7,503	13	11	7,548	10	24	15,051	23
UT	10	9,514	7	4	5,771	3	14	15,285	10
ZE	14	7,522	13	14	7,786	11	28	15,308	24
Total	124	63,513	97	76	41,914	57	200	105,427	154

Table 5b. Distribution of data in the sub-corpus of private letters across region and gender (P2 = 1820–1840).

P2	Male			Female			Total		
	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>
FR	18	10,121	14	9	5,560	6	27	15,681	20
GR	16	7,574	12	20	7,624	12	36	15,198	24
NB	13	4,333	10	18	11,620	10	31	15,953	20
NH	15	8,149	12	12	7,958	11	27	16,107	23
SH	17	7,949	16	14	7,644	9	31	15,593	25
UT	13	7,507	13	13	7,771	7	26	15,278	20
ZE	14	5,622	7	8	5,867	5	22	11,489	12
Total	106	51,255	84	94	54,044	60	200	105,299	144

3.1.2 Diaries and travelogues

In addition to the sub-corpus of private letters (3.1.1), the *Going Dutch Corpus* comprises a second type of handwritten ego-documents, viz. diaries and travelogues. Although these texts are often mentioned in the same breath as private

letters rather than cut-off samples. Whenever texts had to be shortened, transcriptions were continued until the end of the sentence.

letters, it has to be kept in mind that they represent two distinct types of ego-documents, differing in various respects. First and foremost, they represent opposite poles of the monologicity–dialogicity continuum, as pointed out by Elspaß (2012: 162):

Whereas private letters are characterized by dialogue and ‘a social practice’ between the correspondents [...], private diaries are strictly monologic by nature. Such texts may be as informal in style and unplanned in their conception as private letters, but they are usually less ‘oral’.

Compared to private letters, which are characterised by their interactive purpose, diaries and travelogues are generally further away from the side of Koch’s & Oesterreicher’s (1985) ‘language of immediacy’ (van der Wal & Rutten 2013: 2; cf. also Chapter 3). Both terminologically and methodologically, the genre referred to as ‘diaries and travelogues’ needs some further clarification. Elspaß (2012: 163) outlines that

the term ‘diary’ covers different types of monological texts, such as personal diaries (with mostly private content), family books (recording events of family life), account books and private chronicles with irregular entries (thus hardly ‘journals’ in the strict sense) that comprise events of family and village life, interspersed with weather reports and news about wars and accidents.

In addition to these types, there is yet another type of diaries, written in travel settings and fairly inconsistently labelled as *reisdagboeken* ‘travel diaries’, *reisjournalen* ‘travel journals’ or *reisverslagen* ‘travelogues’. In their comprehensive inventory of Dutch travelogues, Lindeman et al. (1994: 10) address the vague character of these categories:

De grenzen met sommige andere genres kunnen vaag zijn. Een dagboek kan bijvoorbeeld overgaan in een reisverslag, en omgekeerd.

‘The boundaries with certain other genres can be vague. A diary, for example, can blend into a travelogue, and the other way round.’

Interestingly, the 3.2 version of ARCHER (*A Representative Corpus of Historical English Registers*), a multi-genre historical corpus of British and American English, introduced a split of the previous single genre ‘journals-diaries’ into two separate genres ‘diaries’ and ‘journals’ (mostly travel journals) (ARCHER website; cf. also Yáñez-Bouza 2011, 2016):

Following the original design of the corpus, the defining criterion for the classification of the materials in ARCHER 3.2 is topic and purpose of the text: diaries record private matters, domestic affairs, everyday activities and routines; journals report on a journey or a task associated with travel (including sea travel and war campaigns) and with political matters. In ARCHER 3.2 there are 122 diaries and 122 journals, of which 105 are travel journals and 17 are political journals.

In the *Going Dutch Corpus*, no such distinction between travel and non-travel settings of diary writing is made. The crucial selection criterion for the categorisation as ‘diaries and travelogues’ was the personal character of these texts, comprising the writer’s own experiences and commentary. Following the approaches in Dekker (1995) and Lindeman et al. (1994), impersonal accounts such as cash account books (*kasboeken*) and ship’s log books (*scheepsjournalen*, *logboeken*) were not included in the *Going Dutch Corpus*, as they cannot be regarded as ego-documents. However, Dekker (1995: 277) admits that it is “not always easy to draw the line, which will come as no surprise for a time when the personal and the public spheres were still strongly intertwined”. What is more, we have to be aware that the ‘personal’ character is very often limited to a fairly factual account of daily activities without a considerably high degree of attention given to introspection and intimacy. In fact, texts from the period under investigation can hardly be compared to our present-day understanding of personal diary writing. Baggerman (2011: 465) rightly remarks that many diaries “provide more thorough information about the outside temperature than about the author’s inner life”.

Apart from their varying terminology, diaries and travelogues also tend to differ in length and layout, ranging from concise telegram-style notes to more comprehensive narrations, as well as from daily to more irregular entries. Generally, text samples of 2,500 words per writer (usually taken from one single document) were randomly selected in order to avoid an overrepresentation of certain writers³¹. For practical reasons, also keeping in mind the limited availability of eighteenth- and nineteenth-century diaries and travelogues, the maximum number of words per writer had to be slightly extended to 2,500 words (as opposed to 2,000 words per writer for private letters) in order to reach the intended corpus size.

As summarised in Table 6, the sub-corpus of diaries and travelogues consists of 140,507 words in all, comprising 71,157 words for the eighteenth-century period and 69,350 words for the nineteenth-century period.

Table 6. General distribution of data in the sub-corpus of diaries and travelogues.

Period	N texts (%)	N words (%)	N writers (%)
1770–1790	26 (52.0)	71,157 (50.6)	25 (50.0)
1820–1840	24 (48.0)	69,350 (49.4)	25 (50.0)
Total	50 (100)	140,507 (100)	50 (100)

All regions are represented by approximately 10,000 words per period. The sub-corpus contains 50 different diaries and travelogues, which were written by 50

³¹ Only in some exceptional cases (i.e. for regions where the amount of suitable texts was limited), 5,000 words per writer were transcribed in order to reach the intended corpus size.

different writers³². The detailed distribution of data across periods, regions and gender is given in Tables 7a and 7b.

Table 7a. Distribution of data in the sub-corpus of diaries and travelogues across region and gender (P1 = 1770–1790).

P1	Male			Female			Total		
	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>
FR	4	10,198	4	0	0	0	4	10,198	4
GR	3	10,144	3	0	0	0	3	10,144	3
NB	3	10,156	2	0	0	0	3	10,156	2
NH	3	7,680	3	1	2,392	1	4	10,072	4
SH	4	10,126	4	0	0	0	4	10,126	4
UT	3	7,662	3	1	2,601	1	4	10,263	4
ZE	3	7,633	3	1	2,565	1	4	10,198	4
Total	23	63,599	22	3	7,558	3	26	71,157	25

Table 7b. Distribution of data in the sub-corpus of diaries and travelogues across region and gender (P2 = 1820–1840).

P2	Male			Female			Total		
	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>	<i>Texts</i>	<i>Words</i>	<i>Writers</i>
FR	4	10,250	4	0	0	0	4	10,250	4
GR	3	10,061	3	0	0	0	3	10,061	3
NB	1	5,009	1	0	0	0	1	5,009	1
NH	3	6,067	3	3	5,101	3	6	11,168	6
SH	3	7,807	3	2	5,120	2	5	12,927	5
UT	1	5,056	1	2	4,727	2	3	9,783	3
ZE	3	10,152	3	0	0	0	3	10,152	3
Total	18	54,402	18	7	14,948	7	25	69,350	25

³² The apparent 1:1 ratio needs some further explanation, though. On the one hand, one nineteenth-century travelogue from North Holland (*Amsterdam_Backker_172_663_dia01*) was actually written by two distinct hands, most probably by a husband (first part) and his wife (second part). A diarist from North Brabant, on the other hand, contributed two different texts (samples) for the eighteenth-century period. Like in the sub-corpus of private letters, writers who contributed data for both time periods were counted as two different persons.

Like the sub-corpus of private letters, this sub-corpus was initially planned as gender-balanced. Unfortunately, research in the visited archives has shown that the distribution of male and female diary writers from the periods under investigation is not balanced at all, which is why the intended gender representation could not be achieved. The final sub-corpus does contain data from at least ten female diarists, though. 22,506 words were written by females, which roughly correspond to 16.0% of the total sub-corpus. However, it must be taken into account that these texts are not equally distributed across periods (i.e. mainly nineteenth century) and regions (i.e. mainly Holland and Utrecht).

3.1.3 Newspapers

In addition to two types of handwritten ego-documents, the multi-genre design of the *Going Dutch Corpus* also incorporates printed and published texts. Unlike private letters (Section 3.1.1) and diaries and travelogues (Section 3.1.2), the genre of newspapers is typically associated with more standardised writing, more closely representing the ‘language of distance’ in Koch & Oesterreicher’s (1985) terms.

With their broad readership and, especially compared to formal and literary texts, a more popular and accessible style of writing, newspapers can certainly be considered as a valuable linguistic source in order to examine variation and change in language practice. Rademann (1998: 49) argues that with regard to the “considerably large target audiences, the language used in newspaper articles is often assumed to be characteristic of the respective period and society they are published in”, which makes this genre particularly suitable for diachronic studies.

Another methodological advantage of newspapers and a decisive factor to include them in the *Going Dutch Corpus* is their geographical spread across the language area. In the late eighteenth and early nineteenth centuries, newspapers were still locally produced and distributed, and thus primarily catered to regional readerships. This makes them a particularly interesting printed genre for a (historical-)sociolinguistic approach. In fact, for each of the seven regions in the corpus, a regional newspaper could be selected. Therefore, the sub-corpus of newspapers is as regionally balanced as the two sub-corpora of ego-documents, covering the same seven regions (cf. Section 3.3).

The use of newspapers for a systematic comparison with ego-documents has also been attested before. Percy (2012: 194) argues that “[t]he register of news reportage has an interesting if indirect relationship with everyday language”. Notably, the documentation of the *GerManC* corpus (Durrell et al. 2012: 1; cf. also Chapter 3) even classifies newspapers as orally oriented registers, alongside personal letters³³. In this respect, newspapers are probably best considered as a genre which, on the one hand, displays a printed, edited and fairly standardised

³³ The orally oriented registers in the *GermanC* corpus comprise drama, newspapers, sermons and personal letters, as opposed to more print-oriented registers like narrative prose, scholarly, scientific and legal texts.

form of writing, but on the other hand, represents authentic ‘everyday’ language and the ‘language of immediacy’ more closely than, for instance, academic prose or literary works. Elspaß & Niehaus (2014: 51-52) suggest a similar corpus design for German, considering regional newspapers as suitable historical data *from above* as opposed to private letters as historical data *from below*.

The selected newspaper texts, as a rule, comprise news reports only, following Niehaus’ (2016: 48) criterion to take into account proper newspaper language, representing the language of editors and correspondents:

Ich habe außerdem darauf geachtet, möglichst nur Texte zu berücksichtigen, die ‚Zeitungssprache‘ i.e.S., also die Sprache der für eine Zeitung schreibenden Redakteure und Korrespondenten, wiedergeben.

Therefore, official government announcements, advertisements as well as extensive lists of names, for instance lists of decedents, were categorically excluded from the corpus.

As summarised in Table 8 below, the sub-corpus of newspapers consists of 70,645 words in all, comprising an equal number of words for both periods (i.e. approximately 35,000 words for 1770–1790 and 1820–1840 each) and all seven regions (i.e. 5,000 words per region)³⁴.

Table 8. Distribution of data in the sub-corpus of newspapers across period and region.

Region	Period 1	Period 2	Total
Friesland	5,025	5,018	10,043
Groningen	5,051	5,064	10,115
North Brabant	5,018	5,036	10,054
North Holland	5,088	5,093	10,181
South Holland	5,048	5,027	10,075
Utrecht	5,040	5,033	10,073
Zeeland	5,053	5,051	10,104
Total	35,323	35,322	70,645

3.2 Diachronic dimension

The diachronic dimension of the *Going Dutch Corpus* is closely linked to the diachronically oriented approach of this dissertation, investigating the possible influence of top-down language policy measures on actual language practice. The

³⁴ The minor deviations from the limit of 5,000 words are due to the methodological decision not to cut off sentences but to fully transcribe them until the next full stop.

historical event of the Dutch *schrijftaalregeling* in the early 1800s, with Siegenbeek's official orthography and Weiland's official grammar being published in 1804 and 1805, respectively (cf. Chapter 2), serves as a starting point for defining the diachronic cross-sections of the *Going Dutch Corpus*.

In order to gain access to language use before and after this landmark in the history of Dutch standardisation, two periods of twenty years each were defined, with a gap of approximately one generation between these cross-sections. The late eighteenth-century period, spanning the years 1770–1790, represents the generation of language users *before* the national language policy was introduced. Symmetrically, the early nineteenth-century period, i.e. *after* the introduction of Siegenbeek (1804) and Weiland (1805), spans the years 1820–1840, representing the generation of language users which had (probably) been exposed to the language policy measures, as envisaged by the government. For the main research objectives of this dissertation, the diachronic dimension is the most important independent variable of the *Going Dutch Corpus*.

3.3 Spatial dimension

Addressing the importance of space as an external factor, Elspaß (2012: 313) argues that when dealing with “languages [...] with considerable regional variation, it is also imperative to consider texts from different regions”. This is certainly the case for late eighteenth- and early nineteenth-century Dutch. Ultimately aiming at a regionally balanced representation of all three genres and both time periods, the *Going Dutch Corpus* comprises data from a variety of regions in the Northern Netherlands.

Previous historical-sociolinguistic research on this language area, most notably the *Letters as Loot* programme (Rutten & van der Wal 2014: 11–13; cf. also Nobels 2013: 28–30; Simons 2013: 104–106;), mainly focused on the regions on the western coast of the Northern Netherlands, viz. Holland and Zeeland with their main cities of Amsterdam, Rotterdam, Middelburg and Vlissingen³⁵. However, as Rutten et al. (2014b: 12) point out, it is important to avoid the emphasis on specific regions, usually demographic and socio-economic centres and their surroundings, and to consider demographically less important regions as well. For this reason, the *Going Dutch Corpus* adds a new layer of four rather under-studied regions to the three westernmost regions of North Holland, South Holland and Zeeland, expanding the previously investigated language area to the north (Friesland, Groningen), to the east (Utrecht) and to the south (North Brabant).

At the same time, the remaining provinces in the eastern part of the Northern Netherlands were not included in the *Going Dutch Corpus*, mainly but not exclusively for practical reasons. While it was not feasible to compile a corpus that

³⁵ The focus of the *Letters as Loot* corpus on the western regions of the Northern Netherlands is due to the prevailing origin of the confiscated letters. The vast majority of letters was sent to and from the provinces of Holland and Zeeland (Rutten et al. 2012: 329; Rutten & van der Wal 2014: 11).

covers the entire language area, the eastern border provinces certainly offer intriguing points of departure for future research.

The spatial dimension of the *Going Dutch Corpus* incorporates two different perspectives. Section 3.3.1 takes into account regional variation on the basis of provincial boundaries. Another distinction is based on demographic and socio-economic differences, focusing on variation between the urbanised centre and the less urbanised periphery, which will be addressed in Section 3.3.2.

3.3.1 Regions

The first variable of the spatial dimension investigates variation across different regions of the Northern Netherlands. For practical purposes, these regional categories were based on present-day provinces and provincial boundaries (cf. also Simons 2013: 104), which, in some cases, deviate from the historical boundaries in the late eighteenth and early nineteenth centuries. The present provinces of North Holland and South Holland, for instance, were part of the province of Holland until its split in 1840. However, previous studies on seventeenth- and eighteenth-century Dutch (e.g. Rutten & van der Wal 2014), have revealed distinct regional patterns in North and South Holland, the latter of which being characterised as a “transitional zone between Holland and Zeeland” (Rutten & van der Wal 2014: 341). Therefore, it seemed both logical and necessary to consider the Holland area before 1840 as two distinct regions.

The following seven regions of the Northern Netherlands are covered in the *Going Dutch Corpus* (listed in alphabetical order): Friesland (FR), Groningen (GR), North Brabant (NB), North Holland (NH), South Holland (SH), Utrecht (UT) and Zeeland (ZE). See Figure 5 for a map of the investigated language area indicating the regions represented in the corpus.

As mentioned before, a balanced representation of all selected regions was envisaged. While this aim was easily achieved in the compilation of the sub-corpus of newspapers (with 10,000 words per region), the compilation of the sub-corpora of handwritten ego-documents (private letters: ideally 30,000 words per region; diaries and travelogues: ideally 20,000 words per region) largely depended on the availability of suitable archival sources.

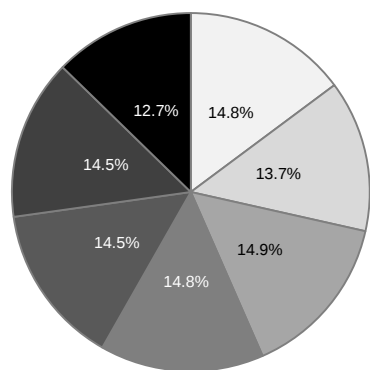
As a consequence, some regions like North Brabant and also Zeeland comprise slightly less words than socio-economically and demographically more dominant regions like North Holland and South Holland with various big cities and, from a practical point of view, more archives to visit.

Figure 5. Map of the Northern Netherlands indicating the regions represented in the *Going Dutch Corpus*.



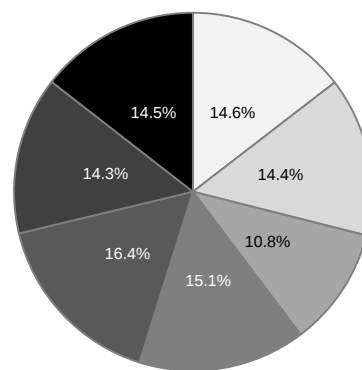
Figures 6 and 7 illustrate the well-balanced distribution of data across regions in the sub-corpora of private letters (cf. Tables 5a and 5b) as well as diaries and travelogues (cf. Tables 7a and 7b).

Figure 6. Distribution of data across region in the sub-corpus of private letters.



□FR □GR ■NB ■NH ■SH ■UT ■ZE

Figure 7. Distribution of data across region sub-corpus of diaries and travelogues.



□FR □GR ■NB ■NH ■SH ■UT ■ZE

Another methodological consideration concerns the regional categorisation of texts. Whereas it was a straightforward task to categorise newspapers according to their place of publication³⁶, ego-documents could not be categorised that easily. The major challenge was to determine the starting point for a reliable categorisation into regions. In the case of private letters, the place from which a letter was sent might serve as an indication, but at the same time, would have been too inaccurate and even misleading. To give an example, a writer who sent a letter from Middelburg (Zeeland) to a close relative in Utrecht was not necessarily a citizen of Zeeland. Moreover, any letter from outside the language area (i.e. outside the seven selected regions and even outside the Northern Netherlands), would have been rejected, even though their writers were actually based in one of the investigated regions. Issues such as travelling, migration and inter-regional marriages further complicate the regional categorisation. Ideally, the so-called 'region of residence', i.e. "the region where a sender was born and raised or where he or she spent most of his or her life" (*Letters as Loot* corpus website) could be traced and identified.

In order to assign letter writers to one of the seven regional categories, the following procedure was applied, listed in descending order of importance:

- (1) Based on the names and information given in the letters, basic genealogical and biographical research was conducted online in order to determine the places of birth and death, also taking into account mobility across the lifespan. It was not possible, though, to trace back the background of every single writer. Generally, in-depth genealogical and biographical research of individual writers was beyond the scope of this project.
- (2) If the background of a writer, and most importantly the region of residence, could not be identified, the general regional association of the family (and the corresponding family archive) was considered.
- (3) Only in a few exceptional cases, i.e. when very little or even nothing was known about a writer, the place from which the letter was sent (as given on the document itself) was used as a tentative indication.

The regional categorisation of diaries and travelogues was based on the first two criteria of the procedure mentioned above.

³⁶ It should be emphasised, though, that this categorisation takes no account of the regional background of the actual writers of news reports. Given the lack of information about correspondents in eighteenth- and nineteenth-century newspapers, it is impossible to determine who contributed the texts, let alone where these writers came from.

3.3.2 Centre versus periphery

While the categorisation of regions described in Section 3.3.1 was based on provincial boundaries, the spatial dimension of the *Going Dutch Corpus* also integrates a second categorisation on the basis of demographic and socio-economic factors (cf. also Vosters 2011: 207-208). Utilising the variety of regions in the corpus, a distinction was made between the centre (i.e. the more urbanised, demographically and economically stronger regions) and the periphery (i.e. the far less urbanised regions outside the centre). Considering the fact that the “biggest sociogeographical contrast in the Republic was [...] not that between town and countryside, but between more and less urbanized provinces” (Kloek & Mijnhardt 2004: 48), the degree of urbanisation was considered as the crucial factor for the grouping of provinces into the two categories of ‘centre’ and ‘periphery’.

Kloek & Mijnhardt (2004: 32) outline the demographic situation in the Northern Netherlands at the turn of the century as follows:

Around 1800, the contours of what would far later come to be known as the “Randstad,” the urban agglomeration of western Holland, were already becoming clear. More than 625,000 people – 30% of the country’s population – lived within the area bounded by Amsterdam, Haarlem, Leiden, The Hague, Rotterdam, and Utrecht. Even beyond this perimeter, the next town was never far away, and the countryside was relatively densely populated. Cities were the natural habitat of the average Dutch man or woman of the day.

Based on this outline, the regions of North Holland, South Holland and Utrecht, in which all of the above-mentioned cities are located, make up the centre of the language area. The regions of Friesland and Groningen (in the north) as well as North Brabant (in the south) can be regarded as peripheral with respect to this centre.

In terms of the binary centre–periphery distinction, the seventh province in the corpus, i.e. Zeeland, takes a more ambiguous position. Historically, it clearly belonged to the demographically and economically leading regions in the sixteenth and seventeenth centuries, i.e. during the Golden Age of the Northern Netherlands (together with Holland). However, Zeeland’s importance declined in the course of the subsequent centuries, ultimately losing its status as a centre. As Kloek & Mijnhardt (2004: 49) point out, “the Republic’s center of economic gravity shifted to the Amsterdam-Rotterdam axis”, which left the once flourishing region of Zeeland as one of the victims of this development (ibid.: 33). Therefore, I decided to leave Zeeland out of consideration and to treat it separately. However, the corpus-based case studies in Chapters 5–12 might shed more light on the position of Zeeland, i.e. whether it is linguistically closer to either the centre or the periphery, or whether the empirical investigation actually confirms the ambiguous intermediate position.

3.4 Social dimension

Studying the relation between linguistic variation and its social significance has always been central to sociolinguistic research ever since the emergence of this academic field of study (Nevalainen & Raumolin-Brunberg 2003: 16). In this section, two major social variables will be briefly discussed: social class (3.4.1) as well as gender (3.4.2), the latter of which will be further investigated in the corpus analyses of this dissertation.

3.4.1 Social class

In both present-day and historical sociolinguistics, the variable of social class has often been regarded as “one of the major – if not *the* major – external constraints” (Nevalainen & Raumolin-Brunberg 2003: 133). Investigating seventeenth- and eighteenth-century Dutch, the findings presented in Rutten & van der Wal (2014), based on the *Letters as Loot* corpus, confirmed social class as one of the central independent variables affecting patterns of language variation and change.

On the basis of the well-established historians’ model of social stratification in the seventeenth- and eighteenth-century Republic of the Seven United Provinces (1581-1795), letter writers were classified into four social categories: the upper class (UC), the upper-middle class (UMC), the lower-middle class (LMC) and the lower class (LC) (cf. Rutten & van der Wal 2014: 9-10). The classification presented in Table 9 was primarily based on the writers’ professions or, in the case of women, on the profession or social position of their husbands or fathers (Rutten & van der Wal 2014: 10; cf. also Nevalainen & Raumolin-Brunberg 2003: 37 for an English example).

Table 9. Social stratification of the seventeenth- and eighteenth-century Republic of the Seven United Provinces in the *Letters as Loot* corpus (cf. Rutten & van der Wal 2014: 10).

	Historians’ stratification	<i>Letters as Loot</i> corpus
(1)	Nobility and the non-noble ruling classes	Upper class (UC)
(2)	Bourgeoisie, e.g. wealthy merchants, ship owners, academics, commissioned officers	
(3)	Prosperous middle class, e.g. large storekeepers, non-commissioned officers, well-to-do farmers	Upper-middle class (UMC)
(4)	Petty bourgeoisie, e.g. petty storekeepers, small craftsmen, minor officials	Lower-middle class (LMC)
(5)	Mass of wage workers, e.g. sailors, servants, soldiers	Lower class (LC)
(6)	Have-nots, e.g. tramps, beggars, disabled	

Although this model needs to be modified according to the changing social stratification in the late eighteenth- and early nineteenth-century period under investigation, I maintain the suggested four-partite division into upper, upper-middle, lower-middle and lower classes. Initially, it was considered to integrate social class variation in this dissertation as well. However, throughout the exploratory preparation phase and the actual collection of data, it became evident that a representative amount of ego-documents written by lower- and lower-middle-class writers in the periods 1770–1790 and 1820–1840 is practically unavailable in Dutch archives. Whereas the eighteenth-century cross-section of private letters, at least to some extent, could have been covered with data from the *Letters as Loot* corpus, suitable material for the nineteenth-century period turned out to be sparse.

These limitations only emphasise the unique character of the collection of Dutch sailing letters used for the *Letters as Loot* corpus. At the same time, they confirm the arbitrariness of written sources preserved and stored in municipal and regional archives. Schneider (2013: 65, originally quoted from Montgomery 1997: 227) describes them as “products of the ‘vagaries and accidents of history (such as which family chose to preserve letters, whether letters survived decay)’”. Not surprisingly, those documents which *have* been preserved and kept in the archives to the present day, are more likely to derive from relatively well-to-do families of the middle to the upper classes rather than from the lower ranks of society.

In order to avoid a far too small and therefore hardly representative sample of lower-class and lower middle-class writing, I preferred to compile a well-balanced and socially more homogeneous corpus of eighteenth- and nineteenth century writers from the (upper-)middle to the upper classes. Most importantly, the very highest rank of Dutch society was excluded from the corpus. In fact, even the upper-middle class has to be regarded as a proper middle class, which is why these texts do not necessarily contradict the historical-sociolinguistic tradition *from below*. Furthermore, the ego-documents and particularly the sub-corpus of private letters represent a wide range of family archives, often comprising more texts than just from the ‘influential’ main branch only. The selected texts also represent less central family members of ‘minor’ or in-law branches of the extended family (Martineau 2013: 141).

With respect to the comparatively homogeneous representation of social ranks in the *Going Dutch Corpus*, the variable of social class will not be considered in this dissertation. Instead, the focus will be on the equally significant social variable of gender (Section 3.4.2).

3.4.2 Gender

Within and across the two sub-corpora of ego-documents (i.e. private letters, diaries and travelogues), it is possible to investigate social variation by focusing on the independent variable of gender. In sociolinguistic research, gender has repeatedly emerged as “one of the most robust social variables” (Nevalainen &

Raumolin-Brunberg 2003: 110) in order to identify and explain patterns in language use by men and women. Even though the categorisation of men and women in the *Going Dutch Corpus* is purely based on their biological sexes, I prefer to use the term *gender* rather than *sex*, taking into account that this social variable primarily focuses on variation based on a social roles and practices rather than on a biologically or physiologically-based distinction (Meyerhoff 2011: 201; Nevalainen & Raumolin-Brunberg 2003: 110). Kielkiewicz-Janowiak (2012: 313) points out that “linguistic patterns distributed according to the sex of the speaker are to be accounted for by reference to the social characteristics of the speakers (their social roles, their attitudes, their preferences) in the larger societal context”. She further argues that the idea of gender as a socio-cultural concept should also be considered when studying language variation in historical contexts (Kielkiewicz-Janowiak 2012: 313):

For historical sociolinguistics too it became obvious that, rather than simply indicating the sex of the speaker, researchers should define gender in terms of a set of social roles and characteristics usually ascribed to, and accepted by, women and men within a given society.

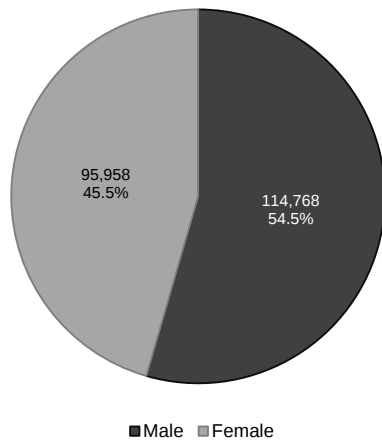
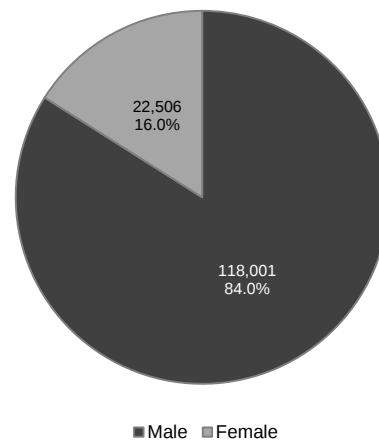
Irrespective of this terminological choice, traditional language histories are almost exclusively based on texts by male writers, mostly from the elite and socio-economically leading regions (cf. Chapter 3). Women, on the other hand, “are, as a rule, under-represented” (Kielkiewicz-Janowiak 2012: 308).

In order to investigate language variation based on gender differences, the *Going Dutch Corpus* initially aimed at a balanced representation of men and women, by including a more or less equal number of words written by male and female writers in the two sub-corpora of ego-documents. Table 10 shows that male writers are overrepresented in the *Going Dutch Corpus* with a total share of two-thirds. However, with respect to the near-absence of female writers in traditional language histories, the gender representation in the *Going Dutch Corpus*, with one-third of the data being written by women, is still a considerable change.

Table 10. General distribution of data across gender and time.

Gender	Period 1	Period 2	Total
Male	127,112 (72.0%)	105,657 (60.5%)	232,769 (66.3%)
Female	49,472 (28.0%)	68,992 (39.5%)	118,464 (33.7%)
Total	176,584 (100%)	174,649 (100%)	351,233 (100%)

Zooming in on the gender representation in the sub-corpus of private letters, Figure 8 shows that gender balance was, in fact, almost achieved for this genre, with a relative distribution of 54.5% (male) versus 45.5% (female) (cf. also Tables 5a and 5b).

Figure 8. Gender distribution in the sub-corpus of private letters.**Figure 9.** Gender distribution in the sub-corpus of diaries and travelogues.

Unfortunately, a similarly balanced gender representation could not be achieved in the sub-corpus of diaries and travelogues, due to the sparsity of diaries and travelogues written by women. Whereas the major part, namely 84.0% of the words, was written by men, merely 16.0% was written by women, as shown in Figure 9 (cf. also Tables 7a and 7b).

4 Individual dimension and the *Martini Buys Correspondence Corpus*

The fifth variational dimension addresses variation and change on the level of individual language users. Whereas the variables presented in Section 3 considered the community at large, or specific groups of language users (e.g. writers from North Holland versus writers from Friesland, or social groups such as men versus women), this dimension zooms in on the behaviour of individual language users, examining variation *between* each other and *within* their own language practices. In modern sociolinguistics in particular, these variables have commonly been referred to as *inter-speaker* and *intra-speaker variation* (e.g. Meyerhoff 2011: 17). However, working with historical and thus written data implies that one can hardly refer to language users as ‘speakers’ in the strict sense. In this dissertation, the modified terms of *inter-writer* and *intra-writer* variation will be used in order to refer to variation between and within individuals, respectively.

By taking a micro-level perspective on the linguistic behaviour of individual writers, a number of (partly interrelated) questions will be addressed. Assuming that language users possibly alternate between different realisations of the same linguistic variable, it will be examined how consistent or inconsistent

individual writers were in the use of particular features, both from a synchronic and diachronic point of view. Furthermore, a close comparison of individual patterns and preferences can shed more light on inter-individual differences between close family members.

Specifically for the study of inter- and intra-individual variation and change, a separate, stand-alone corpus of family correspondence was designed and compiled. The *Martini Buys Correspondence Corpus* comprises approximately 64,000 words, distributed across 102 private letters written by eleven family members. Sixteen of these letters are also included in the *Going Dutch Corpus*. Spanning three generations of male and female informants from the second half of the eighteenth and the first half of the nineteenth century, the correspondence corpus takes into account the additional factors of age and individual lifespan changes. Most interestingly, however, it also allows to take a micro-level perspective on the effects of language policy measures on the linguistic behaviour of individual family members. The *Martini Buys Correspondence Corpus* will be presented in more detail in Chapter 13.

5 The *Normative Corpus of the Northern Netherlands*

In addition to the multi-genre *Going Dutch Corpus* as the main corpus for the linguistic analyses of this dissertation (Sections 2 and 3, cf. also Chapters 5–12), and the *Martini Buys Correspondence Corpus* for the study of individual variation and change (Section 4, cf. also Chapter 13), a third corpus of eighteenth-century normative works was compiled, referred to here as the *Normative Corpus of the Northern Netherlands*. In order to determine the normative influence of spelling and grammatical prescriptions on language practices, the quantitative analyses of orthographic and morphosyntactic variables will be complemented by a qualitative study of contemporary metalinguistic discourse. As outlined in Chapter 2 (Section 2), there was a vivid normative tradition in the Northern Netherlands throughout the eighteenth century, i.e. before the official *schrijftaalregeling* of Dutch. Rather than to focus on the direct influence of these two officialised publications (i.e. Siegenbeek 1804, Weiland 1805) alone, I will also take into account the normative preferences and prescriptions laid down in metalinguistic discourse of the preceding eighteenth century. In fact, the codifying character of Siegenbeek's orthography and Weiland's grammar can certainly be regarded as a 'conclusion' of the eighteenth-century normative tradition (van de Bilt 2009: 192). Based on a wide range of eighteenth-century normative publications, gradually paving the way for the national language policy in the early 1800s, developments in actual language use can be related to the possible influence of norms and prescriptions.

The *Normative Corpus of the Northern Netherlands* compiled for this dissertation comprises 31 normative publications on orthographic and grammatical issues, such as spelling guides, grammar books and more general linguistic treatises. The selection of texts can be considered as a (more or less) exhaustive account of normative works published in the Northern Netherlands in the course of the

eighteenth-century, spanning the period of 1699–1805. The texts listed in Table 11 are available either in print or in digital form.

Table 11. The *Normative Corpus of the Northern Netherlands* (1699–1805).

Year	Author	Title [Place of publication]
1699	Francius, Petrus	<i>Gregorius Nazianzenus, Van de mededeelzaambeidt</i> [Amsterdam]
1700	van Hoogstraten, David	<i>Aenmerkingen over de geslachten der zelfstandige naemwoorden</i> [Amsterdam]
1703	Nylöe, Jakobus	<i>Aanleiding tot de Nederduitsche taal, om goetd en zuiver Nederduitsch te spreken of te schryven</i> [Amsterdam]
1705	Hilarides, Johannes	<i>Nieuwe taalgronden der Neederduitsche taal</i> [Franeker]
1706	Moonen, Arnold	<i>Nederduitsche spraakkunst</i> [Amsterdam]
1707	Verwer, Adriaen	<i>Linguae Belgicae idea grammatica, poëtica, rhetorica</i> [Amsterdam] (Translation <i>Letterkonstige, dichtkonstige en redenkonstige schetse van de Nederduitsche tale</i>)
1708	Sewel, Willem	<i>Nederduitsche spraakkunst</i> [Amsterdam]
1712	Sewel, Willem	<i>Nederduitsche spraakkunst</i> (Second edition) [Amsterdam]
1723	ten Kate, Lambert	<i>Aenleiding tot de kennisse van het verhevene deel der Nederduitsche sprake</i> [Amsterdam]
1730	Huydecoper, Balthazar	<i>Proeve van taal- en dichtkunde</i> [Amsterdam]
1743	van Niervaart, Cornelis	<i>Oprecht onderwijs van de letter-kunst</i> [Purmerend]
1746	Hakvoord, Barend	<i>De nieuwe Nederduitse spel-, lees- en schryf-kunst</i> [Deventer]
1748	van Belle, Jan	<i>Korte wegwijzer, ter spel- spraak- en dichtkonden</i> [Haarlem]
1755	van Belle, Jan	<i>Korte schets der Nederduitsche spraakkunst</i> [Haarlem]
1758	van Rhyn, Leonard	<i>Kort begryp der Nederduitsche spel-kunst</i> [Amsterdam]
1761	Elzevier, Kornelis	<i>Drie dichtproeven [...] benevens een proef van een nieuwe Nederduitsche spraakkunst</i> [Haarlem]
1763	Kluit, Adriaan	<i>Eerste verhoog over de tegenwoordige spelling der Nederduitsche taal</i> [Leiden]
1763	Heugelenburg, Martinus	<i>Klein woordenboek, zijnde een kort en klaar onderwijs in de Nederlandze spel, en leeskonst</i> [Amsterdam]
1764 ³⁷	de Haes, Frans	<i>De nagelaten gedichten, en Nederduitsche spraakkunst</i> [Amsterdam]
1769	van der Palm, Kornelis	<i>Nederduitsche spraakkunst, voor de jeugdt</i> [Rotterdam]
1770	Kunst wordt door arbeid verkreegen	<i>Nederduitsche spraakkunst</i> [Leiden]
1774	Zeydelaar, Ernst	<i>Nederduitsche spelkonst</i> [Dordrecht]

³⁷ De Haes' *Nederduitsche spraakkunst* was published posthumously in 1764, but had probably been written before or around 1740 (Dibbets 1999: 44).

1776	Tollius, Herman	<i>Proeve eener Aanleiding tot de Nederduitsche Letterkunst</i>
1777	Kluit, Adriaan	<i>Vertoog over de tegenwoordige spelling der Nederduitsche taal</i> [Leiden]
1776	Stijl, Klaas & Lambertus van Bolhuis	<i>Beknopte aanleiding tot de kennis der spelling, spraakdeelen, en zinteecken van de Nederduitsche taal</i> [Groningen]
1793	van Bolhuis, Lambertus	<i>Beknopte Nederduitsche spraakkunst</i> [Leiden]
1799	Wester, Hendrik	<i>Bevatlyk onderrys in de Nederlandsche spel- en taalkunde, voor de schooljeugd</i> [Groningen]
1799	Maatschappij tot Nut van 't Algemeen [van Varik, Gerrit]	<i>Rudimenta, of gronden der Nederduitsche spraake</i> [Leiden, Deventer & Utrecht]
1799	Weiland, Petrus	<i>Nederduitsch taalkundig woordenboek</i> (Introduction) [Amsterdam]
1804	Siegenbeek, Matthijs	<i>Verhandeling over de Nederduitsche spelling, ter bevordering van eenparigheid in dezelve</i> [Amsterdam]
1805	Weiland, Petrus	<i>Nederduitsche spraakkunst</i> [Amsterdam]

The overview of normative publications in Table 11 is based on a number of previous studies on eighteenth-century language norms, most notably van de Bilt (2009), Vosters et al. (2010), Rutten (2011) and Simons & Rutten (2014).

6 Procedure and methodological remarks

6.1 Systematic methodological procedure for linguistic analyses

Investigating language norms and language usage in late eighteenth- and early nineteenth-century Dutch, the following Chapters 5–12 present eight corpus-based case studies of five orthographic and three morphosyntactic features, all of which can be considered relevant linguistic issues in the context of the Dutch *schrijftaalregeling*. The official regulations in Siegenbeek's (1804) orthography and Weiland's (1805) grammar, in fact, serve as the starting points for the case studies in this dissertation. Ultimately striving for a sophisticated assessment of the effectiveness of these concrete language policy measures, each linguistic variable will be investigated systematically by following the methodological procedure described below.

In the first part of each chapter, the linguistic variables under investigation will be introduced by providing a summary of the normative discussion by either Siegenbeek (1804) or Weiland (1805). Moreover, this section also introduces the relevant variants that were mentioned and possibly evaluated by Siegenbeek and Weiland. Did they take into account language variation and acknowledge the existence of alternative forms? If so, how explicitly (or implicitly) do they prescribe the officialised variant(s), and on which principles were these choices grounded? Given the high complexity of morphosyntactic variables, the corresponding

chapters briefly outline the variable and its history more generally, before moving on to the discussion of Weiland's (1805) preferences and choices.

In the second part, the officialised norms by Siegenbeek (1804) and Weiland (1805) are placed in the wider context of the eighteenth-century normative tradition. By providing an outline of the preceding discussions and developments in metalinguistic discourse, making use of the *Normative Corpus of the Northern Netherlands* (cf. Section 5), a more fine-grained assessment of Siegenbeek's and Weiland's choices is possible. It will be examined how eighteenth-century variation was represented and commented on, also in comparison to Siegenbeek and Weiland, and which alternative forms were mentioned. Furthermore, this section also discusses whether the officialised choices by Siegenbeek and Weiland were innovative and even radical, or rather grounded on existing preferences, i.e. continuing the eighteenth-century normative tradition.

In the third part, I provide an overview of previous research on the linguistic variable under investigation, establishing links and identifying gaps with regard to the research objectives of the present dissertation.

After having outlined the investigated feature by taking into consideration the corresponding discussions in Siegenbeek or Weiland, as well as in eighteenth-century metalinguistic discourse, the focus shifts to the empirical investigation of actual language usage. Based on the multi-genre *Going Dutch Corpus*, each of the eight linguistic variables will be investigated quantitatively, taking into account the variational dimensions of the corpus (i.e. genre, time, space, gender, cf. Section 3) and, whenever relevant, internal factors potentially conditioning the use and distribution of variants.

In the final section of each chapter, the findings drawn from the corpus analyses will be discussed with reference to the official prescriptions of 1804/1805 as well as the eighteenth-century metalinguistic discourse, aiming to assess the normative influence on variation and change in the use of linguistic variables. By systematically following this methodological procedure in each case study, I seek to measure and determine the overall effectiveness of the national *schrijftaalregeling*.

6.2 Final remarks on statistical methods

Throughout the corpus-linguistic analyses in this dissertation (Chapters 5–12), I will make use of descriptive statistics, presenting quantitative results in the form of cross tabulation and column graphs. With regard to the multidimensionality of the *Going Dutch Corpus*, which takes into account genre, time, space and gender as independent variables, monofactorial statistical tests, such as chi-square tests, t-tests or correlation tests, would hardly do justice to the complex corpus design and the variety of external factors under investigation.

On the other hand, multifactorial approaches, making use of more advanced mixed-effect regression models (e.g. conducted with software tools like R) offer intriguing possibilities, such as the representation of systematic interaction patterns between independent variables. In recent years, these more advanced

statistical models have primarily been applied to modern (socio)linguistic data. In the field of historical sociolinguistics, however, the employability of these methods is still being explored. While quantitative (present-day) sociolinguistic studies commonly rely on perfectly balanced data sets, even well-balanced historical-sociolinguistic corpora have natural inconsistencies, which, in turn, present new challenges for statistical methods. Only in the last couple of years, historical sociolinguists have started to further explore whether and in what ways statistical methods and tests can be applied to historical data. Balancing the advantages and disadvantages of new quantitative methods, a number of corpus-linguistic case studies demonstrate the possibilities for future research (e.g. Tagliamonte & Baayen 2012; Mannila et al. 2013; Krug & Schlüter 2013; Nevalainen & Raumolin-Brunberg 2016: ch. 9). At this point, though, methods of statistical data analysis are yet to be established as obligatory parts of historical-sociolinguistic research.

From a more practical point of view, the amount of time that needs to be invested in statistical data analysis would have considerably reduced the amount and variety of linguistic variables investigated in this dissertation. However, in order to assess the effects of language policy on patterns of actual language usage, a substantial number of both orthographic and morphosyntactic case studies appeared to be essential for a sophisticated assessment. Therefore, I have chosen a wider range of linguistic variables over a statistically more advanced method. In fact, I would argue that a thoroughly designed and compiled corpus, aiming at a well-balanced representation of authentic language use, as well as a systematic procedure of both quantitative and qualitative data analysis, can, to a large extent, counterbalance the lack of a statistically advanced method.

CHAPTER 5

Orthographic variables (1)

Syllable-final /xt/

Following the systematic methodological approach introduced in Section 6 of Chapter 4, the present chapter contains the first of eight linguistic case studies with the *Going Dutch Corpus*, which together form the empirical heart of this dissertation. First of all, Chapters 5–9 focus on five orthographic variables, covering both consonantal and vocalic features, in order to assess the effectiveness of Siegenbeek's (1804) spelling prescriptions. Furthermore, three morphosyntactic variables will be analysed in Chapters 10–12, examining the influence of Weiland's (1805) grammar.

1 Discussion in Siegenbeek (1804)

In his official orthography, Siegenbeek (1804: 160–165) elaborately commented on the differences between the consonants *g* and *ch*, addressing the orthographic representation of the voiceless velar fricative /x/ in syllable-final position as well as before *t* in the /xt/ cluster. Although both positions are interesting and suitable for an examination of the possible normative influence on language practice, only the latter, i.e. the consonant cluster /xt/, will be addressed in this chapter. This choice serves several purposes. First of all, it is my aim to examine a diverse selection of orthographic variables in this chapter and the following. Therefore, I prefer to focus on syllable-final /xt/ rather than to study two very closely related cases. Moreover, it is the orthographic representation of /xt/ that has often been regarded (and debated) as a typical Siegenbeek feature rather than syllable-final /x/.

Siegenbeek introduced a division into two categories of words containing the cluster /xt/, which had to be spelled with either <gt> or <cht>, mainly depending on etymological considerations. To begin with, Siegenbeek mentioned the fundamental phonetic differences of the two *verwantschapte Medeklinkers* 'related consonants' *g* and *ch*, the former of which being *zacht* 'soft' (i.e. voiced) and the latter being *scherp* 'sharp' (i.e. voiceless). Before proposing and illustrating his own spelling norm, Siegenbeek referred to *de Ouden* 'the Old', by which he meant Middle Dutch writers of the thirteenth and fourteenth centuries, who had supposedly based their spelling on pronunciation only. At the end of a syllable and more specifically in syllables ending in *-t*, 'the Old' had suggested <ch> and <cht>, respectively, due to final devoicing (*Auslautverhärtung*)³⁸ (Siegenbeek 1804: 162):

³⁸ In Dutch, voiced obstruents generally do not occur in syllable-final position. Consequently, phonologically voiced obstruents are devoiced in syllable-final position (e.g.

En wat het gebruik der Ouden, die zich op het einde der lettergreep altijd van *ch* bedienden, aangaat, hetzelfde steunt zeker op een' goeden grond, als voldoende de *ch*, vooral wanneer de lettergreep op *t* eindigt, beter aan den natuurlijken toon der uitspraak dan de zachtere *g*, gelijk hetzelfde ook door den voorgang der Grieken, bij wie wij eene soortgelijke verwisseling der γ en χ zagen plaats grijpen, gewettigd wordt.

'And with regard to the use of the Old, who always used *ch* at the end of a syllable, this is certainly grounded on good reasons, as the *ch*, especially when the syllable ends in *t*, meets the natural sound of the pronunciation better than the softer *g*. In the same way, this is also justified by the example of the Greek, where we saw a similar change of γ and χ taking place.'

Although Siegenbeek did not deny the fundamental necessity of a phonetic spelling and, in fact, established the rule *Schrijf, zoo als gij spreekt* 'Write as you speak' as his first principle of the Dutch spelling, he criticised 'the old' Middle Dutch writers for not taking into account the origin of words at all. For this reason, he referred to his second spelling principle, i.e. *Men volge in het schrijven de naaste en zekerere Afleiding* 'One should write according to the closest and certain derivation' (Siegenbeek 1804: 28), taking into the account the supposedly historical form.

The conflict between the two principles of pronunciation and derivation is also central to the orthographic variable studied in this chapter. For the representation of the /xt/ cluster³⁹, Siegenbeek (1804: 163) prescribed the following rule:

Op grond hiervan achten wij het best, overeenkomstig het tegenwoordige gebruik, dezen regel aan te nemen, dat in woorden, welke in het meervoud de *g* hebben, of dezelve om hunnen duidelijken oorsprong vereischen, de *g* altijd hare plaats moet behouden. Men schrijve dus [...] *magt* van *mogen*, *klagt* van *klagen*, *bragt* van *brenge*, *pligt* van *plegen*, *vlugt* van *vliegen*, en desgelijks in de verdere hiertoe behoorende woorden, welker opgave wij voor onze Woordenlijst besparen. Doch, waar geene der gemelde redenen tot de plaatsing der *g* voorhanden is, gebruike men, ter meerdere voldoening aan de uitspraak, bestendig de *ch*, bij de sluiting op eenen harden medeklinker [...], en schrijve dus [...] *lucht*, *achten*, *pracht*, *kocht* van *koopen*, *zacht*, en zoo in meer anderen.

'Due to this we consider it the best way, in accordance with contemporary usage, to adopt this rule, according to which the *g* must always maintain its place in words which have the *g* in the plural, or require the *g* due to their clear origin. One thus writes [...] *magt* from *mogen*, *klagt* from *klagen*, *bragt* from *brenge*, *pligt* from *plegen*, *vlugt* from *vliegen*, and likewise in further words belonging to this type, whose listing we save for the word list. But, where none of the mentioned reasons for the

/ʏ/ becoming /x/). This process is commonly referred to as final devoicing (or *Auslautverhärtung*) (e.g. Booij 1995).

³⁹ The omissions in this quote deal with a different variable, which Siegenbeek addresses in the same paragraph as the /xt/ cluster, viz. the orthographic representation of syllable-final /x/ as either <g> or <ch> as in *oog* and *zich*. As mentioned before, this variable will not be discussed here.

placing of the *g* is present, one must, in order to better meet the pronunciation, consistently use the *ch* when ending in a hard (voiceless) consonant [...], and one thus writes [...] *lucht*, *achten*, *pracht*, *kocht* from *koopen*, *zucht*, and in more others.’

In other words, Siegenbeek’s prescriptive norm for the orthographic representation of /xt/ can be regarded as a hierarchy of two rules. First, the derivation or (clear) etymological origin of a word must be taken into account. Hence, the letter *g* in a root word such as *klagen* ‘to complain’ or *brenge* ‘to bring’ has to be maintained in derived words containing the syllable-final /xt/ cluster, for example *klagt* ‘complaint’ or *bragt* ‘brought’. These words will henceforth be referred to as the category of *gt*-words. However, in cases in which Siegenbeek’s derivation principle does not demand the maintenance of the letter *g*, /xt/ should be consistently spelled <cht> for phonetic reasons (i.e. final devoicing), for example in *lucht* ‘air’ or *kocht* ‘bought’ (< *koopen* ‘to buy’). They form the category of *cht*-words.

While the categorisation into *cht*- and *gt*-words is seemingly clear-cut, Siegenbeek (1804: 165) did acknowledge a few ambiguous cases, in which the etymology of a word may not be entirely transparent to the language user. For instance, *kracht* ‘strength, force’ was sometimes interpreted as a derivation of the verb *krijgen* ‘to get’:

Het laatstgenoemde woord, *kracht*, wordt door eenigen ook *kragt*, geschreven, als afkomstig van het werkwoord *krijgen*. Hoewel deze afleiding zeker niet onwaarschijnlijk is, heeft zij echter, mijns achtens, niet die blijkbaarheid, dat men, uit dien hoofd, van de gebruikelijkste en aan de uitspraak meest voldoende schrijfwijze met *ch* zou hebben af te gaan.

‘The latter word, *kracht*, is also spelled *kragt* by some, deriving from the verb *krijgen*. Although this derivation is certainly not unlikely, I think that it does not have that obviousness to abandon the spelling with *ch* for that reason, which is most commonly used and most closely meets the pronunciation.’

Therefore, Siegenbeek generally suggested that only obvious cases of derivation and etymology should serve as a guideline for spelling (1804: 164):

Doch wij zijn van oordeel, dat [...] de spelling zich niet met de uiterste naauwkeurigheid naar de afleiding behoeft te schikken, daar, gelijk, wij in onze eerste Afdeeling gezien hebben, alleen de duidelijke en naaste afleiding ten rigtsnoer der spelling kan verstrekken.

‘But we are of the opinion that [...] the spelling does not have to go along with the derivation to the most extreme accuracy, because, as we have seen in the first part, only the obvious and close derivation can serve as a guideline for the spelling.’

An extensive word list (*Woordenlijst ter aanwijzing van de spelling*) with more cases of *cht*- and *gt*-words is provided in the appendix of his orthography.

In addition to <cht> and <gt> as the two prescribed variants, Siegenbeek also mentioned a third variant, viz. <ght>, which had been used by seventeenth-century literary authors like P.C. Hooft and Joost van den Vondel. Initially, the

insertion of *b* was an attempt to sharpen the *g* in syllable-final position, for instance in *maght* ‘might, power’. Siegenbeek, however, commented that this historical variant had already been “te regt verworpen” (1804: 162), i.e. rightly rejected by his eighteenth-century predecessors as “ongepast en met den aard der tale geenszins overeenkomende”, i.e. ‘inappropriate and by no means in accordance with the nature of the language’ (ibid.).

Moreover, Siegenbeek discussed the much-debated aspect of homonymy in his paragraph on *gelijkluidende, doch in beteekenis verschillende woorden* ‘homonymous but semantically different words’, where he also addressed two cases of /xt/ words, viz. *agt* ‘eight’ versus *acht* as in *acht geven* ‘take care’, and *jagt* ‘hunt(ing)’ versus *jacht* ‘yacht’ (1804: 56). At yet another point, he revisited the homonymy question (1804: 197-198):

Verder verdient het onderscheid onze opmerking, ‘t welk, sedert eenigen tijd, door sommigen is ingevoerd in de spelling *agt* (*octo*) en *acht*, in *achtgeven*, door het eerste met eene *g*, het laatste met eene *ch* te schrijven. Dat ook dit onderscheid louter willekeurig is, valt van zelf in het oog. Immers doet de uitspraak ons geen het minste verschil tusschen beide deze woorden opmerken. En op de afleiding lettende, vinden wij althans gene reden, om het talwoord, ‘t welk men van het Latijnsche *octo* zou kunnen afleiden, met eene *g*, het andere *acht* met eene *ch* te schrijven.

‘Furthermore, the distinction deserves our comment, which, since some time ago, has been introduced by some in the spelling *agt* (*octo*) and *acht*, in *achtgeven*, by spelling the former with a *g*, the latter with a *ch*. It is obvious that this distinction is also sheerly arbitrary. After all, the pronunciation does not indicate the smallest difference between these two words. And by regarding the derivation, we find no reason to write the numeral, which one might derive from the Latin *octo*, with a *g*, and the other *acht* with a *ch*.’

Siegenbeek thus rejected earlier proposals (cf. Section 2) to apply different spellings in order to distinguish homonymous /x/ words as much as possible, i.e. irrespective of their pronunciation and derivation, which he criticised for being arbitrary. Interestingly, in his word list, Siegenbeek (1804: 199) did differentiate between homonymous *licht* ‘light, bright’ and *ligt* ‘light, not heavy’, referring to the principle of usage:

Andere onderscheidingen zijn in het schrijfgebruik zoo algemeen aangenomen, dat het raadzaamst is dezelve te volgen. [...] Hoewel deze onderscheidingen in den grond louter willekeurig zijn [...] kan men echter en moet men het tegenwoordige gebruik, ‘t welk dezelve heeft ingevoerd, door derzelver inachtneming te wille zijn.

‘Other distinctions are so commonly accepted in the written usage, that it is most well-advised to follow them. [...] Although these distinctions are at bottom sheerly arbitrary [...], one can and must take account of the contemporary usage, which has introduced these distinctions.’

Siegenbeek also distinguished *dicht* ‘poetry’ from *digt* ‘closed; dense’ as well as *nicht* ‘child’ from (*ge*)*wigt* ‘weight’. In these cases, the etymology of the words most

probably resulted in one form being spelled with <gt>, deriving from verbs with *g* (i.e. *wegen* ‘to weigh’ and *dij(g)en* ‘to thrive’). Nevertheless, Siegenbeek’s choice to exclude a very small group of words from his general homonymy rule (i.e. not to distinguish homonyms by spelling) is rather arbitrary. In fact, it does not seem entirely reasonable to refer to the principle of usage in the case of *licht/ligt*, whereas a more frequently used pair of homonyms such as *acht/acht* was not distinguished by spelling.

In sum, Siegenbeek’s official rule for the purely orthographic *cht/gt* issue is remarkably complex, grounded on phonological (<cht>) and etymological (<gt>) considerations, and further touching upon the aspects of homonymy and common usage.

2 Eighteenth-century normative discussion

In eighteenth-century metalinguistic discourse, many grammarians before Siegenbeek commented on the *g/ch* spelling question, also addressing the orthographic representation of the /xt/ cluster. Generally speaking, three recurring main arguments can be identified in the normative tradition.

First of all, the difference in pronunciation between *scherp* ‘sharp’ (voiceless) *ch* and *zacht* ‘soft’ (voiced) *g* is pointed out as a fundamental principle. The main argument is that syllable-final /x/ and more specifically /xt/ should be spelled <ch> and <cht>, respectively, due to final devoicing. The central question discussed in the normative works is whether the voiceless nature of /x/ in syllable-final position and the cluster before *t* should be reflected by spelling (i.e. <cht>) or not (i.e. <gt>).

The second main argument deals with what grammarians variably refer to as the *oorsprong*, *afleiding* or *afkomst* of a word, i.e. taking into account its derivation and/or etymology. As briefly mentioned in Section 1, pronunciation and derivation have often been regarded as two conflicting fundamental spelling principles, widely discussed in eighteenth-century normative works and in Siegenbeek’s (1804) orthography.

The third frequently occurring argument concerns the spelling of homonyms with <cht> and/or <gt>, and the central question whether they should be orthographically distinguished in order to make their semantic differences clear and visible to the language user. Although these three arguments are repeatedly discussed throughout the eighteenth century, grammarians clearly had different opinions and preferences on this spelling issue.

Comparing eighteenth-century comments on the influence of derivation, particularly in relation to the phonological principle, striking differences become apparent. The importance of reflecting a word’s origin by spelling was explicitly highlighted by Sewel (1708), van Belle (1748), van der Palm (1769), Stijl & van Bolhuis (1776), Kluit (1777), van Bolhuis (1793) and the *Rudimenta* (1799). The earliest eighteenth-century attestation of the idea to maintain the <g> in syllable-

final position and before *t* due to the derivation of a word is found in Sewel (1708: 88):

G wil ik liever gebruyken in 't woord *magt* als *ch*, als komende van *mag*, welks onbepaalde wyze [Infinitivus] is *moogen*, en daarom schryf ik *mogt* en niet *mocht*; om dezelfde reden schryf ik *genigtig* en niet *genichtig*, zynde afkomstig van 't woord *weegen*; en daarom wil ik ook liefst spellen *gebragt*, *gezig*, *voorzigtig*, vermids de G in de woorden *brenge*, ik *zag*, wy *zagen*, niet kan achtergelaaten worden.

'I will rather use G in the word *magt* than *ch*, deriving from *mag*, whose infinitive is *moogen*, and therefore I write *mogt* and not *mocht*. For the same reason I write *genigtig* and not *genichtig*, deriving from the word *weegen*. And therefore I also prefer to spell *gebragt*, *gezig*, *voorzigtig*, as the G in the words *brenge*, ik *zag*, wy *zagen*, cannot be left behind.'

Van Belle (1748: 11) more critically pointed at the misspelling of those words, caused by neglecting their characteristics:

Nóg erger is 't wanneer men spelt hy *bracht*, men *mocht*, het *deucht* niet [...], van de Werkwoorden *brenge*, *mogen*, *deuge*, enz: want dusdoende raakt men de eigenschap der woorden zoverre uit het gezigt kwyf, dat nóch Vreemdeling noch Neederlander, die de gronden onzer Taale zoekt magtig te worden, dezelve eenigins reegelmatig kan nagaan.

'It is even worse when one spells hy *bracht*, men *mocht*, het *deucht* niet [...], from the verbs *brenge*, *mogen*, *deuge*, etc, because in doing so one loses the quality of the words to the extent that neither a foreigner nor a Dutchman, who seeks to command the grounds of our language, can reasonably follow it with regularity.'

Stijl & van Bolhuis (1776: 56), and also van Bolhuis (1793), went one step further by presenting a hierarchy of rules, taking into account both derivation (first rule) and pronunciation (second rule) as two conflicting principles:

Hoe zal men spellen *magt* of *macht*? De uitspraak zou het laatste eischen naar den 2 regel; doch de afleiding van *mogen* strijdt er tegen; daarom verkiezen wij *magt* naar den 1 regel. Die reden van afleiding is zoo duidelijk niet in *kracht*, des mag hier *ch* boven de *g* naar den 2 regel gelden.

'How should one spell *magt* or *macht*? The pronunciation would require the latter according to the second rule, but the derivation of *mogen* contradicts with it. This is why we prefer *magt* according to the first rule. That rule of derivation is not that obvious in *kracht*, which is why *ch* rather than *g* may apply here according to the second rule.'

A similar argumentation can be found in the *Rudimenta* (1799: 65), according to which the phonological principle is primary, although exceptions due to derivation have to be acknowledged:

in het algemeen gebruikt men CH in alle woorden, die scherp uitgesproken worden b. v. *gedacht*, *kracht*, *nacht*, *wacht* enz: mits, dat de afleiding daar niet tegen

strijde: want al zijn sommige woorden dan eens scherp van uitspraak, men moet de afleiding niet verwaarloozen b. v. *magt* met *g*, want dat wordt afgeleid van *mogen*; *gebragt* van *brenge*; en meer anderen.

‘Generally, one uses CH in all words which are pronounced sharply, e.g. *gedacht*, *kracht*, *nacht*, *wacht* etc., if the derivation does not contradict it. Because even though some words are sharp in pronunciation, one must not neglect the derivation, e.g. *magt* with *g*, because that is derived from *mogen*, *gebragt* from *brenge*, and more others.’

This hierarchical approach is, in fact, similar to Siegenbeek’s (1804) prescription (Section 1). Undoubtedly, another major source of inspiration for Siegenbeek must have been Kluit (1777), who elaborately discussed the importance of derivation with regard to <gt> and <cht>. Three decades before Siegenbeek, Kluit (1777: 23-24) already criticised the spelling principles of *de Ouden* ‘the Old’ (i.e. Middle Dutch writers) for neglecting the origin of words and rather grounding their spelling on pronunciation only, which had led to syllable-final <ch> and <cht> as the only variant for /xt/:

De oorsprong des woords namelijk deed bij de Ouden er niets toe [...] maar zagen alleen daarna, of de G een sluitletter wierd: zoo ja, dan veranderde hy in CH, en men schreef *vlucht*, *mocht*, *noch*, niet omdat die woorden in zich een CH begrepen (dit doen zij niet; want *vlucht* komt van *vlieGen*, *mocht* van *moGen*); maar omdat de G te zacht was, om een Lettergreep te eindigen, of om op een harde T te stuiten.

‘The origin of the word, in fact, did not matter to the Old [...], but they only took into account whether the G became a closing letter. If so, then it changed to CH, and one wrote *vlucht*, *mocht*, *noch*, not because those words have a CH in them (which they do not have, as *vlucht* derives from *vlieGen*, *mocht* from *moGen*), but because the G was too soft to close a syllable, or to precede a hard T.’

Several other grammarians took into account the impact of derivation as well, even though they presented it more implicitly as exceptions rather than a strict rule. Wester (1799: 29-31), for instance, only mentioned a few examples of <gt> words such as *vlugt*, *bragt* or *klagt*, justified by their “naauwe betrekking” ‘close reference’ to the words *vliegen*, *brenge* and *klaagen*.

The aspect of homonymy is discussed in most eighteenth-century normative works addressing the *cht/gt* issue. For many authors, most notably Sewel (1708), van Belle (1748/1755), van Rhyn (1758), van der Palm (1769), Zeydelaar (1774), van Bolhuis (1793), Wester (1799) and the *Rudimenta* (1799), the two variants <gt> and <cht> were regarded as useful in order to make a visible difference between homonyms. Most publications provided lists of homonyms, illustrating their semantic differences and suggested spellings, like Zeydelaar (1774: 67-69):

Wijl ondertusschen de *ch* en de *g* zoo veel overeenkomst met malkander hebben, zoo verwekt het veeltijd een merkelyk onderscheid in de betekenis der woorden die met de eene of met de andere Letter geschreeven zijn, en om den Leerling van

dat verschil niet onkundig te laten, zal ik de woorden, met hunne onderscheidene betekenissen opgeeven.

DICHT, vaers, maatgezag. Dat is een Bruilofs-*Dicht*.

DIGT, geslooten. De deur is *dig*t. *Digt*, niet ver. Hier *dig*t bij.

LICHT is het tegenovergestelde van duisternis.

LIGT is 't tegengestelde van zwaar of van moeiëlijk.

ACHT komt van 't werkwoord *achten*. Ik *acht* hem niet veel.

AGT is in getale een meer dan *zeven* [...]

WICHT zegt men van een klein kind.

WIGT is gewigt, zwaarte. *Wigtig* is iets dat zijn behoorlijke zwaarte of waarde heeft.

'While in the meantime the *ch* and the *g* have so much in common, it often creates a distinct difference in the meaning of the words which are spelled with either the one or the other letter. And in order not to keep the student unaware of that difference, I will list the word with their different meanings.

DICHT, verse, metrical song. This is a wedding song (*Dicht*).

DIGT, closed. The door is *dig*t 'closed'. *Digt*, not far. Close (*dig*t) by here.

LICHT is the opposite of darkness.

LIGT is the opposite of heavy or of difficult.

ACHT comes from the verb *achten* 'respect'. I do not *acht* 'respect' him much.

AGT, in numbers, is one more than *zeven* 'seven' [...]

WICHT, one says for a little child.

WIGT is weight, heaviness. *Wigtig* 'important' is something that has a considerable weight or value.'

Similarly, Wester (1799: 29-31) highlighted the benefits of differently spelled homonyms:

Eindelyk kan men, door eene oordeelkundige plaatsing van *g* of *ch*, veele woorden, die, of geheel, of byna gelykluidende zyn, doch zeer verschillende in betekenis, duidelyk van elkanderen onderscheiden.

'Finally, through a judicious placing of *g* or *ch*, one can clearly distinguish many words which are either completely or nearly homophonous, but very different in meaning.'

In contrast, Kluit (1777: 24-25) strictly rejected the idea of distinguishing homonyms by using different spellings, which he criticised as being arbitrary and thus uncertain:

Zoo wordt hiermede ten eenemal de bodem ingeslagen aan dat gewaande en naderhand verzonnen onderscheid, om klaarheids halve de woorden *licht* (*levis* en *lux*); *wicht* (*pondus* en *infans*) [...] door bijzondere spelling zoo te onderscheiden, dat in 't ene geval een G, in 't andere geval een CH, gebruikt worde. Want, om niet te reppen van 't klein getal der zoodanigen, de *Analogie* der tale verbiedt dit. Deze zegt, dat ook hier moet gelden: *Similum similis ratio*. Daarbij is zulk ene onderscheiding niet alleen willekeurig, en dus gansch onzeker; maar wordt ook zelf bij velen niet recht gevat, ja gansch verkeerd gebruikt; en blijft veeltijds ook duister, omdat de oorsprong des woords ons dikwils onbekend is.

Wat mij belangt, zoolang het achtbaarste gebruik hieromtrent niet zekers besluit, zal ik mij bij de ouden houden, die tusschen *licht*, en *licht*; *wicht* en *wicht*; *lucht* en *lucht* enz. geen onderscheid gemaakt hebben, en nochtans genoeg te verstaan zijn.

‘Hereby, that alleged and afterwards invented distinction to differentiate the words *licht* (*levis* and *lux*), *wicht* (*pondus* and *infans*) for the sake of clarity [...] by specific spellings, using a G in the one case, and CH in the other case, is completely abolished. Because, without mentioning the small number of such words, the *analogy* of the language prohibits this. It says, which must also apply here: *Similum similis ratio*. Such a distinction is not only arbitrary, and thus entirely uncertain, but it is also wrongly understood by many, even used entirely incorrectly, and mostly remains vague, because the origin of the word is often unknown to us.

As for me, as long as the most respectable use concerning this matter does not clearly decide, I will adhere to the old, who made no distinction between *licht*, and *licht*, *wicht* and *wicht*, *lucht* and *lucht*, and are still sufficiently comprehensible.’

Kluit’s general view on homonymy was later taken up by Siegenbeek (1804), although, unlike Kluit, he did distinguish between *wicht* and *wigt*, but not between *licht* and *licht* (due to the principle of usage).

A radical exception in eighteenth-century metalinguistic discourse is Hakvoord (1746: 36-37), who argued that syllable-final /x/ (and /xt/) should never be spelled <ch(t)> but always <g(t)>:

Om kort en gemakkelijk te gaan, spellen wy in ‘t einde der Lettergrepen, nooit met ch; maar gebruiken daar toe g

‘To be concise and easy, we never spell with ch at the end of syllables, but use g for this.’

Additionally, he provided a list of twenty /xt/ words spelled with <gt>, including *Dagt*, *Wagt*, *Vlugt*, *Magt*, *Klagt* and also *Ligt* (without taking into account homonymy). Two years later, van Belle (1748: 10-11) took a similar, though less radical position as Hakvoord by arguing that <gt> was the only spelling variant needed:

Sommigen spellen ik *zach*, ik *dacht*, van de Werkwoorden *zien*, *denken*, enz: met de CH, in plaats van de G, maar zonder nood: want wie zal iemand, die zelf gelds genoeg hebbende zonder noodzaaklykheid, by eenen anderen daarom te leen vraagt, niet voor dwaas houden? En moet men ze dan ook voor zodaanig niet achten, die, aan de G genoeg hebbende, zonder nood nógtans de CH te leen gebruiken; terwyl dezelve nóg in ‘t Neederduits aan het einde der woorden nooit behoeft óf behoort gebruikt te worden, dan om daardoor eenigsins de onderscheidene beteekenissen van sommige woorden te vertoonen.

‘Some spell ik *zach*, ik *dacht*, from the verbs *zien*, *denken*, etc. with the CH, in place of the G, but needlessly. Because who would not make a fool of someone, who has enough money himself but asks somebody else for a loan without necessity? And do you not need to consider those (words) as such, in which the G is sufficient but which still needlessly use the CH as a loan, although it never needs

nor ought to be used at the end of a word in Dutch, except for showing the different meanings of some words to some extent.’

In contrast to Hakvoord, van Belle thus approved <cht> in some cases, but only in order to highlight the semantic differences of specific homonyms like *Licht* (*dat schynt*) and *Ligt* (*in ‘t weegen*).

Compared to the approaches of his contemporaries, Moonen’s (1706) stance is rather exceptional, as it hardly fits any of the other normative tendencies. In fact, he suggested the use of three (co-existing) variants, viz. <cht>, <gt> and even the widely rejected <ght>, which he did not consider as redundant but necessary. Moonen generally based his choices on phonetic differences, which were orthographically represented by one of the three variants. Even in syllable-final position, Moonen⁴⁰ actually distinguished a sharper (more voiceless) *ch* and a softer (more voiced) *g*, the latter of which he compared to the French pronunciation *gue*:

Deeze Ch, van de Grieken niet alleen door χ , maer eertyts ook door twee merken KH in KHRONOS, KHARON uitgedrukt, en scherper, dan de G, als blykt uit het onderscheit tusschen *lach* en *lag*, *echt*, *matrimonium*, en *egt*, *occat* [...], wordt altyt in het midden en einde der lettergreepe gebruikt, achter de Klinkers in [...] *Acht*, *zacht*, *recht*, *knecht*, *licht*, *sticht*, *klucht*, *lucht*, *zucht* [...] (1706: 5)

Wanneer zy [de G] eene Lettergreep na eenen Klinker besluit, of eene Lettergreep begint, die op eene G volgt, schynt zy een geluit te hebben, dat, zachter, dan het voorgaende, van de Franschen door *gue* wordt uitgedrukt; gelyk in *vlag*, *beg*, *rug* [...] (1706: 7)

‘This Ch, which is not only expressed as χ by the Greek, but formerly also as the two letters KH in KHRONOS, KHARDON, and which is sharper than the G, as appears from the distinction between *lach* and *lag*, *echt*, *matrimonium*, and *egt*, *occat* [...], is always used in the middle and end of the syllable, after the vowels in [...] *Acht*, *zacht*, *recht*, *knecht*, *licht*, *sticht*, *klucht*, *lucht*, *zucht* [...]

When it [the G] closes a syllable after a vowel, or starts a syllable which follows after a G, it appears to have a sound that is softer than the previous, and expressed by the French as *gue*, as in *vlag*, *beg*, *rug* [...]

According to Moonen (1706: 8-9), the third variant <ght> also serves the purpose to differentiate between pronunciations (‘brenge den Woorden in den uitspraek hulp toe’), mainly in (near-) homonymous nouns such as *weg* ‘way’ and *wegh* ‘white bread’⁴¹:

⁴⁰ Moonen was born in Zwolle and later worked as minister in Deventer, both in the eastern province of Overijssel. At least historically, there was less final devoicing in some northeastern dialects of Dutch, with voiced fricatives occurring even in auslaut position (cf. also De Wulf et al. 2005: 252), which probably influenced Moonen’s choice of spelling variants.

⁴¹ The exact phonetic nature of Moonen’s <ght> spelling, however, remains blurry, especially in contrast to words spelled with <gt>. Gledhill (1973: 107-108) critically discusses Moonen’s approach, pointing out that “[h]is fundamental system is to use ‘gh’ in

De H wordt van zommigen achter de G in het einde des woorts, daer zy van outs plagh geschreeven te worden, verworpen onder het voorgeeven van overtolligheit. Maer zy is noodigh, en brengt den Woorden in de uitspraekke hulp toe, zal men den *wegh*, dien man reist, van eene *weg* (wittebroot) den *dagh*, dien wy beleeven, van eene *dag*, dat een wapentuig en een tou te scheepe betekent, [...] behoorlyk onderscheiden.

‘The H after the G at the end of the word, where it always used to be written, is neglected by some with reference to reduncancy. But it is necessary and aids the pronunciation of the words, in order to properly distinguish the *wegh*, which you travel, from a *weg* (white bread), the *dagh*, which we experience, from a *dag*, which means a weapon and a marine rope [...].’

While Moonen discussed a variety of rules, a logical coherence between these rules and their actual application in the (gender-related) word lists is hard to find. Listing nouns ending in *-t*, for instance *zicht*, *vlucht* alongside *klagt* and *draght*, Moonen obviously did not take into account the possible effect of derivation. The only more or less explicitly mentioned case of derivation is <ght> in *braght* and *gebraght* as derived from *brenghen* (1706: 9), as mentioned in the quote above. At the same time, Moonen prescribed *maght* (1706: 9) alongside *magt* (1706: 76), and *Toght* (1706: 75) alongside *Toght* (1706: 90), which strongly implies that his system was anything but consistent.

Summing up the main preferences in eighteenth-century metalinguistic discourse, it appears that most grammarians already acknowledged <cht> and <gt> as co-occurring variants. Consensus about the exact use of these forms was not yet reached, though. It was Siegenbeek (1804) who officialised a fairly complex rule for this spelling issue, categorising words with syllable-final /xt/ into phonologically motivated *cht*-words and etymologically motivated *gt*-words.

3 Previous research

Despite the controversial debate about the orthographic representation of /xt/ as either <cht> or <gt> throughout the eighteenth- and nineteenth centuries, this feature has been hardly been addressed to date, particularly in the context of Siegenbeek’s national spelling prescriptions. Apart from general attestations as one of Siegenbeek’s typical features, for instance as part of Mathijssen’s (1988: 134-135) comparison between Siegenbeek and his competitor Willem Bilderdijk, this orthographic variable has not been investigated from an empirical, let alone historical-sociolinguistic perspective yet.

final position only (never initially), to indicate the hard pronunciation of /ch/. But his rules are very complex, and seemingly arbitrary. [...] In fact Moonen is not very explicit on why he uses ‘gh’ at all”. Van de Bilt (2009: 168) argues that the fear of homonyms in these cases was a crucial point of departure for Moonen. With regard to the focus of this chapter on orthography, this issue will not be discussed in more detail here, though.

However, a comprehensive overview of the orthographic representation of /x/ before *-t* is provided by Gledhill (1973: 117–148) in his PhD thesis on Dutch consonantal spelling in the history of Dutch. In a detailed section on the variants <cht>, <gt> and <ght>, he outlines and critically comments on the main developments from the Middle Dutch period until the twentieth century. While looking at both Northern and Southern normative traditions as well as usage in texts (mainly) written by grammarians and literary figures, empirical evidence of actual language practices remains sparse. On the basis of the *Going Dutch Corpus*, Section 4 will shed more light on usage patterns in the Northern Netherlands, focusing on the late eighteenth- and early nineteenth century.

4 Corpus analysis

4.1 Method

In the following corpus analysis of the orthographic representation of the consonant cluster /xt/, three variants are considered, viz. <cht>, <gt> and <ght>, all of which occur in the *Going Dutch Corpus*. Both <cht> and <gt> were prescribed as national variants in Siegenbeek's (1804) orthography, introducing the categories of *cht*- and *gt*-words. The third variant <ght> was also mentioned by Siegenbeek, but disapproved as an 'inappropriate' and 'rightly rejected' form.

Following Siegenbeek's distinction, the corpus results in this section will be presented as two separate categories for *cht*-words (i.e. prescribed <cht>) and *gt*-words (i.e. prescribed <gt>), respectively. For the sake of clarity and comparability, this type of presentation will be applied for both diachronic cross-sections of the corpus, although it has to be kept in mind that there was no (officialised) distinction into *cht*- and *gt*-words before 1804.

For the analysis of this orthographic variable, the ten most frequent words were selected for <cht> and <gt> each, based on their frequency in the entire *Going Dutch Corpus*. The categorisation into *cht*- and *gt*-words follows the official 1804 spelling as codified in Siegenbeek's orthography and particularly in the *Woordenlijst* 'word list' (in the appendix of his treatise). The following twenty items, comprising both *cht*- and *gt*-words⁴², were selected for the corpus analysis (listed in order of decreasing frequency in the *Going Dutch Corpus*)⁴³:

⁴² The first form of each set represents the prescribed Siegenbeek spelling. For example, the ACHT/AGT set belongs to the category of *cht*-words, whereas REGT/RECHT belongs to the *gt*-category.

⁴³ Surprisingly, some of the most prominent and much-cited examples of the *cht/gt* discussion, such as *klagt* 'complain', *(ge)wigt* 'weight' and *vrucht* 'fruit', are relatively low in frequency in the *Going Dutch Corpus*. Therefore, they will be not taken into account in this analysis.

- **cht-words:**
 ACHT/AGT; WACHT/WAGT; ECHT/EGT; DACHT/DAGT; SLECHT/SLEGT;
 NACHT/NAGT; DOCHTER/DOGT; NICHT/NIGT; ZOCHT/ZOGT;
 KOCHT/KOGT
- **gt-words:**
 REGT/RECHT; ZIGT/ZICHT; BRAGT/BRACHT; RIGT/RICHT;
 MOGT/MOCHT; MAGT/MACHT; JAGT/JACHT; PLIGT/PLICHT;
 TOGT/TOCHT; VLUGT/VLUCHT

These words are best regarded as search queries, covering various word forms as well as derivations and compounds with the same root. To give an example: WACHT/WAGT (from the category of *cht*-words) comprises forms of the verb *wachten* ‘to wait’, *verwachten* ‘to expect’ and *afwachten* ‘to await’, but also derivations like *verwachting* ‘expectation’ or *onverwachts* ‘unexpected’, and compounds like *wachthuis* ‘guard house’ or *erewacht* ‘guard of honour’. Similarly, ZIGT/ZICHT (from the category of *gt*-words) comprises words such as *zichtbaar* ‘visible’, *gezicht* ‘face; view’, *nitzigt* ‘view’, *opzigt* ‘regard, respect’ and *beziggen* ‘to visit’.

It should be noted that some of the above-mentioned queries actually comprise a range of semantically and etymologically different words. For instance, ACHT/AGT includes the numeral *acht* ‘eight’ as well as *acht* in *acht geven* ‘take care’, *achter* ‘behind’ and the adjectival suffix *-achtig* ‘-like’. As pointed out in Section 1, Siegenbeek did not differentiate between these homonyms by spelling, but prescribed <cht> in all cases, which is why they are treated as one item in this case study. A similar example is JAGT/JACHT, which contains tokens of two homonymous words, viz. *jagt* ‘hunt(ing)’ and *jagt* ‘yacht’. In both cases, Siegenbeek prescribed <gt>.

The concordance results were thoroughly filtered, removing all instances of noise such as proper and place names (e.g. *Utrecht* and *Dordrecht* in the results of REGT/RECHT) as well as other undesirable tokens (e.g. *dagteekening* in the results of DACHT/DAGT).

4.2 Results

Investigating the entire *Going Dutch Corpus*, Tables 1a and 1b provide a first general overview of the distribution of variants across time, subdivided into Siegenbeek’s categories of *cht*- and *gt*-words, respectively. The officially prescribed variants in the nineteenth-century period (i.e. <cht> in Table 1a, <gt> in Table 1b) are highlighted in light grey.

As noted before, there was no official categorisation into <cht> and <gt> words in the late eighteenth-century period, i.e. before Siegenbeek’s orthography was introduced. Therefore, it is hardly surprising that the distribution in both categories of words is very similar.

Table 1a. Distribution of variants in the category of *cht*-words across time.

<i>cht</i> -words	Period 1: 1770–1790			Period 2: 1820–1840		
	<cht>	<gt>	<ght>	<cht>	<gt>	<ght>
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Total	241 (17.5)	1,117 (80.9)	23 (1.7)	987 (75.7)	316 (24.3)	0 (0.0)

Table 1b. Distribution of variants in the category of *gt*-words across time.

<i>gt</i> -words	Period 1: 1770–1790			Period 2: 1820–1840		
	<cht>	<gt>	<ght>	<cht>	<gt>	<ght>
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Total	217 (20.3)	838 (78.5)	13 (1.2)	213 (17.7)	992 (82.3)	0 (0.0)

Roughly 80% of all instances in the *Going Dutch Corpus* is spelled <gt>, which makes it the main variant in eighteenth-century usage. The <cht> spelling does occur in the data, but considerably less frequently than <gt>. The third variant <ght> only occurs in less than 2% of all instances. On closer inspection, it turned out that <ght> is a manuscript variant only occurring in ego-documents, whereas there is no single attestation of <ght> in printed texts in the corpus.

In the nineteenth-century period, the officialised spelling variants for both categories of words become the dominant variants in actual language use. In the category of *cht*-words (Table 1a) the share of <cht> increases from 17.5% to 75.7%. In other words, the less frequent variant in the first period becomes by far the predominant variant in the second period. In the category of *gt*-words (Table 1b), <gt> maintains its position as the dominant – and now officially prescribed – variant. The share of <gt> even slightly increases from 78.5% to 82.3%. Although the ‘incorrect’ representations in each category of words do not disappear completely from early nineteenth-century language practice, the remarkable shift in the direction of the prescribed variants, especially <cht>, strongly suggests the impact of Siegenbeek’s official spelling rules.

The <ght> variant does not occur in the nineteenth-century data. With regard to the fact that this third variant plays a very marginal role in period 1 and is completely absent in period 2, I decided to present only the results for <cht> and <gt> in subsequent analyses, excluding the few occurrences of <ght>.

Genre variation

Taking a multi-genre perspective on developments in language use, Figures 1a and 1b display the distribution of variants across the three genres of the *Going Dutch*

Corpus, viz. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW).

Figure 1a. Distribution of variants in the category of *cht*-words across genre and time.

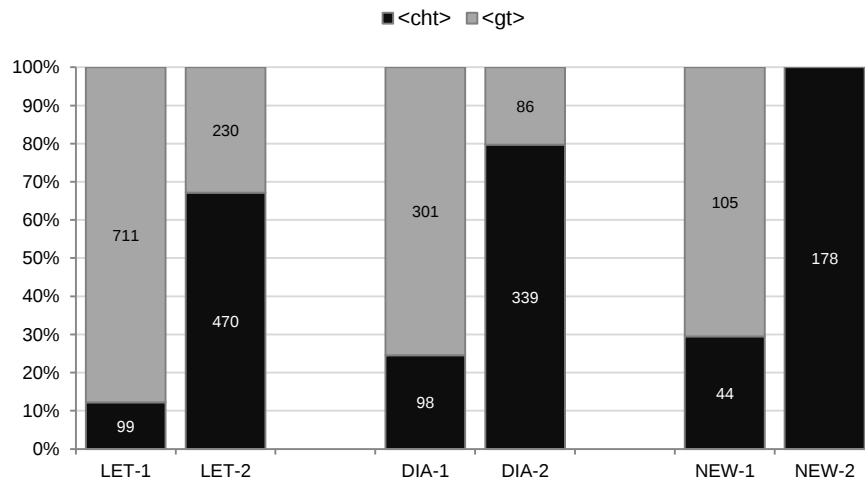
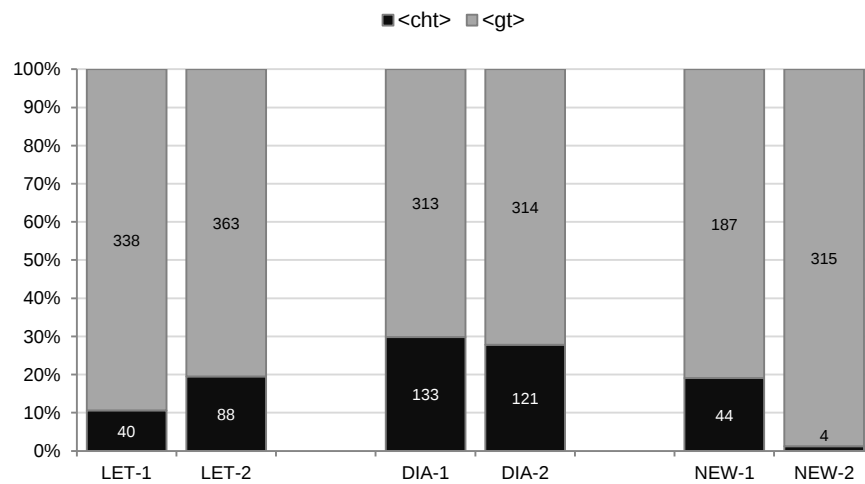


Figure 1b. Distribution of variants in the category of *gt*-words across genre and time.



Diachronically, the category of *cht*-words (Figure 1a) shows similar developments across all three sub-copora. In the late eighteenth century, <gt> is the main variant in private letters, diaries and travelogues as well as newspapers, all of which shift to <cht> as their main variant in the early nineteenth century. At the

same time, considerable differences between the genres become apparent. In the eighteenth-century period, private letters have the highest frequency of <gt> with a share of 87.8%. The same variant is considerably less frequent in diaries and travelogues (75.4%) and least frequent in newspapers (70.5%). In other words, <cht> is less common in private letters (12.2%) than in diaries and travelogues (24.6%) and newspapers (29.5%).

A similar genre gradation can be observed in the nineteenth-century data. Although the use of prescribed <cht> strikingly increases in both types of ego-documents, its share in private letters (67.1%) is still considerably lower than in diaries and travelogues (79.8%). It turns out that letter writers have the strongest preference for <gt> in the first period, and, at the same time, are the slowest in adopting <cht> in the second period. In the newspaper data, prescribed <cht> is used in 100% of all instances, completely replacing <gt>.

The results in the category of *gt*-words (Figure 1b) show that <gt> is the predominant variant across all three genres, both before and after Siegenbeek. From a diachronic perspective, however, some remarkable genre differences can be identified. Like in the category of *cht*-words, nineteenth-century newspapers adopt the prescribed <gt> in practically all instances (98.7%), with only four tokens of <cht>. In diaries and travelogues, the distribution of variants remains stable, with <gt> increasing only marginally from 70.2% to 72.2%. Most strikingly, private letters from the post-Siegenbeek period show an increase in the use of <cht> (for *gt*-words) from 10.6% to 19.5%. These developments testify that different genres follow orthographic norms to different extents. In line with the general expectations, printed and published texts (i.e. newspapers) show less variation in the use of different forms than handwritten ego-documents.

Regional variation

In addition to genre variation, the orthographic representation of syllable-final /xt/ was investigated across the seven regions of the *Going Dutch Corpus* (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

Table 2a presents the distribution of variants in the category of *cht*-words in the entire corpus. In the eighteenth-century period, <gt> is the dominant variant with a share of at least 75% in all seven regions, most notably in Groningen (87.8%) and South Holland (87.4%). On the other hand, the comparatively minor <cht> most frequently occurs in North Brabant, North Holland and Utrecht with a share of more than 20% each. Diachronically, all regions shift to prescribed <cht> as the main variant. Regional differences become apparent, though. Whereas the regions of Utrecht (87.3%) and South Holland (84.8%) apply <cht> most frequently, only in two-thirds of the North Brabant data (66.9%) the prescribed variant is followed.

Table 2a. Distribution of variants in the category of *cht*-words across region and time.

<i>cht</i> -words	Period 1: 1770–1790				Period 2: 1820–1840			
	<cht>		<gt>		<cht>		<gt>	
	N	%	N	%	N	%	N	%
FR	32	15.0	182	85.0	141	72.7	53	27.3
GR	23	12.2	165	87.8	148	70.8	61	29.2
NH	40	22.1	141	77.9	105	66.9	52	33.1
NH	39	21.5	142	78.5	151	71.2	61	28.8
SH	25	12.6	174	87.4	184	84.8	33	15.2
UT	44	24.2	138	75.8	137	87.3	20	12.7
ZE	38	17.8	175	82.2	121	77.1	36	22.9

Table 2b. Distribution of variants in the category of *gt*-words across region and time.

<i>gt</i> -words	Period 1: 1770–1790				Period 2: 1820–1840			
	<cht>		<gt>		<cht>		<gt>	
	N	%	N	%	N	%	N	%
FR	8	7.0	106	93.0	28	17.5	132	82.5
GR	42	22.6	144	77.4	23	12.6	159	87.4
NB	29	21.5	106	78.5	40	29.0	98	71.0
NH	53	34.9	99	65.1	24	11.3	188	88.7
SH	13	9.6	123	90.4	40	21.5	146	78.5
UT	42	25.3	124	74.7	32	16.8	159	83.2
ZE	30	18.1	136	81.9	26	19.1	110	80.9

The distribution of variants across regions in the category of *gt*-words is shown in Table 2b. In the eighteenth century, <gt> is prevalent across all seven regions, particularly in South Holland (90.4%) and Friesland (93.0%). The share of <cht> is remarkably strong in the North Holland data, occurring in more than one-third of all instances. When it comes to the diachronic developments, however, two different tendencies can be witnessed. In the nineteenth-century period, the relative frequency of prescribed <gt> increases in Utrecht (from 74.7% to 83.2%), Groningen (from 77.4% to 87.4%) and most notably in North Holland (from 65.1% to 88.7%). In contrast, there are some regions, in which the use of (not

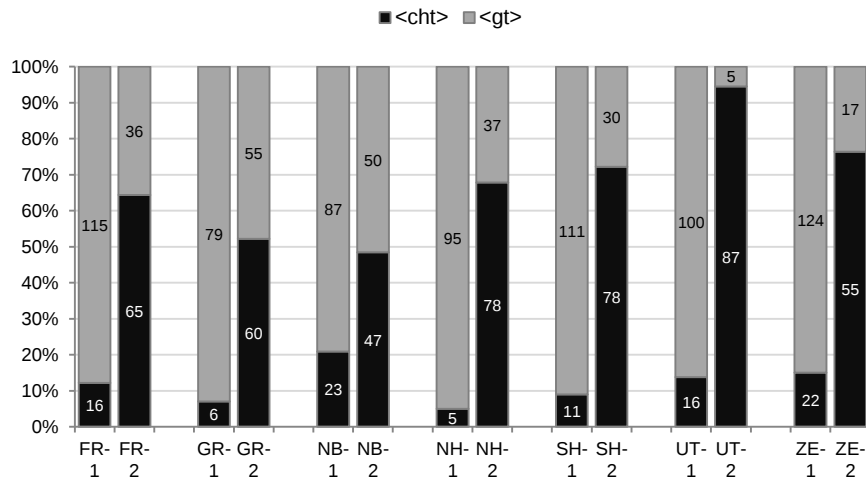
prescribed) <cht> increases, i.e. in Friesland (from 7.0% to 17.5%), South Holland (from 9.6% to 21.5%) and North Brabant (from 21.5% to 29.0%).

Regional variation across genres

The analysis of genre variation revealed that different genres have a considerable effect on the use of either <cht> or <gt>. Therefore, the three genres incorporated in the *Going Dutch Corpus* were analysed individually in order to zoom in on regional variation across genres.

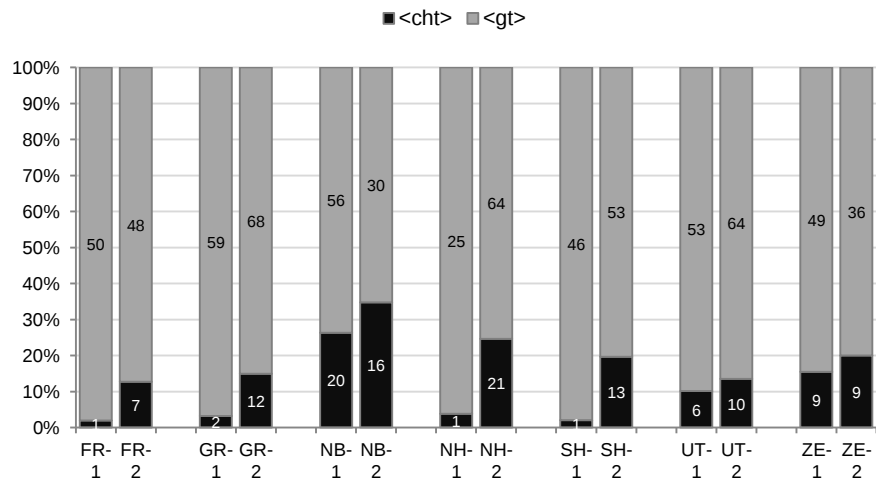
Focusing on the sub-corpus of private letters first, the *cht*-category in Figure 2a reveals some considerable regional differences, particularly in the second period. Whereas the prescribed variant <cht> occurs in 94.6% of the tokens in Utrecht, its share is considerably lower in North Brabant. In fact, North Brabant is the only region in which <gt> outnumbers the official variant <cht> with a share of 51.5%. Moreover, Groningen has a similarly high frequency of <gt> (47.8%), thus co-occurring with prescribed <cht>.

Figure 2a. Distribution of variants in the category of *cht*-words across region and time (private letters).



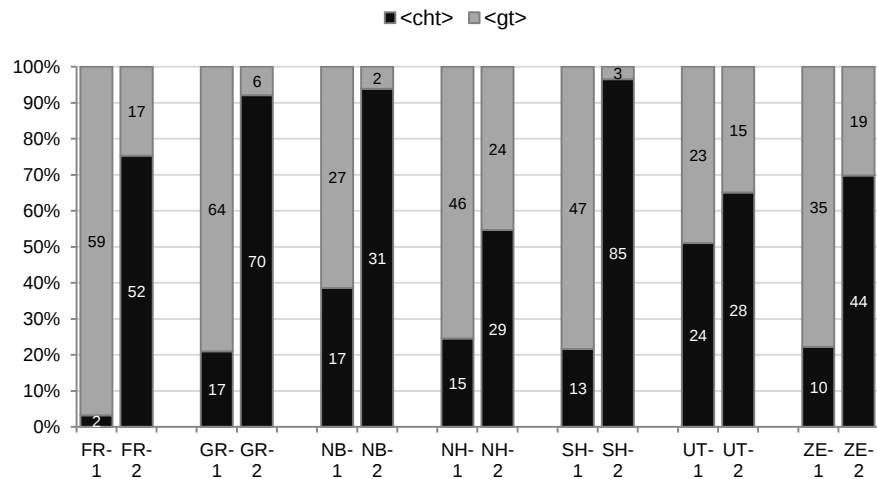
The distribution of words in the *gt*-category (Figure 2b) confirms the prevalence of <gt> across all regions. Diachronically, however, the results reveal that the use of <cht>, i.e. against Siegenbeek's spelling prescription, increases in private letters from all seven regions. The highest percentage of <cht> is found in nineteenth-century letters from North Brabant (34.8%). The lowest shares are attested in the data for Friesland, Groningen and Utrecht (less than 15% each).

Figure 2b. Distribution of variants in the category of *gt*-words across region and time (private letters).



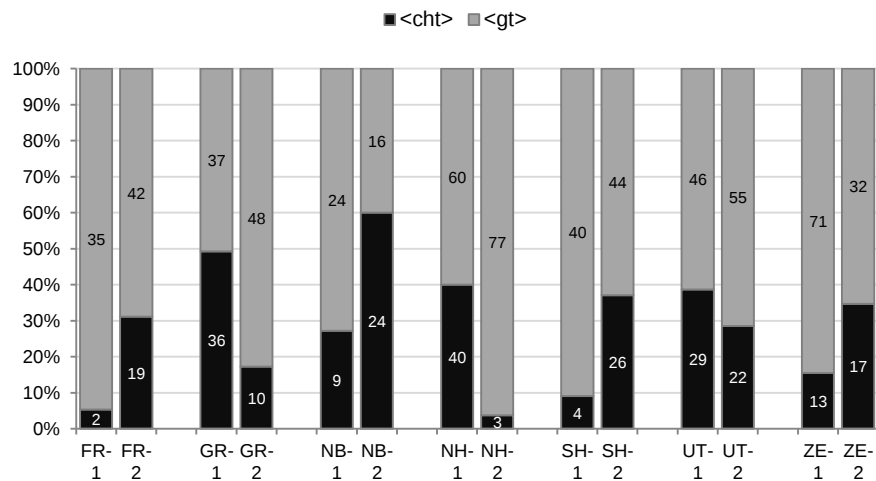
Secondly, the distribution across regions in the sub-corpus of diaries and travelogues was studied. In the category of *cht*-words (Figure 3a) there is a general shift from <gt> as the prevalent variant in period 1 to prescribed <cht> in period 2 in all seven regions, most notably in Groningen, North Brabant and South Holland with a share of more than 90% each.

Figure 3a. Distribution of variants in the category of *cht*-words across region and time (diaries and travelogues).



In the category of *gt*-words (Figure 3b), two different regional developments can be witnessed. On the one hand, the use of prescribed <gt> increases in diaries and travelogues from Groningen, North Holland and Utrecht. On the other hand, in Friesland, North Brabant⁴⁴, South Holland and Zeeland, it is the <cht> spelling which increases in the category of *gt*-words. The variation revealed here helps to explain the surprising regional differences in the entire corpus as presented in Table 2b, viz. the increase of prescribed <cht> for *gt*-words in certain regions. These tendencies can, in fact, only be observed in diaries and travelogues.

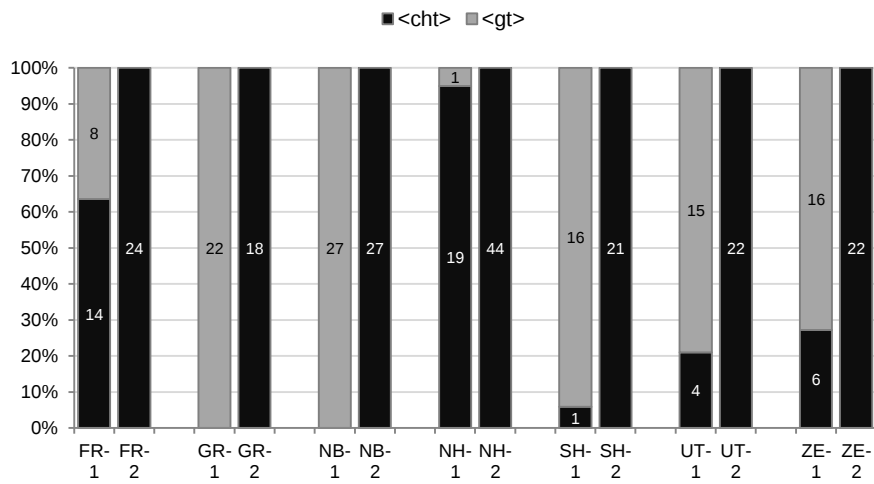
Figure 3b. Distribution of variants in the category of *gt*-words across region and time (diaries and travelogues).



Finally, focusing on regional variation in newspapers, the *cht*-category (Figure 4a) shows that different variants were preferred in the first period, depending on the region. Even though the number of tokens is relatively small, late eighteenth-century newspapers from North Holland clearly favour <cht> in practically all instances. In Friesland <cht> and <gt> are co-occurring variants, whereas <cht> is completely absent in newspapers from Groningen and North Brabant. In the nineteenth-century data, <cht> is the only variant found in newspapers from all regions.

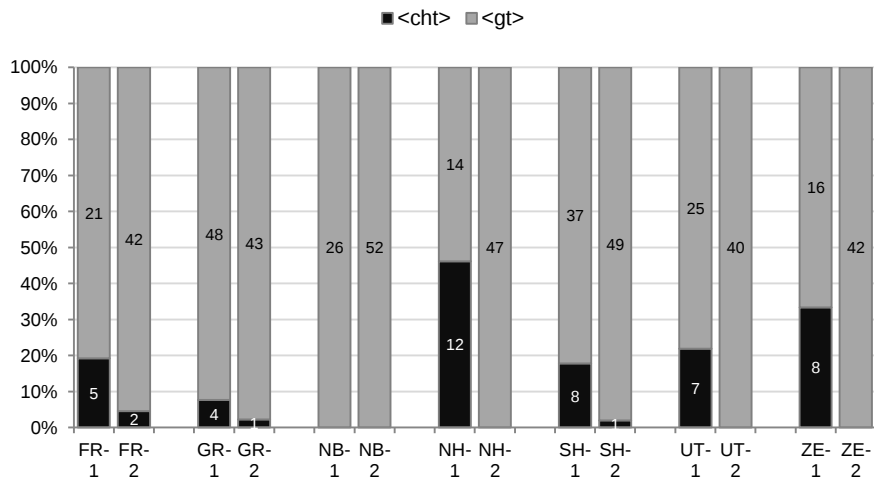
⁴⁴ It has to be noted, though, that the nineteenth-century data for North Brabant is based on one diary only. Therefore, the representativeness of one individual writer from this region has to be considered as limited.

Figure 4a. Distribution of variants in the category of *cht*-words across region and time (newspapers).



In the *gt*-category (Figure 4b), eighteenth-century newspapers from North Holland have the highest share of <cht> (46.2%), which is in line with the distinct preference for <cht> in the *cht*-category in the same period. Other regions such as Zeeland and Utrecht use <cht> in 33.3% and 21.9%, respectively. Again, no single attestation of <cht> is found in the North Brabant data.

Figure 4b. Distribution of variants in the category of *gt*-words across region and time (newspapers).



Despite these regional differences in the first period, nineteenth-century newspapers from all seven regions use <gt> in conformity with the officialised categorisation, which indicates a strong normative influence of Siegenbeek's orthography on newspapers.

Variation across centre and periphery

Building on the results of regional variation, Figures 5a (*cht*-words) and 5b (*gt*-words) take into account the second variable on the spatial dimension, viz. centre (CEN) versus periphery (PER).

Figure 5a. Distribution of variants in the category of *cht*-words across centre–periphery and time.

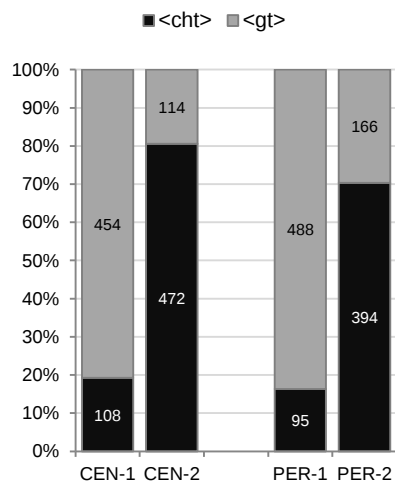


Figure 5b. Distribution of variants in the category of *gt*-words across centre–periphery and time.

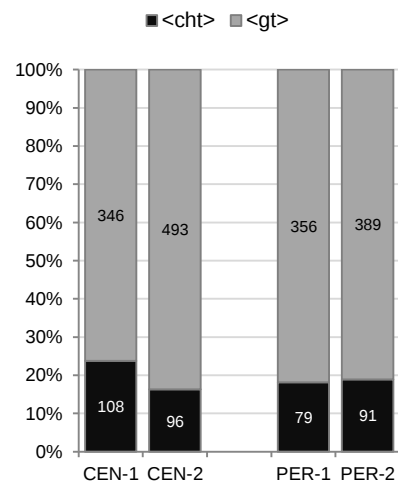


Figure 5a shows similar tendencies in the category of *cht*-words in both centre and periphery: <cht> as the marginal variant in period 1 becomes the dominant variant in period 2, but the prevalence of prescribed <cht> is more prevalent in the centre (80.5%) than in the periphery (70.4%). The category of *gt*-words in Figure 5b reveals a more remarkable difference in the diachronic development of the variants. Whereas the use of prescribed <gt> increases from 76.2% to 83.7% in the centre, there is hardly any change in the distribution of variants in the periphery. In fact, <cht> even increases marginally from 18.2% to 18.6%.

Gender variation

In order to shed light on possible gender variation, the two sub-corpora of ego-documents, viz. private letters as well as diaries and travelogues, were analysed.

Figures 6a and 6b display the distribution of variants across men (M) and women (F) in the categories of *cht*- and *gt*-words, respectively.

Figure 6a below shows that eighteenth-century men and women predominantly use <gt> in the group of words which Siegenbeek later categorised as *cht*-words. In the results for both genders, <cht> only occurs marginally, particularly among women (11.3%). In the nineteenth-century period, the majority of both men and women adopts the prescribed variant <cht>. Women, however, appear to follow the officialised spelling considerably less frequently than men. In fact, almost 40% of the female writers still use <gt> for *cht*-words. In contrast, the vast majority of male writers (almost 80%) use <cht> in conformity with Siegenbeek's prescription.

Figure 6a. Distribution of variants in the category of *cht*-words across gender and time.

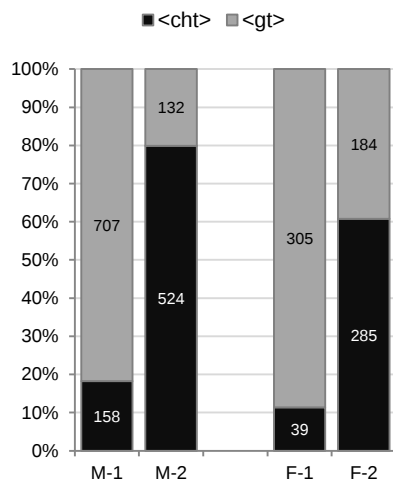


Figure 6b. Distribution of variants in the category of *gt*-words across gender and time.

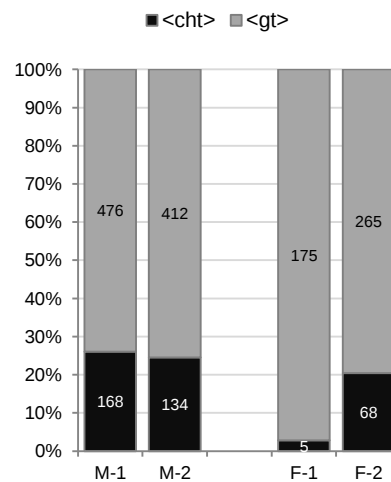


Figure 6b presents the results in the category of *gt*-words. The distribution of variants in the late eighteenth-century period is similar to that presented in Figure 6a. The <gt> spelling is by far the most dominant variant among men and women of the pre-Siegenbeek generation. Whereas male writers use <cht> in 26.1% of all instances, it is striking that <cht> is practically absent in texts by female writers (in both categories of words actually). Interestingly, in the period after Siegenbeek, the use of <cht> for *gt*-words among women increases from 2.8% to 20.4%. No such development is visible in the results for men: The distribution of the two variants is stable across time. Despite the striking increase of <cht> among nineteenth-century women, the relative frequency of <gt> (79.6%) is still slightly higher than in ego-documents written by men (75.5%).

5 Discussion

Taking into account both language norms and language usage, the orthographic representation of the consonant cluster /xt/ in syllable-final position can be considered a fairly complex variable. As discussed in Section 1, Siegenbeek (1804) officially introduced a division into two categories of words, viz. words prescribed to be spelled with either <cht> (e.g. *dacht* ‘thought’ in line with pronunciation, i.e. final devoicing) or <gt> (e.g. as *bragt* ‘brought’ < *brenge* ‘to bring’, due to their derivation and/or etymology). Furthermore, Section 2 outlined the eighteenth-century normative discussion, showing that despite the limited consensus regarding the spelling of individual words, both <cht> and <gt> were already acknowledged as (co-existing) variants before the official codification of the Dutch spelling. In that sense, Siegenbeek’s selection of variants as such was not that innovative. Some of his choices, however, could be considered as progressive in relation to the eighteenth-century normative tradition. Especially his view on homonymy, according to which he rejected the widespread idea to distinguish homonyms by spelling (e.g. *agt* ‘eight’ versus *acht* as in *acht geven* ‘take care’), deviated from most eighteenth-century normative works. The only exception was Kluit (1777), on which Siegenbeek heavily relied.

Analysing the corpus results of actual language usage in Section 4, it becomes apparent that the representation of /xt/ as <gt> is clearly prevalent among late eighteenth-century language users. Even though <cht> had already been an acknowledged and prescribed variant in many eighteenth-century normative works, this is hardly reflected in language practice, where <cht> only marginally occurs. In this respect, the increase of <cht> after Siegenbeek’s prescription is even more striking. The successful shift from <gt> to <cht> as the dominant variant in the category of *cht*-words most probably indicates the top-down effect of Siegenbeek’s orthography on language practices.

The expected genre differences between printed published texts and handwritten ego-documents were generally confirmed. It was observed that nineteenth-century newspapers adopted the prescribed variants in practically all instances, without any regional variation. The same analysis in the sub-corpus of private letters revealed considerably more variation, though. Here it was particularly striking that, in addition to the major shift from <gt> to <cht> in the category of *cht*-words, the relative frequency of <cht> in the *gt*-category also increased. This development, in fact, appeared to be a national phenomenon, as it was attested across all seven regions of the corpus. One might assume that the official 1804 spelling increased the awareness of the previously marginal <cht> variant among language users. On the one hand, it led to an increase of <cht> and ultimately its predominance in the category of *cht*-words – as envisaged in Siegenbeek’s orthography. On the other hand, it could be argued that <cht> was also overgeneralised by language users to words which did not belong to the category of *cht*-words, possibly leading to cases of hypercorrection in the *gt*-category.

Especially with regard to the use of variants deviating from Siegenbeek’s spelling rule, the results for North Brabant were particularly interesting. Of all

regions, nineteenth-century letters from North Brabant had the highest share of the <gt> spelling for *cht*-words (more than half of all instances), but also the highest share of <cht> for *gt*-words (more than two-thirds). Furthermore, <cht> was completely absent in eighteenth-century newspapers of this region, possibly indicating regional conventions favouring <gt>. On the contrary, newspapers from North Holland already used <cht> before 1804. With regard to the fact that Siegenbeek was a native of Amsterdam, it might even be suggested that (printed) language practices from the (North) Holland area had some influence on Siegenbeek's choice of variants.

The results on the centre–periphery level also supported these regional tendencies. The centre (including the Holland area) was generally more successful in adopting the prescribed variants for both *cht*- and *gt*-words than the periphery (including North Brabant).

With respect to gender variation, it was particularly noticeable that nineteenth-century women increasingly used <cht> for *gt*-words. The surprising development from almost non-existent <cht> before Siegenbeek to 20% after Siegenbeek can probably be interpreted as instances of hypercorrection as well.

To sum up, the normative effects of Siegenbeek's prescription for the *cht/gt* issue are certainly visible in early nineteenth-century language usage. Despite the complexity of the officialised rule, the categorization into *cht*- and *gt*-words was established across all genres, regions and genders. Both <cht> and <gt> became the most frequently used variants in their respective categories of words. Looking at the diachronic developments more generally, the striking increase of the previously marginal variant <cht> (in the category of *cht*-words) probably reflects an evolution in the relation between language norms and language usage. Even though <cht> had been an acknowledged and prescribed spelling (alongside <gt>) in metalinguistic comments before Siegenbeek, it only played a marginal role in eighteenth-century usage, whereas <gt> was clearly dominant in language usage. In principle, Siegenbeek's choice of variants did not deviate much from eighteenth-century grammarians, but, in contrast to those normative works, his 1804 orthography had such an impact that normative effects became visible in actual usage. One might argue that the official regulation of the Dutch orthography was the decisive step needed to 'reach' the language user.

Some questions about the remarkable success of the *cht/gt* categorisation remain, though. Which aspect(s) of Siegenbeek's norm actually reached the language users? Which part(s) of the official rule were they aware of? Two main scenarios are possible: First of all, language users were aware of the exact spelling rule prescribed by Siegenbeek, dividing words with syllable-final /xt/ into two categories and consequently spelled with either (phonologically motivated) <cht> or (etymologically motivated) <gt>. Secondly, and probably more realistically, language users were aware of the mere existence of <cht> as part of the official spelling norm, but probably independent of its exact rules, i.e. on how and when to apply either the phonological or etymological principle. In fact, the attested cases of hypercorrect <cht> in the category of *gt*-words support the second scenario. In private letters from all regions, an increase of probably overgeneralised <cht> for

gt-words was observed. Particularly striking were the developments among female letter writers, from a practically non-existent use of <cht> in the eighteenth-century data to around 20% in the early nineteenth century. For these language users, the <cht> spelling must have been perceived as the newly promoted ‘Siegenbeekian’ spelling, as it was considerably less common in language practice in the late eighteenth century and only began to be used more frequently (and more consciously) after Siegenbeek’s orthography.

Orthographic variables (2)

Final /t/ in *d*-stem verbs

1 Discussion in Siegenbeek (1804)

The case study in this chapter investigates the orthographic representation of final /t/ in second and third person singular and second person plural present tense indicative forms of verbs with *d*-stems, such as *worden* ‘to become’ or *vinden* ‘to find’. It is one of the consonantal features codified in Siegenbeek’s (1804) orthography, which touches upon the levels of both orthography and morphology. Even though Siegenbeek (1804: 156) did not elaborately comment on this spelling issue⁴⁵, he unambiguously prescribed <dt> as the standard variant for verb-final /t/:

Ten aanzien der *dt*, welker vereeniging zeker op zich zelve iets vreemds en wanstaltigs heeft, zij nog met een woord aangemerkt, dat men dezelve, ter voldoening aan het tegenwoordige gebruik, en ter bevordering der duidelijkheid, die hoofdwet der tale, alleen dan te gebruiken heeft, wanneer zij voorkomt als eene verkorting van *det*, dat is, met andere woorden, in den tweeden en derden persoon van den tegenwoordigen, en den tweeden van den onvolmaakt verledenen tijd der aantoonende wijze, in de werkwoorden *binden*, *vinden*, en meer dergelijke. Immers is het te voren reeds aangemerkt, dat men oudtijds deze, gelijk alle andere werkwoorden, op de volgende wijze vervoegd heeft:

Ik *binde*, gij *bindet*, hij *bindet*.

Ik *bonde*, gij *bondet*, hij *bonde*.

waaruit bij de weglating der zachte *e*, ik *bind*, gij, *bij bindt*, ik, *bij bond*, gij *bondt*, geboren wordt.

‘With regard to the *dt*, whose combination certainly has something odd and malformed in itself, it should be noted that, in order to meet the contemporary usage and to enhance clarity, the main principle of language, one has to use it (*dt*) only when it occurs as a shortening of *det*, that is, in other words, in the second and third person of the present tense, and the second of the imperfect of the indicative, in the verbs *binden*, *vinden*, and the like. After all, it has already been noted that in the olden days, these verbs, as all other verbs, were conjugated in the following way:

Ik *binde*, gij *bindet*, hij *bindet*.

Ik *bonde*, gij *bondet*, hij *bonde*.

Through the omission of the soft *e*, ik *bind*, gij, *bij bindt*, ik, *bij bond*, gij *bondt*, are born.’

⁴⁵ As van der Velde (1956: 92) rightly remarks, Siegenbeek generally discussed the spelling of verbs only incidentally and very concisely.

To begin with, Siegenbeek remarked that the combination of the letters *d* and *t* has something ‘odd’ and ‘malformed’ (“iets vreemds en wanstaltigs”) to it. However, he argued that it is necessary to use the <dt> spelling in order to comply with contemporary practices, as well as to enhance the clarity, which he regarded as the fundamental law of the language. Siegenbeek further explained that <dt> is a contracted form of the historical ending *-det*. Schwa syncope (“de weglating der zachte *e*”) in these forms gave rise to new forms with <dt>, for instance *gij bindt* < *gij bindet* or *hij bindt* < *hij bindet*. Therefore, <dt> should only be applied in forms which originally had an ending in *-det*. This means that Siegenbeek’s main argument in favour of <dt> is chiefly based on etymology, whereas morphologically motivated arguments are not mentioned explicitly. In this respect, he deviated from many eighteenth-century grammarians, who argued that verbs with a *d*-stem should be inflected like any other (non-*d*-stem) verb, which will be further outlined in Section 2.

There is yet another aspect of Siegenbeek’s prescription, which remains implicit. In fact, it is remarkable that Siegenbeek did not specifically address the number of second person forms, i.e. whether *gij* ‘you’ only refers to second person singular, or to both second person singular and plural. However, as second person plural verb forms with *gij* or *gijlieden* ‘you (pl.)’ historically had a *-det* ending as well, Siegenbeek most likely included both singular and plural forms of the second person.

Finally, it has to be mentioned that Siegenbeek’s prescription also covered simple past forms ending in *-dt*, viz. second person indicative such as *gij bondt* (< *gij bonde*). In eighteenth-century metalinguistic discourse, however, past tense forms were commonly treated in separate paradigms (i.e. not combined with present tense forms), which is why the spelling of final /t/ in simple past forms of *d*-stem words is best regarded as a variable in its own right. Aiming to investigate a diverse range of orthographic features, the focus on present tense indicative forms seems sufficient to examine patterns of variation and change in the representation of verb-final /t/. For this reason, verb-final /t/ in second person simple past indicative forms will not be considered in this dissertation.

2 Eighteenth-century normative discussion

Already in the eighteenth century, metalinguistic comments on the orthographic representation of final /t/ in second and third person singular and second person plural present tense indicative forms of *d*-stem verbs were remarkably homogeneous in the Northern Netherlands (cf. also Vosters et al. 2014: 80). Like Siegenbeek (1804), the vast majority of eighteenth-century normative works advocated the <dt> spelling. In fact, this preference is represented throughout the eighteenth century, from Moonen (1706), Verwer (1707) and Sewel (1708/1712) in the early 1700s, to Huydecoper (1730), Elzevier (1761), Zeydelaar (1774), Stijl & van Bolhuis (1776) and Kluit (1777), to the *Rudimenta* (1799), Weiland (1799) and Wester (1799) just before the turn of the century.

It is surprising to see that some eighteenth-century comments on this spelling issue were considerably more elaborate than Siegenbeek's official prescription. Whereas Siegenbeek only motivated the prescribed <dt> spelling with reference to the historical endings in *-det*, thus focusing on etymological reasons (Section 1), several eighteenth-century grammarians discussed the spelling of verb-final /t/ more explicitly. While the general eighteenth-century preference was in favour of <dt>, three different approaches can be identified, which are based on (1) morphological motivations, (2) etymological motivations, or (3) a combination of those two.

One of the earliest eighteenth-century attestations of <dt> as the preferred variant for verb-final /t/ can be found in Moonen (1706: 13), whose approach was morphologically motivated:

De T is ook de merkleetter en nootwendigh in het spellen der tweede persoonen van beide Getallen der Aentoonende Wyze, en des derden persoons in het Eenvouwige Getal: als *Gy, Hy, Neemt, geeft, leeft, hoort, zingt* &c.

Dit houdt zynen regel, schoon het Wortelwoort in een D eindigt, wanneer de T niet achtergelaeten wordt, noch de D in de T verandert. Want men schryft *Gy, Hy houdt, wordt, vindt, landt, grondt*, &c. alle afgeleidt van de Wortelwoorden in de Gebiedende Wyze, *Houd, word, bind, vind, land, grond*.

'The T is also the characteristic letter and necessary for the spelling of the second person of both numbers of the indicative case, and of the third person in the singular, as *Gy, Hy, Neemt, geeft, leeft, hoort, zingt*, etc.

This rule also applies when the root word ends in a D, when neither the T is left behind, nor the D is changed into T. Because one writes *Hy houdt, wordt, vindt, landt, grondt*, etc., all derived from the root words in the imperative, *Houd, word, bind, vind, land, grond*.'

Moonen pointed out that in second and third person singular and second person plural forms of *d*-stem verbs like *bouden* 'to hold' and *vinden* 'to find', neither the inflectional suffix *-t* must be omitted, nor must stem-final *d* be orthographically altered into <t>. In other words, final /t/ has to be represented as <dt>, not as <d> or <t>.

Sewel (1712: 17-18), also prescribing <dt>, explicitly disfavoured the <d> spelling for second and third person verb forms:

Wyders kan de D niet gemist, maar behoort echter met eene T getemperd te worden, om de Tweeden en derde persoon te betekenen der Werkwoorden, die in de Onbepaalende wyze een D hebben, als *Bidden, Myden, Leyden, Lyden, Kleeden*. Want men behoort te schryven: *Gy bidt, Hy mydt haar niet, De weg leydt ten verderve, Hy leedt veel ongemaks, zy kleedt het kind*: maar in woorden waarin geen D komt, als *Beminnen, komen, vermaanen*, is het ten hoogste wanschickelyk eene D in den Tegenwoordigen tyd te brengen, alhoewel veele zich niet ontzien te schryven, *Hy bemind haar niet, Zy komd straks* [...] doch zulks is een quaaide gewoonte. Nog inschickelyker is het te schryven: *Hy arbeyd sterk, Hy bloed uyt de neus, Men vind: omdat deeze woorden de d niet kunnen missen: maar nogtans is een t daarby beter.*

Further, the D cannot be missed, but has to be tempered with a T, in order to denote the second and third person of verbs, which have a D in the infinitive, as *Bidden, Myden, Leyden, Lyden, Kleeden*. Because one has to write *Gy bidt: Hy mydt haar niet: De weg leydt ten verderve: Hy leedt veel ongemaks: zy kleedt het kind*. But in words, in which there is no D, as *Beminnen, komen, vermaanen*, it is highly irregular to add a D to the present tense, although many do not shy away from writing *Hy bemind haar niet: Zy komd straks [...]* but this is a bad habit. It is even more obliging to write *Hy arbeyd sterk; Hy bloed nyt de neus; Men vind*, because these words cannot go without the *d*. However, it is better together with a *t*?

Like Siegenbeek (1804), Sewel also included simple past forms in his spelling rule, although he did so rather implicitly by giving the example of *Hij leedt* 'he suffered', alongside present tense forms.

In the late 1700s, the *Rudimenta* (1799: 62), for example, still advocated <dt> in a similar way as Moonen (1706) in the beginning of the century. The (morphologically) consistent use of final *-t* in second and third person singular and second person plural verb forms was central to the rule:

Wat de spelling met DT aangaat, die vloeit uit dezelfde regelen voort, en wordt alleen veroorzaakt als er in de werkwoorden ééne D in de *onbepaalde* wijze is, b. v. *gij, hij, zij* of *men, andwoordt, vindt, wordt, houdt, doodt*, enz. in den tegenwoordigen tijd [...] komende van de *onbepaalde* wijze *andwoorden, vinden, worden, houden, dooden*; en wanneer men twijfelt, of men noodig hebbe de T te gebruiken, neeme men, ter beproeving, een werkwoord, in welks *onbepaalde* wijze geen D noch T gevonden wordt gelijk b. v. *loopen, beminnen, vertellen*, en vergelijke dat met zodanig een in welks *onbepaalde* wijze D of T voorkomen, zo als *lijden, bieden*, en bevindende, gelijk men zal, dat in den tegenwoordigen tijd, in den tweeden en derden persoon enkelvoudig, en den tweeden meervoudig, gelijk hier voor gezegd is, de T noodig is, zal men bevinden dezelve in die laatst genoemde woorden aldaar ook niet te kunnen missen.

'As regards the spelling with DT, this arises from the same rules, and is only caused when there is a D in the verbs in the infinitive, e.g. *gij, hij, zij* or *men, andwoordt, vindt, wordt, houdt, doodt*, etc. in the present tense [...] coming from the infinitive *andwoorden, vinden, worden, houden, dooden*. And when one doubts whether one needs to use the T, one should test by taking a verb, in which neither D nor T is found, like e.g. *loopen, beminnen, vertellen*, and compare it to such a verb in whose infinitive D or T occur, as in *lijden, bieden*. If one finds, as one should, that the T is necessary in the present tense, in the second and third person singular, and the second person plural, as said before, one will find that one cannot go without it (the T) in the latter words either.'

In the same year as the *Rudimenta*, Weiland (1799: 49-50) added an etymological dimension to his primarily morphologically motivated spelling rule:

daar alle de werkwoorden in den tweeden en derden person van den tegenwoordigen tijd der aantoonende wijs, in het enkelvoudige getal, eene T hebben, als *gij en hij zegt, leest* enz.; zoo is het natuurlijk, dat diezelfde tweede en derde persoon van de werkwoorden, welken eene D in hun zaaklijk deel hebben,

als *zenden, randen, branden, binden, vinden* enz. eene DT ontvange, als *gij* en *hij zendt, randt, brandt, bindt, vindt*, enz., waarvoor de Ouden schreven, *gij zendest, hij zendet – randest, randet – brandest, brandet – bindest, bindet – vindest, vindt* enz.

‘As all verbs in the second and third person of the present tense of the indicative, in the singular, have a T, as *gij* and *hij zegt, leest* etc., it is natural that the same second and third person of the verbs which have a D in their essential part (root), as *zenden, randen, branden, binden, vinden* etc. receive a DT, as *gij* and *hij zendt, randt, brandt, bindt, vindt*, etc., for which the Old wrote, *gij zendest, hij zendet – randest, randet – brandest, brandet – bindest, bindet – vindest, vindt* etc.’

Weiland and, of course, Siegenbeek in his 1804 spelling were not the first to establish a link between the <dt> spelling and the historical verb forms ending in *-det*. Seven decades earlier, Huydecoper (1730: 288) already argued the following:

Overal dan, daar wy nu de spelling van DT behouden, moet men denken dat de Ouden geschreeven zouden hebben DET.

‘Everywhere that we now maintain the spelling of DT, one must think that the Old would have written DET.’

Some eighteenth-century choices in favour of <dt> were also purely etymologically motivated, probably serving as a source of inspiration for Siegenbeek (1804). Apart from Zeydelaar (1774: 82-83), this approach is found in Kluit (1777: 33), who elaborately commented on the development from verb forms ending in *-det* to the <dt> spelling:

zoo de D en T letteren van een natuur waren, men gevolglijk niet, dan walglijk en ongerijmd, dezelve kan te zamen voegen, en de een door de andere besluiten; wanneer men schrijft, *hij zemDT, hij ranDT, hij branDT*. Maar, die dit tegenwerpen, bedenken niet, dat zij iets opwerpen, ‘t welk zelf in de oudheid geenen grond heeft, en dus, dat, als er walglijkheid en ongerijmdheid in deze spelling is, deze ongerijmdheid uit de later opgerezen en nu in zwang gaande spelling voortgevloeid, en niet uit de oudheid herkomstig is. Deze kende zulke spelling met *dt* niet; in tegendeel, wanneer men daar *hij brandt, hij wordt*, enz. schrijft, zoo vindt men of *hij brandET, hij worDET*, of eenvoudig *hij brantT, hij wortT*.

‘if the D and T were letters of one nature, one cannot consequently combine them other than disgustingly and absurdly, and close the one by the other, when one writes *hij zemDT, hij ranDT, hij branDT*. But those, who object this, do not consider that they bring something up which does not even have ground in ancient times, and that, when there is disgustingness and absurdity in this spelling, this absurdity arose from the spelling that emerged later and is now coming in use, and does not derive from the ancient times. They did not know such a spelling with *dt*. On the contrary, when one writes *hij brandt, hij wordt*, etc. there, then one finds either *hij brandET, hij worDET*, or simply *hij brantT, hij wortT*.’

Moreover, Siegenbeek’s evaluation of <dt> as something *vreemds* ‘odd’ and *wanstaltigs* ‘malformed’ was obviously influenced by Kluit’s *walglijk* ‘disgusting’ and *ongerijmd* ‘absurd’.

Although the strong preference for <dt> is attested in metalinguistic discourse throughout the eighteenth century (morphologically and/or etymologically motivated), there were a few grammarians who still advocated the <d> spelling. One of these remarkable exceptions was the approach by van Belle (1748: 8-9), who expressed his sympathy for the ‘poor’ first person singular verb forms for not having a -t ending (unlike second and third person forms):

In allen gevalle, ik beklaag den armen eersten persoon in ’t enkelvoudig, dat die (schoon de liefde, gelyk men zegt, van zig zelven eerst komt) naer zulker Dryveren spellinge [T-naa D-Dryvers], zo veele eer, van met eene T vermeerderd te worden, niet mag genieten, dat men zo wel schryve, ik *redt*, als gy hy, gyl: *redt*; terwyl hy doch met dezelve personen gelyk staat, in zig door het Voornaamwoord bekend te maaken.

Kortom, ik ontken dat ‘er nader reegel zy, dan om de Werkwoorden, die D óf DD in de onbepaalende Wyze hebben, op den zelven leeft te schoeijen als die ‘er T óf TT hebben, en zo wel te schryven: ik, gy, hy, gyl: *red, bid, brand, zend*, als ik, gy, hy, gyl: *bet, zit, laat, weet*, enz:

‘In any case, I pity the poor first person in the singular, that this (although the love, as they say, comes from oneself first), according to the spelling of such pushers [pushers of T after D], may not enjoy so much honour of being augmented with a T, so that one can equally write ik *redt*, as gy hy, gyl: *redt*, while it (the first person) is equal to the other persons in that it reveals itself through the pronoun.

In short, I deny that there is a more precise rule than to lump the verbs, which have D or DD in the infinitive, together with those which have T or TT, and to write: ik, gy, hy, gyl: *red, bid, brand, zend*, as ik, gy, hy, gyl: *bet, zit, laat, weet*, etc.’

In slightly different words, van Belle (1755: 7) discussed the ‘unequal treatment’ of singular present indicative forms, referring to Sewel’s (1712) <dt> spelling in second and third (but not first person) forms:

Voorwaar de eerste person staat by hem [Sewel] zeerwel op voeten, zo dat de D in die woorden alleen kan gaan; maar hoe de tweede en derde persoon daar in zo zwak zyn, datze de T tót eene kruk gebruiken moeten is buiten myn boekje en ook buiten réden.

‘Indeed, in Sewel the first person stands on its own feet very well, so that the D can go alone in those words. But why the second and third person are so weak in that respect, that they have to use the T as a crutch, oversteps my bounds and is also without a reason.’

Like van Belle, ten Kate (1723) preferred the <d> spelling. Although he did not propose an explicit rule, his preference for <d> occasionally shines through, e.g. in *Ik, Gij, en Hij word Gemerkt* (1723: 678). Furthermore, ten Kate was consistent in his use of <d> for second and third person singular and second person plural present indicative forms of *d*-stem verbs (van der Velde 1956: 66).

Yet another approach is found in de Haes (1764) and van der Palm (1769). Although they did not explicitly comment on the conjugation of *d*-stem verbs

either, their paradigms of the auxiliary verb *worden* ‘to become’ revealed a striking difference with most eighteenth-century normative works. Both de Haes (1764: 59) and van der Palm (1769: 70) prescribed the <dt> spelling only for second person present indicative forms (both singular and plural), but <d> for third person singular forms:

Eenvouwig.	Meervouwig.
Ik word.	Wy worden.
Gy wordt.	Gylieden wordt.
Hy, zy, het word.	Zy worden.

Apart from these few exceptions, eighteenth-century metalinguistic discourse largely promoted <dt>, paving the way for Siegenbeek’s logical choice in his 1804 orthography. However, compared to the approaches in most normative works, most notably the elaborate discussion by Kluit (1777), Siegenbeek’s (1804) prescription is surprisingly concise, particularly with regard to the morphological motivations in favour of <dt>, which he did not address at all.

3 Previous research

The orthographic representation of verb-final /t/ in *d*-stem verbs has been the subject of numerous linguistic studies. As Vosters et al. (2014: 84) rightly note, the “history of <d> and <dt> spellings is well-documented”. Indeed, van der Velde (1956), Gledhill (1973) and Daems (2002), among others, provide overviews of the development of the spelling of verb forms, including final /t/ in second and third person present tense indicative forms of *d*-stem verbs, as investigated in this chapter.

Van der Velde (1956) summarises and comments on the metalinguistic perspectives on verbal spelling in a number of eighteenth-century normative works such as Moonen (1706), ten Kate (1723), Kluit (1763/1777), van der Palm (1769) and Zeydelaar (1781), as well as in the officialised reference works by Siegenbeek (1804) and Weiland (1805).

The concise overview in Daems (2002) focuses on the major changes in verbal spelling from the Middle Dutch period until the twentieth century, mainly addressing the different principles and motivations (i.e. phonetic, etymological and morphological spelling).

Gledhill (1973) dedicates an entire chapter to *-d* in second and third person singular verb endings, and also discusses the special case of *d*-stem verbs, which he described as “probably the most discussed in the developments of verbal spelling” (1973: 255). For each of the three orthographic variants, viz. <t>, <d> and <dt>, Gledhill outlines the developments from Middle Dutch until 1850 (and beyond).

More recently, the spelling of verbal endings of *d*-stem verbs has also been investigated from a historical-sociolinguistic perspective, most notably in the works by Vosters (2011), Vosters et al. (2010), Vosters et al. (2012) and Vosters et al. (2014). While the present dissertation examines norms and usage in the Northern

Netherlands, previous research mainly focused on the situation in the Southern Netherlands. Generally speaking, the <d> spelling has prototypically been perceived as a characteristic spelling feature of Southern Dutch, whereas <dt> has been regarded as its prototypically Northern counterpart.

Investigating nineteenth-century norms and usage in the Southern Netherlands, Vosters et al. (2014) select the spelling of verbal endings in *d*-stem verbs as one of their case studies. They point out that <d> was prescribed in almost all Southern normative works throughout the eighteenth century until 1815, whereas the Northern (Siegenbeek) variant <dt> became “by far the most dominant form in prescriptions in the 1820s” (Vosters et al. 2014: 95). A similar development can be witnessed in the nineteenth-century usage data based on a corpus of (handwritten) judicial and administrative texts, showing that in the relatively short time period between 1823 and 1829, “a radical shift from a strong predominance of traditional Southern <d> spellings to a majority prevalence of Northern <dt> spellings” (ibid.) took place.

The more contrastive case study presented in Vosters et al. (2010) gives some first insights into late eighteenth-century language norms and practice in both the Southern and Northern Netherlands⁴⁶. Whereas the prototypically Southern <d> spelling was indeed dominant in normative works as well as in actual usage in the Southern Netherlands, the paper also shows that the situation in the Northern Netherlands was different. It is true that the Northern normative tradition mainly prescribed the <dt> spelling (cf. Section 2), but at the same time, actual language practice did not comply with the normative tradition. Despite a relatively low number of tokens, Vosters et al. (2010: 104-105) reveal the surprising tendency that the letter writers in the corpus mainly use <d> (84%; 16 tokens) rather than the ‘Northern’ <dt> spelling (16%; 3 tokens) as widely favoured in the normative tradition. Based on data from the newly compiled *Going Dutch Corpus*, Section 4 follows up on these exploratory results by taking a more thorough look at the late eighteenth-century situation and, most importantly, the developments in actual language use after Siegenbeek’s (1804) officialised prescription in favour of <dt>.

4 Corpus analysis⁴⁷

4.1 Method

In the corpus analysis of the orthographic representation of final /t/ in second and third person single and second person plural indicative forms of verbs with a dental root ending in *-d*, three variants will be considered, viz. <dt>, <d> and <t>. These variants occur in the history of Dutch and also in the *Going Dutch Corpus*. It is

⁴⁶ The corpus used in the analysis of Southern Dutch, again, comprises handwritten texts from the judicial and administrative domains (Vosters 2011). The analysis of Northern practices, however, is based on a selection of 100 private letters from the 1780s, taken from the *Letters as Looft* corpus (Rutten & van der Wal 2014).

⁴⁷ Parts of this case study were also presented in Krogull (2018).

important to highlight that there is no phonetic difference between <dt>, <d> and <t>, all of them being pronounced as /t/ due to final devoicing.

The first and most complex variant <dt>, as prescribed by Siegenbeek and most eighteenth-century grammarians before him, is primarily grounded on morphological reasons. As a rule, the ending *-t* is attached to second and third person singular and second person plural present indicative verb forms (e.g. *hij neemt* < *neem* + *t* 'he takes'). <dt> is thus an analogical spelling in conformity with the stem + <t> principle. Moreover, <dt> can also be considered an etymological variant, implying that <dt> is a contracted form of the historical verb ending *-det* (e.g. Siegenbeek 1804).

Secondly, the variant <d> is based on the principle of uniformity (*gelijkvormigheid*), according to which all forms of *d*-stem verbs, irrespective of pronunciation, are spelled <d>. In other words, <d> in final position is analogous to those forms in which /d/ is pronounced, such as the infinitive (e.g. *vinden*).

Finally, the variant <t> represents the voiceless pronunciation of verb-final /t/, i.e. due to final devoicing in Dutch. This phonetic spelling, which dates back to the Middle Dutch period, is "based on the principle of a complete grapheme–phoneme correspondence" (Vosters et al. 2014: 84). While previous studies, including Vosters et al. (2014), only considered <dt> and <d> as the two main variants, I also take into account <t> as a third variant, given its occurrence in the *Going Dutch Corpus*.

By consulting a retrograde dictionary of Dutch (Nieuwborg 1969), searching for dental-root verbs ending in *-den* (infinitive forms) and subsequently checking their frequency in the *Going Dutch Corpus*, the following fifteen verbs⁴⁸ were selected (listed in order of decreasing frequency in the corpus):

- WORDEN; VINDEN; HOUDEN⁴⁹; MELDEN, ZOUDEN; RIJDEN; ZENDEN; BIEDEN; LIJDEN; WENDEN; TREDEN; ANTWOORDEN; RADEN; SCHEIDEN; BIDDEN

These verbs have to be regarded as search queries (e.g. *VINDT, *VIND, *VINT), as they also cover second and third person forms of derived verbs with prefixes, for instance *bevinden* and *ondervinden* in the case of VINDEN, and *aanhouden* and *behouden* in the case of HOUDEN. Furthermore, possible spelling variation such as *s/z* variation in *souden/zouden* and *senden/zenden* was also taken into account.

In terms of frequency, it has to be noted that the selected verbs considerably differ from each other. The auxiliary WORDEN 'to become' is by far the most frequent *d*-stem verb (456 tokens in the entire corpus), followed by

⁴⁸ It has to be noted that ZOUDEN, unlike the other fourteen *d*-stem verbs, is not the infinitive but an inflected form of the verb ZULLEN. However, as the second and third person forms end in /t/ (search queries: ZOUDT, ZOUD, ZOUT), it was decided to include ZOUDEN in this analysis.

⁴⁹ For some of these verbs, *d*-syncope might be relevant, for instance in *bouden* > *bou(w)en* > *ik bou*. Therefore, the use of <t> as in *hij bout* could also represent the stem + <t> rule. In the *Going Dutch Corpus*, these instances of <t> are rare, though.

VINDEN ‘to find’ (140 tokens) and HOUDEN ‘to hold’ (54 tokens). The frequency of most other verbs is relatively low in the *Going Dutch Corpus* (<50 tokens each).

4.2 Results

Giving an overview of the orthographic representation of final /t/ in second and third person singular and second personal plural present tense indicative forms of *d*-stem verbs, Table 1 shows the general distribution of variants across the two periods, based on data from the entire *Going Dutch Corpus*. The officially prescribed variant after 1804 (i.e. <dt>) is highlighted in light grey.

Table 1. Distribution of variants across time.

	Period 1: 1770–1790			Period 2: 1820–1840		
	<dt>	<d>	<t>	<dt>	<d>	<t>
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Total	62 (17.1)	258 (71.1)	43 (11.8)	293 (66.3)	138 (31.2)	11 (2.5)
WORDEN	34 (15.5)	157 (71.4)	29 (13.2)	158 (66.9)	71 (30.1)	7 (3.0)
–WORDEN	28 (19.6)	101 (70.6)	14 (9.8)	135 (65.5)	67 (32.5)	4 (1.9)

In the late eighteenth-century period, <d> turns out to be by far the most common variant, occurring in 71.1%. <dt> is considerably less frequent, with a share of no more than 17.1%. Against the background of a strong preference for <dt> in eighteenth-century normative works (Section 2), this is certainly a surprising result, indicating a discrepancy between norms and usage. The third variant <t> occurs in 11.8% of all instances.

In the early nineteenth-century, a striking increase of the prescribed spelling <dt> can be witnessed, becoming the main variant in actual usage with a share of 66.3%. The previously dominant <d> considerably loses ground, dropping from 71.1% to 31.2%. Nonetheless, its rather strong share of almost one-third shows that <d> does not disappear completely from language use. Furthermore, there are still a few remnants of <t> (2.5% of all instances), although it can no longer be considered a relevant variant in the nineteenth century.

While the orthographic analyses in this dissertation mainly focus on the effects of external variables, this case study also considers two internal factors as potential sources of influence on the distribution of variants, viz. (1) frequency and (2) grammatical person (second versus third person). To begin with, it was noted in Section 4.1 that the fifteen selected *d*-stem verbs differ in terms of frequency in the *Going Dutch Corpus*. Forms of auxiliary WORDEN ‘to become’ are by far the most frequent, and occur considerably more often than any other *d*-stem verb under

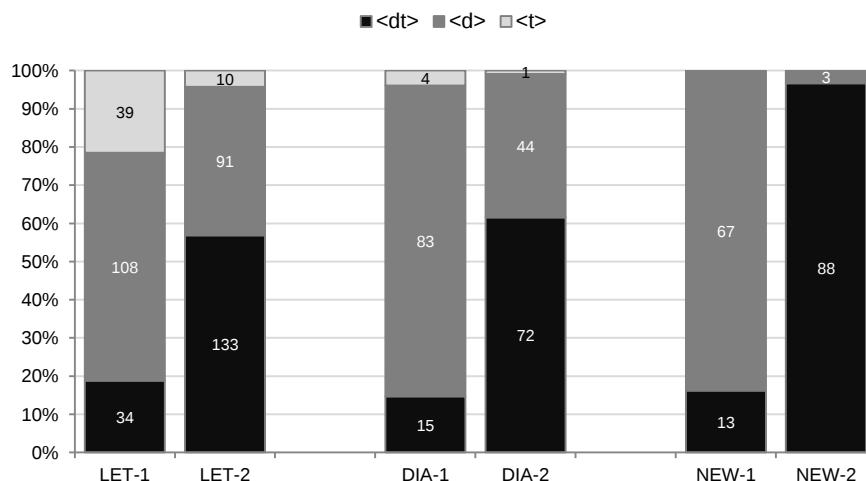
investigation. In fact, more than half of all 805 tokens (56.6%) are forms of WORDEN, whereas the remaining 43.4% comprise forms of the other fourteen *d*-stem verbs. In order to see whether the exceptionally high frequency of WORDEN has an effect on the distribution of variants, the overall results in Table 1 were split into two groups according to their frequency, viz. the highly frequent WORDEN versus the other *d*-stem verbs (referred to as –WORDEN). However, as the results in Table 1 show, frequency does not have an effect on the distribution of /t/ variants. In both periods, the variants are distributed very similarly, with <d> as the main variant in the eighteenth century (71.4% WORDEN vs. 70.6% other verbs) and <dt> in the nineteenth century (66.9% WORDEN vs. 65.5% other verbs).

The second internal factor that was initially considered is the difference between grammatical persons, viz. second person (singular/plural) versus third person (singular) forms. With regard to the multi-genre design of the *Going Dutch Corpus*, this distinction has its drawbacks, though. Generally, second person verb forms are most typically found in private letters (as a dialogical genre) rather than in diaries and newspapers, where they rarely occur given the monological character of these texts. Even in private letters, second person forms of *d*-stem verbs turn out to be relatively low in frequency. As a consequence, grammatical person as a supposedly internal factor would be strongly influenced by genre as an external factor. Therefore, the effect of grammatical person will not be investigated any further at this point.

Genre variation

Focusing on possible gender differences, Figure 1 presents the distribution of variants across the three genres of the *Going Dutch Corpus*, i.e. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW).

Figure 1. Distribution of variants across genre and time.



The eighteenth-century results display a striking similarity across all three genres, revealing that <dt>, as widely preferred in metalinguistic discourse, plays an equally marginal role in each of the sub-corpora, i.e. private letters (18.8%), diaries and travelogues (14.7%) and newspapers (16.3%). <d> is the main variant across all genres, particularly in diaries and travelogues (81.4%) and in newspapers (83.8%). The somewhat lower share of <d> in private letters (59.7%) can be explained by a relatively high frequency of the third variant <t> in these texts. In fact, <t> occurs in 21.5% in eighteenth-century private letters (which is, in fact, slightly higher than <dt>), but only very marginally in diaries and travelogues (3.9%) and not at all in newspapers. It seems that this phonetic spelling was merely a manuscript variant in the late eighteenth-century.

In the nineteenth-century period, the distribution of variants changes in the direction of Siegenbeek's prescription, as <dt> becomes the main variant across all three genres. Most remarkably, a complete shift from <d> to <dt> can be witnessed in newspapers, adopting the prescribed variant in 96.7% of all instances (previously 16.3%). In the two types of ego-documents, the increase of the prescribed variant is less drastic but still remarkable, increasing from less than 20% to 56.8% in private letters and to 61.5% in diaries and travelogues. At the same time, <d> remains a fairly strong second variant in ego-documents, both in private letters (38.9%) and diaries and travelogues (37.6%). The phonetic spelling <t>, on the other hand, practically disappears from usage with only a few remaining tokens in private letters (4.3%) and one single token in diaries and travelogues (0.9%).

Regional variation

Moving on to regional variation, Table 2 presents the distribution of variants across the seven regions (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

Table 2. Distribution of variants across region and time.

	Period 1: 1770–1790						Period 2: 1820–1840					
	<dt>		<d>		<t>		<dt>		<d>		<t>	
	N	%	N	%	N	%	N	%	N	%	N	%
FR	5	12.2	26	63.4	10	24.4	48	80.0	11	18.3	1	1.7
GR	10	15.4	51	78.5	4	6.2	22	44.0	19	38.0	9	18.0
NB	1	2.3	40	93.0	2	4.7	41	49.4	41	49.4	1	1.2
NH	21	47.7	16	36.4	7	15.9	47	64.4	26	35.6	0	0.0
SH	13	21.0	44	71.0	5	8.1	41	73.2	15	26.8	0	0.0
UT	10	15.6	43	67.2	11	17.2	55	76.4	17	23.6	0.0	0.0
ZE	2	4.5	38	86.4	4	9.1	39	81.3	9	18.8	0.0	0.0

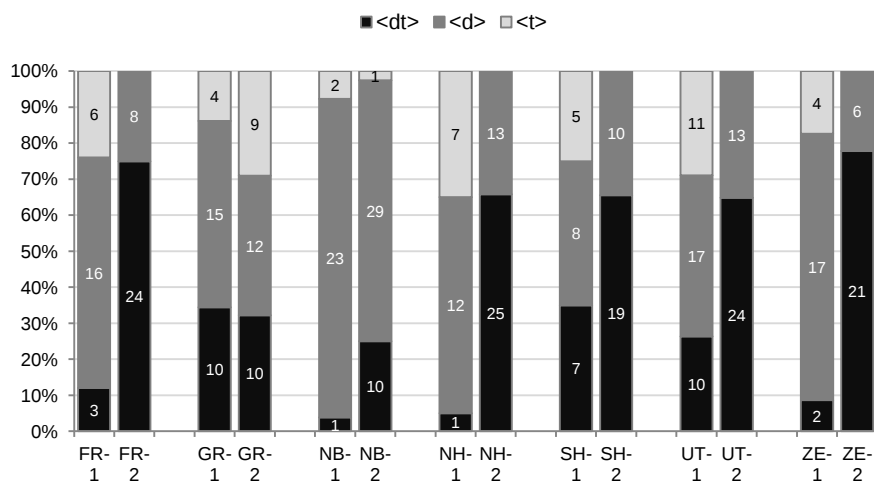
In all seven regions, the eighteenth-century usage data reveal the co-occurrence of three variants. The prevalent variant is <d>, except for the region of North Holland, where <dt> is more frequent (47.7%) than <d> (36.4%). Notably, <dt> is practically absent in North Brabant (2.3%) and also very marginal in Zeeland (4.5%). One might argue that these two border regions in the south of the Northern Netherlands were more strongly oriented towards Southern writing conventions, typically favouring <d> (Section 3), than the rest of the language area.

In the nineteenth-century results, prescribed <dt> gains ground across all regions, although its share in Groningen and North Brabant is below 50% each. It is particularly striking that <d> and <dt> are equally frequent in North Brabant (both 49.4%). It seems as if the orientation of North Brabant towards either the Southern or the Northern normative tradition was rather unclear in this period, resulting in a competing coexistence of prototypically Northern (<dt>) and prototypically Southern (<d>) norms⁵⁰ in usage. The phonetic <t> practically disappears across all regions, except for nine tokens in the Groningen data.

Regional variation across genres

Regional variation was further investigated by zooming in on genres. It has to be noted, though, that the absolute numbers of tokens are relatively low when we subdivide the occurrences of <dt>, <d> and <t> into two periods, three genres and seven regions. Therefore, all results should be regarded as rough tendencies.

Figure 2. Distribution of variants across region and time (private letters).

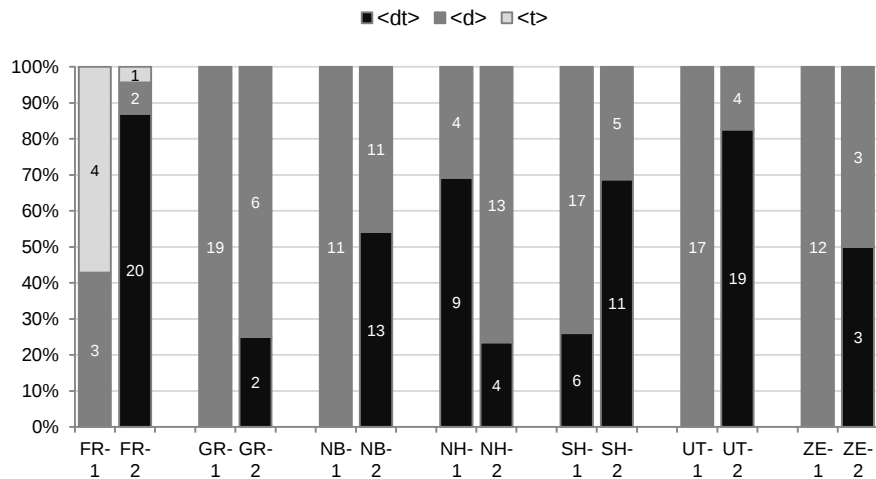


⁵⁰ As the overview provided by Vosters et al. (2014: 83-84) shows, <d> was the traditionally preferred variant by most eighteenth-century grammarians in the Southern Netherlands. Even after 1815, <d> was still prescribed in most Southern normative works, often alongside the official Northern variant <dt>.

Figure 2 presents the distribution of variants across regions in the sub-corpus of private letters. In eighteenth-century private letters, all three variants occur in usage in all seven regions, with <d> being the most frequently used variant across the entire language area. After Siegenbeek (1804), all regions shift to <dt> as the main variant in private letters with a share of at least 60% each, except for the regions of Groningen, where <d> remains stable, and most notably of North Brabant, where <d> maintains its prevalence with a share of 72.5%. It is striking that the only nineteenth-century remnants of the <t> spelling are also in these two regions.

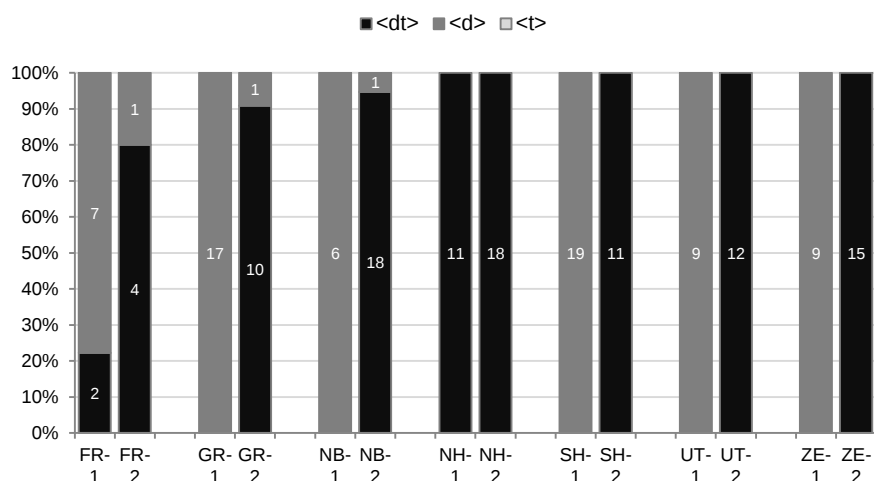
Next, the distribution across regions in the sub-corpus of diaries and travelogues is presented in Figure 3.

Figure 3. Distribution of variants across region and time (diaries and travelogues).



The results nicely illustrate that <dt> only occurs in the Holland area in the eighteenth-century period, whereas <d> is the only variant used in texts from the remaining regions, or co-existing with <t> in the case of Friesland. In the nineteenth-century period, prescribed <dt> emerges as a 'new' variant in diaries and travelogues from all regions, particularly in Friesland (87.0%) and Utrecht (82.6%), although its share considerably differs per region. In fact, the relative decrease of <dt> in North Holland is unexpected, but can also be a result of the low number of tokens.

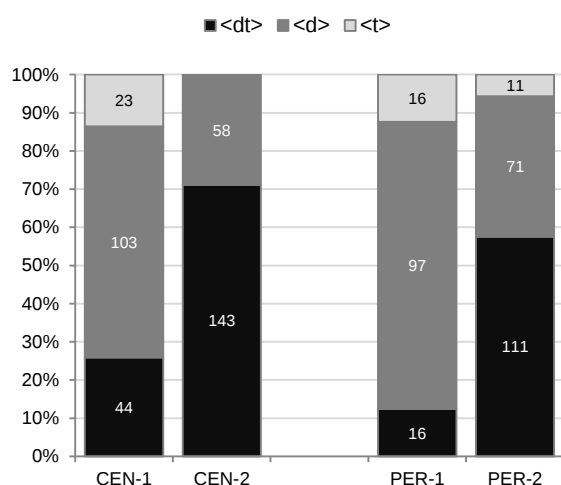
Finally, Figure 4 displays the distribution of variants across regions in the sub-corpus of newspapers. These results, in fact, provide a good example of a case, in which a remarkable tendency can be observed despite a low number of tokens. Whereas most late eighteenth-century newspapers have <d> as the only variant, newspapers from North Holland already use <dt> in all instances. In the newspaper data from Friesland, <d> occurs alongside a few instances of <dt>.

Figure 4. Distribution of variants across region and time (newspapers).

Diachronically, it is striking that newspapers from all regions shift from <d> to prescribed <dt> in the early nineteenth century, whereas North Holland maintains <dt> in both periods.

Variation across centre and periphery

Figure 5 shows the distribution of variants across the centre (CEN) and the periphery (PER), based on the entire *Going Dutch Corpus*.

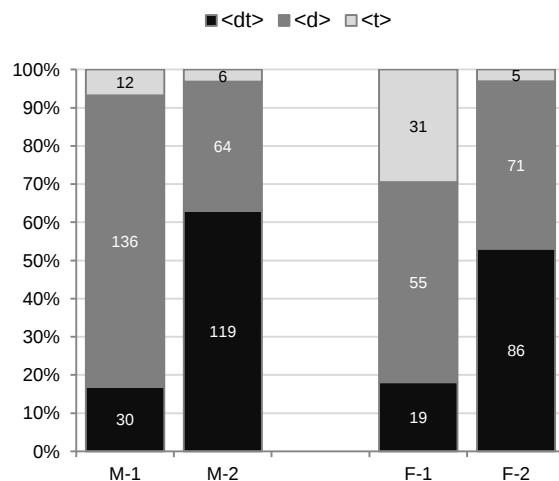
Figure 5. Distribution of variants across centre–periphery and time.

The results reveal quite some variation in both periods. In the eighteenth century, <dt> already appears to be more common in the centre (25.9%) than in the periphery (12.4%). The <d> variant is the main variant in the centre (60.6%), but particularly prevalent in the periphery with a share of 75.2%. Interestingly, there is no centre–periphery difference in the use of <t>, which occurs in approximately 13% in each of the two categories. In the nineteenth century, both the centre and the periphery shift to <dt> as the main variant in usage, although the centre adopts the prescribed spelling to a greater extent (71.1%) than the periphery (57.5%). Remarkably, all remnants of the phonetic spelling <t> are found in data from the periphery.

Gender variation

Figure 6 presents the distribution of variants across gender (M = male writers, F = female writers), based on data from the two sub-corpora of ego-documents.

Figure 6. Distribution of variants across gender and time.



The eighteenth-century results show that there is no gender difference in the use of <dt>, occurring almost equally low in ego-documents written by men (16.9%) and women (18.1%). Irrespective of gender, <d> is the preferred variant among male writers (76.4%) and, to a considerably lesser degree, among female writers (52.4%). Furthermore, the use of <t> reveals interesting gender differences, as it is much more common in texts by women (29.5%) than by men (6.7%). In other words, eighteenth-century women use <t> even more often than <dt>.

In the nineteenth century, the majority of both men and women shift to <dt>, i.e. in conformity with Siegenbeek's prescription. However, the shares of 63.0% and 53.1%, respectively, show that there is still a considerable degree of variation in terms of co-existing variants. <d> is still used by men (33.9%) and

particularly women (43.8%). Apart from a few remnants of <t> in both male and female data, this variant no longer plays a significant role in the nineteenth century.

5 Discussion

In this chapter, I investigated the orthographic representation of verb-final /t/ in second and third person singular and second person plural present tense indicative of *d*-stem verbs as either <dt>, <d> or <t>. Most notably, this case study revealed major changes in the effectiveness of language norms on actual language usage in the eighteenth and nineteenth centuries.

As outlined in Section 1, Siegenbeek's (1804) orthography officially prescribed <dt> as the standard variant for verb-final /t/, referring to this spelling as contracted forms of the historical verbal ending *-det*. His choice, however, was no innovation but rather a continuation of the more or less coherent eighteenth-century normative tradition. As discussed in Section 2, there had been a strong preference for <dt> in most metalinguistic comments and normative works throughout the eighteenth century (e.g. Moonen 1706; Verwer 1707; Sewel 1708/1712; Huydecoper 1730; Elzevier 1761, Zeydelaar 1774, Stijl & van Bolhuis 1776; Kluit 1777; van Bolhuis 1793; *Rudimenta* 1799; Weiland 1799; Wester 1799). Only a few grammarians either explicitly (van Belle 1748/1755) or implicitly (ten Kate 1723) advocated the <d> spelling, whereas a few others (de Haes 1764; van der Palm 1769) prescribed <dt> for second person and <d> for third person forms.

Even though the individual choices were motivated differently (morphologically and/or etymologically), the widely promoted and prescribed representation of verb-final /t/ was thus <dt>, both in the eighteenth-century normative tradition and in Siegenbeek's (1804) orthography. However, the corpus results (Section 4.2) revealed a completely different picture in actual language usage. Against the normative preference for <dt>, <d> appeared to be by far the most frequently used variant in the late eighteenth century – both in (handwritten) ego-documents and in (printed) newspapers. On the other hand, <dt> was merely one of the minor variants, alongside <t>. In other words, the corpus results signalled a clear discrepancy between language norms and language use in the eighteenth century. This is in line with earlier observations by Vosters et al. (2010). Based on a comparatively small corpus of 100 private letters from the 1780s, they also point out that language practice in the Northern Netherlands (84% <d> vs. 16% <dt>) did not coincide with the <dt>-dominated Northern normative tradition.

Keeping in mind that there was no considerable evolution in metalinguistic discourse on the orthographic representation of verb-final /t/ from the early 1700s until Siegenbeek (1804), the shift from <d> as the common variant in the late eighteenth century to the prevalent use of <dt> in the early nineteenth century is striking. It can be assumed that this change in spelling practices must have taken place under the influence of the top-down implementation of Standard Dutch and Siegenbeek's officialised spelling norms. More generally, these developments

indicate that the normative awareness and influence on actual language use were probably limited in the eighteenth century. Otherwise, the general normative preference for <dt> would have been more clearly reflected in the corpus results from that period.

In fact, the shift from <d> to <dt> as the main variant in language use was witnessed across all genres in the *Going Dutch Corpus*, both handwritten and printed. However, some genre-related differences could be observed. In nineteenth-century newspapers, <dt> was adopted almost invariably, much more than in ego-documents. In private letters, diaries and travelogues, <d> continued to be used as an alternative variant well into the nineteenth century. Similarly to the previous orthographic case study in Chapter 5, these findings indicate a specific genre gradation from newspapers (i.e. following the prescription most successfully) to private letters (i.e. highest degree of variation of both prescribed and alternative variants).

With regard to regional variation, some interesting tendencies emerged. In the eighteenth-century, <dt> was practically absent in the southernmost regions of North Brabant and Zeeland, while it already occurred in the rest of the language area. Even in the nineteenth century, North Brabant still maintained <d> in almost 50% of all instances, competing with the equally frequent <dt>. As North Brabant is a border region to the Southern Netherlands, the co-existence of <d> and <dt> even after 1804 can possibly be explained by a stronger orientation towards the <d>-promoting Southern normative tradition. In Zeeland, the other border region to the Southern Netherlands, however, the striking change from <d> to <dt> suggests a shift from a more Southern-oriented region in the eighteenth century to a more Northern-oriented region in the nineteenth century. Another interesting finding was revealed in the North Holland data. Here, <dt> was already found in late eighteenth-century newspapers and also in some diaries and travelogues, which is in sharp contrast to almost all other regions, still using <d> exclusively.

These regional tendencies were further supported by the results focusing on possible differences on the centre–periphery level. In the eighteenth century, <dt> was already more common in the centre, suggesting that the normative awareness in these parts of the language area was probably higher than in the periphery. After the introduction of Siegenbeek’s orthography, <dt> was also more successfully adopted in the centre.

In terms of gender variation, male writers of the post-Siegenbeek generation appeared to be faster in adopting the prescribed variant than their female contemporaries. This is interesting as there were no gender differences in the use of <dt> in the late eighteenth century, possibly indicating that the official spelling norm ‘reached’ female language users to a lesser extent (or less successfully) than male users.

To sum up, the corpus results for this case study give clear evidence of the effectiveness of Siegenbeek’s 1804 orthography on the representation of final /t/ in second and third person singular and second person plural present tense indicative forms of *d*-stem verbs. More generally, the developments in language practice suggest that language norms only became effective after they had been

codified officially and disseminated on a national level, given the discrepancy between eighteenth-century norms and usage. Nevertheless, the results also reveal a considerable degree of nineteenth-century variation with regard to genres as well as on the regional and gender dimensions, which disproves that spelling was entirely homogeneous in the Northern Netherlands after the *schrijftaalregeling* of 1804/1805.

Orthographic variables (3)

Word-medial and word-final /s/

1 Discussion in Siegenbeek (1804)

The variable discussed in this chapter is characterised by a discrepancy between pronunciation and spelling. In fact, the orthographic representation of word-medial and word-final /s/ (< Wgm. *sk) as <sch> or <s> in words like *men(sch)* ‘man, person’ and *tuss(ch)en* ‘between’ has been described as “a thorny problem right up to the present day, when the last relics of the ‘-sch’ spelling are being tidied away” (Gledhill 1973: 421). Historically, the spelling variable under investigation goes back to the West Germanic consonant cluster *sk, which originally occurred in all positions of a word, for instance initially in *skijp (> *schip* ‘ship’), medially in *waskean (> *was(ch)en* ‘to wash, to clean’), and finally in *buske (> *bos(ch)* ‘wood(s), forest’). It is generally assumed that the pronunciation as /sk/ was still very common in the Middle Dutch period. However, the variation found in spelling practices probably reflected an ongoing sound change. Whereas the <sc> spelling was used in initial and medial position, possibly still reflecting the /sk/ pronunciation, <sch> frequently occurred in final position⁵¹. The <sch> spelling had also been used in initial and medial position from the thirteenth century onwards. With regard to pronunciation, <sch> was maintained in initial position as the cluster /sx/ (< /sk/), but further reduced to /s/ in medial and final position (Rutten & van der Wal 2014: 49).

De Wulf et al. (2005: 18) argue that this position-dependent differentiation in pronunciation must have taken place in most Middle Dutch dialects. Although the remnants of this split can be found in present-day Standard Dutch (i.e. initial /sx/, but medial and final /s/), there is some more dialectal variation. In various dialects of North Holland and Friesland, for instance, /sk/ still occurs at least in word-initial position. Word-medially, /sk/ is incidentally found in some places in North Holland, too, whereas both /sk/ and /sx/ occur in the north-eastern regions (De Wulf et al. 2005: 24)⁵². In word-final position, clusters (either /sk/ or /sx/) are practically absent in the entire language area (De Wulf et al. 2005: 28-29).

⁵¹ In contrast, De Wulf et al. (2005: 18) argue that <sc> was the typical spelling only in initial position (but not in medial position, cf. also Rutten & van der Wal 2014: 49), as opposed to <sch> for both medial and final position. This actually suggests variation in medial position, probably signalling a change in both pronunciation and spelling practices.

⁵² The Wgm. cluster *sk in word-medial position has also been maintained in several Flemish dialects (cf. e.g. Taeldeman 2013). However, as this dissertation focuses on the Northern Netherlands, the situation in the Southern Netherlands will not be discussed here.

The phonological variation across positions still found in Dutch dialects also supports the claim that the <sch> spelling was preserved longer in word-medial position than in word-final position (Gledhill 1973: 426).

In his official orthography of Dutch, Siegenbeek (1804: 228-232) also addressed this orthographic feature, discussing “het gebruik der enkele *s*, of van den zamengestellten klank van *sch* in sommige woorden” ‘the use of the single *s*, or of the combined sound *sch* in some words’. Generally, he prescribed <sch> in these positions. In reference to the mistakes even made by ‘otherwise neat writers’, Siegenbeek (1804: 228) remarked the necessity to comment on this spelling issue:

Daar het niet vreemd is, ten aanzien der voorgestelde bijzonderheid, ook anderszins keurige schrijvers te zien mistasten, zal het noodzakelijk zijn, daarover in deze Verhandeling ook met een enkel woord te spreken.

‘As it is not unusual to see otherwise neat writers miscalculating with regard to the presented special case, it will be necessary to also say a few words about it in this treatise.’

To begin with, it has to be noted that Siegenbeek did not prescribe word-final <sch> in all cases. In fact, he discussed the crucial – and etymologically correct – difference between the spelling of adverbs and adjectives⁵³. According to Siegenbeek (1804: 229), adverbs like *dagelijks*, *bedendaags* or *vergeefs* (originally genitive forms) had to be spelled with word-final <s>, the homonymous adjectives required the <sch> (derived from the historical suffix *-isch*):

Dus behoort men ook *dagelijks*, *bedendaags*, *ginds*, *regts*, *links*, in den zin van bijwoorden, met eene enkele *s* te schrijven, als eigenlijk tweede naamvallen zijnde van de onderscheidene bijvoegelijke of zelfstandige naamwoorden, waarvan zij komen; doch diezelfde woorden, als bijvoegelijke gebezigd, vereischen de *sch*. Deze uitgang namelijk, bij verkorting voor *isch*, is bij ons een zeer gemeenzame uitgang van bijvoegelijke naamwoorden, het zij van andere bijvoegelijke, het zij van zelfstandige naamwoorden afkomstig, als *grootsch* van *groot*, *trotsch* van *trots* (*hoogmoed*), *aardsch* van *aarde*, *Haarlemsch* van *Haarlem*, *Amsterdamsch* van *Amsterdam* en honderd andere.

‘Therefore one also has to write *dagelijks*, *bedendaags*, *ginds*, *regts*, *links*, in the sense of adverbs, with a single *s*, as they are actually second cases of the distinct adjectives or nouns, where they derive from. But the same words, used as adjectives, require the *sch*. This ending, shortened for *isch*, is a very common ending for adjectives in our language, either deriving from other adjectives or nouns, as *grootsch* from *groot*, *trotsch* from *trots* (pride), *aardsch* from *aarde*, *Haarlemsch* from *Haarlem*, *Amsterdamsch* from *Amsterdam* and hundred others.’

⁵³ The adjectival suffix *-s* (here: Siegenbeek’s <sch>) derives from the Old Dutch *-sc/-sk*, e.g. in *himilisc* ‘heavenly’, and later *-sc/-sch/-s* in Middle Dutch. The unstressed *i* in this suffix had been dropped very early in Dutch (compared to other Germanic languages). The adverbial suffix *-s* (here: Siegenbeek’s <s>), on the other hand, was originally a genitive suffix, which can still be seen in fossilised temporal adverbials like ‘*s morgens*’ ‘in the morning’ (< *des morgens*) (cf. EWN).

The question arises whether there were any differences in pronunciation between *dagelijks* (adv.) and *dagelijksch* (adj.), or more generally, between word-final <s> and <sch>. Referring to the example of inflected forms of adjectives, Siegenbeek's (1804: 229) comment gives an idea of the differences found in contemporary (everyday) spoken language and writing:

Men vindt hiervoor, vooral bij vroegere schrijvers, in de verbuiging wel eens, overeenkomstig de dagelijksche uitspraak, *aardse*, *grootse*; doch deze schikking naar de spreektaal werd in den schrijfstijl, sedert lang, met regt verworpen.

'Especially among earlier writers, one sometimes finds *aardse*, *grootse* in the declension, in line with the daily pronunciation, but this compliance with the spoken language had been rightly rejected in the writing style for a long time'

This indicates that <s(ch)> in inflected forms such as *aards(ch)e* or *groots(ch)e* (i.e. in word-medial position) must have been pronounced as /s/ rather than /sx/. Although Siegenbeek acknowledged these changes in spoken language, his (conservative) spelling choice did not reflect them. It even contradicts his first spelling principle *Schrijf, zoo als gij spreekt* 'Write as you speak'. As can be seen in the quote above, at least in the case of adjectives, the variant <sch> is etymologically motivated. Siegenbeek explicitly referred to the original and very common (*zeer gemeenzame*) adjectival suffix *-isch*, shortened *-sch*.

A closer inspection of Siegenbeek's word list in the appendix of his orthography reveals that <sch> is the default spelling for both word-medial and word-final position – except for the group of adverbs mentioned above, as well as to make a distinction between homonyms such as *wassen* (*crescere*) 'to grow' and *wasschen* (*lavare*) 'to wash, to clean'⁵⁴ (1804: 231). The word-medial and word-final <sch> words cover practically all parts of speech such as nouns (*mens(ch)*, *vlees(ch)*), verbs (*wens(ch)en*), adjectives (*Hollands(ch)*) and prepositions (*tuss(ch)en*). This range will also be reflected in the corpus analysis in Section 4.

2 Eighteenth-century normative discussion

Despite the striking differences between spoken and written language, orthographic variation between <sch> and <s> in word-medial and word-final position was rarely discussed in eighteenth-century metalinguistic discourse. The widespread preference among grammarians for <sch> becomes evident in their use of this variant, while normative comments or even explicit rules on this spelling issue were surprisingly sparse.

In the early eighteenth century, Sewel (1712) at least briefly mentioned his preference for <sch> in both word-medial and word-final position:

⁵⁴ Etymologically, Siegenbeek's different spellings of these homonyms are indeed grounded on two different Old Dutch words: *wassen* < *wassan* 'to grow', and *wasschen* < *waskan* 'to wash, to clean' (cf. *EWN*).

Ook voegtze wel in [...] eysch, mensch, aardsch, wereldsch, Duytsch, Engelsch” (1712: 13)

Hierom behoort men te spellen, [...] tusschen, menschen” (1712: 41)

‘It is also used well in [...] eysch, mensch, aardsch, wereldsch, Duytsch, Engelsch
Therefore one has to spell [...] tusschen, menschen’

Interestingly, Sewel (1712: 33) attested that there was no difference in pronunciation between <sch> and <s>, when he discussed the case of homonymous *bosch* ‘wood(s), forest’ and *bos* ‘bunch’:

Het woord *Bosch* spreekt men gemeenlyk maar uyt, *bos*; evenwel is het best zich aan de oude spelling te houden; want behalve dat *bos* zoude kunnen genomen worden voor *bus*, zo kan met ‘t onderscheyd der woorden door een verscheelende spelling gevoeglyk betekenen, als *Bosch* [een woud], en *bos* [een bondel], alhoewel men zou moogen zeggen dat het laatste zynen oorsprong heeft van ‘t eerste, omdat een *Bos* pennen schynt te zyn een *bosch* van pennen [sylva pennarum].

‘The word *Bosch* is commonly pronounced as *bos*. However, it is best to maintain the old spelling, because apart from the fact that *bos* could be taken for *bus*, one can properly signify the difference of the words by a different spelling, as *Bosch* [woods], and *bos* [bundle], although one would say that the latter derives from the former, because a *Bos* pens seems to be a *bosch* of pens [sylva pennarum].’

In these cases, the ‘old spelling’ <sch> was thus (also) maintained to avoid the identical spelling of homonyms written with <s>. The issue of homonyms with <sch> and <s>, in fact, reoccurs several times throughout the eighteenth-century tradition. Stijl & van Bolhuis (1776: 58), for instance, also illustrated the *bosch*/*bos* example. According to van Rhyn (1758: 25f.), the <sch> spelling helps to distinguish between the two homonymous verbs *wassen in de zĳn van groeijen* ‘to grow’ and *wasschen, in de zĳn van rĳnigen* ‘to wash, to clean’. Three years later, Elzevier (1761: 127) mentioned the same example.

A remarkable comment on the contemporary pronunciation in word-final position is found in van der Palm (1769: 23), who explicitly rejected the (frequently occurring) spelling with single <s>:

Vr. Moeten de woorden *visch*, *mensch*, enz. niet enkel geschreven worden *vis*, *mens*, enz.

Antw. Men mag de *ch* in de woorden *visch*, enz. niet verwerpen, schoon zulks van vele onkundigen geschiedt: want zulks zoude tegen den aert der woorden en de rechte uitspraek stryden.

Q. Don’t the words *visch*, *mensch*, etc. just have to be written *vis*, *mens*, etc.?

A. One must not reject the *ch* in the words *visch*, etc., although this happens among many unknowing people. Because that would contradict the nature of the words and the right pronunciation.’

Unfortunately, van der Palm did not elaborate on the *aert der woorden* 'nature of the words' and particularly the *rechte uitspraak* 'right pronunciation' in word-final position.

A more elaborate discussion of the phonetic nature of <sch> was suggested by Zeydelaar (1774). Generally, he advocated <sch> for all positions and explicitly rejected the single <s> in word-final position. Whereas Zeydelaar (1774: 69-70) acknowledged that final <sch> had been realised as /s/ by many language users, he still heard a *blaasend geluid* 'blowing sound' in medial position:

Aan het einde der woorden klinkt *sch* niet anders dan eene enkele *s* en men moet ze ook niet anders uitspreken, in

mensch	boersch	hollandsch
wensch	helsch	zweedsch
hemelsch	visch	Fransch
aardsch	rinsch	duitsch
trotsch	rusch	Engelsch.
gottisch		

Veelen hebben de bovenstaande woorden, omdat *sch* daar als eene enkele *s* in klinkt, reeds begonnen met eene enkele *s* te schrijven; 't geen men niet naarvolgen moet.

Niettegenstaande de *sch*, in de opgegevene woorden, als eene enkele *s* wordt uitgesproken, zoo bekomt de *sch* wederom haar blaasend geluid, zoo draa die woorden meervoudig gemaakt of in derzelver buiging geschreeven worden, op de volgende wijze.

Menschen	boersche	hollandsche
wenschen	helsche	zweedsche
hemelsche	visschen	fransche
aardsche	rinsche	duitsche
trotsche	russchen	engelsche.
gottische		

'At the end of words *sch* does not sound different from a single *s*, and one should not pronounce it differently either, in [...]

Many have already started to write the abovementioned words with a single *s*, because *sch* sounds like a single *s* here, which one should not follow.

Notwithstanding the fact that *sch* is pronounced as a single *s* in the listed words, the *sch* receives its blowing sound again, as soon as those words are in the plural or written in their inflection, in the following way [...]

Zeydelaar (1774: 70-71) further argued that despite the same pronunciation in word-final position, <sch> must not be confused with <s> in homonymous words, illustrated by the often-cited examples *bosch/bos*, *wasch/was* and so forth.

What becomes evident is that in contrast to the coherent preference for <sch>, eighteenth-century comments on the phonetic background of this spelling were more heterogeneous. Five decades before van der Palm and Zeydelaar, ten Kate (1723 II: 74) attested the /s/ pronunciation in both word-medial and word-final position:

Onder Zagstaertigen hebben wij nog tweederhande Uitgangen, om een Bijnaemliken of *Adjectivalen* Zin aen een Woord te geven; naemlijk, (I) ISCH of ISCHE (bij inkrimping SCHE of SCH, en, na 't gemak van de uitspraek, in Spreek- en Lees-tael slegts SE of S), als GROOTSCH of GROOTS, *Magnificus*, *Superbus*, van GROOT *Magnus*; AERDSCH *terrenus*, van AERD, *terra*.

‘Among the softly-tailed we still have two kinds of endings to give an adjectival meaning to a word, namely (I) ISCH or ISCHE (in the case of reduction SCHE or SCH, and, for the ease of pronunciation, in spoken and reading language only SE or S), as GROOTSCH or GROOTS, *Magnificus*, *Superbus*, from GROOT *Magnus*; AERDSCH *terrenus*, from AERD, *terra*.’

At least in the specific case of adjectival suffixes, ten Kate’s approach seems remarkably tolerant, which is characteristic of his so-called common orthography (as opposed to his so-called critical orthography, see van der Wal 2002a). In fact, he acknowledged – and possibly accepted – both the etymological <sch> spelling (shortened from *-isch(e)*) and the <s> spelling derived from the *Spreek- en Lees-tael* ‘spoken and reading language’. Ten Kate mentioned the example of *mens(ch)*, acknowledging <sch> and <s> as possible contemporary variants, i.e. *Ménsche*, *Ménsch* or *Méns*. His stance on <s> in writing remains ambiguous, though. Gledhill (1973: 433f.) argues that ten Kate “has no desire to reflect this pronunciation in the spelling (‘schrijf-tael’), though he does not condemn it in the spoken language”. He adds that <sch> is the most common form in ten Kate’s publication, although his usage of variants is rather erratic, with co-occurring forms such as *tussen* and *tussen*, or *Nederduitsche* and *Hoogduitsen* (Gledhill 1973: *ibid*).

The most striking exception to the <sch>-promoting majority of eighteenth-century normative works is van Belle (1748; 1755), who was the only grammarian in this century to explicitly reject the <sch> spelling in word-final position:

[...] nooit naa de S in *Wensch*,
Mensch, *Valsch* óf *Fransch*: het is genoeg *Vals*, *Mens*.
Schaap, *Schep*, *Schip*, *Schop* zyn ligtlyk uit te spreken,
 Maar *hémelSCH Goed* blyft in de keele steeken:
Duitsch, *Hollandsch Geld*, *Moorsch Goud*, *Helsch Spel*, *Aardsch Guit*
 Spreekt nimmermeer een Neederlander uit. (1748: 12)

[...] never after the S in *Wensch*,
Mensch, *Valsch* or *Fransch*: *Vals*, *Mens* is enough.
Schaap, *Schep*, *Schip*, *Schop* are easy to pronounce,
 But *hémelSCH Goed* sticks in one’s throat:
Duitsch, *Hollandsch Geld*, *Moorsch Goud*, *Helsch Spel*, *Aardsch Guit*
 Are no longer pronounced by a Dutchman.’

Here, van Belle unambiguously referred to the changes that had taken place in spoken language. In contrast to word-initial position, /sx/ in word-final position had no longer been pronounced and, moreover, was much more difficult to realise. Not only was van Belle the first grammarian since Winschoten’s *Letterkonst* of

1683 (cf. Gledhill 1973: 431) to reject <sch> in non-initial position – he remained a progressive exception throughout the eighteenth century. As Gledhill (1973: 436) concludes, “tradition was still too well-set for the public to adopt this spelling on a large scale”.

In sum, the vast majority of eighteenth-century grammarians preferred the historical spelling as <sch> in both word-medial and word-final position. Late eighteenth-century grammarians such as Wester (1799) and Weiland (1799) continued to advocate <sch> rather than <s>, paving the way for Siegenbeek (1804), who followed his predecessors and officialised the conservative variant in his national orthography. Eighteenth-century comments on the contemporary pronunciation remain somewhat vague. As outlined in this section, there was more dissent on this issue, ranging from /s/ in both positions (ten Kate 1723) to /s/ in final but a ‘blowing sound’ in medial position (Zeydelaar 1774). However, as this case study focuses on word-medial and word-final <sch>/<s> as an orthographic variable, possible variation and change in pronunciation will not be investigated further at this point.

3 Previous research

Although *sch/s* variation can be regarded as a controversial orthographic variable, illustrating the tension between pronunciation and spelling, only very little (socio)linguistic and/or corpus-based research has been conducted so far. Gledhill (1973), however, has a comprehensive chapter on this feature. Providing a useful, critical outline of <sch>/<s> in different positions, his chapter comprises the long history of this spelling feature from the Middle Dutch period until the second half of the twentieth century. It is particularly interesting to see that although attempts to get rid of non-initial <sch> had been made since 1683 (Winschooten), the traditional <sch> spelling was only abolished as late as 1934 “when Marchant put the first official nail in its coffin” (Gledhill 1973: 426).

Based on data drawn from the *Letters as Loot* corpus, Rutten & van der Wal (2014: 49–54) investigate the orthographic representation of reflexes of Wgm. **sk* in seventeenth- and eighteenth-century private letters. Their case study focuses on the regions of North Holland and Amsterdam, where the /sk/ pronunciation was still found word-initially. Thus, the <sc> and <sk> spellings can be considered as (unconventional) localisable variants reflecting the dialectal /sk/ pronunciation, as opposed to <sch> as the (conventional) supraregional variant. However, Rutten & van der Wal’s (2014) case study reveals that supraregional <sch> was by far the most frequently used spelling in initial position, even in the dialect areas maintaining the /sk/ pronunciation. In other words, the localisable <sc>/<sk> spellings as possible dialect interference with written language could only be attested in a minority of seventeenth- and eighteenth-century letters, mainly from less-experienced writers (i.e. lower and lower middle class, as well as women).

4 Corpus analysis

4.1 Method

For the analysis of this orthographic variable, two variants will be considered, which appear as the main variants in the normative discussion as well as in the *Going Dutch Corpus*, viz. <sch> and <s>. Alternative spelling variants such as <s>/<sse>, <s>/<sze>, <sg>/<sg(h)e> (cf. Gledhill 1973: 447-453) occur only very marginally in the corpus data (less than 15 occurrences altogether) and will therefore be left out of consideration in the corpus analysis. Furthermore, there are no attestations of the historical spellings <sc> or <sk> in word-medial or word-final position in the corpus at all.

In order to assess the actual use of variants in late eighteenth- and early nineteenth-century texts, the ten most frequent words containing <sch>/<s> in word-medial and/or word-final position were selected according to their occurrences in the *Going Dutch Corpus* (listed in order of decreasing frequency in the corpus):

- WENS(CH); TUSS(CH)EN; MENS(CH); FRANS(CH); ENGELS(CH); VIS(CH);
HOLLANDS(CH); DUTTS(CH); TURKS(CH); VLEES(CH)

These items are best regarded as sets of words with <s(ch)> in word-medial or word-final position, covering search queries such as *MENS(CH) (i.e. word-final position with occurrences of, e.g., *mens*ch and *medemens*ch), *MENS(CH)E* (i.e. forms with <s(ch)> in word-medial position, e.g. *mens*chen or *mens*chelij*ke*) or *TUSS(CH)EN* (i.e. *tuss*(ch)*en* as well as *ondertuss*(ch)*en* and *tuss*(ch)*entyd*). Possible spelling variation, for instance *ij/y* variation in the DUTTS(CH) queries, was also taken into account. In line with Siegenbeek's discussion on the <sch>/<s> spelling, the corpus analysis also comprises various parts of speech, including nouns (*mens*(ch)), verbs (*wens*(ch)*en*), adjectives (*Hollands*ch) and prepositions (*tuss*(ch)*en*).

All selected words (except for FRANS(CH)) are, at least in some form, explicitly mentioned in Siegenbeek's word list in the appendix of his 1804 orthography, prescribed to be spelled with <sch>. FRANS(CH) spelled with <sch> occurs several times throughout Siegenbeek's treatise.

4.2 Results

Table 1 shows the distribution of variants across the two diachronic cross-sections and in the entire *Going Dutch Corpus*. The prescribed Siegenbeek variant (i.e. <sch>) is highlighted in light grey.

In the late eighteenth-century period, <sch> is the most frequently used variant, occurring in 72.7% of all instances. The alternative <s>, however, can hardly be regarded as marginal with a share of 27.3%. Despite the general

eighteenth-century normative preference for <sch>, the ‘phonetic’ spelling <s> does occur relatively frequently in actual language usage.

Table 1. Distribution of variants across time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<sch>		<s>		<sch>		<s>	
	N	%	N	%	N	%	N	%
Total	607	72.7	228	27.3	690	95.0	34	5.0
Word-medial	487	78.0	137	22.0	518	96.1	21	3.9
Word-final	120	56.9	91	43.1	172	92.0	15	8.0

In the early nineteenth-century period, i.e. after <sch> had officially been prescribed by Siegenbeek (1804), this traditional spelling further consolidates its position as the dominant variant, increasing to 95.0%. This means that the <s> spelling considerably loses ground to 5.0% in the entire corpus – regardless of its more accurate representation of the contemporary pronunciation. Generally speaking, this orthographic variable undoubtedly reveals an immediate effect of the national spelling regulation on language practice.

With respect to the diverse eighteenth-century comments on the phonetic nature of <sch> in different positions in a word, one might assume that this internal factor also affects the distribution of spelling variants. Historically, the position had an influence on the use of spelling variants (cf. Section 1). The ‘old’ spelling <sch> was preserved much longer in word-medial position (e.g. in *tusschen* and *menschen*) than in word-final position, “especially in the Southern dialects” (Gledhill 1973: 426).

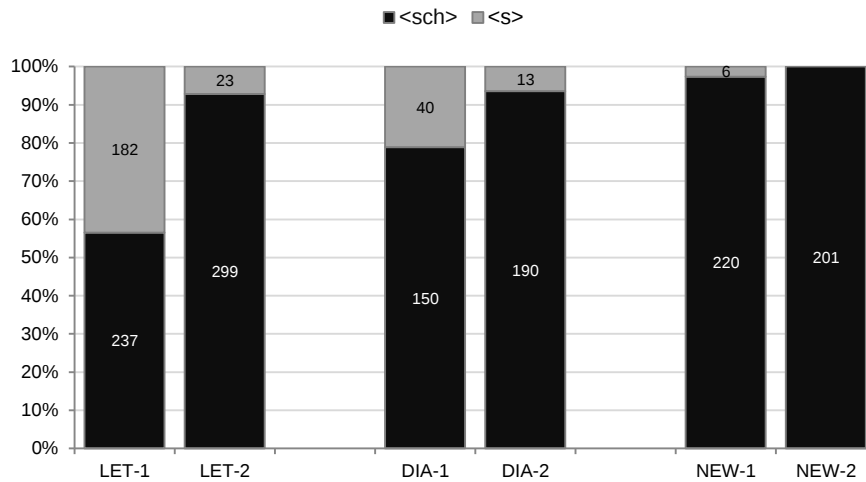
In order to take into account the possible effects of this internal factor, Table 1 also shows the results for both positions separately. It turns out that in the period before Siegenbeek, there is indeed a difference in distribution. Whereas <sch> is clearly the main variant in word-medial position with a share of 78.0%, the two co-occurring variants <sch> and, to a lesser extent, <s> are more evenly distributed in word-final position, viz. 56.9% versus 43.1%. These position-dependent differences level out after Siegenbeek, although <s> is still somewhat higher in word-medial position (96.1% as opposed to 92.0% word-finally).

One might carefully argue that the eighteenth-century findings confirm the previous observation that word-medial <sch> was maintained longer than word-final <sch>. Or do they even reflect a possible difference in eighteenth-century pronunciation, as argued by Zeydelaar (1774)? A more fine-grained analysis is needed, which is why I return to this internal factor by taking a closer look at the position-dependent distribution across genres.

Genre variation

Figure 1 shows the relative distribution of variants across the three genres of the *Going Dutch Corpus*, viz. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW).

Figure 1. Distribution of variants across genre and time.



Genre differences are most evident in the eighteenth-century data. The results for private letters display the highest degree of variation between the two variants. While <sch> (56.6%) is slightly more frequent than <s> (43.4%), both variants coexist in letter writing. In diaries and travelogues, the prevalence of <sch> is more pronounced with a share of 78.9%, whereas <s> is a less frequent variant, compared to private letters. In newspapers from this period, <sch> is already used in 97.3%, with only six tokens of the alternative <s>. It appears that <sch> had probably been established as the default variant in printed, published texts by the late eighteenth century.

The nineteenth-century results show that the diachronic developments from the pre-Siegenbeek to the post-Siegenbeek period lead to the same situation in all three genres. With very little room for variation, <sch> is established as the (only) prevalent variant in the period from 1820–1840. Whereas the distribution remains stable in newspapers (i.e. <sch> in all instances), the change is much more visible in the two types of ego-documents. The share of <sch> increases from 78.9% to 93.6% in diaries and travelogues, and, most strikingly, from 56.6% to 92.9% in private letters. In other words, <s> practically vanishes from language usage, even in the handwritten sources.

As shown in Table 1 before, the position in a word reveals some interesting differences between word-medial and word-final <sch>/<s>. With regard to the remarkable degree of genre variation, particularly in the eighteenth century, I return to this internal factor by zooming in on the genre-specific distribution of

<sch>/<s> in word-medial and word-final position separately. The results are presented in Table 2.

Table 2. Position-dependent distribution of variants across genre and time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<sch>		<s>		<sch>		<s>	
	N	%	N	%	N	%	N	%
LET medial	172	59.5	117	40.5	211	92.1	18	7.9
LET final	65	50.0	65	50.0	88	94.6	5	5.4
DIA medial	122	87.8	17	12.2	142	97.9	3	2.1
DIA final	28	54.9	23	45.1	48	82.8	10	17.2
NEW medial	193	98.5	3	1.5	165	100	0	0.0
NEW final	27	90.0	3	10.0	36	100	0	0.0

According to the eighteenth-century results, word-medial position is more likely to have the <sch> spelling than word-final position – across all genres. However, while the distribution of variants across positions is relatively similar in private letters (59.5% medial versus 50.0% final) and newspapers (98.5% medial versus 90.0% final), there is a striking difference in the sub-corpus of diaries and travelogues. In these sources, <sch> occurs in 87.8% of all instances in word-medial position, but only in 54.9% in word-final position.

In the nineteenth-century results, the discrepancy between word-medial and word-final position in diaries and travelogues largely levels out. Its remnants are still noticeable, though. Whereas <sch> has a share of 97.9% in word-medial position, it is used in ‘only’ 82.8% in word-final position. No such difference can be attested in private letters and newspapers from the same period.

The question arises why the position of <sch>/<s> in a word only seems to affect the distribution of variants in diaries and travelogues, but not in the other two genres. First of all, one can presume that the position-dependent differences in diaries and travelogues are not purely based on differences in pronunciation. If this was the case, one would expect a similar or even more remarkable difference in private letters, i.e. the genre closest to authentic spoken language. Instead, variation across positions in a word turns out to be genre-specific, only occurring in diaries and travelogues. As briefly discussed in Section 1, the <sch> spelling (as well as the pronunciation of clusters /sx/ and regionally /sk/) has generally been assumed to be preserved longer in word-medial position, whereas <s> (reflecting the reduced /s/ pronunciation) became conventional in word-final position. With regard to the corpus results, this situation of position-dependent variation and change in spelling practices (and possibly pronunciation) is actually only visible in the eighteenth-century diaries and travelogues. Remarkably, diarists seem to maintain the position-dependent distinction between medial <sch> and final <s>, perfectly in line with

traditional historical phonologies of Dutch, at least for earlier periods. Based on the corpus data and contemporary grammarians, however, it is unlikely that late eighteenth-century Dutch still maintained this distinction in phonology. In writing, neither private letters nor newspapers appear to reflect or maintain such a position-related distinction. The results for diaries and travelogues are thus unexpected and potentially indicate a more conservative writing tradition that was (genre-) specifically preserved in these sources.

In any case, these quantitative findings once again support the assumption that the genre of diaries and travelogues takes a special intermediate position in the corpus design between private letters and newspapers. In the eighteenth century data, the distribution of variants in word-final position is actually very similar to that in the private letters. Word-medially, on the other hand, it is relatively close to that in the newspapers. One might conclude that the use of the traditional spelling <sch> (regardless of the contemporary pronunciation) was somehow maintained longer in diaries and travelogues than in private letters, at least in word-medial position.

Diachronically, however, the developments towards <sch> as the (only) dominant spelling in nineteenth-century usage are witnessed for both word-medial and word-final position. Therefore, this internal factor will not be considered in the investigation of the remaining external variables.

Regional variation

The relative distribution of variants across the different regions in the *Going Dutch Corpus* is presented in Table 3 (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

Table 3. Distribution of variants across region and time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<sch>		<s>		<sch>		<s>	
	N	%	N	%	N	%	N	%
FR	86	72.3	33	27.7	109	94.8	6	5.2
GR	66	66.0	34	34.0	89	87.3	13	12.7
NB	84	74.3	29	25.7	91	91.9	8	8.1
NH	96	68.1	45	31.9	98	94.2	6	5.8
SH	101	75.9	32	24.1	103	99.0	1	1.0
UT	84	82.4	18	17.6	111	100	0	0.0
ZE	90	70.9	37	29.1	89	97.8	2	2.2

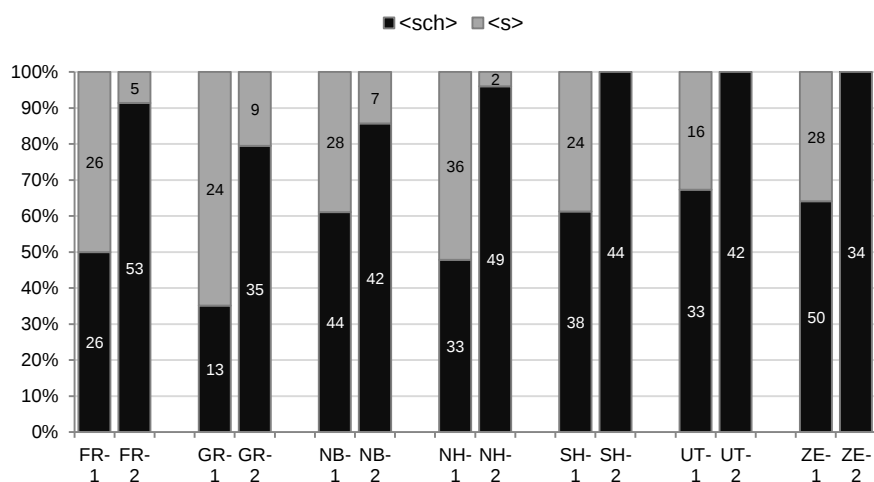
Generally, <sch> is the most frequently used variant in the late eighteenth century in all seven regions. Some degree of regional variation can certainly be attested, though. The highest share of 82.4% is found in Utrecht, whereas Groningen has a comparatively low share of 66.0%. Despite these differences, no clear regional patterns become evident.

The same is true for the early nineteenth-century results, in which <sch> occurs in approximately or even more than 90% in all seven regions. Again, the region of Utrecht leads with <sch> being used in 100% of all instances. Like in the first period, the most instances of the alternative <s> spelling are still found in the Groningen data (12.7%).

Regional variation across genres

Taking into account the genre differences attested before, the regional distribution of variants was also investigated across the three genres. Zooming in on private letters first, Figure 2 reveals some more regional variation.

Figure 2. Distribution of variants across region and time (private letters).

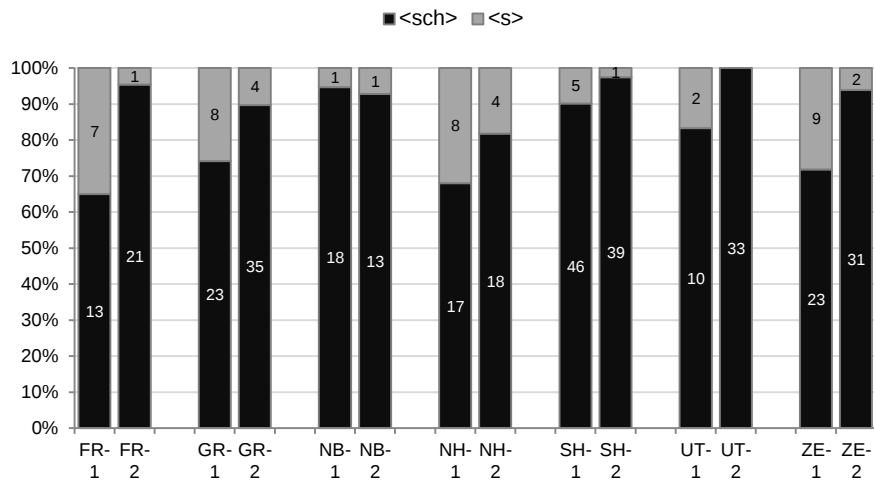


The eighteenth-century results show that <sch> is the most frequently used spelling in private letters from North Brabant (61.1%), South Holland (61.3%), Zeeland (64.1%) and particularly Utrecht (67.3%), but not in all regions. Most strikingly, <s> has a relatively high share of 64.9% in Groningen, where it outnumbers <sch>. Moreover, the two variants are (almost) equally frequent in Friesland and North Holland. This suggests a tendency that <s> is more common in private letters from the northern regions than from the southern ones. In the nineteenth-century period, all seven regions shift to <sch>. In South Holland, Utrecht and Zeeland, the prescribed variant even reaches 100%, and 96.1% in North Holland. The remnants of <s> primarily occur in letters from the peripheral

regions of Friesland, Groningen, and North Brabant. Once again, the highest share of <s> is attested in Groningen (20.5%).

Figure 3 shows the regional distribution of variants in the sub-corpus of diaries and travelogues.

Figure 3. Distribution of variants across region and time (diaries and travelogues).



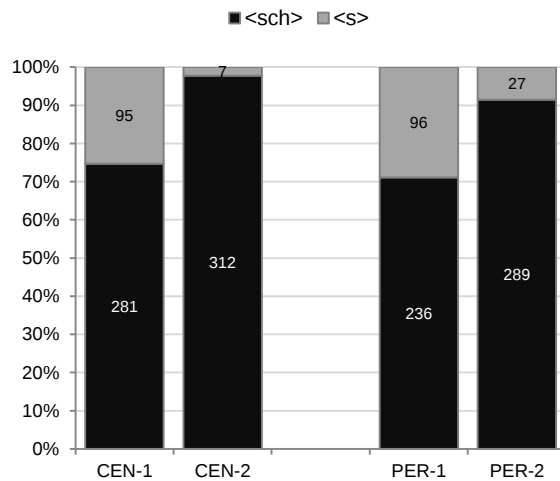
Unlike the results for private letters, <sch> is the most frequently used variant in eighteenth-century diaries and travelogues from all seven regions. It is particularly prevalent in North Brabant (94.7%) and South Holland (90.2%). On the other hand, the share of the alternative spelling <s> is comparatively high in Friesland (35.0%) and North Holland (32.0%) and, to a slightly lesser extent, Zeeland (28.1%) and Groningen (25.8%). In the nineteenth-century period, diaries and travelogues from all seven regions have <sch> as their prevalent variant, ranging from 100% in Utrecht to 81.8% in North Holland.

As shown in Figure 1, no variation could be attested in the newspaper data, except for a very low number of <s> in the eighteenth-century period (i.e. six occurrences). Therefore, it is redundant to present the distribution across regions and time in the newspaper data.

Variation across centre and periphery

Spatial variation in the distribution of variants was further investigated with regard to the centre–periphery distinction (CEN = centre, PER = periphery), as presented in Figure 4.

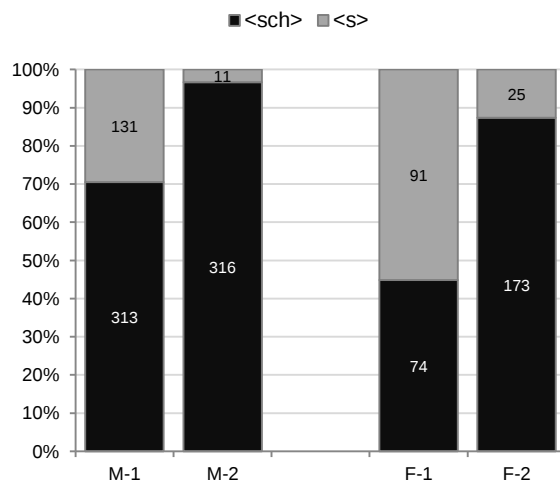
The differences between the centre and the periphery turn out to be fairly marginal in both periods. In the eighteenth-century data, <sch> is the dominant variant both in the centre (74.7%) and in the periphery (71.1%).

Figure 4. Distribution of variants across centre–periphery and time.

In the nineteenth century, the share of the prescribed variant <sch> increases to 97.8% in the centre, being practically the only variant in usage. In the periphery, we find more or less the same distribution, although the predominance of <sch> is slightly weaker than in the centre with a share of 91.5%.

Gender variation

Figure 5 shows the distribution of variants across gender (M = male writers, F = female writers), based on data from the two sub-corpora of ego-documents.

Figure 5. Distribution of variants across gender and time.

The results actually reveal striking gender differences, particularly in the late eighteenth-century period. Men use <sch> as the most frequent variant, occurring in 70.5% of all instances. In contrast, <s> is the most frequently used variant among women of the same period with a share of 55.2%. In fact, the ‘phonetic’ spelling co-occurs with the slightly less frequent <sch> (44.8%) in ego-documents produced by women. In the early nineteenth century, both genders increasingly use <sch> in conformity with Siegenbeek, i.e. male writers in 96.6% and, slightly less pronounced, female writers in 87.4%. It seems evident that Siegenbeek’s prescription in favour of <sch> affected both genders. Even though the spelling norm was adopted more frequently by men than by women, its effect is most visible in the strong increase of <sch> in the female data.

5 Discussion

The orthographic variable investigated in this chapter can certainly be regarded as a good example of an effective spelling norm. In Section 1, it was pointed out that Siegenbeek (1804) prescribed the historical <sch> spelling for Wgm. *s~~k~~ in word-medial (e.g. *tusschen*) and word-final position (e.g. *Hollandsch*), rejecting the alternative variant <s>. This was a conservative choice in two respects. First of all, Siegenbeek followed the vast majority of grammarians, who used and advocated the <sch> spelling in their normative works throughout the eighteenth century (Section 2). Secondly, by selecting <sch> as the standard variant, Siegenbeek also chose not to reflect the changes that had taken place in spoken language. As acknowledged by several grammarians including Sewel (1712), ten Kate (1723), van Belle (1748; 1755) and even Siegenbeek (1804) himself, word-medial and word-final <sch> had actually been pronounced as /s/ rather than the fricative cluster /sx/ still found in word-initial position. This is interesting as the preference for the traditional, etymologically motivated <sch> spelling actually contradicts Siegenbeek’s first fundamental spelling principle, viz. *Schrijf, zoo als gij spreekt* ‘Write as you speak’.

Regardless of the officialised discrepancy between pronunciation and spelling, the corpus analysis in Section 4 revealed that the prescribed variant <sch> was almost invariably adopted in early nineteenth-century usage. In the late eighteenth-century data, <s> still occurred in more than one-fourth of all instances. In private letters from that period, the share of <s> was even higher. In fact, genre proved to be a strong external factor for this orthographic variable. Whereas <sch> had already been established as the default variant in late eighteenth-century newspapers, i.e. in printed and published texts, there was more room for the alternative ‘phonetic’ <s> spelling in handwritten ego-documents. In diaries and travelogues, but most notably in private letters, <sch> coexisted alongside the almost equally frequent <s> spelling.

Furthermore, the position in a word, i.e. word-medial or word-final <sch>/<s>, was taken into consideration as an internal factor. Although earlier observations (e.g. Gledhill 1973), arguing that <sch> was longer preserved in

medial than in final position, were confirmed to some extent, the differences in distribution appeared to be mainly genre-specific. In fact, a striking position-dependent difference became evident in eighteenth-century diaries and travelogues, but not in the other two genres. It could be argued that this highlights the special intermediate position of these sources in the *Going Dutch Corpus* design, suggesting that the spelling in diaries and travelogues was more conservative and ‘written’ than in private letters (word-medial) but more ‘oral’, i.e. more closely reflecting the contemporary pronunciation than in newspapers (word-final). Possibly, diaries and travelogues reflect an older writing tradition with two position-dependent spellings in line with the historical-phonological distinction of medial /sx/ (<sch>) and final /s/ (<s>).

Regional variation was mainly attested in ego-documents, most notably in private letters. In the eighteenth century, not all regions used <sch> as their main variant. In fact, <s> was the most frequent variant in Groningen and also occurred equally frequent as <sch> in Friesland and North Holland. In other words, letter writers from the northern regions used <s> relatively more often than in the rest of the investigated language area. With regard to the dialectal variation briefly addressed in Section 1, this might be somewhat surprising. Keeping in mind that Wgm. **ske*, at least word-medially, has been maintained as a cluster (/sk/ and/or /sx/) in some dialects in North Holland and the north-eastern regions including Groningen, one might assume that <sch> was interpreted as a possible representation of /sk/ (Rutten & van der Wal 2014: 51) or /sx/. However, the reduced form <s> appears to be relatively frequent in these regions, thus not reflecting a possible (dialectal/regional) cluster pronunciation. This is largely in line with Rutten & van der Wal’s (2014) case study on initial <sc>/<sk> in private letters from North Holland/Amsterdam, also showing that dialect interference on spelling in seventeenth- and eighteenth-century letter writing rarely occurred. In the nineteenth-century data of the *Going Dutch Corpus*, regional differences levelled out. Furthermore, spatial variation between the centre and the periphery turned out to be marginal in both periods.

In contrast, gender was another strong external variable affecting the distribution of variants. Particularly striking in the late eighteenth century, male writers primarily used <sch>, whereas women used <s> slightly more often than <sch>. In line with the general increase of <sch> after 1804, these gender differences were less striking in the early nineteenth century. However, the shift in the female data from a slightly dominant usage of <s> (alongside <sch>) in the first period towards <sch> as the clearly prevalent variant in the second period emphasise Siegenbeek’s normative influence.

To sum up, the case of word-medial and word-final <sch> gives evidence of a remarkably effective spelling norm, officialised by Siegenbeek in the context of the *schrijftaalregeling*, and widely adopted in both printed and handwritten texts, across all regions and both genders.

Orthographic variables (4)

Long *e*'s in open syllable

1 Discussion in Siegenbeek (1804)

Comprising almost forty pages, the spelling of long vowels is undoubtedly one of the most extensively discussed features in Siegenbeek's (1804) national orthography of Dutch. Siegenbeek paid particular attention to the orthographic representation of long *e*'s and *o*'s, treating these etymologically distinct vowels as an exception to the general rule for long vowels, which had to be spelled with a single grapheme in open syllables. In this chapter, the special case of the long *e*'s will exemplify this spelling issue, although the orthographic representation of long *o*'s could have served as a possible case study, too (cf. also Rutten 2011: ch. 5; Rutten 2009b).

In Dutch historical linguistics, two long *e*'s are traditionally distinguished based on their etymologies (Rutten & van der Wal 2014: 34-44). First, the so-called soft-long *ē* evolved through lengthening of the originally short vowels [e] and [i] in open syllables, as in *geven* 'to give' and *nemen* 'to take' (also compare German *geben*, *nehmen*). Secondly, the so-called sharp-long *ê* derives from the West Germanic diphthong **ai*, as in *steen* 'stone' and *deelen* 'to share' (German *Stein*, *teilen*).

In North Holland, particularly in the Amsterdam area, these two phonemes had merged into one long [e:] already by the end of the sixteenth century, which is also the situation in present-day Standard Dutch. However, the historical-phonological distinction between the rather monophthongal soft-long *ē* and the diphthongal sharp-long *ê* has been maintained in various dialect areas of the Northern Netherlands until today⁵⁵, for example in Groningen, Zeeland, as well as in parts of South Holland and North Brabant, primarily along the river Meuse (Goossens et al. 2000a, maps 21/128).

With respect to the orthographic representation of the two long *e*'s, it is generally assumed that supraregional (originally Southern) writing practices had developed, distinguishing between <e> for soft-long *ē* in open syllables, and <ee> for sharp-long *ê* in open syllables (Rutten & van der Wal 2014: 36). In his spelling treatise, Siegenbeek's (1804: 134) rule for long *e*'s in open syllables was, in fact, mainly founded on this phonology-based system:

Bedien u in lettergrepen, niet op eenen medeklinker sluitende, ter aanwijzing van den langen klank, altijd van eenen enkelen klinker, met uitzondering slechts van die woorden, welke volgens hunne oorspronkelijke eigenschap, de harde lange *e* of

⁵⁵ With regard to the situation in Southern Dutch, the difference in pronunciation has also been maintained in most Flemish, Brabantian and Limburgish dialects (Rutten 2011: 85).

o, met de *ei* of *au* vermaagschapt, hebben, en dus eene verdubbeling der vokaal vereischen.

'In syllables which do not end in a consonant, you should always use a single vowel to indicate the long sound, with the exception of those words, which, according to their original quality, have the hard long *e* or *o*, related to the *ei* or *au*, and thus require a doubling of the vowel.'

Summarised in this concise and clear rule, Siegenbeek (1804) thus prescribed the single grapheme <e> for soft-long (*zachte lange*) *ē* in open syllables (i.e. in line with his general rule for long vowels), but the digraph <ee> for sharp-long (*harde lange*) *ē* in open syllables.

Largely following his influential eighteenth-century predecessors ten Kate (1723) and Kluit (1763) (cf. Section 2), Siegenbeek (1804: 118f.) aimed to substantiate his decision by highlighting the differences between the long *e*'s:

ook de lange *e* en *o* [hebben] een' zachten, en een' harden, scherpen of hoogen klank [...], meer of min zwemende naar eenen tweeklank; welke laatste in *heelen* (heilen), *leenen* (*mutuo dare*) [...] enz. plaats vindt; terwijl de eerste, of zachte klank eigen is aan *belen* (verbergen), *lenen* (leunen) [...] en meer dergelijken.

'the long *e* and *o* also have a soft, and a hard, sharp or high sound [...], more or less resembling a diphthong, the latter of which is found in *heelen* (heal), *leenen* (*mutuo dare*) [...] etc, whereas the former, or soft sound is characteristic of *belen* (hide), *lenen* (lean) [...] and the like.'

Siegenbeek (1804: 119) acknowledged that the differences between the two long *e*'s had been lost in the 'corrupted' (*verbasterd*) pronunciation of many Dutch speakers, especially of the *Amstellanderen* (i.e. Amsterdammers). On the other hand, he also emphasised that the difference in pronunciation had been preserved in the dialects found in the *Maaskant* (by which he meant Rotterdam and other places along the River Meuse), Zeeland, Groningen, and others:

Het is waar, dat hetzelfde [= onderscheid van klank in de lange *e*] thans in de uitspraak van vele Nederlanders, bijzonderlijk der Amstellanderen, is verloren gegaan; doch daarentegen laat het zich op de tong der Maaskanters, Zeeuwen, Groningers, en andere bewoners van ons Vaderland, duidelijk hooren. Ja, hoe verbasterd in dit opzigt de uitspraak der Amstellanderen ook zijn moge, heeft echter die van het lage gemeen het kenmerk der harde scherpe *ē* nog bewaard, zeggende, volgens een Vriesch dialect, *ien* voor *één*, *bien* voor *béén*, *stien* voor *stéén*, *wiek* voor *week* (*mollis*), *gien* voor *geen* (*nullus*); terwijl de zachte *ee* in *week* (*hebdomas*), *geen* (*hic, ille*), bij de uitspraak nimmer in *ie* overgaat.

'It is true that the difference has nowadays been lost in the pronunciation of many Dutch people, especially the Amsterdammers. But in contrast, it can clearly be heard in the dialects of the people in the Maaskant region, Zeeland, Groningen, and other inhabitants of our fatherland. Yes, however corrupted the pronunciation of the Amsterdammers may be in this respect, the pronunciation of the rabble has still preserved the feature of the hard sharp *ē*, saying, according to a Frisian dialect, *ien* for *één*, *bien* for *béén*, *stien* for *stéén*, *wiek* for *week* (*mollis*), *gien* for *geen* (*nullus*),

whereas the soft *ee* in *week* (*hebdomas*), *geen* (*hic, ille*) never changes into *ie* in the pronunciation.⁵

What follows is a demonstration of the etymological differences by presenting a number of cognate words in languages related with Dutch, such as Gothic, Old Saxon and Old Franconian. For instance, Siegenbeek (1804: 120-121) illustrated the (diphthongal) sharp-long *ē* by the verb *deelen* 'share', and the (monophthongal) soft-long *ē* by *geven* 'give':

DEELEN, M-G.⁵⁶ *dailjan*, Al. *teilan*, *teilen*, A-S. *dēlan*.

GEVEN, M-G. *giban*, Fr-D. *giban*, *geban*, A-S. *gyfan*, *gyfan*, *gefan*. [...]

Bij dit bewijs, uit de oude, met ons Nederduitsch verwantschapte, talen ontleend, voegt zich een andere niet min klemmende bewijsgrond, welken het gebruik van velen onzer achtbaarste schrijveren en de uitspraak in verscheidene streken van Nederland ons aanbieden.

'In addition to this evidence derived from the old languages related to our Dutch language, there is another, not less convincing evidence, which the use of many of our most respectable writers and the pronunciation in various regions of the Netherlands suggest to us.'

In addition to the different etymologies illustrated by cognates, Siegenbeek mentioned the usage in the writing of many respectable authors as well as the pronunciation in various Dutch regions as convincing evidence. It is particularly the latter argument which was pointed out by Siegenbeek (1804: 130) as the most crucial factor, demonstrating that the difference between the two long *e*'s was neither 'whimsical' (*grillig*) nor 'imaginary' (*ingebeeld*), but grounded on the *gemeenlandsche*, i.e. common 'national' pronunciation:

Zie daar dan, op voorgang van TEN KATE en andere taalkenners, zoo wij meenen, onwederlegbaar betoogd, dat het onderscheid van klank tusschen de harde en zachte lange *e* en *o* niet grillig en ingebeeld, maar wezenlijk en op de gemeenlandsche uitspraak gegrond is. En zal men, dit erkennende, niet tevens moeten toestemmen, dat dit verschil van klank door een onderscheidende schrijfwijze dient aangeduid te worden? [...]

Het is derhalve volstrekt noodzakelijk in de spelling zorg te dragen, dat een zoo wezenlijk en belangrijk taaleigen, door de verbastering der uitspraak, niet eindelijk geheel onkenbaar worde en verloren ga.

'Following the example of TEN KATE and other language experts, which, as we think, irrefutably demonstrates that the distinction in sound between the hard and soft long *e* and *o* is not whimsical and imaginary, but essential and grounded on the common 'national' pronunciation. And, acknowledging this, should one not also agree that the difference in sound has to be indicated by distinct spellings? [...]

⁵⁶ The abbreviations of languages used in this quote refer to Moeso-Gothic (*Moesogothisch*, M-G.), Alemannic (*Alemanisch*, Al.), Anglo-Saxon (*Angelsaxisch*, A-S.), and Franconian (*Frankduitsch*, Fr-D.).

It is therefore completely necessary in spelling to take care that such an essential and important feature of the language does not become entirely unrecognisable and lost in the end, through the corruption of the pronunciation.’

Not only did Siegenbeek emphasise that such a phonological distinction had to be reflected in spelling, he also concluded that it was necessary to take this essential idiomatic feature of the Dutch language into account in order to prevent its decay and loss.

2 Eighteenth-century normative discussion

A comprehensive account of the orthographic representation of etymologically distinct long *e*'s in eighteenth-century metalinguistic discourse is provided by Rutten (2011: ch. 5), also serving as the basis for this section. In general terms, he argues that the normative discussion in the Northern Netherlands from the late seventeenth century until Siegenbeek (1804) was divided into two camps (Rutten 2011: 94). On the one hand, there was an originally Southern tradition, often associated with the sixteenth-century lexicographer Cornelis Kiliaen and the *Statenbijbel* (1637), which was grounded on the historical-phonological difference between soft-long *ē* (spelled <e> in open syllable) and sharp-long *ê* (spelled <ee> in open syllable). On the other hand, there was a more morphologically oriented tradition with North Holland as its centre, which followed the linguistic practices of the influential poet and writer Joost van den Vondel. As opposed to the phonology-based system from the South, this Northern tradition had a strong focus on the principle of uniformity (*gelijkvormigheid*), according to which sharp-long *deelen* occurred alongside soft-long *weeten* (both spelled with <ee> by analogy with the root words *deel* and *weet*), but also sharp-long *hemel* alongside soft-long *beter* (no analogy involved) (Vosters 2011: 275).

Francius (1699) was one of the first grammarians to follow and codify Vondel's morphology-based system with regard to long vowels. Uniformity and analogy were the crucial factors for his spelling rule, arguing that the double vowel spelling <ee> should be derived either from the singular form or the root word:

De verdubbeling der klinkers is somtijds noodig, somtijds niet, en 't gaat mijns oordeels niet altijd door, dat de tweede niet noodzaakelijk is. Op een lettertje meer of min zal 't niet aankomen, als die verdubbeling maar meer klaarheids byzet, en uit het minder getal, of uit het wortelwoordt haren oorspronk heeft. (1699: 65)

‘Sometimes the doubling of the vowels is necessary, sometimes it is not, and in my opinion it does not always happen that the second is not necessary. It will not matter whether it is one letter more or less, as long as the doubling adds more clarity, and derives from the singular or from the root word.’

A more detailed discussion grounded on morphological considerations is found in Moonen (1706). He generally advocated the spelling with a single vowel in open syllables, as in *edel*, *even* and *hemel*. However, under specific morphological

conditions, primarily related to the root word, the spelling with a double vowel was required in open syllables. Moonen (1706: 27-29) listed three groups of words to be spelled with <ee>: (1) plural forms of nouns with a long vowel in the singular (e.g. *beer* – *beeren*), (2) verbs with a long vowel in their root or stem, i.e. the second person singular imperative (e.g. *leer* – *leere*, *leeren*), and (3) derivations of words with a long vowel (e.g. *eenigh*, *eenigzins*, *eenigheit* < *een*, also *leeraer* < *leere*):

Gelyk nu hier toe het verdubbelen of verlengen der Klinkeren niet noodigh is, zoo meene ik, dat deeze verdubbeling omtrent veele andere woorden noodigh en dienstigh is: te weeten, eerst in Naemwoorden, van alle de drie Geslachten, die in hun Eenvouwich Getal eenen verlengden Klinker hebben, dien zy, myns bedunkens, in het Meervouwich moeten behouden, gelyk te zien is in de volgende en diergelyke woorden, [...] *beer*, *beeren*, [...] *oor*, *ooren*, *spoor*, *spooren* [...]

Daer na in Werkwoorden, die in hun Wortelwoort, den tweeden persoon des Eenvouwichen Getals in de Gebiedende Wyze, eenen langen Klinker gebruiken; als in *Haet*, *leer*, *stier*, *hoor*, *schuur*, waer van afkoomen *Ik haete*, *leere*, *stiere*, *hoore*, *schuure*, *ny haeten*, *leeren*, *stieren*, *booren*, *schuuren*.

In welke woorden de helfte van den langen Klinker alzo weinig magh uitgeworpen worden, en dus de Wortelletter verminkt, om te spellen, *bate*, *lere*, *bore*, *schure*, *baten*, *leren*, *boren*, *schuren*, als men spellen magh in den Onvolmaekten Tyt der Aentoonende Wyze, *Ik hatte*, *lerde*, *borde*, *schurde*, *ny hatten*, *lerden*, *borden*, *schurden*. [...]

Dus behoort men ook andere woorden te spellen, die afkoomen van andere, die eenen Langen Klinker in den oorsprong hebben; gelyk *eenigh*, *eenigzins*, *eenigheit*, die gesprooten zyn van *een*, en *eenwigh* met *eenwigheit*, die van *een* komen. Breng hier ook toe *leeraer* van *leere*, *groot* van *groot*, *boozger* van *boos*, in de plaetse van *leraer*, *groter*, *bozer*.

'While the doubling or lengthening of the vowels is not necessary here, I think that this doubling is necessary and useful in many other words, namely, first in nouns of all three genders, which have a lengthened vowel in their singular, which, in my opinion, they must maintain in the plural, as can be seen in the following and similar words, [...] *beer*, *beeren*, [...] *oor*, *ooren*, *spoor*, *spooren*.

Moreover, in verbs which use a long vowel in their root word of the second person singular imperative, as in *Haet*, *leer*, *stier*, *hoor*, *schuur*, from which *Ik haete*, *leere*, *stiere*, *hoore*, *schuure*, *ny haeten*, *leeren*, *stieren*, *booren*, *schuuren* derive.

In these words, one half of the long vowel may be thrown out, thus mutilating the root letter, in order to spell *bate*, *lere*, *bore*, *schure*, *baten*, *leren*, *boren*, *schuren*, as little as one may spell in the imperfect tense of the indicative, *Ik hatte*, *lerde*, *borde*, *schurde*, *ny hatten*, *lerden*, *borden*, *schurden*. [...]

In such a way, one must also spell other words which derive from others, which originally have a long vowel; like *eenigh*, *eenigzins*, *eenigheit*, which derive from *een*, and *eenwigh* with *eenwigheit*, which derive from *een*. Furthermore, *leeraer* from *leere*, *groot* from *groot*, *boozger* from *boos*, in place of *leraer*, *groter*, *bozer*.'

Although Moonen's system is chiefly grounded on morphological considerations, he also mentioned the differences in pronunciation with regard to *e*'s and *o*'s at some point, referring to the poet Jeremias de Decker. However, in his grammar, oriented to the North Holland dialects, he seemed to have no idea about the historical-phonological distinction and its maintenance in other dialect regions,

as Rutten (2011: 96) remarks: “In zijn op het Noord-Hollands georiënteerde spraakkunst zijn scherp- en zachtlang niet alleen samengevallen, hij lijkt ook geen idee te hebben van het voormalige onderscheid ervan noch van het regionale voortbestaan”.

Sewel (1708: 12-13) did not comment on etymologically and phonologically distinct long *e*'s (Rutten 2011: 97). He advocated the spelling of single <e> in open syllable, for example in *ézel*, *hémel* and *lépel*, where the use of accent marks served the purpose to distinguish the long *e*'s [e:] and [ɛ] from [ə] (Rutten 2011: 96). Nevertheless, Sewel (1708: 12) prescribed *steen* rather than *stenen* in order to retain the vowel spelling of the root word. In his discussion of long *a* (A4), Sewel (1708: 7) also highlighted the principle of uniformity:

men krenkt dan de eygenschap van 't woord, en zo doet men ook als men schryft *Zaken*, *benen*, *poten*, *raden* enz. Deeze redenen, hoewel sommige zich daartegen verzetten, zyn by my nógtans van groot gewigt, hoewel ik anders tegen alle overtóllige letteren ben.

‘then one harms the quality of the word, and one also does so when one writes *Zaken*, *benen*, *poten*, *raden* etc. Although some offer resistance, these reasons are still of great importance to me, although I am against all redundant letters otherwise.’

Morphological aspects were also central to van Rhyn's (1758) choices, arguing that the spelling of vowels needs to take into account the *Oorspronkelykheid der Woorden* (1758: 6), by which he referred to the root word. The necessity to use the double vowel spelling is exemplified by *Steenen* (plural) and *Steen* (singular), illustrating that “wanneer de Woorden in 't Eenvoud twee Vocaalen hebben, zy dezelve in 't Meervoud moeten behouden” (van Rhyn 1758: 7)

During the last decades of the eighteenth century, the merger of sharp-long *ê* and soft-long *ē* in (North) Holland was addressed in the grammar by *Kunst wordt door arbeid verkreegen* (1770: 33), raising the question who could actually hear a difference by the mere pronunciation of *steen* ‘stones’ and *stenen* ‘moan’: “Wie kan in het enkel uitspreken van *steen* (*lapides*) en *stenen* (*zuchten*) onderscheid horen?”.

Just before the turn of the century, the *Rudimenta* (1799: 49-50) still followed Vondel, Moonen (1706) and Sewel (1707, 1708), summarising the morphology-based Northern tradition:

De letter E heeft drieërlei geluid of uitspraak; als 1. helder, 2. dof, of zacht, 3. lang of zwaar.

E is helder in *vel*, *melk*, *zelf*, *zet*, *net*, enz.

E is dof of zacht in *zadel*, *fabel*, *aarde*, *bedrijf*, enz.

E is lang of zwaar in *Eva*, *Lea*, *Eland*, enz. [...]

Dan moet men ook in het gebruik van één E of twee EE weder letten op de afleiding, of verschillende betekenis [...] b. v. men schrijft:

Smeeden, *smeeken*, *geeven*, *speelen*, [...] Als deeze woorden in het Enkelvoudige beteken, of afgeleid worden van *smeed*, *smeek*, *geef*, *speel* [...] (1799: 49-50)

‘The letter E has three kinds of sound or pronunciation, as 1. clear, 2. dull, or soft, 3. long or heavy.

E is clear in *vel, melk, zelf, zet, net*, etc.
 E is dull or soft in *zadel, fabel, aarde, bedrijf*, etc.
 E is long or heavy in *Eva, Lea, Eland*, etc. [...]
 Then one must also consider the derivation, or different meaning in the use of one E or two EE [...] e.g. one writes:
Smeeden, smeecken, geeven, speelen, [...] When these words in the singular mean or derive from *smeed, smeeke, geef, speel* [...]

In opposition to the morphologically oriented system from North Holland, the alternative eighteenth-century tradition from South Holland was grounded on the phonological difference between sharp-long *è* and soft-long *ē*.

With regard to the spelling of long vowels, Verwer (1707) took the *Statenbijbel* as his main point of reference, making a strict distinction between soft <e> (*epsilon*) and sharp <ee> (*èta*) (Verwer 1707: 2). Terminologically, in fact, this was the first time that *zachtlange* en *scherplange* *e*'s appeared in the literature (cf. Rutten 2011: 98). Taking into account regional differences, Verwer (1707: 97) acknowledged that the phonological difference had been 'fatefully confused' (*rampzalig verward*) in Amsterdam and, more generally, in North Holland, whereas it had been preserved in South Holland:

Epsilon en *èta* [...], hebben bij ons een verschillende klank, gebaseerd op het onderscheid zelf der zaken en die met de juiste helderheid gehoord wordt in het genoemde zuidelijke gebied van Holland. Want leder (*leder*), "corium", is iets anders dan leeder (*leder*), "scala" [...], ofschoon dit alles door het gepeupel en zijn nalopers in Amsterdam en Noord-Holland rampzalig wordt verward.

'*Epsilon* and *èta* [...] have a different sound in our language, based on the distinction itself and which can be heard with the right clarity in the mentioned southern part of Holland. Because leder (leather), "corium" is something different than leeder (ladder), "scala" [...], although this is all fatefully confused by the rabble and its followers in Amsterdam and North Holland.'

Verwer's comment on the lower-class people (*het gepeupel*) from Amsterdam and North Holland is particularly remarkable when compared to Siegenbeek's remark on the lower class (*het lage gemeen*) from the same area almost one century later (cf. Section 1), arguing that the phonological distinction had been preserved in their pronunciation.

In order to facilitate learning the difference between long *e*'s, Verwer (1707: 97) even provided mnemonics, referring to the German cognates spelled with <ei> or <ä> (for sharp-long *èta* <ee>): "Woorden die in het Duits geschreven worden met ei of met ä en door ons met e, worden uitgesproken als met *èta*, bijv. Theilen/deelen, geist/geest, steinen/steen, gemein/gemeen; gelährter/geleerd" 'Words which are written with ei or with ä in German, and with e in our language, are pronounced as with *èta* [...]'.

Verwer's (1707, 1708) observations with regard to the etymologically distinct long *e*'s were further systematised by ten Kate (1723; cf. also van der Wal 2002a). In his so-called critical orthography, which aimed at a one-to-one

correspondence between sign and sound, ten Kate (1723 I: 129) distinguished four *e*'s: (1) the sharp-short <é>, (2) the soft-short <e>, (3) the sharp-long <ée>, and (4) the soft-long <ee>. Ten Kate suggested accent marks (*bovenstreping*, cf. 1723 I: 163) to indicate sharp vowels, and argued that the phonological difference between sharp and soft vowels had to be represented by spelling. At the same time, ten Kate (1723 I: 118) also acknowledged that this distinction had no longer been perceived in the area between North Holland and the river Rhine, which not only violated the *Gemeene-lands Dialect* (common 'national' dialect), but also led to ambiguous homonyms:

Gelijk het onderscheid van Spelling tussen de Langklinkende EE en ÉÉ, OO en ÓÓ, veeltijds bij de teegenwoordige Schrijvers word naagelaaten, zo word zelf het onderscheid in de Uitspraak bij ons, en de geenen die tussen Noord-Holland en den Rijn woenen, niet waargenoomen, als gebruikende alleen de zagte lange EE en OO; waar door wij niet alleen zondigen teegens de Gemeene-lands Dialect, maar ook vervallen in een' Dubbelzinnigheid van woorden, van 't welke veele Zuid-Hollandsche Steeden, en andren van onze Nederduitsche Provinciën, die dit onderscheid in agt neemen, vrij zijn. De Volmaaktheid vereist onderscheid in de Klanken, en dit weederom onderscheidene Letter-Teekenen.

'Although the spelling difference between the long-sounding EE and ÉÉ, OO and ÓÓ is often neglected by contemporary writers, even the difference in pronunciation is not recognised by us and those who live between North Holland and the river Rhine, only using the soft-long EE and OO. In doing so we not only sin against the common 'national' dialect, but also fall into an ambiguity of words, of which many cities in South Holland, and others of our Dutch provinces, which consider this distinction, are free. Perfection requires a distinction in the sounds, and this, in turn, requires different letters.'

He further remarked that many places in South Holland and other provinces had been immune to the merger of the two long *e*'s. Later, ten Kate (1723 I: 157) returned to the aspect of regional variation, adding that inhabitants of Zeeland, Flanders, parts of Brabant and Friesland (*Landfriezen*), among others, had also maintained the difference in pronunciation:

Ten opzigte van 't behoorlijke Onderscheid van Uitspraak tussen EE en ÉÉ vond ik onder onze Nederlanders geen andere makkers in ons verzuim dan die van Over-Yssel; dog de Zuid-Hollanders, Stigtenaers, Zeeuwen, Vlamingen & eenige Brabanders staen 'er beter bij; zo ook onze overbueren de Zaenlanders: de Vriezen voornaemlijk de land-Friezen onderhouden 't desgelijks, hoewel op eene andere wijze; want voor de ÉÉ spreken zij een klank als IE

'With respect to the considerable difference in pronunciation between EE and ÉÉ, I found no other fellows in our carelessness among our Dutch people than those from Overijssel. But the people from South Holland, Utrecht, Zeeland, Flanders and some from Brabant are doing better, just like our neighbours from Zaanland. The Frisians, especially the rural Frisians, maintain it, too, though in a different way, because for the ÉÉ they pronounce a sound like IE'

While his critical orthography thus distinguished sharp-long <ée> and soft-long <ee>, ten Kate also admitted that the (desirable) use of accent marks for sharp vowels had not been established, which is why he followed the Southern writing practices (*het Agtbare gebruik*) in his common orthography, spelling double <ee> for sharp-long *ê* in open syllable, and single <e> for soft-long *ē* in open syllable.

Four decades later, Kluit (1763) largely continued ten Kate's system, regarding the distinction between sharp-long *ê* (<ee>) and soft-long *ē* (<e>) as a characteristic of the common 'national' language ("de gemeenelantsdialekt, die ons eenig richtsnoer weyzen moet", 1763: 343), as opposed to the 'corrupt Amsterdam dialect' ("de bedorven Amstellantsche Dialekt", 1763: 329):

alle onze Schrijvers [hebben] zich eenparig [...] toegeleit, om, den gemeenelantschen spraakvorm volgende, dit zo merklijk klankverschil der scherpe langklinker *ée* behoorelijk in acht te nemen, zodat zy alle die woorden, waaraan die scherpe klank verbonden was, naaukeurig van de *zachte e* onderscheidden door het dubbellaten dezer letter; schrijvende dus nooit anders dan *deelen, heelen, speenen, teeken, meenen, verkleenen*; behoudende voor het overige de *zachte lange e* in *leven, beven, steken, bevelen, gezwegen, verleden*. (1763: 328-329)

'all of our writers have unanimously focused on properly taking account of this considerable difference in the sound of the sharp-long vowel *ée*, following the common 'national' form of the language, so that they accurately distinguish all those words, to which the sharp sound was related, from the *soft e* by doubling of this letter. Thus never write other than *deelen, heelen, speenen, teeken, meenen, verkleenen*, for the rest maintaining the *soft long e* in *leven, beven, steken, bevelen, gezwegen, verleden*.'

Referring to the Alemannic, Franconian and (Moeso-)Gothic dialects, which had preserved the diphthong *ei/ai* (< West Germanic *ai), Kluit (1763: 328-329) emphasised that the distinction between the two long *e*'s is not arbitrary but grounded on the etymology of words:

in alle die woorden, waar de EE voor de EI gebezicht wordt, [wordt] meer dan een enkele klank gehoord [...] En dat deze EE waarlijk de kracht van een *Diphthong* bezitte, en dit dus geen willekeurig onderscheid zy; bewijzen ons de Alemanische, Franktheutsche en Moesogotische Dialekten; de eersten door alle die scherp- en langklinkende woorden [...] stantvastig met de *ei*, de laatste met de *ai* te schrijven

'in all those words, where the EE is used for the EI, more than a single sound is heard [...] And that this EE truly has the power of a diphthong, and that this is thus no arbitrary distinction, is proven by the Alemanic, Franconian and Moeso-Gothic dialects, writing the former in all those sharp and long-sounding words constantly with the *ei*, and the latter with the *ai*'

Like Kluit (1763) as well as de Haes ("De *e* verdubbelen wy niet dan waer zy gehoord word", 1764: 13) and van der Palm (<e> "voor den lagen klank" but <ee> "voor het hooge geluidt", 1769: 14-15), Stijl & van Bolhuis (1776) also followed the phonology-based tradition from South Holland, with ten Kate as "de beste leidsman" 'the best leader':

In lettergrepen, die op een medeklinker eindigen, zou men een bovenstreping gebruiken kunnen, en de harde *ee* in *wéék* (zacht) daar door onderscheiden van *week* (7 dagen). Dit zou echter wat veel moeite verwekken, en is nog weinig in gebruik. Wij zullen er daarom ook niet op aandringen, maar sterk aanprijzen, om in lettergrepen, die op een klinker eindigen, in het hooge geluid de *e* en *o* te verdubbelen, en in het zachte geluid één klinker te gebruiken [...] Dan zou *weeken* (zacht maken) van *weken* (7 dagen) zoo wel in spelling verschillen, als in klank en beteekenis (1776: 35-36)

'In syllables which end in a consonant, one could use accent marks, and thereby distinguish the hard *ee* in *wéék* (soft) from *week* (7 days). However, this would require quite some effort, and it is still hardly in use. For syllables which end in a vowel, we will therefore not insist but strongly recommend to double the *e* and *o* in the high sound, and to use one vowel in the soft sound [...] Then *weeken* (make soft) would differentiate from *weken* (7 days) in spelling, as well as in sound and meaning'

The accent spelling, first and foremost proposed by ten Kate (1723), was presented as more or less nonobligatory, as it required some effort and had, in fact, hardly been adopted in language usage. This was also commented on by van der Palm (1769: 15) a few years earlier: "Sommigen hebben geoordeelt, dat het niet ondienstig zoude zyn de hooge of scherpe E telkens met een klankteeken van boven te merken; 't welk echter van zeer weinigen nagevolgt wordt" 'Some have argued that it would not be useless to mark the high or sharp E with accents, which, however, was followed by very few'.

In addition to the two opposing (main) traditions outlined in this section, there were a few alternative approaches in the eighteenth century. Nyløe (1703: 13-17), for instance, rigorously suggested that long vowels in open syllables should always be spelled with a single vowel.

ik zie geen reden ter werelt die die spelling met twee vocalen kan verdedigen; wat taal is'er, van die in enige achting zijn, daar een lange sillabe of lettergreep met twee klinkletteren wort geschreven? [...] Of zijn de Nederlanders minder bequaam dan andere volken om te kunnen onderscheiden wat sillaben in hunne tale lank of kort zijn, ten zy hun dit met twee vocalen worde aangewezen? Het is zeker dat de ene vocaal hier te veel is, naardienze niets ter werelt uitrecht, *va* in *vader*, *le* in *leven*, *ko* in *koning*, zijn met ene *a*, *e*, en *o*, zo lank, als ofze met tien vocalen geschreven waren, en die daar meer dan ene zijn, zijn overtollig. (1703: 14)

'I see no reason in the world which can justify the spelling with two vowels. Which respectable language is there, in which a long syllable is written with two vowel letters? [...] Or are the Dutch people less competent than other people at distinguishing which syllables in their language are long or short, unless they are indicated to them by two vowels? It is certain that the one vowel is too much here, as it does nothing in the world. *Va* in *vader*, *le* in *leven*, *ko* in *koning* with one *a*, *e* and *o* are as long as if they were written with ten vowels, and those which are more than one, are redundant.'

Another dissident spelling system was suggested by Elzevier (1774: 13-33), who distinguished no less than six different *e*-like sounds. Their orthographic representations, however, were fairly inconsistently grounded on phonological and/or morphological considerations. As Rutten (2011: 103) remarks, Elzevier (1774) was probably not fully aware of the etymological difference between sharp-long *ê* and soft-long *ē*, as words like *eeten* 'eat' and *keelen* 'throats' (soft-long) were mentioned in the same list of <ee> spellings alongside *eeden* 'oaths' and *steen* 'stones' (sharp-long).

Against the background of two strong opposing eighteenth-century traditions as well as a few alternative approaches, Siegenbeek (1804) clearly followed the spelling choices of his predecessors ten Kate (1723) and Kluit (1763) in his official orthography of Dutch. In other words, he eventually codified the Southern system, grounded on the phonological distinction between sharp-long *ê* and soft-long *ē*, on a national level (cf. Section 1).

3 Previous research

The orthographic representation of etymologically distinct long *e*'s as discussed in metalinguistic discourse has been addressed in a number of publications. Rutten (2011: ch. 5; cf. also 2009b) dedicates an entire chapter to the spelling of long *e*'s (as well as long *o*'s), outlining the normative traditions of both the Northern and the Southern Netherlands with a particular focus on the eighteenth century. Section 2 of this chapter, in fact, summarises the main developments in the eighteenth-century normative discussion in the Northern Netherlands, viz. the division into a originally Southern tradition based on the phonological difference between sharp-long *ê* and soft-long *ē*, and a more morphologically-oriented Northern tradition following Vondel's practices.

Vosters (2011; cf. also Rutten & Vosters 2010) also builds on the results presented in Rutten (2011), shifting the focus to the situation in the Southern Netherlands, mainly in the previously understudied nineteenth century. He points out that the spelling of sharp- and soft-long *e*'s (and *o*'s) is closely linked to the typically Southern accent spelling. The latter, however, occurs only marginally in actual language usage in the nineteenth century, as is revealed in the quantitative analysis based on a corpus of handwritten judicial and administrative texts (Vosters 2011: 306-309). Instead, the official Northern norms codified in Siegenbeek's (1804) orthography appear to be widely adopted in the Southern Netherlands.

Focusing on eighteenth-century metalinguistic discourse in the Northern Netherlands, van de Bilt (2009: 175-177) addresses the spelling of long *e*'s as one of the features discussed in the normative works by Verwer (1707) and Kluit (1763), also touching upon their relevance for the choices in Siegenbeek's (1804) national orthography.

From a historical-sociolinguistic perspective, Rutten & van der Wal (2014) present a corpus-based study of long *e*'s in seventeenth- and eighteenth-century private letters. Their analysis focuses on the region of Zeeland, where the

difference between soft-long *ē* and sharp-long *ê* had been preserved as a salient dialectal feature. For the seventeenth century, Rutten & van der Wal (2014: 36) observe that the phonological distinction is fairly well represented by spelling, and also in line with supraregional writing practices. In the eighteenth century, however, the phonological spelling in Zeeland is primarily replaced by a syllabic system, in which the grapheme is chosen based on syllable structure rather than on phonological differences. Generally referring to the *graphemisation* of the writing system, Rutten & van der Wal (2014: 67) conclude that phonology became a less important factor for spelling practices, giving way to orthographic choices grounded on syllabic or morphological aspects. In a comparable case study on letters linked to Amsterdam, where the two long *e*'s had merged, Rutten & van der Wal (2014: 67-72) show that in contrast to the phonology-based system used in seventeenth-century Zeeland, the Amsterdam results are chiefly variable, reflecting the merger of sharp- and soft-long *e*'s in spoken language. As the regional differences level out in the eighteenth century and the distribution in the Zeeland data converge to the Amsterdam results (despite the phonological distinction maintained in spoken dialects), Rutten & van der Wal (2014: 72) provide “solid evidence for the classic view of a supraregional variety that expands from Amsterdam into other regions”, i.e. replacing the previous writing practices grounded on the phonological difference.

In the corpus analysis presented in Section 4, I also build on these previous findings by examining whether and to what extent regional differences (i.e. merger or maintenance of etymologically distinct long *e*'s) can still be identified in the spelling practices in late eighteenth- and early nineteenth-century Dutch.

4 Corpus analysis

4.1 Method

In the following corpus analysis of the orthographic representation of etymologically distinct long *e*'s in open syllables, I consider <ee> and <e> as the two main variants, which occur in the *Going Dutch Corpus* as well as in the eighteenth- and early nineteenth-century normative discussion. Rutten & van der Wal (2014: 36-40) note that <ej> is another grapheme used for sharp-long *ê*, underscoring its realisation as a diphthong, although this variant rarely occurs in the *Letters as Loot* corpus. Given the fact that only three attestations of <ej> (<eÿ>, <ey>) for sharp-long *ê* in open syllable can be found in the *Going Dutch Corpus* (e.g. *teÿken*), I will exclude this variant from the analysis. Similarly, the alternative spelling variants with accent marks, i.e. <ée> and <ê>, as suggested in ten Kate's (1723) critical orthography, will not be taken into account here. Apart from the special case of stressed *éene(n)*, there are, in fact, no relevant occurrences of these accent spellings in open syllable in the entire *Going Dutch Corpus*.

Following the officially codified phonology-based distinction in Siegenbeek's (1804) national orthography (cf. Section 1), all results in this case

study will be presented as two separate categories for sharp-long *ê* and soft-long *ē*, respectively. In order to assess the effectiveness of the official spelling norms of 1804, the categorisation of words, either spelled with <ee> for sharp-long *ê* or <e> for soft-long *ē* in open syllable, is first and foremost based on the unambiguous prescriptions in Siegenbeek's (1804) word list in the appendix of his spelling treatise. For the sake of clarity and comparability, this distinction will not only be applied for data from the post-Siegenbeek period, but for both diachronic cross-sections.

For the corpus analysis, I selected the fifteen most frequent words with sharp-long *ê* in open syllable, and the ten most frequent words with soft-long *ē* in open syllable⁵⁷. The selected words are listed below in order of decreasing frequency⁵⁸ in the *Going Dutch Corpus*:

▪ **Sharp-long *ê***

EENE*; EENIGE*; HEERE*⁵⁹; *HEELE*; GEENE*; *TEEKE*; *DEELE*;
MEENE; *TWEDE*; *KEERE*; *STEENE*; *KLEEDE*; *VREEZE*;
LEERE; *BEENE*

▪ **Soft-long *ē***

DEZE*; *WEDER*; *MEDE*; *GEVE*; *TEGEN*; *NEME*; *ZEKER*; LEVE*;
VELE; *BETER*

These words are best regarded as search queries (e.g. used with *WordSmith Tools*), as they comprise different word forms, derivations and compounds with the same root or stem. For example, the query *DEELE* comprises *deelen*, *mededeelen*, *veroordeelen*, and so forth. Similarly, the query *ZEKER* includes words such as *verzeker*, *zekerheid*, and *onzeker*. In the case of *WEDER*, two semantically unrelated homonyms were actually taken into account, viz. *we(d)er* 'again' and *we(d)er* 'weather', both of which have a soft-long *ē*, prescribed by Siegenbeek (1804) to be spelled with <e> in open syllables. All undesired occurrences were filtered out by hand.

⁵⁷ With regard to the generally higher token frequency of words with soft-long *ē* (compared to a more restricted amount of frequent words with sharp-long *ê*), I decided to limit the number of investigated words with soft-long *ē* to the ten (rather than fifteen) most frequent in the corpus. Regarding the treatment of specific homonyms with different spellings in Siegenbeek's (1804) word list, *reede(n)/rede(n)* 'roadstead'/'reason; speech' and *weezen/wezen* 'orphans'/'be; being' were excluded from the analysis.

⁵⁸ The (by far) most frequent word of each category comprises 1495 (i.e. EENE*) and 1583 tokens (i.e. DEZE*), respectively.

⁵⁹ In the case of HEERE* (plural of *beer* 'lord; (gentle)man'), Siegenbeek (1804) prescribed the spelling with <ee> in open syllable, which serves as the point of reference for the categorisation in this chapter. Rutten & van der Wal (2014: 39) categorise *beere* as a word with soft-long *ē*, though. In fact, various etymological explanations have been suggested, including a reconstruction possibly going back to the diphthong **ai* (Old High German *ber* 'noble, sublime' might evolve from the meaning 'grey, grey-haired' < Proto-Germanic **haiṣa-* 'grey', cf. EWN).

Furthermore, spelling variation apart from the long *e*'s was taken into account, for instance *s/z* in *ze(e)ker/se(e)ker* and *de(e)ze/de(e)se*, and so forth.

4.2 Results

First of all, the general distribution of variants was investigated in the entire *Going Dutch Corpus*. Table 1a displays the relative distribution in the category of words with sharp-long *e* in open syllables, and Table 1b in the category of words with soft-long *e* in open syllables. The variants prescribed by Siegenbeek (1804), i.e. <ee> for sharp-long *e* and <e> for soft-long *e*, are highlighted by light-grey shading in the nineteenth-century period.

Table 1a. Sharp-long *e*: Distribution of variants across time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
Total	1,364	90.1	150	9.9	1,806	91.5	167	8.5

Table 1b. Soft-long *e*: Distribution of variants across time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
Total	1,133	40.5	1,664	59.5	246	7.5	3,043	92.5

With regard to the orthographic representation of sharp-long *e* in open syllables, the results in Table 1a show that <ee> is by far the most frequently used variant in the late eighteenth century, occurring in 90.1% (as opposed to only 9.9% spelled <e>). This distribution remains stable in the early nineteenth century, with the officially prescribed <ee> spelling increasing slightly from 90.1% to 91.5%. In this case, it seems that Siegenbeek (1804) followed the widespread eighteenth-century writing practices for sharp-long *e*, strongly favouring <ee>, rather than vice versa.

As presented in Table 1b, there is much more variation with regard to the spelling of soft-long *e* in open syllables. In fact, <ee> and <e> turn out to be co-existing variants in eighteenth-century usage, with <e> being somewhat more frequent with a share of 59.5%. The <ee> spelling also occurs fairly frequently in 40.5%, though. This means that, in sharp contrast to the results for sharp-long *e*, no clear preference for one specific variant had been established by the late eighteenth century. Only after Siegenbeek (1804) had officialised the <e> spelling for soft-

long *e* in open syllables, one can witness a steep increase of <e> from 59.5% to 92.5%, clearly becoming the predominant variant in usage. The <ee> spelling loses considerable ground, from 40.5% in the first period to a comparatively marginal 7.5% in the second period.

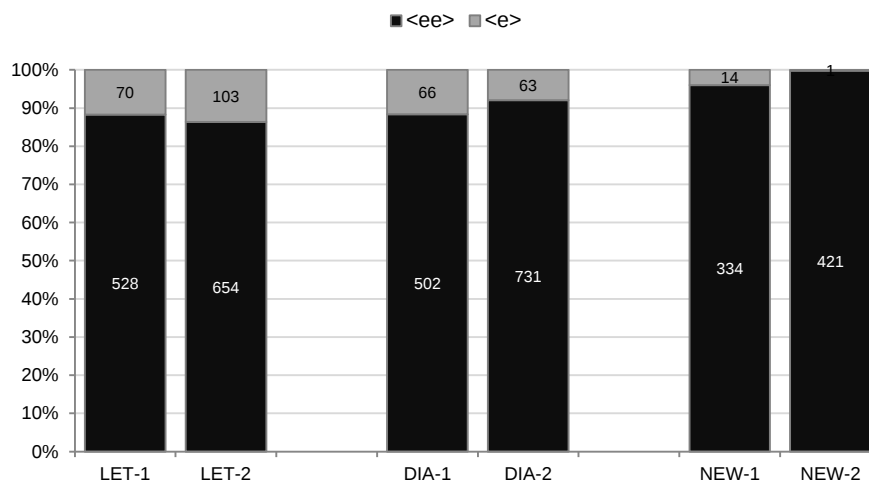
Summing up the general results for sharp-long *e* and soft-long *e* in open syllables, the two etymologically distinct long *e*'s reveal strong differences, both synchronically and diachronically. While the preference in favour of <ee> for sharp-long *e* is already established in the eighteenth century and remains stable in the nineteenth century, we can see striking diachronic changes with respect to the spelling of soft-long *e*. Here, <e> is widely established as the prevalent variant after the introduction of Siegenbeek's orthography. In fact, the prescribed variants have a share of more than 90 per cent in both categories, strongly suggesting that the officialised distinction of sharp-long *e* and soft-long *e* by two different spellings (i.e. <ee> and <e>, respectively) was successfully adopted in early nineteenth-century language practice.

Genre variation

Variation and change was also examined across the three genres of the *Going Dutch Corpus*, viz. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW). The results for sharp-long *e* and soft-long *e* (in open syllables) are displayed in Figures 1a and 1b, respectively.

As already indicated by the general results for sharp-long *e* in open syllables (Figure 1a), <ee> is by far the prevalent variant in eighteenth-century usage. Therefore, it is not surprising that the degree of genre variation is also relatively limited. In the first period, the <ee> spelling has a share of around 88% in both types of (handwritten) ego-documents, whereas it is even stronger in (printed) newspapers, occurring in 96.0%.

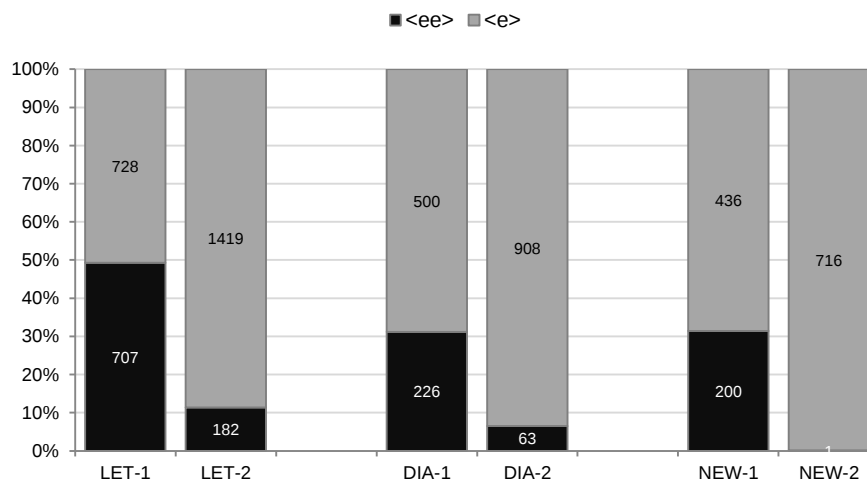
Figure 1a. Sharp-long *e*: Distribution of variants across genre and time.



Following Siegenbeek's prescription, nineteenth-century newspapers fully adopt <ee> as the only spelling variant for sharp-long $\bar{e}$. In diaries and travelogues from the same period, the use of <ee> also increases from 88.4% to 92.1%. Interestingly, the share of prescribed <ee> in private letters loses some ground from 88.3% to 86.4%, whereas <e> increases slightly from 11.7% to 13.6%.

There is considerably more genre variation in Figure 1b, showing the results for the spelling of soft-long $\bar{e}$ in open syllables. In the eighteenth-century period, <e> turns out to be the preferred variant in newspapers (68.6%) as well as in diaries and travelogues (68.9%). Although the <ee> spelling occurs in almost one third of the instances, it is a considerably less common option in these two genres. In private letters, however, no such preference in favour of <e> is visible, with <e> (50.7%) and <ee> (49.3%) occurring as equally frequent competing variants. In line with the observations from previous orthographic case studies, the most 'oral' genre of the *Going Dutch Corpus* shows the highest degree of spelling variation, certainly in the eighteenth century.

Figure 1b. Soft-long $\bar{e}$: Distribution of variants across genre and time.



In the nineteenth century, after Siegenbeek (1804) had prescribed <e> for soft-long $\bar{e}$ in open syllables, <e> becomes the prevalent spelling across all three genres. In the newspaper data, like in the case of sharp-long $\bar{e}$, the use of the prescribed variant for soft-long $\bar{e}$ (i.e. <e>) increases to practically 100%. Furthermore, there is a strong increase of <e> in diaries and travelogues, from 68.9% to 93.5%. Again, the highest proportion of variation is attested in private letters. Although the rise of <e> from 50.7% in period 1 to 88.6% is undoubtedly striking, the <ee> spelling for soft-long $\bar{e}$ can be attested in 11.4%.

What is interesting about the nineteenth-century results across genres is the almost perfectly mirrored distribution of variants for both sharp-long $\bar{e}$ and soft-long $\bar{e}$. The share of the prescribed variants is the highest in newspapers ($\bar{e}$

99.8%; *ē*: 99.9%), and the lowest in private letters (*ê*: 86.4%; *ē*: 88.6%), whereas diaries and travelogues once again take an intermediate position (*ê*: 92.1%; *ē*: 93.5%) between the two other genres.

Regional variation

The analysis of regional variation is particularly interesting with respect to the historical-phonological distinction between sharp-long *ê* and soft-long *ē*. As outlined in Section 1, this difference in pronunciation had disappeared in regions such as North Holland and particularly Amsterdam, while it had been preserved in most dialects of, for instance, Groningen, Zeeland, and parts of South Holland (Rotterdam, *Maaskant*) and North Brabant. The situation in Friesland is more complex, with *Stadsfries* in towns (like Leeuwarden, Franeker, Harlingen, Dokkum, Bolsward, Sneek⁶⁰), as opposed to Frisian or *Landfries* in the countryside, which neither show an unambiguous merger nor an unambiguous maintenance of the phonological distinction (cf. also van Bree & Versloot 2008: 108–112). In fact, the categories of ‘merger’ or ‘non-merger’ regions discussed in this section are best treated as tentative generalisations, as we have to take into account more dialectal variation as well as specific phonological conditions even in seemingly ‘clear-cut’ regions like Zeeland and North Holland (cf. Goossens et al. 2000a, maps 21/128; Weijnen 1966: 216).

The relative distribution of variants across region for sharp-long *ê* and soft-long *ē* is presented in Tables 2a and 2b, respectively (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

Table 2a. Sharp-long *ê*: Distribution of variants across region and time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
FR	162	86.2	26	13.8	256	91.4	24	8.6
GR	236	87.1	35	12.9	286	84.4	53	15.6
NB	161	97.6	4	2.4	226	92.2	19	7.8
NH	190	94.1	12	5.9	314	96.6	11	3.4
SH	210	94.2	13	5.8	282	91.3	27	8.7
UT	249	87.1	37	12.9	231	90.6	24	9.4
ZE	156	87.2	23	12.9	211	95.9	9	4.1

⁶⁰ The vast majority of texts from the region of Friesland in the *Going Dutch Corpus* is, in fact, linked to writers from these towns, in which *Stadsfries* was the dominant dialect.

Table 2a shows that regional variation in the orthographic representation of sharp-long $\bar{e}$ in open syllables is fairly limited. In fact, <ee> is the prevalent eighteenth-century spelling across all seven regions, ranging from 86.2% in Friesland to 97.6% in North Brabant. In the nineteenth century, <ee> is used in more than 90% in practically all regions. Only Groningen lags somewhat behind with a comparatively low share of 84.4%. The share of <e> even increases from 12.9% in period 1 to 15.6% in period 2.

With regard to the orthographic representation of soft-long $\bar{e}$ in open syllables, Table 2b shows that there are regional differences in the eighteenth-century period.

Table 2b. Soft-long $\bar{e}$: Distribution of variants across region and time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
FR	192	50.8	186	49.2	14	3.2	422	96.8
GR	118	30.3	271	69.7	19	4.0	461	96.0
NB	149	43.3	195	56.7	64	14.5	377	85.5
NH	209	50.5	205	49.5	45	8.7	474	91.3
SH	186	46.3	216	53.7	68	12.2	489	87.8
UT	149	33.6	295	66.4	17	4.1	398	95.9
ZE	130	30.5	296	69.5	19	4.3	422	95.7

It is remarkable that the <e> spelling is most prevalent in Zeeland (69.5%) and Groningen (69.7%). These two regions, in fact, have maintained the difference between sharp-long $\bar{e}$ and soft-long $\bar{e}$ in their dialects until the present day, and it is possible that this dialectic distinction is, at least to some extent, still reflected in eighteenth-century writing practices. In contrast, North Holland, where the long $\bar{e}$'s had merged in most parts, as well as in Friesland, the share of <e> is some 20% lower. In these regions, the distribution of variants is at chance level.

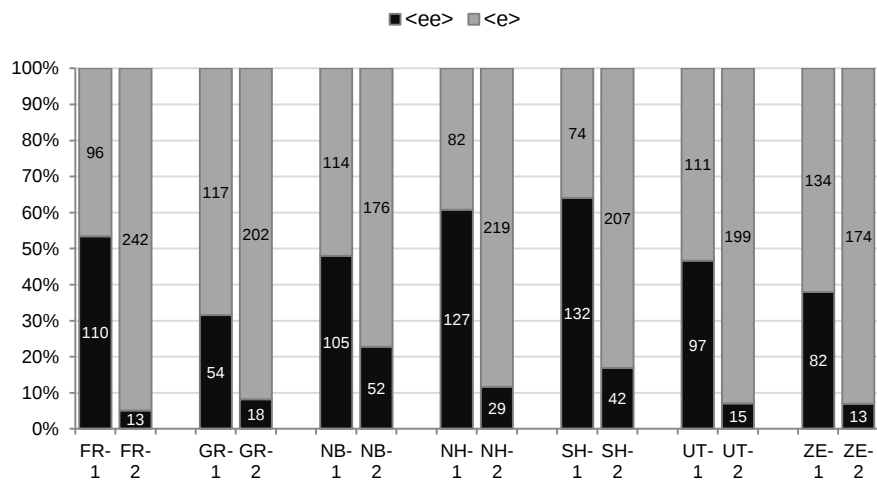
In the nineteenth-century period, all regions shift to <e> as the predominant variant in up to more than 95% in Zeeland, Utrecht, Groningen and Friesland. The highest proportions of <ee> for soft-long $\bar{e}$ are attested in South Holland (12.2%) and especially North Brabant (14.5%).

Regional variation across genres

As Table 2a has shown, the distributional patterns of sharp-long $\bar{e}$ is fairly homogeneous across regions and diachronically stable, whereas Table 2b has attested a considerable amount of variation and change for soft-long $\bar{e}$. This is

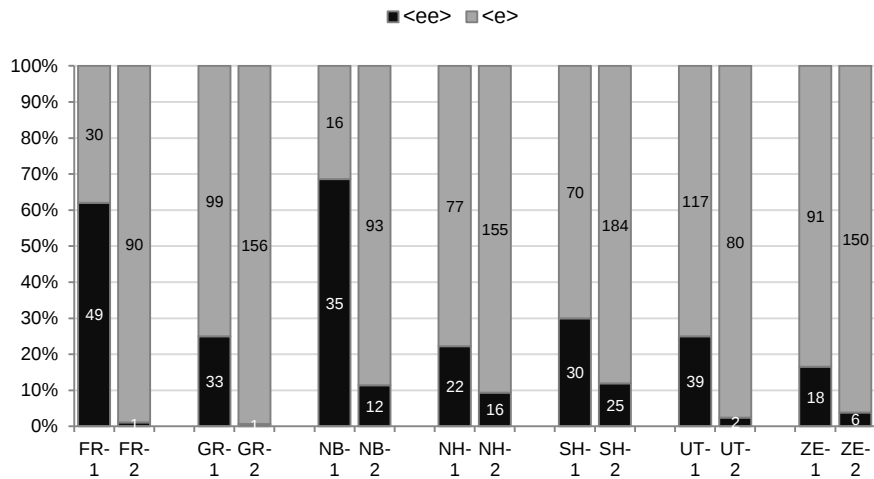
actually in line with earlier observations for seventeenth- and eighteenth-century Dutch based on the *Letters as Loot* corpus (Rutten & van der Wal 2014: 42). Therefore, only the case of soft-long *e* will be examined in more detail by zooming in on regional variation across the three genres. Figure 2a presents the relative distribution across regions in the sub-corpus of private letters.

Figure 2a. Soft-long *e*: Distribution of variants across region and time (private letters).

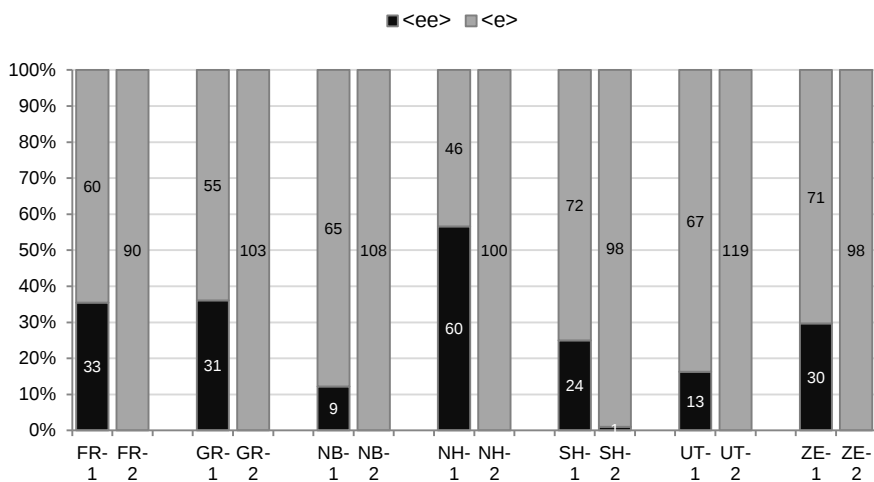


In the eighteenth-century period, one can see that <e> is most frequently used in Zeeland (62.0%) and Groningen (68.4%), both of which had preserved the phonological distinction, whereas the <ee> is particularly prevalent in the Holland area (around 60%). In the nineteenth-century period, the rise of the prescribed <e> can be witnessed in all seven regions. However, <e> remains a fairly common option in North Brabant (22.8%) and, to a lesser extent, South Holland (16.9%).

Figure 2b presents the results in the sub-corpus of diaries and travelogues. Again, the regional differences are limited to the eighteenth century. Friesland and North Brabant stand out with a strikingly high share of <ee>, occurring in 62.0% and 68.6%, respectively. In sharp contrast, the <e> spelling is the most frequently used variant in Zeeland, occurring in 83.5%. However, it is unlikely that this can solely be linked to the phonological distinction maintained in Zeeland, as North Holland, where the two long *e*'s had already merged, also has a high share of <e> (77.8%). In nineteenth-century diaries and travelogues, regional variation largely levels out, as prescribed <e> becomes the (only) dominant variant, ranging from around 88% in North Brabant and South Holland up to practically 100% in Friesland and Groningen.

Figure 2b. Soft-long $\bar{e}$: Distribution of variants across region and time (diaries and travelogues).

Finally, Figure 2c shows that some regional variation can also be attested in the newspaper data, at least in the eighteenth century. Whereas <e> is the prevalent variant in most regions, particularly in Utrecht (83.8%) and North Brabant (87.8%), <ee> is remarkably strong in North Holland (56.6%). However, with the introduction of Siegenbeek's (1804) official spelling rule for soft-long $\bar{e}$ in open syllables, variation completely disappears in nineteenth-century newspapers, as prescribed <ee> becomes the only spelling variant used in these printed sources.

Figure 2c. Soft-long $\bar{e}$: Distribution of variants across regions and time (newspapers).

Variation across centre and periphery

The results in Figures 3a and 3b give evidence that there is hardly any variation between the centre (CEN) and the periphery (PER).

Figure 3a. Sharp-long *ē*: Distribution of variants across centre–periphery and time.

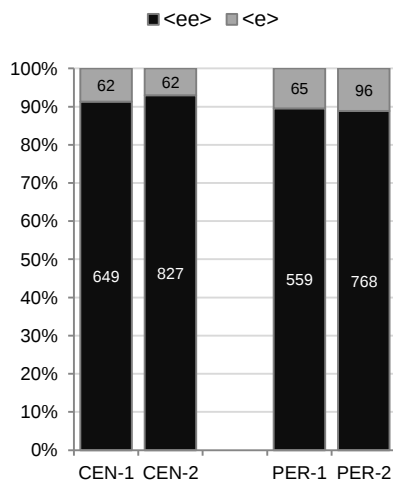
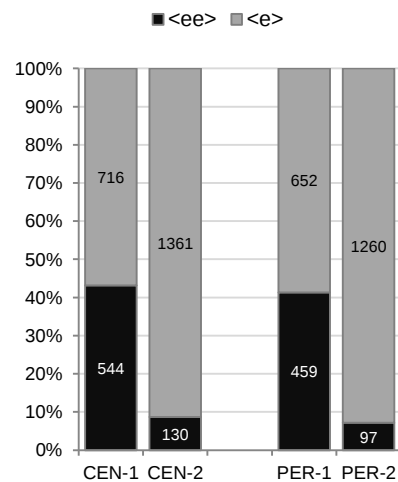


Figure 3b. Soft-long *ē*: Distribution of variants across centre–periphery and time.



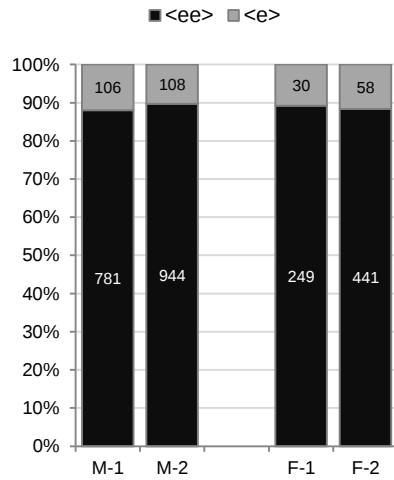
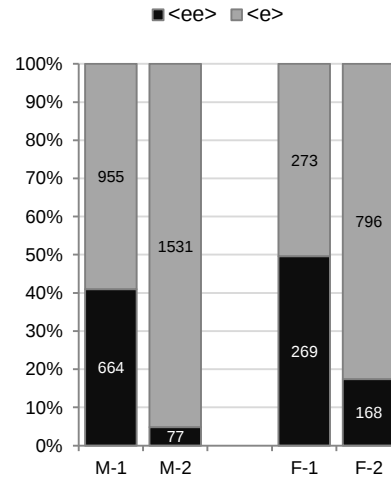
In the case of sharp-long *ē* in open syllables (Figure 3a), the proportion of <ee> in the eighteenth century is approximately 90% in both the centre and the periphery. Diachronically, the distribution for both categories remain stable in the nineteenth-century data.

With respect to soft-long *ē* in open syllables (Figure 3b), the distribution of variants is almost identical in both periods. In the eighteenth century, <e> is slightly more frequent (56.8% centre; 58.7% periphery) than <ee>. In nineteenth-century usage, the prescribed <e> spelling clearly dominates in the centre (91.3%) as well as in the periphery (92.9%).

Gender variation

Possible gender variation was investigated with the two sub-corpora of handwritten ego-documents (i.e. private letters, diaries and travelogues), produced by male (M) and female (F) writers. The results for sharp-long *ē* and soft-long *ē* in open syllables are displayed in Figures 4a and 4b, respectively.

With regard to the spelling of sharp-long *ē* in open syllables (Figure 4a), there is no considerable gender variation. In the eighteenth century, <ee> is by far the most frequently used variant among both men (88.1%) and women (89.3%). Apart from minimal fluctuations, this distribution remains stable in the early nineteenth century, with a share of 89.7% among men, and 88.4% among women.

Figure 4a. Sharp-long $\bar{e}$: Distribution of variants across gender and time.**Figure 4b.** Soft-long $\bar{e}$: Distribution of variants across gender and time.

More interesting patterns emerge in the spelling of soft-long $\bar{e}$ (Figure 4b). In the eighteenth-century period, male writers appear to use <e> slightly more often, occurring in 59.0%. In the texts by female writers, however, the two variants are evenly distributed (49.6% <ee> vs. 50.4% <e>). These differences are also reflected in the nineteenth-century data. While the prescribed <e> spelling increases drastically in texts written by men (from 59.0% to 95.2%), the rise of <e> somewhat lags behind in texts written by women. Although the share of <e> increases from 50.4% to 82.6%, a comparatively high proportion of the rejected <ee> spelling remains (17.4%).

In the following, these distributional patterns are examined on a more detailed level, cross-tabulating gender and genre. As the results for sharp-long $\bar{e}$ in Figure 4a do not reveal any interesting patterns, only the spelling of soft-long $\bar{e}$ will be considered here. Table 3a displays the genre-specific distribution of variants in private letters.

Table 3a. Soft-long $\bar{e}$: Distribution of variants across gender and time (private letters).

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
Male	454	46.9	514	53.1	33	4.1	778	95.9
Female	253	54.2	214	45.8	149	18.9	641	81.1

Roughly speaking, the two spelling variants are co-occurring options in eighteenth-century letters by both men and women. Nevertheless, male letter writers tend to use <e> somewhat more often (53.1%), whereas female writers have a slight preference for <ee> (54.2%). In nineteenth-century letters, men adopt the prescribed <e> spelling in 95.9% of all instances. Although the use of <e> also increases considerably up to 81.1% among women, <ee> still occurs in 18.9%. In other words, a fair amount of gender variation can be attested in private letters, even after the Dutch orthography had been regulated by Siegenbeek (1804).

Table 3b shows that the genre-specific distribution in diaries and travelogues differs from that in the letter data.

Table 3b. Soft-long *e*: Distribution of variants across gender and time (diaries and travelogues).

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
Male	210	32.3	441	67.7	44	5.5	753	94.5
Female	16	21.3	59	78.7	19	10.9	155	89.1

First and foremost, the share of <e> among eighteenth-century female diarists (78.7%) is higher than among their male contemporaries (67.7%). After Siegenbeek's (1804) prescription, however, the increase of <e> in diaries by men (from 67.7% to 94.5%) is more pronounced than in diaries by women (from 78.7% to 89.1%). Diachronically, both types of ego-documents attest the same genre-related tendencies, namely that male writers adopt the prescribed spelling more often than female writers.

5 Discussion

This chapter investigated variation and change in the orthographic representation of etymologically distinct long *e*'s in open syllable, traditionally referred to as sharp-long *e* and soft-long *e*.

To begin with, eighteenth-century metalinguistic discourse in the Northern Netherlands was dominated by two opposing traditions (cf. Section 2 and Rutten 2011 for a detailed outline). On the one hand, the originally Southern tradition, referring back to the *Statenbijbel* and based on the historical-phonological difference between sharp-long *e* and soft-long *e*, was promoted by influential figures such as Verwer (1707, 1708), ten Kate (1723), Kluit (1763), van der Palm (1769) and Stijl & van Bolhuis (1776). On the other hand, early language commentators like Francius (1699), Moonen (1706), but also van Rhyn (1758) and the *Rudimenta* (1799)

followed Vondel's morphologically oriented system, which was largely grounded on the principles of uniformity and analogy.

In the national orthography of Dutch, Siegenbeek (1804) officialised the orthographic distinction between soft-long *ē* and sharp-long *ê* in open syllables (cf. Section 1) by following the phonology-based Southern tradition of his eighteenth-century predecessors ten Kate (1723) and Kluit (1763). Arguing that the difference between two long *e*'s was an essential characteristic of the Dutch language, which had, in fact, been maintained in various dialect regions, Siegenbeek (1804) prescribed the single grapheme <e> for soft-long *ē* in open syllables, and the digraph <ee> for sharp-long *ê* in open syllables.

The corpus analysis, presented in Section 4, revealed that <ee> must have been established as the main spelling variant for sharp-long *ê* in open syllables by the late eighteenth century, occurring in more than 90%. Considerably more variation was attested for soft-long *ē* in open syllables. In the eighteenth-century period, both <e> and <ee> were commonly used variants, although <e> turned out to be more frequent in the overall corpus data.

After Siegenbeek's (1804) phonology-based distinction had been introduced, the distribution of variants for sharp-long *ê* remained stable. It seemed that Siegenbeek followed the established eighteenth-century writing practices for sharp-long *ê* (i.e. <ee>), which is why no visible effect on the distribution of variants was attested. With respect to the spelling of soft-long *ē*, however, one could observe a drastic change. In contrast to the competition between <e> and <ee> in eighteenth-century usage, the officially prescribed <e> spelling was adopted for soft-long *ē* in more than 90% after 1804. Given the clear predominance of both <ee> for sharp-long *ê* (91.5%) and <e> for soft-long *ē* (92.5%), it can be concluded that Siegenbeek's (1804) phonology-based distinction was successfully diffused in early nineteenth-century usage.

Examining possible genre variation, it was shown that the corpus results for sharp-long *ê* did not involve interesting patterns. For soft-long *ē*, however, the distribution in eighteenth-century newspapers but also diaries and travelogues indicated a preference for the <e> spelling, occurring in more than two-thirds of the instances. In private letters from the same period, there was a strong competition between the equally frequent variants <e> and <ee>. In line with Siegenbeek's (1804) prescription, all three genres adapted <e> as the (only) nineteenth-century main variant for soft-long *ē*. This also includes private letters, even though the amount of variation was still most marked in this conceptually 'oral' genre.

With regard to regional variation, the results for sharp-long *ê* did not reveal striking patterns. There was more variation in the results for soft-long *ē*, though. Against the background that the historical-phonological distinction between the two long *e*'s had been preserved in most dialects of, for instance, Zeeland and Groningen, while they had merged in other regions like North Holland and Amsterdam in particular, it might be assumed that these differences in spoken language were, at least partly, also reflected in the writing practices.

Comparing eighteenth-century letters from Zeeland (maintenance) and Amsterdam (merger), Rutten & van der Wal (2014) no longer observe a striking influence of this distinction in pronunciation on the writing practices. Instead, they witness a so-called *graphemisation* of the writing system, i.e. “the reduction of phonological considerations and the increase in choices directly linked to the written code” (Rutten & van der Wal 2014: 41). For the late eighteenth- and early nineteenth-century period under investigation, the results based on data from the *Going Dutch Corpus* at least indicate some tendencies. Two regions, in which the distinction had been maintained, i.e. Zeeland and Groningen, showed a considerably higher share of the <e> spelling for soft-long *ē* (as opposed to <ee> for sharp-long *ê*) than, for instance, North Holland, where the phonemes had merged. While it is difficult to establish a direct link between these distributional preferences in writing and the preserved phonological distinction in dialectal pronunciation, the relative tendencies in the corpus data are certainly noticeable.

Investigating gender variation in handwritten ego-documents, it was once again shown that sharp-long *ê* did not involve much variation in either of the two periods. Some interesting patterns were revealed for soft-long *ē*, though. Eighteenth-century men appeared to use <e> more often than their female contemporaries. This pattern was also attested for the nineteenth-century period, in which the rise of the prescribed <e> spelling was considerably more marked among male writers than among females, the latter of which still used a fairly high proportion of <ee> for soft-long *ē*. Especially in private letters, the dominance of <e> in texts written by men was some 15% higher than in texts by women.

In general, however, the officialised orthographic distinction between <e> and <ee> for etymologically distinct long *e*'s in open syllables can certainly be assessed as an effective spelling norm in the context of the Dutch *schrijftaafregeling*. Despite the ongoing graphemisation of the writing system, according to which localisable phonological features became less important for spelling practices (Rutten & van der Wal 2014: ch. 2), Siegenbeek's (1804) phonology-based system for sharp-long *ê* and soft-long *ē* was widely adopted in early-nineteenth century practice, across all genres, regions and genders.

Orthographic variables (5)

West Germanic * $\bar{i}$

1 Discussion in Siegenbeek (1804)

As the very first spelling issue discussed and regulated in the second part (*Tweede afdeeling*) of his 1804 orthography, Siegenbeek addressed the letters *ij* and *y*, which also covers the orthographic representation of the vowel derived from West Germanic * $\bar{i}$ ⁶¹ as <ij> or <y> in words such as *wij/ny* ‘we’ or *mijn/myn* ‘my’. Historically, Wgm. * $\bar{i}$ was pronounced as a monophthong [i:] probably up until the end of the Middle Dutch period. In the sixteenth and seventeenth centuries (or in initial stages even earlier, cf. Willemys 2013: 76), the diphthongisation of the long *i* into [ei] took place. Traditionally considered to have its roots in the Brabant dialects, diphthongisation ultimately spread from the culturally and politically dominant cities in Holland to the surrounding areas. This sound change is typically associated with Kloeke’s (1927) theory of the *Hollandsche expansie* ‘Hollandic expansion’. More recently, however, it has been argued that diphthongisation is “a polygenetic phenomenon, having started almost simultaneously in Brabant and Holland” (Willemys 2013: 76).

Irrespective of the exact regional or social genesis, it is important to highlight that the diphthongisation of the long *i* did not spread to the entire language area. In fact, whereas the central part of the language area (including Holland, Utrecht and Brabant) realise [ei] (or the more open [æi]), Wgm. * $\bar{i}$ has remained undiphthongised in several other dialect regions until the present day. In large parts of Zeeland as well as in the eastern parts of the Netherlands, from Groningen to (eastern) Limburg, the historical monophthong has been maintained (cf. Goossens et al. 2000a: 123–145).

Focusing on the orthographic representation of Wgm. * $\bar{i}$, Siegenbeek (1804) both elaborately and critically outlined the historical developments, from its origins in medieval manuscripts as the doubled <ii> spelling to the two central representations as <ij> and <y>. The rise of the latter was criticised by Siegenbeek (1804: 76) for being unnecessary and against the nature of Dutch language:

⁶¹ Vosters et al. (2012: 263) refer to this variable as the spelling of the diphthongised [ei]. However, since the vowel under investigation was not diphthongised throughout the whole language area (van der Wal & van Bree 2008: 262), I more abstractly refer to the vowel derived from Wgm. * $\bar{i}$.

Uit het voorgestelde laat zich nu ook gemakkelijk het antwoord opmaken op de vraag, of men de *y* in de spelling van Nederduitsche woorden tegenwoordig al dan niet zal blijven behouden. Immers, daar dit klankteeken, buiten eenige noodzake, en tegen de oorspronkelijke inrigting onzer tale, door eene ongepaste navolging der Franschen, in dezelve is doorgedrongen, daar men de *y* reeds overlang uit Nederduitsche woorden heeft begonnen uit te monstren, en latere taalkundigen en de netste schrijvers het te dezen aanzien eens zijn, zoo kan er geene reden worden uitgedacht, waarom men deze letter uit het Nederduitsch *alphabet* niet geheel zou uitwisschen, en daarvoor het echte Nederduitsche klankteeken *ij* in de plaats stellen.

'From the presented, one can easily give the answer to the question whether or not one should maintain the *y* in the contemporary spelling of Dutch words. After all, as this letter has penetrated into our language, needlessly and against its original design, by inappropriately following the French, and as one had already started to reject the *y* in Dutch words long time ago, which later linguists and the most respectable writers agreed on, one cannot think of any other reason, why one should not entirely erase this letter from the Dutch *alphabet*, and substitute it by the real Dutch letter *ij* instead.'

In line with 'later linguists and the most respectable writers', Siegenbeek (1804: 79-80) ultimately rejected the foreign letter <y> in favour of the indigenous Dutch <ij>:

Het besluit van al het verhandelde [...] is derhalve, dat men het klankteeken *y* alleen gebruiken moet in woorden, welke, door middel der Latijnsche, uit de Grieksche taal genomen zijn, en waarin hetzelfde den waren klank van *u* heeft, als *Cyrus*, *Cyprus*, *Assyrie* en meer dergelijke vreemde benamingen; doch in echte Nederduitsche woorden, met geheele verwerping der aan onze taal niet eigene *y*, alleen het Nederduitsche klankteeken *ij* behoort te bezigen.

'The conclusion of the whole discussion [...] is therefore that one must only use the letter *y* in words which were borrowed from the Greek language via Latin, and in which it has the true sound of *u*, as *Cyrus*, *Cyprus*, *Assyrie* and more of these foreign names. In proper Dutch words, however, one must only use the Dutch letter *ij*, completely rejecting the *y*, which is alien to our language.'

In other words, Siegenbeek prescribed <ij> as the only orthographic representation for *echte Nederduitsche woorden* 'proper Dutch words' with Wgm. **i*. Since he regarded *y* as a letter which was not part of the Dutch language but had rather been derived from the Greek language via Latin, Siegenbeek completely rejected the foreign letter in these words⁶². As an exception, foreign proper names (of places, persons, etc.) such as *Cyprus* or *Abysinie* still had to be spelled with <y> and, importantly, also represents a different pronunciation.

⁶² In a similar way as the Greek *y*, Siegenbeek also rejected three other (foreign) letters, viz. *c*, *q* and *x*. In Dutch words, they should be replaced by the 'indigenous' representations <s> or <k> (for <c>), <kw> (for <q>) and <ks> (for <x>), whereas they were maintained in foreign proper names such as *Cyrus*, *Quintus* or *Xerxes* (cf. Siegenbeek 1804: 80-81).

2 Eighteenth-century normative discussion

The orthographic representation of Wgm. **i* as <ij> or <y> can certainly be regarded as a widely discussed spelling issue in the Northern Netherlands throughout the 1700s. In fact, there is hardly any consensus in metalinguistic discourse, with most eighteenth-century grammarians either preferring undotted <y> or double-dotted <ij>. A few others rather indifferently acknowledged <y> and <ij> as coexisting variants, or even introduced a spelling difference in open and closed syllables. This section aims to provide an outline of the normative discussion, focusing on the major tendencies in the eighteenth century.

In the first decade, grammarians such as Moonen (1706) and Sewel (1708) acknowledged the *y* as a legitimate letter of the Dutch alphabet and favoured the <y> spelling for the vowel derived from Wgm. **i*. Moonen (1706: 19), for instance, argued that the lengthening of the *i* by adding *j* is ‘not so good or natural’, as the *j* in the <ij> spelling is actually a consonant:

De Y, de leste en zeste Klinkletter, alleen lang, als wy voorhene gezegt hebben, en uit den langen klank der woorden *Vry, myn, pyn, tyt*, waer in zy gevonden wordt, blykt, kan niet verlengt worden.

‘The Y, the last and sixth vowel, only long, as we have said before, and from the long sound of the words *Vry, myn, pyn, tyt*, in which it can be found, it appears that it cannot be lengthened.’

Similarly, Sewel (1708) also preferred <y> over <ij>, remarking that the latter is frequently used, although it inappropriately consists of a vowel and a consonant:

Ij wordt zeer veel gebruykt in plaats van de Y, zonder dat men aanmerkt dat zy een klinker en een medeklinker is; maar moogelyk heeft de gelykheyd der Duytsche letteren *ij* en *y* aanleyding tót deeze dwaalinge gegeven [...] (1708: 21)

Y, by sommigen afgekeurd, omdat die by de aaloude Grieken voor eene U gebruykt wierdt, heeft eechter by ons, gelyk ook by de Engelschen, een’ klank dien wy niet derven konnen; en dient in de woorden *Ys, ydel, yver, yzer, ysl, myden, ryden, pryzen, vry, slaaverny, spotterny, schildery* [...] (1708: 31)

‘Ij is very often used in place of the Y, without considering that it is a vowel and a consonant. But possibly the resemblance of the German letters *ij* and *y* has caused this mistake. [...]

Y, rejected by some because it was used for a U by the ancient Greek, however, has a sound in our language, like in English, which we cannot lack, and which is used in the words *Ys, ydel, yver, yzer, ysl, myden, ryden, pryzen, vry, slaaverny, spotterny, schildery* [...]

Less explicitly than Moonen and Sewel, other early eighteenth-century grammarians like Nylöe (1703) or Verwer (1707) also seemed to support the <y> spelling (Vosters 2011: 240), although these preferences were only vaguely touched upon in

their publications. Furthermore, Nylöe (1703: 10) used both <y> and <ij> in his writing, despite arguing that “de *y* is een dubbelde *i*” (and not *ij*).

Hakvoord (1746: 46) accepted <y> as some kind of compromise. He actually preferred the old <i> spelling, but, at the same time, admitted that it might be too ‘odd’ and ‘poor’ from the user’s perspective:

Met deze i behoorde men te spellen by, gy, hi, wi, zi⁶³, enz. want het is een volkomen letter die de woorden haren behooryken klank geeft. Maar indien we dit volstrektelyk zo wilden gedaan hebben en zelfs deden dat zou den Lezer al te vreemd en mager voorkomen en niet nagevolgt worden want:

Men kan in ‘t oud misbruik, en lang verloop der zaken,

Eer alles stukken slaan, als ‘t kwade beter maken.

En daarom volgen wy hier in de sleur, en spellen by, gy, hy, wy, zy, met een dubbelde y hoewel het ander beter is.

‘With this i one had to spell by, gy, hi, wi, zi, etc., because it is a complete letter, which gives the words their proper sound. But if we wanted to do this completely and even did, then it would appear far too odd and poor to the reader, and would not be followed because:

In the old misuse, and long course of things,

One can rather smash everything into pieces, than make the wrong better.

And thus we follow the routine here, and spell by, gy, hy, wy, zy, with a double y although the other is better.’

In general, it becomes evident that a number of eighteenth-century grammarians were more tolerant towards alternative spelling variants. Ten Kate (1723: 116), for example, described the choice between <y> and <ij> as arbitrary⁶⁴, since it is usage which gives a particular value to the character and the words (cf. also van de Bilt 2009: 185):

onze scherpe *Kort-klinker* I, als bij MIN; en deeze tot omtrent op het dubbeld verlangt zijnde onze *Lang-klinker* Y (of IJ), als bij MYN (*Meus*) en LYDEN (*pati*) [...] De Latynsche Y, schoon die van de Grieksche *Ypsilon* ontleent is, hebben onze Drukkers seedert eenige jaeren het zelfde doen gelden als onzen *Langklinker* IJ. Ik twist hier niet oover de Gedaente der Letters, maer handele van onze Klanken, en welke Letterteikens daar voor gangbaer zijn. ’t Gebruik geeft de Waerde zo wel aen de Characters als aen de Woorden.

⁶³ It remains unclear why Hakvoord (1746) mentioned two different spellings here, i.e. <y> in *by, gy* alongside <i> in *hi, wi, zi*.

⁶⁴ It is important to keep in mind that ten Kate (1723) described two kinds of orthography. As pointed out by van der Wal (2002: 59), he made a distinction between “the ‘burgerlijke’ (civil) or ‘gemeene’ (common) orthography, based on custom, that is mainly on the usage of prestigious authors versus the ‘natuerkundige’ (physical) or ‘critique’ (critical) orthography, based on the principle of representing one sound by only one symbol”. Therefore, ten Kate’s ‘tolerance’ or ‘indifference’ towards alternative variants in common usage (e.g. <y> or <ij> as an ‘arbitrary’ choice) has to be interpreted against the background of his two spelling systems.

‘our sharp short vowel *i*, as in MIN, and, approximately doubled in length, our long vowel *Y* (or *IJ*), as in MYN (*Meus*) and LYDEN (*pati*) [...] Since a few years, our printers have used the Latin *Y*, although borrowed from the Greek *Ypsilon*, in the same way as our long vowel *IJ*. I do not argue about the form of the letters, but deal with our sounds, and which letters are commonly accepted for them. The use gives the value to both the characters and the words.’

Van der Palm’s (1769: 4-5) view is remarkable as he advocated (and consistently used) the Greek-derived <y> spelling, but also acknowledged alternative variants. He argued that <ij> is not qualified to lengthen the *i* as it contains the consonant *j* (see the argumentation by Moonen and Sewel above). Therefore, <ii> should be used, which, he admitted, did not happen in practice:

Vr. Hoe kunnen dan de woorden *gy*, *by*, *myn*, *zyn*, enz. gespelt worden?

Antw. Zy zeggen dat men de woorden, die men voorhene met eene *y* geschreven heeft, met *ii*, of sieraedshalve met *ij* moet schryven, als *ziin* of *zijn*, *miin* of *mijn*, enz.

Vr. Wat is ‘er dan van de letter *y* te houden?

Antw. Ons dunkt dat men de letter *y* niet zoo ligt uit onze Klinkeren moet verbannen: vooreerst, om dat onze Schryvers deze Grieksche Letter meest altydt zo veel als den langklinker *ij* hebben doen gelden, en dat dus hunne lettergrepen, met *y* gespelt, eenen anderen klank zouden verkrygen, indien men deze letter geheel wilde uitmonsteren: ten andere, om dat de *j*, als een Medeklinker, niet bekwaem is om de *i* te verlengen; waerom men genoodzaekt zoude zyn de spelling met *ii* intevoeren, ’t geen echter van niemant geschiedt.

Q. How can the words *gy*, *by*, *myn*, *zyn*, etc. be spelled then?

A. They say that one must write the words which had previously been written with a *y*, with *ii* or for the sake of decoration with *ij*, as *ziin* or *zijn*, *miin* or *mijn*, etc.

Q. What should one think about the letter *y* then?

A. It seems to us that one should not ban the letter *y* from our vowels so easily. First of all, because our writers have mostly used this Greek letter as the long vowel *ij*, and their syllables spelled with *y* would thus get a different sound, if one wants to reject this letter completely. On the other hand, because the *j* as a consonant is not able to lengthen the *i*, which is why one would be forced to introduce the spelling with *ii*, which is done by nobody, though.’

Although he was certainly aware of the importance of norms, van der Palm (1769: 5) did not care about the spelling as either <y> or <ij> in usage, as long as the selected grapheme represents the right pronunciation:

Vr. Is dit geschil van eenig belang?

Antw. Het is van zeer weinig belang hoedanig men deze letter schryve, mits men dezelve den rechten klank mededeelde.

Q. Is this difference of any importance?

A. It is of very low importance how one writes this letter, as long as it represents the right sound.’

A similar stance towards the orthographic representation of Wgm. **i* can be found in Heugelenburg (1763: 18). Unlike van der Palm, Heugelenburg preferred <ij>, but also acknowledged the alternative spelling (the *uitlandzē* ‘foreign’ <y> in this case):

Maar, aangemerkt dat de dubbelde ij mede een zeer helder geluit slaat, en aan veele woorden, in haar uitspraak kragt en klem bij zet, zo en zoude ik niet afkeerig zijn om dezelve voor een zesde Klinkletter, in het Staatendom van Nederlandse Spraake in te huldigen, schoon die uit een I en J word te zaamen gesteld. [...] Dog al de gene die in deszelvs plaats de uitlandze Y, met die waardigheid willen vereeren ik kan verdraagen, en laat een ieder Beminnaar van de Spelkonst, daar in zijn eige bevatting en verkiezing opvolgen.

‘But noting that the double ij also has a very clear sound, which adds power and emphasis to the pronunciation of many words, I should not be averse to inaugurating it as a sixth vowel letter in the State of the Dutch language, although it is composed of a I and J. [...]

But I can tolerate everyone who wants to honour the foreign Y instead with that dignity, and I let every lover of spelling follow his own opinion and choice.’

In the course of the century, another tendency emerged in metalinguistic comments, according to which <ij> should be used as the only spelling variant. Huydecoper (1730: 644) was one of the first grammarians to prescribe <ij> instead of <y>, the latter of which was exclusively used for foreign proper names:

Wegens de Letter Y.

De Leezer zal deeze Letter in de twee volgende Bladwijzers niet vinden, dan in Eigen Naamen van vreemden oorsprong; in Duitsche woorden, overall *ij*. [...] De *y* is geen Duitsche Letter, maar een Latijnsche; vervangende, in die taale, de Grieksche *υ* of *Υ*, als in *Pythagoras*, *Cyprus*, *Tyrannus*, *Polydemon*, enz.

‘With regard to the letter Y.

The reader will not find this letter in the two following tables of contents, only in proper names of foreign origin: in Dutch words, everywhere *ij*. [...] The *y* is not a Dutch letter, but a Latin one, replacing the Greek *υ* or *Υ* in this language, as in *Pythagoras*, *Cyprus*, *Tyrannus*, *Polydemon*, etc.’

Particularly in the second half of the eighteenth century, the use of <ij> was increasingly advocated in normative works. The 1770 grammar by the society *Kunst wordt door arbeid verkreegen* explicitly prescribed the use of the double-dotted spelling, while rejecting the <y>, which had sloppily and wrongfully ‘intruded’ into many Dutch words:

De Grieksche *y* tellen wij daer niet onder, omdat zij niet tot ons behoort, en welke wij, op het voetspoor van den grooten HUYDECOPER, en andere Vraegbakens onzer Taelkunde, uitzonderen, gebruikende haer alleen in woorden, waerī zij volstrekt wezen moet, en den vollen klank der *i* heeft, zonder t’samenvoeging met de *e*, als in *Cyrus*, *Cyprus*, en diergelijken. (1770: 8)

Door deze *ij* verstaen wij niet de *y*, die zoo slordig in velen onzer Nederduitsche woorden, op een gansch onrechtmaetige wijze is ingedrongen; [...] Wij bezigen dan voorts onveranderlijk niet de *y* maer de *ij*, met twee punten, en voornamelijk in deze woorden: *bedijden*, *bedrijven*, *belijden*, *benijden*, *bedijken*, *beklijven*, *bezwijmen*, en ontelbaere meer, genoegzaam aen den klank te kennen. (1770: 21-22)

‘We do not count the Greek *y* as such, because it does not belong to us, and which we exclude, following the great HUYDECOPER and other handbooks of our language, only using it in words, where it must absolutely be, and where it has the full sound of *i*, without the combination with the *e*, as in *Cyrus*, *Cypres*, and the like.

By this *ij* we do not mean the *y*, which has so sloppily intruded into so many of our Dutch words in an entirely wrongful way [...] In the following we invariably use not the *y* but the *ij*, with two dots, and especially in these words: *bedijden*, *bedrijven*, *belijden*, *benijden*, *bedijken*, *beklijven*, *bezwijmen*, and countless more, sufficiently recognisable by the sound.’

The argumentation here already shows strong similarities to the officialised rule by Siegenbeek, referring to the idea that the Greek and thus foreign *y* was alien to the Dutch language and must therefore be replaced by <ij> (except in proper names). Zeydelaar (1774) and Stijl & van Bolhuis (1776) followed this approach, too.

Towards the turn of the century, metalinguistic comments on <ij> versus <y> had become increasingly coherent. In fact, the rule in favour of <ij>, ultimately prescribed in Siegenbeek’s 1804 orthography can already be found in most normative works published in the late 1700s, most notably van Bolhuis (1793), Weiland (1799) and, as quoted below, the *Rudimenta* (1799: 53-55):

Van de letter Y heeft men drie zaaken⁶⁵ op te merken als:

1. Dat de oprechte Y geene Nederduitsche maar eene vreemde letter is: – en daarom ook niet dan in vreemde woorden mag gebruikt worden, als in *Cyrus*, *Syllabe*, *Synode*, *Cyprus*, *Ivoor*, *Egypte*, *Hyssop*, *Pyrrus*, *Pyramide*, en dan klinkt zij als of er stont *Cirus*, *Egipte*, *Sinode*, *Hissop*, *Ivoor* enz.
2. Dat de IJ, welke in het Nederduitsch gebezigd en voor eene letter deezer taale aangetekend wordt, zoo zeer geen’ Klinkletter, als wel eene verlengde I is, en dus als twee II behoorde geschreeven en uitgesproken te worden [...]

‘About the letter Y one has to note three things, namely:

1. That the true Y is not a Dutch but a foreign letter: and thus it must not be used apart from in foreign words, as in *Cyrus*, *Syllabe*, *Synode*, *Cyprus*, *Ivoor*, *Egypte*, *Hyssop*, *Pyrrus*, *Pyramide*, and then it sounds as if it was *Cirus*, *Egipte*, *Sinode*, *Hissop*, *Ivoor* etc.

⁶⁵ The third point, which is not quoted here, refers to another spelling issue, viz. the difference between <ij> and <ei>: “3. Dat vermits deeze letter bijna overal in ons Vaderland, als *ei* wordt uitgesproken, men naauwkeurig behoort opteletten, dat men de IJ dan ook niet met EI verwarre” ‘Since this letter is pronounced as *ei* almost everywhere in our fatherland, one has to mind carefully that one does not confuse the IJ with EI’ (*Rudimenta* 1799: 54).

2. That the IJ, which is used in Dutch and denoted as a letter of this language, which is not primarily a vowel letter but rather a lengthened I, and thus has to be spelled and pronounced as two II [...]

The spelling choices by Kluit, who was undisputably a major source of inspiration for Siegenbeek, deserve special attention. In his 1763 and 1777 treatises, he actually suggested two different approaches. In Kluit's first *Vertoog* (1763), he prescribed both <ij> and <y> as coexisting variants, though used for Wgm. **i* in different positions. This led to a comparatively complex spelling rule (1763: 345):

Daar wy thans de *i* gebruiken, in *bereyd*, *goetheyt* enz., daar schreven zy de *y*. Diezelfde *y* was by hun gangbaar op het slot eens woorts, als *gerhy*, *my*, *hy*; in tegendeel vinden wy altijd *ij*, wanneer de sluitletter een dubbele *i* vereischte in *wijn*, *schijn*, *schrijft*; en deze *ij* treffen wy ook meest aan, ingeval die opgenoemde woorden verlengt worden, als *wijne*, *schijnen*, *schrijven*.

‘Where we use the *i* nowadays, in *bereyd*, *goetheyt* etc., they wrote the *y*. The same *y* was common at the end of a word, as *gerhy*, *my*, *hy*. In contrast, we always find *ij* when the final letter required a double *i* in *wijn*, *schijn*, *schrijft*; and this *ij* we also mostly find in case these mentioned words were lengthened, as *wijne*, *schijnen*, *schrijven*.’

For Wgm. **i* in open syllables as in *my* ‘me’, Kluit suggested the <y> spelling, whereas <ij> should be used in closed syllables as in *wijn* ‘wine’. In derivations of words with Wgm. **i* in closed syllables, <ij> should also be used when the vowel occurs in open syllable, for instance in *schrijf* (closed syllable) – *schrijven* (open syllable, but not spelled as *schryf* due to derivation).

In his second *Vertoog*, Kluit (1777: 6) no longer distinguished <ij> and <y> in different syllabic positions, but invariably prescribed double-dotted <ij>:

onder welke letters er ook bij ons *een* is, de *i*, namelijk, die thans ook door toeval in hare verlenging een teeken op zich zelf (de *ij*) heeft aangenomen, en daardoor zeer verkeerdlijk in het getal der Vocalen als een zesde Vocaal geplaatst is: waarbij ook komt, dat door zekere taalverbastering deze dubbele *ii*, (thans *ij*, of door de Drukkers lomper *y*, geschreven) in sommige Provincien of Dialecten den wanklank van de Diphthong *ei* gekregen heeft.

‘among our letters there is also *one*, namely the *i*, which, in its lengthening, has now adopted a letter in its own right (the *ij*), also by coincidence, and thus placed very wrongfully among the vowels as a sixth vowel. In addition, this double *ii* (now *ij* or more clumsily written *y* by the printers) has received the cacophony of the diphthong *ei* in some provinces or dialects through a certain language corruption.’

Kluit remarked that the representation as <y> had been used ‘clumsily’ by printers. Interestingly, he did not refer to his earlier choices as discussed in 1763 (van de Bilt 2009: 191). In that sense, Kluit's shift from syllable-dependent <ij> and <y> in 1763 to <ij> only in his 1777 also illustrates the general development towards <ij> in normative works of the late 1700s (cf. also Vosters 2011: 240). Despite a general lack of uniformity in eighteenth-century metalinguistic

comments, the increasingly strong tendency in favour of <ij> paved the way for Siegenbeek's official norm prescribing the double-dotted spelling.

3 Previous research

It is not surprising that a controversially discussed spelling feature like the orthographic representation of Wgm. **i* as either (undotted) <y> or (double-dotted) <ij> has attracted the interest of several linguists. Matthijsen (1988: 133–134), for example, presents an overview of Siegenbeek's spelling choices in contrast to those of his well-known critic Willem Bilderdijk, referring to the *ij/y* controversy as “[h]et meest in het oog lopende verschil, dat direct als herkenningspunt gebruikt kan worden” ‘the most striking difference, which can directly be used as a distinct feature’.

In the broader context of the standardisation of the Dutch spelling, Molewijk (1992: 113) also addresses *ij/y*, claiming that the <ij> spelling was very quickly adopted in the nineteenth century (i.e. after Siegenbeek's 1804 orthography). The corpus study of this variable in this chapter will show whether this was really the case.

The eighteenth-century normative tradition in the Northern Netherlands is at the heart of van de Bilt (2009). In his PhD thesis, he discusses the metalinguistic comments by various influential grammarians like Moonen, Verwer, ten Kate, Huydecoper, Kluit as well as Siegenbeek, also addressing the *ij/y* spelling issue.

Previous studies from a historical-sociolinguistic perspective have mainly focused on the situation in the Southern Netherlands, most notably Vosters (2011) and Vosters et al. (2012). With respect to norms and usage, they point out that eighteenth-century Flemish orthographers exclusively prescribed the undotted <y> spelling. In fact, <y> became a typically Southern spelling feature as opposed to double-dotted <ij> as the (seemingly) typical counterpart of the Northern Netherlands (Vosters et al. 2012: 263–264). In one of their case studies, based on a corpus of nineteenth-century manuscripts from the judicial and administrative domains, Vosters et al. (2012: 268) show that <y> is the dominant variant in Southern usage, occurring in roughly three quarters of all instances in the data sets from both 1823 and 1829. Given the fact that the Northern (Siegenbeek) variant <ij> was also increasingly prescribed in Southern normative works after 1815, this is a remarkable result, supporting the idea of the ‘Southernness’ of this variant.

Vosters et al. (2010) take a more comparative perspective, also investigating language variation in the Northern Netherlands. They present an overview of Northern norms as well as an exploratory case study on *ij/y* variation in a corpus of 100 private letters from the 1780s. It turns out that late eighteenth-century language practice (<ij>: 63%; <y>: 37%) was mainly in line with the heterogeneous character of eighteenth-century metalinguistic comments, which promoted both <ij> and <y> (cf. Section 2). Vosters et al. (2010: 105) further argue that this result is particularly striking from the perspective of the Southern normative tradition, as <ij> had frequently been evaluated as a typical Northern

feature since the eighteenth century. Building on these exploratory findings, the present chapter systematically examines variation and change in the use of variants, focusing on the late eighteenth and early nineteenth century.

4 Corpus analysis

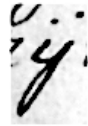
4.1 Method

In the corpus analysis of the orthographic representation of Wgm. **i̇*, four different variants were considered. Whereas both eighteenth-century normative works and Siegenbeek (1804) make a distinction between <ij> and <y>, there is actually much more variation in the handwritten texts of the *Going Dutch Corpus*, which had to be taken into account as well. Therefore, the following four (rather than just two) variants were distinguished during the transcription process of handwritten ego-documents (as pointed out earlier in Chapter 4):

- (1) <ij> i.e. double-dotted <ij> with <i> and <j> written as two separate characters
- (2) <ÿ> i.e. double-dotted <y>
- (3) <y> i.e. (undotted) <y>⁶⁶
- (4) Other e.g. single-dotted <y>, <y> with accent marks or other diacritics, etc.



(1) <ij>



(2) <ÿ>



(3) <y>



(4) Other

Admittedly, the boundary between variants (1) and (2) is not always clear-cut in handwriting. In fact, they are best regarded as (sub-)variants both representing the double-dotted spelling, as opposed to the undotted variant (3). The fourth variant, transcribed in the *Going Dutch Corpus* as °y, actually comprises various forms, which are neither double-dotted nor undotted. Referred to as the variant 'Other' in this chapter, (4) comprises variants such as the single-dotted <y>, <y> with different accent marks, diacritics, and so on. None of these forms was actually mentioned or discussed in the eighteenth- and nineteenth-century

⁶⁶ For the sake of clarity, I refer to <y> as the 'undotted' variant in this chapter. It should be noted, though, that this is mainly a perception from Siegenbeek (1804) onwards, with <y> being the 'undotted' variant of the prescribed spelling norm <ij> (or <ÿ>). Terminologically, 'undotted' might not be entirely accurate in the context of the more heterogeneous eighteenth-century normative tradition.

normative tradition, but they do occur in the *Going Dutch Corpus*. However, as they are relatively marginal in usage compared to the double-dotted and undotted spellings, it was decided to merge them into one category. In fact, the crucial aspect of the spelling choices in both language norms and language usage is the presence or absence of the (two) dots, i.e. the double-dotted spelling(s) <ij>/<ÿ> versus the undotted <y> spelling.

In order to be able to assess the actual use of variants in late eighteenth- and early nineteenth-century texts, the ten most frequent words containing Wgm. *ī were selected according to their occurrences in the *Going Dutch Corpus* (listed in order of decreasing frequency in the corpus):

- WIJ; ZIJN/SIJN; MIJ; BIJ, MIJN; ZIJ/SIJ; HIJ; GIJ; TIJD/ALTIJD;
SCHRIJVEN/SCHRIJF

In the cases of ZIJN/SIJN and ZIJ/SIJ, orthographic variation between *s/z* was taken into account. Furthermore, TIJD and ALTIJD were combined into one set of words, as well as SCHRIJVEN/SCHRIJF, i.e. with two different verb forms, mainly in order to increase the number of tokens.

4.2 Results

In order to provide a general overview of the distribution of variants, the orthographic representation of Wgm. *ī was investigated in the entire *Going Dutch Corpus*, as shown in Table 1. The officially prescribed spelling (i.e. double-dotted <ij>/<ÿ>) is highlighted in light grey.

Table 1. Distribution of variants across time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ij>	<ÿ>	<y>	Other	<ij>	<ÿ>	<y>	Other
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Total	322 (3.6)	3,156 (35.3)	4,668 (52.2)	800 (8.9)	989 (10.1)	3,369 (34.4)	4,934 (50.4)	497 (5.1)
Open syll.	121 (3.0)	1,525 (37.9)	2,042 (50.7)	341 (8.5)	277 (5.7)	1,776 (36.8)	2,508 (52.0)	266 (5.5)
Closed syll.	151 (4.5)	1,158 (34.3)	1,707 (50.6)	359 (10.6)	395 (16.4)	802 (33.2)	1,108 (45.9)	111 (4.6)

In the eighteenth century, the undotted <y> is prevalent in usage (52.2%), thus considerably more frequent than double-dotted <ij>/<ÿ> (38.9%). On a more graphological level, it turns out that the double-dotted realisation as <ÿ> (35.3%) clearly outnumbers the realisation as <ij> (3.6%), i.e. with <i> and <j> as

neatly separated characters. In addition to the undotted/double-dotted distinction, there is even room for more variation, collected in the category ‘Other’ (8.9%).

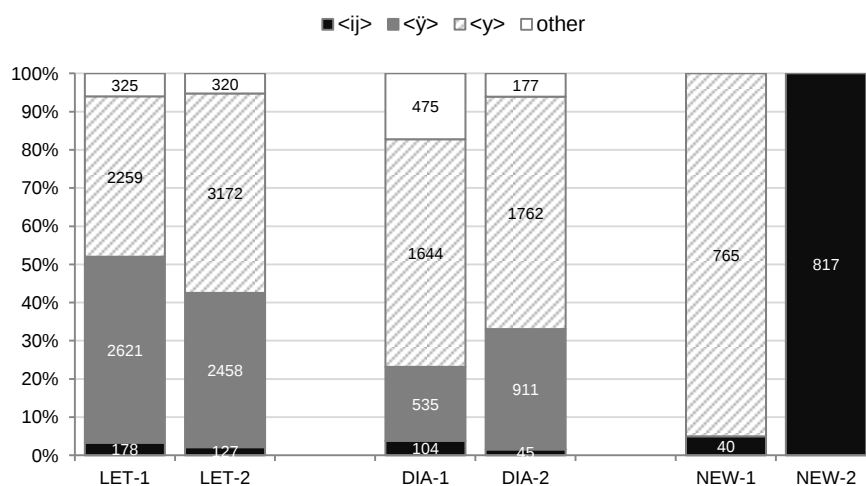
In the nineteenth-century period, i.e. after Siegenbeek’s orthography was introduced, the overall distribution of variants remains surprisingly stable. The use of the prescribed double-dotted variants does increase from 38.9% to 44.5%, but with a share of 50.4%, the rejected undotted <y> is still most frequently used. Furthermore, the use of the other forms drops to an even more marginal 5.1%.

As pointed out in Section 2, a couple of eighteenth-century grammarians, including Kluit (1763), distinguished between Wgm. **i* in open and closed syllables, suggested to be spelled with <y> and <ij>, respectively. In order to take this internal factor into account as a possible source of influence on (particularly eighteenth-century) variation, the results for the three most frequent words with Wgm. **i* in open syllable (WIJ; MIJ; BIJ) and the three most frequent words with Wgm. **i* in closed syllable (ZIJN/SIJN; MIJN; TIJD/ALTIJD) are also presented separately in Table 1. However, it turns out that there are hardly any syllable-related differences between the items under investigation. Especially in the eighteenth-century data, the variants are very similarly distributed across both open and closed syllables. It can therefore be assumed that the syllable-related distinction does not explain the high degree of variation in the distribution of spelling variants. In the discussion of external variables, this internal factor will no longer be considered.

Genre variation

Focusing on genre differences, the distribution of variants was investigated across the three sub-corpora, i.e. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW). Figure 1 reveals major differences in the distribution of variants across genres.

Figure 1. Distribution of variants across genre and time.



First of all, in eighteenth-century private letters, the double-dotted variants are slightly dominant (52.0%), but undotted <y> also occurs frequently (42.0%). The remaining 6.0% comprise the other variants. In diaries and travelogues, it is the undotted spelling which is predominantly used as the main eighteenth-century variant (59.6%). The double-dotted spellings are realised in only 23.2% of all instances. Furthermore, we can see a comparatively high share of the ‘Other’ category (17.2%). In the newspaper data from the same period, the use of variants is a rather clear-cut choice. Undotted <y> is more or less exclusively used (95.0%), except for a number occurrences of <ij> (5.0%), which all derive from the *Utrechtse courant* (representing the region of Utrecht).

In the nineteenth century, after Siegenbeek prescribed <ij> as the national variant, the use of the double-dotted spellings surprisingly decrease from 52.0% to 42.5% in private letters. The undotted spelling <y> becomes the dominant variant (52.2%). In contrast, the use of the double-dotted spelling increases from 23.2% to 33.0% in diaries and travelogues. <y> remains by far the most frequently used variant, though, with a stable share of 60.9%. More generally, this means that <y> is the main variant in nineteenth-century ego-documents. As in the first period, nineteenth-century newspapers consistently use one single variant. However, a radical change of variants took place in these texts, involving a complete shift from <y> in period 1 to <ij> in period 2. Newspapers thus adopt the prescribed spelling in 100%.

In general, genre variation reveals two different types of distribution: On the one hand, much variation in the use and distribution of variants can be attested in the sub-corpora of private letters as well as diaries and travelogues, i.e. in handwritten ego-documents. On the other hand, there is a clear-cut, consistent choice of variants in printed, published newspapers. This strongly suggests that genre or, more precisely, the medium of the genre (i.e. print versus handwriting, cf. also Rutkowska & Rössler 2012) is an important factor with regard to the orthographic representation of Wgm. **ī*.

Regional variation

Examining possible regional variation in the investigated language area, Table 2 presents the distribution of variants across regions in the entire *Going Dutch Corpus* (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

To begin with, there is much regional variation in the distribution of variants. In the eighteenth-century period, undotted and double-dotted spelling generally co-occur in actual language usage, although the degree to which the main variants dominate differs per region. In the eighteenth-century period, the double-dotted variants are prevalent in usage in the two northernmost regions of Friesland (51.7%) and Groningen (53.9%). In all other regions, undotted <y> is most frequently used, particularly in North Holland, which has by far the highest <y> share (71.8%).

Remarkably, the category of variants other than undotted or double-dotted is comparatively strong in the southern regions of North Brabant (16.1%) and Zeeland (16.9%), both of which are border regions to the Southern Netherlands⁶⁷.

Table 2. Distribution of variants across region and time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	<ij>	<ÿ>	<y>	Other	<ij>	<ÿ>	<y>	Other
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
FR	124 (11.2)	450 (40.5)	455 (41.0)	81 (7.3)	115 (8.2)	451 (32.1)	737 (52.5)	102 (7.3)
GR	36 (3.2)	580 (50.7)	482 (42.1)	46 (4.0)	127 (8.4)	655 (43.5)	662 (43.9)	63 (4.2)
NB	13 (0.9)	382 (27.5)	769 (55.4)	223 (16.1)	229 (19.8)	296 (25.7)	527 (45.7)	102 (8.8)
NH	19 (1.6)	292 (25.0)	837 (71.8)	18 (1.5)	167 (11.0)	465 (30.6)	853 (56.0)	37 (2.4)
SH	34 (2.6)	472 (35.8)	679 (51.5)	134 (10.2)	117 (7.1)	545 (32.9)	953 (57.4)	44 (2.7)
UT	78 (6.6)	454 (38.3)	622 (52.4)	33 (2.8)	117 (8.2)	356 (24.9)	907 (63.5)	48 (3.4)
ZE	18 (1.2)	526 (33.5)	760 (48.4)	265 (16.9)	117 (10.5)	601 (54.0)	295 (26.5)	101 (9.1)

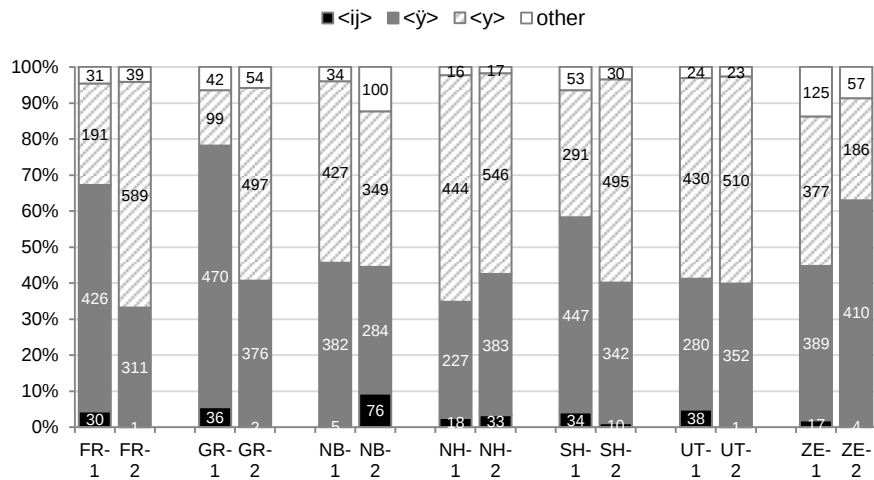
In the nineteenth-century period, undotted <y> remains the main variant in nearly all regions, most notably in Utrecht (63.5%) and, to a lesser extent, the Holland area. The use of Siegenbeek's double-dotted spelling even drops considerably in Friesland (from 51.7% to 40.3%) and Utrecht (from 44.9% to 33.1%). The prescribed spelling, on the other hand, gains ground in North Holland (from 26.6% to 41.6%), North Brabant (from 28.4% to 45.5%) and especially Zeeland, which shows the strongest increase of <ij>/<ÿ> from 34.7% to a 64.5%. At the same time, the use of undotted <y> decreases from 48.4% to 26.5% here. In fact, Zeeland is the only region in which the double-dotted spelling is established as the main variant in conformity with Siegenbeek's prescription.

⁶⁷ It might be argued that the strikingly high share of alternative variants (including <y> with accent marks) in the two border regions North Brabant and Zeeland is related to their possible orientation towards the Southern norms and/or practices. In fact, accent marks (at least for *ɛ*'s and *o*'s) were typical of Southern usage and widely discussed in Southern normative works (cf. ch. 5 and 6 in Rutten 2011). Even though they were not intended for *y*'s, it might well be that the salience of accent marks in general led to an increasing use of <y> variants with accent marks in (hand)writing.

Regional variation across genres

Taking into account the major differences in the distribution of variants across genres (as shown in Figure 1), regional variation will also be looked at in the three genres individually. First, Figure 2 displays the results in the sub-corpus of private letters.

Figure 2. Distribution of variants across region and time (private letters).

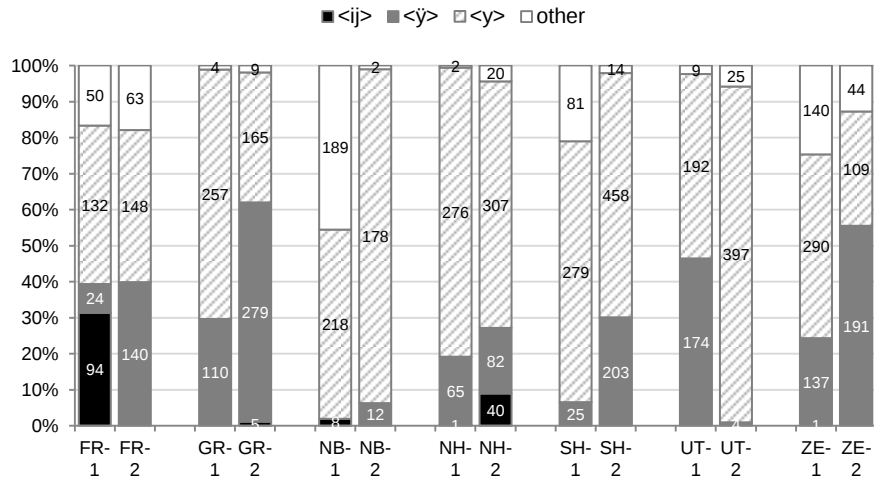


The eighteenth-century data reveal that the double-dotted variants <ij>/<y> are dominant in private letters from the northern regions of Friesland (67.3%) and particularly Groningen (78.2%), as well as in South Holland (58.3%). Surprisingly, after Siegenbeek's officialised norm, the use of the double-dotted spelling considerably drops in these three regions, while undotted <y> becomes the main variant in the nineteenth-century. A rather stable coexistence of undotted and double-dotted variants across both periods can be found in North Brabant and Utrecht. <y> is most frequently used in eighteenth-century North Holland (62.9%). In the nineteenth-century data of this region, the use of the double-dotted spellings increases from 34.8% to 42.5%. However, it is only in private letters from Zeeland in which the officialised double-dotted spelling considerably gains ground as the predominant variant(s) in nineteenth-century usage (from 44.7% to 63.0%).

Next, the results across regions drawn from the sub-corpus of diaries and travelogues are presented in Figure 3. In the eighteenth-century period, <y> is the most frequently used variant across all seven regions. The extent to which it dominates in usage differs, though. It is clearly the main variant in Groningen (69.3%), South Holland (72.5%) and particularly North Holland (80.2%). In Utrecht, <y> (51.2%) co-occurs with the similarly frequent double-dotted spelling (46.4%). In several regions, the presence of alternative variants from the 'Other' category is also remarkable: In North Brabant, they have an extraordinarily high share of 45.5%, co-occurring with <y>. Moreover, the alternative options are fairly

frequent in Zeeland (24.7%) and South Holland (21.0%), and, to a lesser extent, in Friesland (16.7%).

Figure 3. Distribution of variants across region and time (diaries and travelogues).



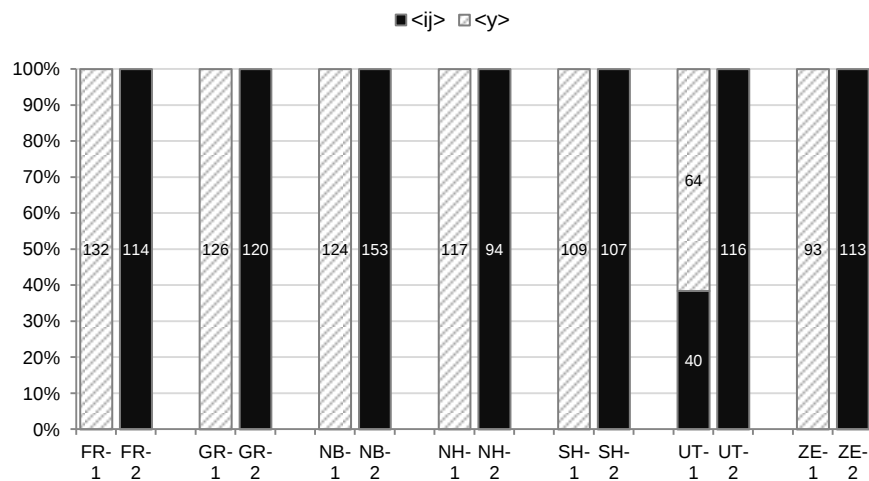
In the nineteenth-century period, the prescribed double-dotted spelling becomes the main variant in diaries and travelogues from Groningen (62.0%) and Zeeland (55.5%). It also gains considerable ground in South Holland (from 6.5% to 30.1%), although <y> remains the predominantly used variant, which is actually the case in most regions. Particularly from a diachronic perspective, some results in this sub-corpus are surprising. In nineteenth-century texts from Utrecht, the prescribed double-dotted spellings practically disappear, whereas <y> becomes by far the dominant variant (93.2%). The same distribution can be found in North Brabant (<y>: 92.7%), although the development across periods is different than in Utrecht. In fact, it seems that North Brabant diarists do not apply the double-dotting spelling, neither before nor after 1804. It has to be noted, though, that there is only one nineteenth-century diary from North Brabant in the *Going Dutch Corpus*, which is why the results might also be due to an idiosyncratic preference for <y> rather than a regional pattern.

Finally, Figure 4 shows the distribution of variants across regions in newspapers. In contrast to the regional results for handwritten ego-documents, the use of variants in these printed, published texts can be described as clear-cut and invariable across all regions. As noted before, there is a complete shift from <y> in eighteenth-century newspapers to <ij> in nineteenth-century newspapers, perfectly in line with Siegenbeek (1804).

The only deviation here is that eighteenth-century newspaper texts from Utrecht (all from the *Utrechtse courant*) use both <y> (61.5%) and <ij> (38.5%). On closer inspection, however, it turns out that this distribution is not based on a

systemic use of variants for open and closed syllables, as suggested by several eighteenth-century grammarians like Kluit (1763). In fact, there is no evidence that this syllable-related rule is reflected in eighteenth-century language practice at all.

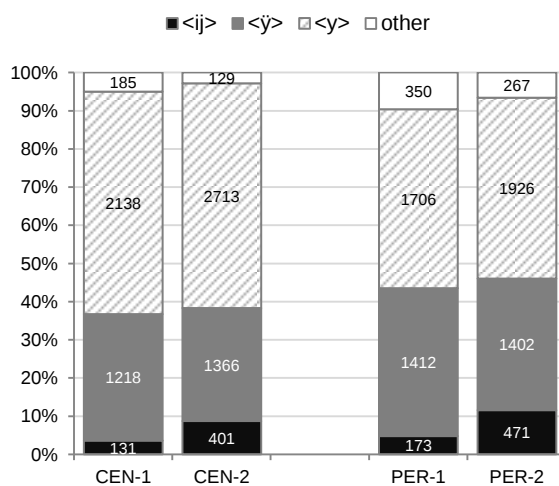
Figure 4. Distribution of variants across region and time (newspapers).



Variation across centre and periphery

Spatial variation was also investigated on the centre–periphery level (CEN = centre, PER = periphery), based on the entire *Going Dutch Corpus* and shown in Figure 5.

Figure 5. Distribution of variants across centre–periphery and time.

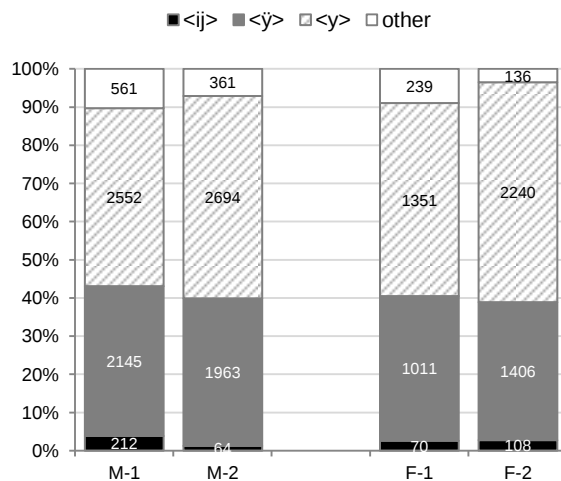


The results indicate that the changes in the distribution of variants across centre and periphery are rather marginal. In the first period, undotted <y> (58.2%) is the main variant in the centre, whereas <ij> and <ÿ> together have a share of only 36.7%. In the periphery, undotted (46.9%) and double-dotted (43.5%) spellings coexist as similarly strong variants. In the second period, the use of the double-dotted spelling minimally increases both in the centre (38.3%) and the periphery (46.1%). The share of <y> also remains stable in the centre (58.9%) and the periphery (47.4%).

Gender variation

Figure 6 displays the distribution of variants across gender (M = male writers, F = female writers), based on the ego-document data.

Figure 6. Distribution of variants across gender and time.



The results reveal practically no gender differences. Both male and female writers use double-dotted variants <ij> and <ÿ> in roughly 40% of all instances across both time periods. In the late eighteenth-century period, undotted <y> is the most frequently used variant written by men (46.7%) and women (50.6%). In the early nineteenth century, <y> even gains ground in ego-documents by both male (53.0%) and female writers (57.6%). In other words, the use of the double-dotted spelling slightly decreases after it was officially prescribed: from 43.1% to 39.9% among men, and from 40.5% to 38.9% among women. Generally speaking, the distribution across genders can be described as rather stable, as there are no considerable diachronic changes.

5 Discussion

In this chapter, the orthographic representation of Wgm. **i* was investigated, taking into account metalinguistic comments in eighteenth and early nineteenth-century normative works as well as actual language usage in the *Going Dutch Corpus*. As pointed out in Section 1, Siegenbeek (1804) prescribed the double-dotted spelling <ij> as the national variant, while vehemently rejecting the Greek-derived and thus foreign <y> in ‘proper Dutch’ words with Wgm. **i*. Siegenbeek ultimately codified this variant in his orthography after a controversial normative discussion about the spelling as <y> or <ij> throughout the eighteenth century.

In fact, eighteenth-century grammarians had different views on this spelling issue, promoting either <y> or <ij>, or even both of them (cf. also Table 3 in Vosters et al. 2010: 101). In Section 2, it was outlined that early eighteenth-century grammarians preferred <y> over <ij>, either implicitly (Moonen 1706) or explicitly (Sewel 1708). Not all grammarians favouring <y> heavily disapproved <ij>, though. Van der Palm (1769), for example, did prefer <y> but was rather indifferent about the use of alternative spellings. Heugelenburg (1763), on the other hand, preferred <ij> but did not mind the use of <y> either. Ten Kate (1723) even acknowledged the use of both <y> and <ij> as two entirely equal, coexisting variants, whereas Kluit (1763) introduced a more complex rule, prescribing <y> for open syllables and <ij> in closed syllables. Although Huydecoper already advocated <ij> as the only variant as early as 1730, it was only in the last decades of the eighteenth-century that normative works such as *Kunst wordt door arbeid verkreegen* (1770), Zeydelaar (1774), Stijl & van Bolhuis (1776), Kluit (1777), the *Rudimenta* (1799) and Weiland (1799) more coherently promoted <ij> – paving the way for Siegenbeek’s choice in his official orthography.

The corpus results of eighteenth-century language practice (Section 4) were largely in line with the heterogeneous normative discussion, in which both <y> and <ij> were promoted (cf. also Vosters, et al. 2010: 103). It was shown that both undotted and double-dotted spellings occurred frequently in actual language usage and were, in fact, two coexisting main variants (alongside a few other, more marginally occurring variants). This distribution was at least typical of handwritten ego-documents. In eighteenth-century newspapers, there was a clear preference for <y>, which was invariably used in practically all texts from this sub-corpus.

After Siegenbeek’s prescription in favour of <ij>, thus rejecting <y>, the use and distribution of variants remained surprisingly stable. Keeping in mind the results from the previous case studies in Chapters 5–8, one might assume that the official spelling must have gained ground in actual language use, but this was not the case – in ego-documents at least. Undotted <y> turned out to be the prevalent nineteenth-century variant in these handwritten texts. In private letters, the use of the double-dotted spelling even dropped, whereas it slightly increased in diaries and travelogues. In contrast, nineteenth-century newspapers consistently adopted the officialised <ij> spelling, completely shifting from pre-Siegenbeek <y> to post-Siegenbeek <ij>.

On the spatial dimension, a considerable amount of regional variation was attested. Although distinct patterns were difficult to find, some results were striking. First of all, there was a more or less balanced coexistence of various spellings in most eighteenth-century regions, whereas <y> clearly dominated in the North Holland data of the first period. In the nineteenth century, the use of <y> decreased in favour of prescribed <ij>/<ÿ>, but maintained its dominant position. However, Zeeland turned out to be the only region, in which the prescribed spelling notably gained ground and clearly became the main variant in nineteenth-century usage. In line with previous observations, these regional differences were also limited to handwritten ego-documents. In newspapers, the choice of variants was consistent across the entire language area (except for some variation in eighteenth-century newspapers from Utrecht).

The gender dimension did not reveal any interesting variation patterns as male and female writers used undotted and double-dotted variants similarly across both periods.

With regard to (seemingly) unsystematic regional variation and practically no gender variation at all, it can be concluded that variation and change in this orthographic variable are primarily genre-dependent. More specifically, the medium of the genre, i.e. printed and handwritten, turned out to be the most crucial factor. While the overall development in the entire *Going Dutch Corpus* was surprisingly stable and did not reveal any considerable changes from the eighteenth to the nineteenth century, the results in the individual sub-corpora showed major differences, mainly between printed (i.e. newspapers) and handwritten (i.e. ego-documents) texts. Referring back to Molewijk's (1992: 113) bold claim that “de *ij* [zou] zeer snel algemeen worden aanvaard” ‘the *ij* would be adopted very quickly in general’ in the nineteenth century, this study clearly shows that this was not the case – if we also take into account handwritten texts.

As mentioned before, printed and published texts like newspapers did adopt the prescribed norm invariably. But how can we explain the minimal changes in eighteenth- and nineteenth-century ego-documents, maintaining the officially rejected <y> as the most frequently used variant? Did Siegenbeek's prescription fail to reach the language users, or is it the very nature of this orthographic variable that prevented a noticeable change in handwriting? Vosters et al. (2010: 99) witness a similar tendency in their case study and suggest a possible explanation:

Alleen in de keuze tussen <y> en <ij> treedt nauwelijks een verandering op, wat goed kan samenhangen met het minimale verschil tussen de varianten in handschrift: de letter wordt altijd hetzelfde gevormd, onderscheidend is alleen de aan- dan wel afwezigheid van de puntjes.

‘Only in the choice between <y> and <ij>, hardly any change occurs, which can very well be connected with the minimal difference between the variants in handwriting. The letter is always formed in the same way. Only the presence or absence of the dots is distinctive.’

Indeed, the close similarities between the variants in handwriting might result in a relatively limited awareness of differences among language users. It is

questionable whether the presence or absence of the two dots was really as salient for the early nineteenth-century language user as it was for Siegenbeek (and many of his eighteenth-century predecessors). Judging from the corpus results, it seems as if many post-Siegenbeek writers of private letters, diaries and travelogues were hardly or not aware of the double-dotting of the <y> (as the officially prescribed variant) in handwriting.

Another question that arises with regard to the particularly high degree of variation in this orthographic variable concerns individual behavior. Is the variation attested in both eighteenth- and nineteenth-century ego-documents based on groups of writers with a clear spelling preference, consistently using either <ij>/<ÿ> or <y>? Or did these writers inconsistently use various variants in their texts, without any awareness of the double- or undottedness of their <y> spelling? This issue will be addressed separately in Chapter 13, zooming in on variation and change in inter- and intra-individual spelling practices.

Morphosyntactic variables (1)

Neuter relative pronouns

The analyses of five orthographic variables presented in Chapters 5–9 generally signalled a strong normative influence of Siegenbeek's (1804) spelling prescriptions on actual language practices. Shifting the focus to morphosyntactic issues, Chapters 10–12 examine the effectiveness of Weiland's (1805) official grammar norms. First, two central aspects of the Dutch relativisation system will be taken into account. The present chapter focuses on neuter relative pronouns, followed by an analysis of masculine and feminine singular and plural pronouns in Chapter 11.

While the complexity and variability of relative pronouns certainly justify the split into two separate, though closely related variables, the history of Dutch relativisation is best treated as a whole, as the major developments in the relativisation system affected neuter as well as masculine and feminine forms. Therefore, Section 1 of this chapter provides a general outline of relativisation in Dutch, covering the forms of both case studies under investigation. Moreover, grammarians and other language commentators did not strictly distinguish between different types of relativisers in their discussions, but typically addressed relativisation strategies under the same heading. This is why the present chapter includes a general outline of metalinguistic comments for both neuter and masculine/feminine forms, discussing norms and preferences in Weiland's (1805) grammar (Section 2) and in the preceding eighteenth-century normative tradition (Section 3). Bringing together the findings drawn from the two case studies, a comprehensive conclusion on variation and change in the use of Dutch relative pronouns will be drawn at the end of Chapter 11.

1 Relativisation in Dutch

Both synchronically and diachronically, relativisation in Dutch can be characterised as a highly complex morphosyntactic issue. Like other West Germanic languages such as English and German, Dutch has an extensive system of relative pronouns, which are generally grammaticalised forms of the demonstrative and interrogative pronouns. The situation in present-day Dutch, in fact, reflects the age-long competition between these two subsystems of originally demonstrative, so-called *d*-forms on the one hand, and originally interrogative, so-called *w*-forms on the other.

Historically, Dutch relativisers derive from demonstrative pronouns. In the Old and Middle Dutch periods, the *d*-forms *die* and *dat* were the default forms for masculine and feminine (i.e. present-day common gender) and neuter referents, respectively (van der Horst 2008: 172–173, 377). These pronouns were sensitive for

case marking, resulting in additional declined forms *diens* in the masculine and neuter genitive singular, *dier* in the genitive plural and the feminine genitive and dative singular, and *dien* in the dative plural.

From the late Middle Dutch period onwards, however, the competing *w*-forms *wie* and *wat* have increasingly replaced *die* and *dat* in specific contexts, for instance as free relatives. Moreover, the declined genitive forms *wiens* in the masculine and neuter genitive singular, and *wier* in the feminine singular and the plural have been used since the Middle Dutch period (van der Horst 2008: 602). These genitive forms are, in fact, still used in present-day Dutch, although they are commonly considered to be restricted to written and/or formal language (cf. ANS, Section 5.8.6).

Table 1 presents a diachronic overview of the main variants of masculine/feminine and neuter relative pronouns in the nominative singular. For reasons of simplification, the spelling of all forms was normalised.

Table 1. Main variants of relative pronouns across periods (nominative singular).

	Masculine/feminine	Neuter
Old Dutch before 1150	<i>die</i>	<i>dat</i>
Middle Dutch 1150–1500	<i>die, wie, welke, dewelke</i>	<i>dat, wat, welk, betwelk</i>
Modern Dutch 1500–1900	<i>die, wie, welke, dewelke</i>	<i>dat, wat, welk, betwelk, hetgeen</i>
Present-day Dutch 1900–present	<i>die, wie</i> (formal/archaic: <i>welke</i>)	<i>dat, wat</i> (formal/archaic: <i>welk</i> , <i>betwelk, hetgeen</i>)

In addition to the competition between *d*-forms and *w*-forms, the history of Dutch relativisation also saw the rise of alternative pronominal forms such as masculine and feminine *welke* and *dewelke*, as well as neuter *hetgeen*, *welk* and *betwelk*. These additional forms will also be a crucial part of the corpus analysis presented in the present and following chapter.

Already in the Middle Dutch period, *welc* or *welk*, derived from the originally interrogative *welk(e)*, was used as a relative pronoun alongside *die* and *dat* (van der Horst 2008: 380). Moreover, the extended forms *dewelke* for masculine and feminine referents, and *betwelk* for neuter referents also emerged as early as the thirteenth century, probably under influence of Latin and/or French *lequel* (van der Horst 2008: 381). Both *dewelke* and *betwelk* were particularly frequent in usage in the sixteenth and seventeenth centuries (van der Horst 2008: 830).

Another neuter form is *hetgeen*, originally a demonstrative pronoun. The relative use of *hetgeen* is usually dated to the sixteenth century, when the common combination of demonstrative *hetgeen* + relative *dat* was probably reinterpreted as relative *hetgeen* + (optional) subordinating conjunction *dat*, ultimately giving rise to the new stand-alone relative pronoun *hetgeen* in the seventeenth century (van der Horst 2008: 1115–1116, 1396). Schoonenboom (1997) shows that in Bible

translations from the fourteenth to the twentieth centuries, *hetgeen* was absent to begin with, but then became a relevant variant – and strong competitor of *wat* (van der Horst 2008: 1116) – in the seventeenth to the nineteenth centuries, before it declined again in the twentieth century. Generally, *develke*, *hetwelk* and *hetgeen* have become archaic or restricted to highly formal language from 1900 onwards (van der Wal 2002b: 34; van der Horst 2008: 1686).

Apart from the emergence (and subsequent decline) of these additional pronominal forms, the Dutch relativisation system has undergone a major shift from *d-* to *w-*forms⁶⁸. This change affects both pronominal and adverbial relativisers, occurring in both restrictive and non-restrictive relative clauses, as well as in dependent and independent (i.e. free or headless) relative clauses (Rutten & van der Wal 2014: 290). The change from *d-* to *w-*relativisation in Dutch comprises the following three developments:

- *die* > *wie*
- *dat* > *wat*
- *daar* > *waar*

The third change affects the relative adverbs *daar* and *waar*, as well as relative pronominal adverbs, consisting of a relative adverb and a preposition, for instance *daarin* and *waarin*.

It is striking that these three similar changes have a different chronology (van der Wal 2002b: 32; van der Wal 2003: 364). In the case of relative adverbs and relative pronominal adverbs, the development from *d-* to *w-*forms was completed by around 1900. Compared to the relatively rapid change from *daar* to *waar*, the pronominal relativisers have progressed far less. The change of the neuter relative pronoun is currently nearing its completion, at least in subject and object position, where *wat* is replacing *dat* (van der Horst 2008: 1683). With regard to the masculine and feminine relative pronouns, the change from *die* to *wie* is still ongoing. For free relatives, *wie* has become the current form, but it has not yet replaced *die* in other contexts, at least not in accepted Standard Dutch (van der Wal 2002b: 32). In fact, *wie* is still relatively rare in the common gender in subject and object position, although examples can be found easily (van der Horst 2008: 1683).

In sum, all three major developments in the Dutch relativisation system are slow and gradual processes, resulting in an age-long (and historical-sociolinguistically intriguing) situation of variation and change. The occurrence of alternative relativisers such as *welk(e)*, *develke*, *hetwelk* and *hetgeen* diversify this situation even more.

Due to the complexity of relativisers and their competing forms, I decided to focus on the two types of relative pronouns. The two case studies presented here and in Chapter 11 investigate variation and change in the neuter as well as in

⁶⁸ The change from *d-* to *w-*relativisers appears to be a broad West Germanic development, also attested for English (from invariant *that* to originally interrogative *who*, *what*, *which*), for instance.

the masculine and feminine relative pronouns, respectively. This means that the relative adverbs and relative pronominal adverbs will not be taken into consideration. For a detailed corpus-based study focusing on these adverbial relativisers, see Rutten & van der Wal's (2014: ch. 8) analysis of relative clauses in seventeenth- and eighteenth-century Dutch private letters. The central aim of the two case studies in this dissertation is to shed light on variation and change in the use of relative pronouns in late eighteenth- and early nineteenth-century Dutch. Moreover, by taking into account the normative discussion, both in Weiland's (1805) national grammar (cf. Section 2) and in the works of his eighteenth-century predecessors (cf. Section 3), the possible normative influence on actual language practice will be examined and assessed.

2 Discussion in Weiland (1805)

With regard to relativisation, Weiland's (1805) national grammar can be considered as a turning point in the Dutch normative tradition. As pointed out by van der Wal (2003: 367-369), the elaborateness of Weiland forms a sharp contrast to his eighteenth-century predecessors (cf. Section 3). In fact, he provided relatively detailed information on Dutch relativisers⁶⁹: not only a definition and paradigms, but also explicit norms and rules for usage, observations on stylistic or register differences, and numerous examples, from which further (more implicit) norms and rules can be deduced.

Starting with a general definition of relative pronouns, Weiland (1805: 119-120) listed *welke*, *denwelke*, *die* and *wie* as possible forms:

De *betrekkelijke* voornaamwoorden zijn zulke, welke betrekking hebben op personen of zaken, van welke te voren gesproken is. Hiertoe behooren *welke*, *denwelke*, *die*, *wie*.

'The relative pronouns are those, which refer to persons or things of which it was spoken before. *Welke*, *denwelke*, *die*, *wie* belong to this.'

The order in which Weiland listed the forms is interesting and might indicate his preferences for the more 'written' forms *welke* and *denwelke*, which are mentioned first, followed by the more 'common' *die* and *wie*. Although this list only comprises the masculine and feminine forms, more relativisers, i.e. their neuter counterparts (viz. *dat*, *wat*, *welk*, *betwelk*), were presented in the subsequent description and examples. Surprisingly, a more comprehensive inventory of relative pronouns – including the neuter forms – had been presented in Weiland (1799: 123):

⁶⁹ Given the focus on relative pronouns in this chapter, metalinguistic comments with regard to relative adverbs and relative pronominal adverbs will be left out of consideration (both in Sections 2 and 3).

De voornaamwoorden, welken betrekking hebben op personen, of zaken, waarvan te voren gesproken is, worden *betrekkelijke voornaamwoorden* genoemd. Zij zijn *welke, welk, dewelke, betwelk, die, dat, wie, en wat*.

‘The pronouns, which relate to persons or things of which it was spoken before, are called *relative pronouns*. They are *welke, welk, dewelke, betwelk, die, dat, wie, and wat*.’

Like in his 1805 grammar, Weiland listed the typically written forms *welke, welk, dewelke, betwelk* before the more common *d-* and *w-*forms.

Furthermore, it is striking that Weiland (1805) did not provide information on the alternative neuter pronoun *betgeen*, whereas we can find a comment on its use in Weiland (1799: 123):

Doorgaands wordt ook het onzijdige *betgene*, onder de betrekkelijke voornaamwoorden gesteld; doch verkeerdlijk, dewijl het eigenlijk zoo veel zegt, als *dat, betwelk*; en dus past het nergens wel, zegt HUYDECOPER, [...] dan waar het deze omschrijving kan lijden. Gelijk *degene*, zoo ook behoort *betgene* tot de *aanwijzende*, en niet tot *betrekkelijke* voornaamwoorden; want men zegt wel *degene, die mij eert*, maar nimmer, *die, degene mij eert*; en dus ook wel *betgene, dat ik wil*, maar niet *dat, betgene ik wil*. [...] Intusschen is *betgene*, in de beteekenis van *betgene dat*, reeds door het gebruik gewettigd, b. v. *betgene ik zeg, is waarheid*.

‘Generally, the neuter *betgene* is placed among the relative pronouns, but wrongfully, while it actually means the same as *dat, betwelk*. Therefore it fits nowhere, says HUYDECOPER, [...], except for where it meets this description. Like *degene, betgene* belongs to the *demonstrative*, and not to the *relative* pronouns, because one says *degene, die mij eert*, but never *die, degene mij eert*; and thus *betgene, dat ik wil*, but not *dat, betgene ik wil*. [...] In the meantime, *betgene*, in the meaning of *betgene dat*, is already sanctioned by usage, e.g. *betgene ik zeg, is waarheid*.’

Referring to Huydecoper (1730), Weiland thus initially rejected the neuter *betgene* (equivalent to *degene*) as a relative pronoun. However, he did acknowledge (and accept) its function as free relative as legitimised in language practice.

Either explicitly from norms and rules, or implicitly from examples, it is possible to deduce preferences for specific forms in different contexts. First of all, in Weiland’s (1805: 120-121) view, only *w-*forms – but not *betgeen* (or *betgene*), as argued in Weiland (1799) – can function as free relatives:

Het betrekkelijke voornaamwoord *wie* wordt dikwerf zoo gebruikt, dat het betrekking heeft op iets, dat volgt; doch daar dit volgende zich gevoeglijk vooraan laat plaatsen, zoo blijft *wie* een waar betrekkelijk voornaamwoord, en slaat eigenlijk op het voorgaande. Zoo zegt men, bij voorbeeld: *wien ik mijn woord geef, dien zal ik niet misleiden*; het welk men ook dus kan omkeeren: *dien zal ik niet misleiden, wien ik mijn woord geef*. Hetzelfde heeft plaats ten aanzien van het onzijdige *wat*: *wat mij gebeurd is, dat zal ik u verhalen*; waarvoor men ook kan zeggen: *dat zal ik u verhalen*, of: *ik zal u verhalen, wat mij gebeurd is*.

‘The relative pronoun *wie* is often used that it relates to something that follows. But since what follows can be simply placed in front, *wie* remains a proper relative pronoun, and actually refers to the preceding. Thus one says, for example: *wien ik*

mijn woord geef, dien zal ik niet misleiden, which one can thus also turn around: *dien zal ik niet misleiden, wien ik mijn woord geef*. The same happens with regard to the neuter *wat*: *wat mij gebeurd is, dat zal ik u verhalen*, for which one can also say *dat zal ik u verhalen*, or: *ik zal u verhalen, wat mij gebeurd is*?

In the case of a sentence or clause as antecedent, Weiland (1805: 244) referred to the neuter forms *dat* and *hetwelk* as possible options:

Wanneer eene uitdrukking naar eene geheele rede te rug gevoerd wordt, dan wordt het betreffelijke voornaamwoord in het onzijdige geslacht gebezigd: *zij spraken over deugd en godsvrucht, dat mij zeer aangenaam was. Die zaak heeft eenen slechten keer genomen, het welk ik wel gevreesd had.*

‘When an expression refers back to a complete sentence, the relative pronoun in the neuter gender is used: *zij spraken over deugd en godsvrucht, dat mij zeer aangenaam was. Die zaak heeft eenen slechten keer genomen, het welk ik wel gevreesd had.*’

This choice in favour of *dat* and *hetwelk* actually raises the question about Weiland’s attitude towards the remaining forms *wat*, *hetgeen* and *welk* in these cases.

In the indefinite pronoun combination with *alles*, the neuter form *wat* is the preferred choice, as can be inferred from the examples *alles, wat van hem gezegd wordt, is waar* ‘all that is said about him, is true’, and *Alles, wat ik daarvan weet, zal ik u verhalen* ‘all that I know thereof, I will tell you’ (Weiland 1805: 121). For nominalised adjectives as antecedents, Weiland preferred *dat* or *hetwelk* (and *welk*), although the *n*-form *wat* may occur as well: “*Het goede, wat gij mij bewezen hebt, beter dat, of het welk enz.*” (1805: 246). Weiland (1805: 246) was even more prescriptive in the case of relative pronouns referring to noun phrases, either indefinite or definite, explicitly rejecting the use of *wat*:

Is echter het voorwerp, waarop het betreffelijke voornaamwoord slaat, een zelfstandig naamwoord, dan wordt ook, wanneer het van het onzijdige geslacht is, *welk*, of *het welk*, in plaats van *wat* gebezigd, als: [...] *Het huis dat, of het welk gij gekocht hebt, nooit wat.*

‘If the object, which the relative pronoun refers to, is a noun, then one also uses, if it is of the neuter gender, *welk*, or *het welk*, is used in place of *wat*, as: [...] *Het huis dat, of het welk gij gekocht hebt, never wat.*’

Furthermore, we can find a clear rule for the use of the genitive forms *niens* and *welks*, both of which were frequently used for neuter antecedents. Weiland (1805: 120) explained that *niens* is the appropriate form for genitive singular masculine, whereas *welks* is used for genitive singular neuter:

Dikwerf worden *niens* en *welks* onverschillig in het onzijdige geslacht gebruikt, schoon *niens* alleen de tweede enkelvoudige naamval van het mannelijke, en *welks* die van het onzijdige geslacht is. Men zegge derhalve: *de man, niens geleerdheid enz; het land, welks uitgestrektheid enz.*

‘Oftentimes *wiens* and *welks* are indifferently used in the neuter gender, although *wiens* is only the second case singular of the masculine, and *welks* that of the neuter gender. Therefore one says: *de man, wiens geleerdheid* etc., *het land, welks uitgestrektheid* etc.’

Weiland (1805: 245) also commented on the case-dependent use of *die*, *wie* and *welke* after personal pronouns: *die* in the nominative case (e.g. *hij, die mijn vriend wil zijn*), and *wie* or *welke* in the other cases (e.g. *Hij, wien ik dit gezegd heb*) and after prepositions (e.g. *Zij, van welke ik dit gehoord heb*). Weiland (1805: 246) further clarified that the choice between *wie* and *welke* is due to the common use of *wie* (rather than *welke*) in the genitive and dative masculine singular and after prepositions (e.g. *hij was de man, wiens vriend ik wilde wesen, wien ik zoo veel verschuldigd was, van wien ik zoo veel goed ontvangen had*).

Finally, a remarkable part of Weiland’s discussion on relativisation concerns his awareness of stylistic or register differences. Referring to the terminology introduced by his eighteenth-century predecessor ten Kate (1723; cf. also van der Wal 2002a), Weiland (1805: 244) distinguished between forms either used in the solemn style (*deftig*) or in the plain style (*gemeenzaam*):

Welke, of dewelke, wordt, als het eigenlijkste betrekkelijke voornaamwoord, meest in den deftigen stijl, het kortere die, dat ook voor een ander voornaamwoord gebezigd wordt, in den gemeenzamen stijl gebruikt, als: de gelukzaligheid des tegenwoordigen en toekomstenden levens, welke langs verschillende wegen gezocht word. Hij woont in het huis, dat zijn vader gebouwd heeft.

‘*Welke, or dewelke, as the most proper relative pronoun, are mostly used in the solemn style, the shorter die, which is also used for another pronoun, in the plain style, as de gelukzaligheid des tegenwoordigen en toekomstenden levens, welke langs verschillende wegen gezocht word. Hij woont in het huis, dat zijn vader gebouwd heeft.*’

This indicates that, according to Weiland, *welke* and *dewelke* belonged to a higher register, whereas *die* was mostly used in common writing (van der Wal 2003: 369). Previously, Weiland (1805: 121) had noted that *dewelke* was a less frequently used variant. Although he did not refer to the neuter counterparts explicitly, the example with *het huis, dat* (illustrating the ‘plain’ style) signals a similar stylistic difference, i.e. *dat* used in the ‘plain’ style, as opposed to *welke* and *hetwelke* in the more ‘solemn’ style.

In sum, Weiland’s (1805) national grammar provided remarkably detailed information on the contemporary use of relativisers, also indicating preferences and even some explicit rules for particular forms. In the corpus analyses of neuter (Section 5 of this chapter) as well as masculine and feminine relative pronouns (Chapter 11), I will investigate whether and to what extent Weiland’s normative influence on early nineteenth-century language usage can actually be attested.

3 Eighteenth-century normative discussion

In comparison with the elaborate discussion found in Weiland's (1805) national grammar, relativisation did not appear to be a core topic in the eighteenth-century normative tradition (cf. van der Wal 2003). Although a number of grammarians did refer to relativisers more or less sporadically, mainly in the form of short descriptions or paradigms, specific rules and/or guidance for their use were generally lacking.

To begin with, the inventory of relative pronouns listed in eighteenth-century grammars appears to be highly variable. Sewel (1708: 121; 1712: 236), for instance, only mentioned *wie* and *welk*. Only *die*, *dat*, *wie* and the genitive forms *wiens* and *wier* were mentioned in the grammar published by *Kunst wordt door arbeid verkregen* (1770: 140).

Moonen (1706: 125), one of the leading grammars in the eighteenth century, presented *die*, *dat*, *wie* in his definition of relative pronouns:

De Betreklyke hebben hun opzicht op iemant of iet waer van voorhene gesprooken is; en zyn *Die*, *Dat*, *Wie*, qui, quae, quod; als in deeze redenen, *Hy slaapt nu, die altyt waekte; het zwaert, dat ik ontkoomen ben; de vyant, wiens maght te vreezen is; de vrou, wiens man gestorven is.*

'The relative [pronouns] refer to someone or something of which it was spoken before, and which are *Die*, *Dat*, *Wie*, qui, quae, quod, as in these sentences, *Hy slaapt nu, die altyt waekte; het zwaert, dat ik ontkoomen ben; de vyant, wiens maght te vreezen is; de vrou, wiens man gestorven is.*'

Additionally, *welke* (for masculine and feminine) and *welk* (for neuter) were also mentioned as both interrogative and relative pronouns (Moonen 1706: 131).

The same inventory of relative pronouns recurred in Elzevier (1761: 64, 69, 71) and van der Palm (1769 II: 46, 49–50) more than half a century later. In addition to *die*, *dat*, *wie*, *welk(e)*, a few grammarians also mentioned the additional forms *dewelke* (e.g. *Rudimenta* 1799: 27), *betgeen* (Tollius 1776: 70), or both *dewelke* and *betgeen* (Stijl & van Bolhuis 1776: 96; van Bolhuis 1793: 51).

The neuter form *betgeen* in particular attracted some more prescriptive commentary throughout the eighteenth-century. Argued to be used exclusively as a free relative, Verwer (1707: 50a) rejected instances of *het gene* in other contexts, for instance with nominal antecedents, as unacceptable mistakes:

Zoals die, dewelke terugslaat op wat voorafgaat, zo hebben de Nederlanders een bijzonder relatief, uitsluitend voor wat volgt – geen enkele andere taal heeft dit, voor zover ik weet –, te weten het gene in het onzijdig: “Sy pleegden alle ongebondenheit, ende (’t gene selfs alle menschelijkheid te buiten gaet) sy en spaerden geenen suigelingen aen ’s moeders borsten”. Het gene verwijst hier naar wat volgt, en dat is altijd het geval. Bijgevolg is de fout geenszins te duldten van hen die het gene gebruikten als relatief voor wat voorafgaat in het onzijdig geslacht, zeggende “het hart, ’t gene reikhalst”, “het swaert, ’t gene geslepen was”. En dit zult ge ook wel nergens door de ‘usus’ bevestigd vinden.

Like *die*, *dewelke* refer back to what precedes, the Dutch people have a specific relative pronoun, exclusively for what follows – not a single other language has this, as far as I know –, namely *het gene* in the neuter: “Sy pleegden alle ongebondenheit, ende (’t gene selfs alle menschelijckheit te buiten gaet) sy en spaerden geen en suigelingen aen ’s moeders borsten”. Here *het gene* refers to what follows, and that is always the case. Consequently, the mistake by those who used *het gene* as a relative pronoun for what precedes in the neuter gender, saying “het hart, ’t gene reikhalst”, “het swaert, ’t gene geslepen was”, is by no means tolerable. And you will also not find this confirmed by the use anywhere.’

Like Verwer (1707), Huydecoper (1730: 620) also prescribed *dat* or *’t welk* instead of *’t geen* with a nominal antecedent (here: *het werk*):

Omtrent het Onzydige *het gene*, merken wy hier kortelyk aan, dat men niet wel zegt, *het werk*, ’T GEEN *ik begonnen heb*. moet zyn, DAT, of ’T WELK *ik beg*. Maar zeer wel zegtmē, ’T GEEN, *ik doe, is een zwaar werk*: daar alleenlyk *’t geen*, zo veel zegt, als *dat het welk*, in ’t Latyn, *id quod*.

‘With regard to the neuter *het gene*, we briefly comment that one does not say well, *het werk*, ’T GEEN *ik begonnen heb*. It must be DAT, or ’T WELK *ik beg*. But one says very well, ’T GEEN, *ik doe, is een zwaar werk*: because only *’t geen* means as much as *dat het welk*, in Latin *id quod*.’

Similarly, Tollius (1776: 74-75) considered *’t geen* as a less acceptable option referring to nominal antecedents:

het geen by ’t onzydig geslacht dikwerf by enige onzer schryvers voorkomt voor het relativum *dat* of *’t welk*, daar echter deszelfs oud gebruik aan anderen schijnt te vereischen, dat het alleenlyk op het volgende en niet op het voorgaande gepast worde, b.v. “Zy pleegden ongebondenheit, en *’t geen* alle palen van menschelijckheit te buiten gaat, zy spaerden selfs geen zuigelingen.” Minder goed zou men zeggen: “Het kwaad, *’t geen* alle begrip overtreft.”

‘Among some of our writers, *het geen* in the neuter gender often occurs in place of the relative pronoun *dat* or *’t welk*, because its old usage seems to demand from others, that it only refers to the following and not to the preceding, e.g. “Zy pleegden ongebondenheit, en *’t geen* alle palen van menschelijckheit te buiten gaat, zy spaerden selfs geen zuigelingen.” One would say less well: “Het kwaad, *’t geen* alle begrip overtreft.”

Ten Kate (1723 I: 489-492) was more tolerant towards the use of *hetgeen*, suggesting it as a possible neuter form alongside *dat* and *hetwelk* in his paradigms. Generally, ten Kate (1723 I: 489) treated relativisation in a relatively fine-grained and elaborate way, listing *die*, *welke* and *dewelke* as masculine and feminine forms, and *dat*, *welk*, *het welke* en *’t gene* as neuter forms. According to ten Kate, the *n*-form *n*ie was not used in the nominative case, but restricted to forms declined for other cases such as *niens* or *nien* (cf. 1723 I: 491).

Ten Kate distinguished between three styles or rather registers of language, viz. the sublime style (*hoogdravend* or *verbeven*), the solemn style (*deftig* or *statig*) and

the plain style (*gemeenzaam*) (cf. also van der Wal 2002a: 56-59). Interestingly, the solemn and the plain style later recurred in Weiland's (1805) national grammar (Section 2). Despite his detailed paradigms, ten Kate did not make clear correspondences between particular relativisers and these styles, though. In fact, most relativisers are listed as options for all three styles. For the nominative case, for instance, we find *die*, *welke*, *de welke* (for masculine and feminine referents), and *dat*, *het gene*, *het welke* (for neuter referents) for all three styles. For other cases, forms of *develke* and *betwelke* seem to belong mainly, but not exclusively, to the two higher (i.e. sublime and solemn) styles. Moreover, it is striking that forms of the neuter pronoun *hetgeen* occur in all three registers, whereas *wat* is completely absent. Generally, it appears that ten Kate's differentiation with regard to style or register mainly implies different degrees of case marking rather than the choice of exclusive relativisers. In other words, the higher the register, the more case endings we find.

With regard to relative pronouns functioning as free relatives, not all forms of the inventory were acceptable according to ten Kate (1723 I: 492):

Want ons WELKE *M* & *F*, en Welk *N*, en DE WELKE *M* & *F*. erkent men nu in dat ampt niet; dog het laetste alleen in *Neutro*. Dus zeid men niet WELKE of DEWELKE, *het ons overgaf, was de Man of Vrouw*; dog DIE 'T ONS OVERGAF, WAS DE MAN OF VROUW; gelijk ook DAT of 'T WELK, of 'T GENE ONS QUELDE, WAS HET WATER, enz.

'Because one does not acknowledge our WELKE *M* & *F*, and Welk *N*, and DE WELKE *M* & *F*. in that position, but the latter only in neuter. Thus one does not say WELKE or DEWELKE, *het ons overgaf, was de Man of Vrouw*; but DIE 'T ONS OVERGAF, WAS DE MAN OF VROUW; also like DAT or 'T WELK, or 'T GENE ONS QUELDE, WAS HET WATER, etc.

Thus, he gave preference to masculine and feminine *die*, rejecting the use of *welke* and *develke*. The neuter pronouns *dat*, *'t welke* and *'t gene* could all function as free relatives.

There are some more interesting comments on relativisers in usage. Ten Kate (1723 I: 491) pointed out that the neuter pronouns *het welke* and *'t gene* have reduced forms, viz. *'t welke* and *'t geen*, which even occur in the sublime and solemn styles. Another observation concerns the use of *welke* for masculine referents, which had become quite uncommon, in contrast to the still very common *het welke* (ibid.), indicating differences in frequency in contemporary usage.

Whereas ten Kate presented relative pronouns according to different styles or register, van Bolhuis (1793: 51-52) distinguished between common and less common relative pronouns:

Gewoonlijk gebruikt men daar toe *welke*, *de welke*, *die*, en de verbogene naamvallen van *wie*, en *wat*: - minder gebruikelijk is *wie*, *wat*, *het* [sic], *hetgeen*, en de verbogene naamvallen van *die*.

'Normally one uses *welke*, *de welke*, *die*, and the declined cases of *wie* and *wat* for this. Less common are *wie*, *wat*, *het* [sic], *hetgeen*, and the declined cases of *die*.'

From this distinction, it might be concluded that the oblique *d*-forms *diens* and *dien* were giving way to their *n*-counterparts *wiens* and *wien*. For some reason, *wie*, *wat* and *hetgeen* in subject and object position were less commonly used. Van Bolhuis (1793: 51-52) presented a fairly elaborate discussion of relative pronouns, illustrating the best usage of these forms by examples for each of the three genders and for the various cases. Nonetheless, it is difficult to deduce clear norms or rules from these examples. In the nominative case, for instance, van Bolhuis seemed to prefer *die* over *welke* and *dewelke*: “God, (*dewelke*, of liever *welke*, en liefst) *die groot is*” (1793: 51). In other cases, forms of *welke* were the favoured options. In fact, these examples are probably best considered as suggestions rather than prescriptive rules. Van Bolhuis (1793: 52) himself added: “Om te weten, wat best vloeit, plege men raad met zijn gehoor” – literally: ‘in order to know what flows best, one should consult one’s ear’.

Van Bolhuis (1793: 52) also addressed the question of free relatives, allowing a wide range of forms in this function, viz. masculine/feminine *wie*, *die* and neuter *wat*, *dat*, *het welk* and *het geen*:

Enige van deze voornaamwoorden kunnen ook zo gebruikt worden, dat zij *betrekkelijk schijnen* te zijn, niet op het voorgaande, maar *op het volgende*. Deze zijn *wie*, *die*, *wat*, *dat*, *het welk*, *het geen*, doch *in de daad* wijzen zij, gelijk alle betrekkelijke voornaamwoorden *op het voorgaande*

‘Some of these pronouns can also be used in a way that they *seem* to be *relative*, not referring to the preceding, but *to the following*. These are *wie*, *die*, *wat*, *dat*, *het welk*, *het geen*, but *indeed* they refer to the preceding, like all relative pronouns.’

Recall that Weiland (1805) presented a more limiting choice of free relatives, only listing the *n*-forms *wie* and *wat*.

To sum up, there is little evidence for (explicit) norms, rules or guidelines for the use of relative pronouns in the eighteenth-century normative tradition. Only a few grammarians, for instance ten Kate (1723) and van Bolhuis (1793), addressed the topic on a comparatively detailed level. Generally, it appears that normative works published before Weiland (1805) did not comment on the choice between *d*- and *n*-forms. As van der Wal (2002b: 33) argues, “it is not until the nineteenth century that explicit prescriptive rules are given which could have either stimulated or delayed the ongoing D->W-developments”. Probably the most explicitly discussed aspect of relativisation concerns the use of *hetgeen*, which was restricted to free relatives and mostly rejected in other contexts.

Given the general scarcity of norms and rules for relative pronouns, it is doubtful whether the eighteenth-century grammar could have an influence on the use of relativisers in actual language practice at all⁷⁰. Making use of the *Going Dutch Corpus*, the analyses of neuter as well as masculine and feminine relative pronouns

⁷⁰ With regard to the lack of explicit norms, van der Wal (2003: 372) concludes that if there was normative influence before 1800, it must have been restricted to actual language use of prestigious authors.

will shed light on the situation in actual language use before and after Weiland (1805).

4 Previous research

The history of Dutch relativisation has gained a considerable amount of attention in the literature. To begin with, van der Horst (2008) provides a diachronic overview of relative pronouns from the Old Dutch period to the situation in nineteenth- and twentieth-century Modern Dutch, illustrated with plenty of examples for each period.

Apart from this chronological outline, a number of pioneering articles, for instance by van der Horst (1988), Schoonenboom (1997, 2000), de Schutter & Kloots (2000) and van der Wal (2002b, 2003), have addressed different aspects of Dutch relativisation from various theoretical and methodological perspectives. Van der Horst's (1988) seminal study on the neuter pronouns *dat* and *wat* introduces the generalisation that the (in)definiteness of the antecedent plays a crucial role in the change from *d*-relativisers to *w*-relativisers. He claims that *w*-forms have spread from the most indefinite to the most definite contexts. In a modified form, this internal conditioning also recurs in Schoonenboom's (1997, 2000) diachronic studies on the Dutch neuter relative pronouns *dat*, *wat* and *hetgeen*, based on a corpus of Bible texts from the fourteenth to the twentieth centuries. Adding the possible influence of genre as an external factor, de Schutter & Kloots (2000) investigate relativisers, both pronominal and adverbial forms, in seventeenth-century literary texts by prestigious Dutch writers.

A more exploratory approach to relativisation in the history of Dutch is taken by van der Wal (2002b), who particularly focuses on the major shift from *d*-forms to *w*-forms. Furthermore, van der Wal (2003) presents another study on relativisation in the Dutch normative tradition from the sixteenth to the twentieth centuries, pointing out the elaborateness of Weiland (1805) as compared to his predecessors. One of the central questions, which will also be addressed here, concerns the possible influence of normative publications and writing conventions. With respect to the absence of explicit rules before Weiland's grammar, van der Wal (2003) suggests that normative influence on the change from *d*- to *w*-relativisers can only be attested from the nineteenth century onwards.

Whereas the previously mentioned articles generally support van der Horst's (1988) claim with respect to the definiteness of the antecedent, Rutten (2010) critically reassesses this generalisation by taking an alternative approach based upon construction grammar. One of his main arguments is that *w*-relativisers first occurred in specific constructions, before they were generalised to a more abstract level. From a historical-sociolinguistic perspective⁷¹, the most extensive

⁷¹ The inherent variability of relativisation has also attracted considerable interest from historical sociolinguists working on other languages than Dutch. See, for instance, Romaine

corpus-based analyses on Dutch relativisation so far are reported in Rutten & van der Wal (2014: chapter 8) and the follow-up study in Rutten & van der Wal (2017). Based on the *Letters as Loot* corpus of seventeenth- and eighteenth-century Dutch, they take into account sociolinguistic factors and constructional diffusion. First of all, Rutten & van der Wal (2014: chapter 8) point out that writing experience is a relevant factor affecting distributional patterns, showing that the change from *d-* to *w-*forms was a change from above, led by men from the upper ranks. Moreover, by investigating different epistolary formulae with relative clauses, they also confirm Rutten's (2010) preliminary conclusion with regard to the diffusion via specific constructions.

The primary focus of both Rutten & van der Wal (2014: chapter 8) and Rutten & van der Wal (2017) is on the change from *d-* to *w-*forms in the case of relative adverbs and relative pronominal adverbs. For relative pronouns in Late Modern Dutch, much less research has been conducted, certainly from a historical-sociolinguistic perspective. It is the purpose of the two corpus-based case studies presented in this dissertation to fill this gap in the research literature by investigating neuter relative pronouns in Section 5, followed by masculine and feminine relative pronouns in Chapter 11.

5 Corpus analysis⁷²

5.1 Method

This case study investigates the Dutch neuter singular relative pronouns in subject and object position. The focus on these positions is mainly due to the fact that in other paradigms, particularly in combination with prepositions, the pronominal paradigm mixes with forms of the pronominal adverbial paradigm (e.g. pronominal *met wie* versus adverbial *waarmee*).

The occurrences of the following neuter forms (in nominative and accusative) were extracted:

- *dat, wat, hetgeen, welke, hetwelk.*

Possible spelling variation was also taken into account. In the case of *hetwelk*, for instance, the pronoun might also occur as *het welk*, *het welke*, *'t welk* or *twelk*.

All non-relative occurrences were filtered out by hand, including tokens of *dat* (e.g. functioning as a subordinating conjunction or demonstrative pronoun) and *wat* (e.g. as an indefinite pronoun meaning 'something' or in the construction *wat x betreft* 'as regards x'). As discussed by Schoonenboom (2000: 91-108) and Rutten (2010: 8-9), the distinction between the relative and the interrogative interpretation

(2009 [1982]), Bergs (2005: chapter 5) and Nevalainen & Raumolin-Brunberg (2003) for English, and Negele (2012) for German.

⁷² Parts of this case study were also presented in Krogull, Rutten & van der Wal (2017).

of *wat* is not always clear-cut and thus difficult to keep apart (cf. also Fischer 1992: 297-298 for the English case).

For the purpose of this case study on relative pronouns, all unambiguous instances of interrogative *wat* were removed. An example taken from the *Going Dutch Corpus* is (1), in which the construction of the type *niet weten wat* ‘not knowing what’ expresses the unknowingness of the writer, typically indicating the interrogative interpretation of *wat* (Rutten 2010: 8). Another example is (2), in which the writer asks for unknown information (viz. ‘what other young ladies do’). I did include those occurrences of *wat* in which both relative and interrogative interpretations are theoretically possible. Examples (3) and (4) illustrate two more ambiguous cases, in which some degree of indefiniteness or unknowingness is present, but an interpretation of *wat* as free relative is (also) possible. In fact, *wat* can be replaced by, for instance, *datgene wat* or *hetgeen* here. This is also true for example (5), in which *wat* has first and foremost a relative interpretation.

- (1) *termyl ik niet weet **wat** er gebeuren zal*
‘while I do not know what will happen’
- (2) *schrjft mij eens **wat** andere juffw doen*
‘write to me what other young ladies do’
- (3) *de Heere weet **wat** hy an mijn zjfel gedaan heeft*
‘the Lord knows what he has done to my soul’
- (4) *dat bet onmooglyk is te beschryven **wat** men hier gevoelt*
‘that is impossible to describe what one feels here’
- (5) *verstond ik direct **wat** den Inlander my gezeid had*
‘I directly understood what the native had said to me’

Ultimately, 1,009 occurrences of neuter relative pronouns in subject and object position were extracted from the *Going Dutch Corpus*. Before the corpus results will be presented and discussed in Sections 5.2 and 5.3, I give an overview of the occurring forms in different contexts, taking into account the concept of definiteness.

The (in)definiteness of the antecedent has traditionally been regarded as a crucial internal factor affecting variation and change in the Dutch relativisation system. An early generalisation, mainly by van der Horst (1988: 196), claims that relative *w*-forms entered the language from indefinite to definite contexts, depending on the definiteness of the antecedent (cf. also van der Wal 2002b). In this generalisation, different types of antecedents are located on a cline from the most indefinite to the most definite contexts:

- (I) free relatives > (II) clauses or sentences > (III) indefinite pronouns > (IV) nominalised adjectives > (V) indefinite noun phrases > (VI) definite noun phrases.

These different contexts, depending on the definiteness of the antecedent will be illustrated by authentic language usage examples taken from the *Going Dutch Corpus*.

The first and most indefinite context (I) comprises free (or headless) relatives, i.e. which lack an antecedent altogether (*ingesloten antecedent*). In the corpus data, the *w*-form *wat* (6) as well as *hetgeen* (7) occur as neuter free relatives:

- (6) ***Wat*** *de Montenegrijnen willen, is bekend*
 ‘What the Montenegrins want, is known’
- (7) ***hetgeen*** *ik u gezegd heb herhaal ik u*
 ‘What I have told you, I repeat to you’

In examples (8–10), illustrating context II, the antecedent is a sentence or clause. All five variants of the neuter relative pronoun occur in the corpus data:

- (8) *Ik hebbe reets alhier het plain ront gewandelt, **dat** my volmaakt wel bekoomen is*
 ‘I have already walked around the square here, which I completely enjoyed’
- (9) *hier is het ten minsten geducht koud, **wat** het ergste is voor de armen*
 ‘here it is at least terribly cold, which is the worst for the poor’
- (10) *Hij diende by eene weduwe voor knecht, **‘tgeen** ons aanleiding gaf hem te raaden zijn best te doen*
 ‘He worked as servant for a widow, which caused us to advise him to do his best’
- (11) [...] *mooije gezichten op het grazende vee door lieten **welk** een en ander door eene heerlyke avond begunstigd werd*
 ‘let through beautiful views on the grazing cattle, which was also favoured by a beautiful evening’
- (12) *deze menschen spraken gedurig van de slechte wegen **hetwelk** ons eenigsinds bevreesd maakte*
 ‘these people talked continually about the bad roads, which somewhat frightened us’

Context III comprises a fairly heterogeneous group of indefinite pronouns, including the idiomatic combination of *al(les)* + *wat* (cf. Schoonenboom 2000: 35–46; Rutten 2010: 9–10). Hence, I propose a split of context III into IIIa (*alles*) and IIIb (remaining indefinite pronouns such as *iets*, *niets*, *veel*, *weinig*). The former (sub-)context is illustrated by examples (13–16), attesting the use of *wat*, *dat*, *hetgeen* and *hetwelk*. In (sub-)context IIIb (see examples (17–19)), *dat*, *hetgeen* and *hetwelk* – but not *wat* – occur in the corpus data:

- (13) [*een lysje van*] *alles **dat** gy hebt*
 ‘[a list of] everything that you have’

- (14) *alles, **wat** *nij* gezien en genoten hadden*
'everything that we had seen and enjoyed'
- (15) *al **het geen** verder tot kindergoed behoort*
'all that further belongs to baby's clothes'
- (16) [*naar*] *al **het welke** ik zeer verlangende*
'all which I much longed [for]'
- (17) *iets **dat** na een lichtje van een kaars lykende*
'something that resembled the light of a candle'
- (18) *iets '**t geen** byna nooit gebeurd is*
'something which almost never happened'
- (19) *iets **hetwelk**, naar zijn oordeel, ook bij de onafhankelijkheid der geestelijken*
volstrekt geene zwaarigheid zou baren
'something which, in his opinion, would cause no difficulty at all, also
in the case of the independence of the clergymen'

In context IV (see examples 20–22), the antecedent is a nominalised adjective, mostly a superlative. In the corpus data, occurrences of *dat*, *wat* and *hetwelk* are found:

- (20) *'t liefste **dat** ik ooit in dien aart bygewoond heb*
'the loveliest that I have ever attended of this sort'
- (21) *het edelste **wat** den mensch bezit*
'the noblest that the human possesses'
- (22) *het voornaamste **hetwelk** er bij mijne ziekte is voorgevallen*
'the main thing which happened during my illness'

The full inventory of all five neuter relative pronouns occurs in context V (examples 23–27), referring to an indefinite noun phrase:

- (23) *een steigertje **dat** Papa en ik reeds gezien hadden*
'a small jetty that Dad and I had already seen'
- (24) *boumland **wat** zeer vruchtbaar is*
'farmland that is very fertile'
- (25) *een zeer oud gebouw, '**t geen** geschikt is tot den koophandel*
'a very old building, which is suitable for the trade'
- (26) *een paard, **welk** hen de nacht naar Coblenz mocht brengen*
'a horse, which had to bring them to Coblenz that night'
- (27) *een oud kasteel '**t welk** voormaals tot gevangenis diende*
'an old castle, which formerly served as prison'

Moreover, all five possible relative pronouns also occur in the most definite context VI (see examples 28–32), in which the antecedent is a definite noun phrase:

- (28) *het beelderige mooye mandje **dat** zy voor my gemaakt heeft*
‘the lovely beautiful basket that she has made for me’
- (29) *Ons rytuig **wat** wy om 8 Uur besteld hadden*
‘our coach that we had ordered at eight o’clock’
- (30) *het gure en regenachtig weder, **het geen** reeds den geheelen dag had gebeurd*
‘the biting and rainy weather which had already lasted the whole day’
- (31) *het geheim **welk** ‘er gaande was*
‘the secret that was happening there’
- (32) *Het schoon gezicht **‘t welk** men nog lang van deze zyde op de stad’ blijft genieten*
‘the beautiful view of the city that one keeps enjoying from this side for a long time’

The main purpose of this overview is to illustrate eighteenth- and nineteenth-century usage of the various neuter relative pronouns in different contexts, as they occur in the *Going Dutch Corpus* data. The quantitative analysis will be presented and discussed in Sections 5.2 and 5.3.

5.2 Results

To begin with, the general results presented in Table 2 give evidence of a high degree of variation in terms of frequently occurring neuter relative pronouns in both diachronic cross-sections (P1 = 1770–1790, P2 = 1820–1840).

In the entire *Going Dutch Corpus*, the two most frequent variants are *hetwelk* (32.5%) and the *d*-form *dat* (30.8%), whereas *hetgeen* (17.5%) and the *w*-form *wat* (15.6%) are somewhat less frequent in both diachronic cross-sections. The fifth variant, viz. *welk*, turns out to be a comparatively marginal form.

Table 2. Distribution of neuter relative pronouns across time.

	<i>dat</i>		<i>wat</i>		<i>hetgeen</i>		<i>welk</i>		<i>hetwelk</i>		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
P1	177	32.9	77	14.3	102	19.0	22	4.1	160	29.7	538	100
P2	134	28.5	80	17.0	75	15.9	14	3.0	168	35.7	471	100
Total	311	30.8	157	15.6	177	17.5	36	3.6	328	32.5	1,009	100

Diachronically, the results indicate a remarkably stable distribution of variants across periods. Some tendencies with regard to the most frequently used variants are notable, though. In the late eighteenth-century data, the *d*-form *dat* (32.9%) is slightly more frequent than *hetwelk* (29.7%), while the latter becomes the prevalent variant in the early nineteenth century (35.7%), outnumbering *dat* (28.5%). Further developments are less notable. The relative frequency of the *w*-form *wat* only slightly increases from 14.3% to 17.0%. The additional pronominal forms *hetgeen* and *welk*, on the other hand, lose ground in the second period.

In sum, *dat* and *hetwelk* prove to be the two predominant neuter relative pronouns, both in late eighteenth- and early nineteenth-century usage. The corpus data do not show evidence of the rise of the *w*-form *wat*. As outlined in Section 2, Weiland (1805) explicitly rejected the use of *wat* in certain contexts, depending on the definiteness of the antecedent. In the following, a closer look will be taken at the supposed influence of this internal factor.

Definiteness of the antecedent

This part of the case study has two aims: First, van der Horst's (1988) generalisation with regard to different types of antecedents, claiming that *w*-forms entered the language from indefinite to definite contexts (i.e. from left to right on the definiteness cline), will be tested on corpus data. Secondly, the corpus results will also be compared with the prescriptions in Weiland's (1805) national grammar. The preferred forms per context in Weiland (1805) are as follows: I: *wat*; II: *dat*, *hetwelk*; IIIa: *wat*; IIIb: no indications); IV: preferably *dat*, *hetwelk*, *welk* (*wat* may occur in usage); V/VI: *dat*, *hetwelk*, *welk* (cf. also Section 2).

Figure 1a. Distribution of neuter relative pronouns across context (1770–1790).

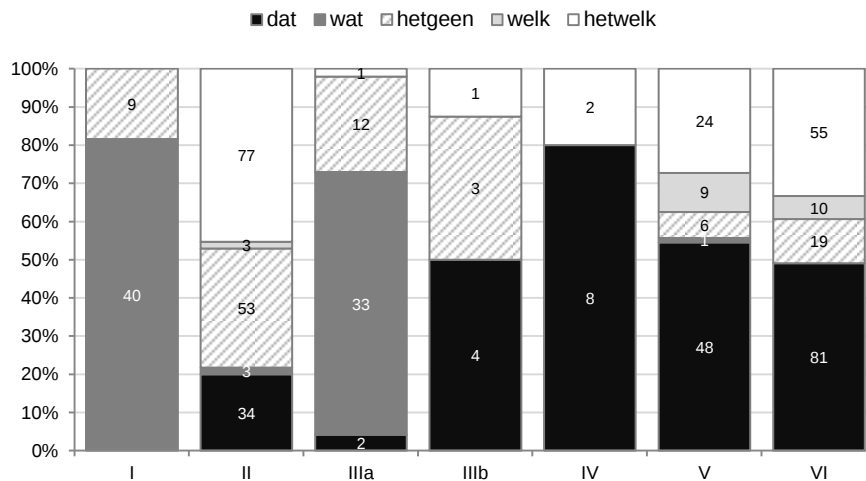
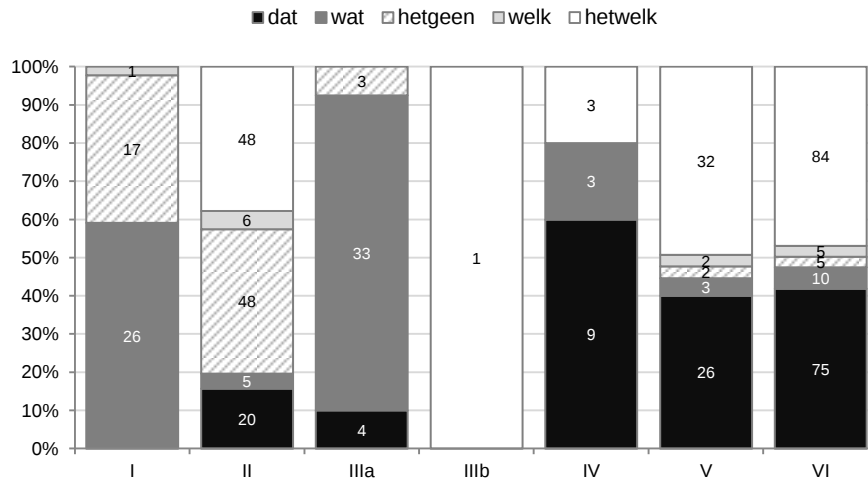


Figure 1b. Distribution of neuter relative pronouns across context (1820–1840).

In general, the results presented in Figures 1a and 1b certainly show similarities between the most frequently occurring variants per context and the variants preferred by Weiland (1805). However, as already indicated in the overview of forms in Section 5.1, both late eighteenth- and early nineteenth-century language practice reveal much more variation than Weiland acknowledged.

In both periods, the *w*-form *wat* turns out to be the most frequently used free relative (context I), occurring in 81.6% in the late eighteenth century and in 59.1% in the early nineteenth century. This largely corresponds to Weiland's (1805) preference for *wat* as a free relative. In the corpus data, however, the alternative relative pronoun *hetgeen* also occurs in 18.4% in the first period and even increases to 38.6% in the second period. As pointed out earlier, Weiland's stance on *hetgeen* remains implicit in his 1805 grammar, although he did acknowledge it as a legitimate variant in usage in Weiland (1799: 124). The *d*-form *dat*, on the other hand, is completely absent in context I, where it had been replaced by the *w*-form by that time.

In the case of a sentence or clause as antecedent (context II), the main variants in both periods are *hetwelk* and *hetgeen*, although *dat* also occurs to a lesser extent. In the nineteenth century, *hetwelk* and *hetgeen* have an equally high share of 37.8%, outnumbering *dat* (15.8%). Weiland (1805) suggested the use of *dat* and *hetwelk* in these cases, but not the *w*-form *wat*, which is a minor variant in the results. Moreover, the surprisingly high share of *hetgeen* in context II contradicts the eighteenth-century normative tradition, according to which *hetgeen* is only acknowledged as a free relative (I).

The combination with the indefinite pronoun *al(les)* (context IIIa) typically appears with *wat*, apart from a few attestations with *hetgeen* or *dat*. Particularly in the nineteenth-century period, *wat* consolidates its dominant position in this context, increasing from 68.8% to 82.5%. Weiland (1805) also illustrated this context by two

examples with *alles wat*. Unfortunately, there are too few tokens for context IIIb, but what these tentative results indicate is that the antecedents *iets*, *niets*, *veel* and *weinig* do not occur with the *n*-form *wat*, but with either *dat*, *hetgeen* or *hetwelk* – in contrast to *al(les)*. This supports the decision to split context III on the definiteness cline into two subcontexts.

Similarly, the limited number of tokens for context IV (nominalised adjectives) does not allow for a detailed evaluation. What becomes apparent, though, is that *dat* occurs in most instances (80.0% in period 1; 60.0% in period 2), with some additional attestations of *hetwelk* and *wat*. Despite the low number of tokens, it is notable that the distribution of variants in this context is rather stable.

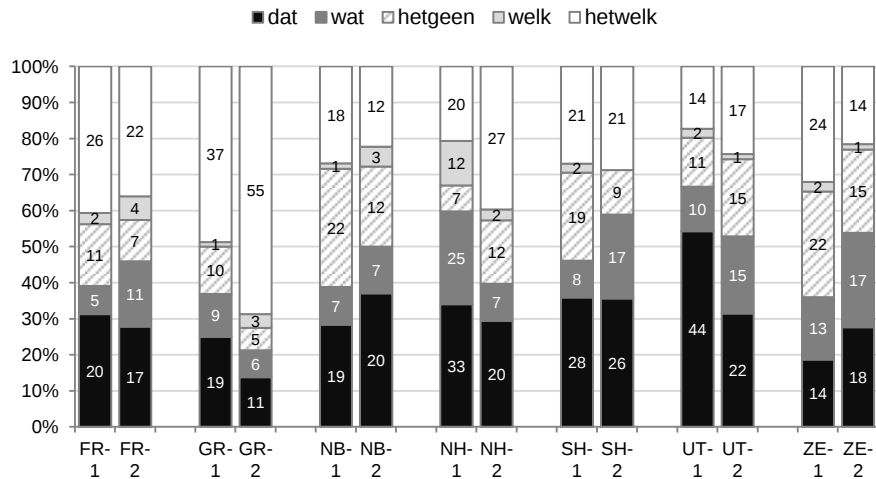
In many respects, the corpus results for contexts V (indefinite noun phrases) and VI (definite noun phrases) are similar. In the eighteenth century, *dat* is the main variant with a relative frequency of 54.6% (V) and 49.1% (VI), respectively. In both cases, *hetwelk* turns out to be the second most frequent variant in usage: 27.3% (V) and 33.3% (VI). Except for one single attestation of *wat*, the *n*-form does not occur in the eighteenth-century data. In the nineteenth century, one can witness a change in the distribution of variants in these two contexts: *dat* drops from 54.6% to 40.0% (V) and from 49.1% to 41.9% (VI). In contrast, the use of *hetwelk* increases from 27.3% to 49.2% in context V and from 33.3% to 46.9% in context VI. There are a few more nineteenth-century attestations of relative *wat* referring to indefinite and definite noun phrases than in the eighteenth century, but they are still relatively marginal. Weiland (1805) explicitly rejected the use of *wat* in contexts V/VI, only allowing for *dat*, *hetwelk* and *welk* as ‘legitimate’ options.

Although the *n*-form does not gain ground in early nineteenth-century usage, the results indicate that *wat* occurs in more contexts other than contexts I (free relatives) and III(a) (grammaticalised *al(les) wat*) – in both periods. In general, the distribution of variants in most contexts is stable across time, except for the increase of *hetwelk* in contexts V/VI.

The corpus data do not give evidence that *n*-relativisers enter the language from the most indefinite to the most definite context. Relative *wat* does occur in contexts I and IIIa, but not in II, which disproves the assumption of a linear spread on the definiteness cline from left to right. Furthermore, the *n*-form seems to appear, at least as a marginal variant, in contexts IV, V and VI in the nineteenth-century data. This suggests that *wat* spread to these contexts simultaneously rather than successively. With respect to the diachronic stability across contexts, I will leave this internal factor out of consideration in the corpus analyses of the external variables, i.e. spatial variation, gender variation and genre variation.

Regional variation

Figure 2 presents the relative distribution of Dutch neuter relative pronouns across the seven selected regions of the Northern Netherlands (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

Figure 2. Distribution of neuter relative pronouns across region and time.

In the late eighteenth century, *dat* occurs by far most frequently in the region of Utrecht (54.3%), but is also common in South Holland (35.9%) and North Holland (34.0%). It is least frequent in Zeeland (18.7%). In the two northernmost regions of Groningen and Friesland, *hetwelk* is the dominant variant in late eighteenth-century usage with a share of 48.7% and 40.6%, respectively. The other alternative form *hetgeen* is the prevalent variant in North Brabant (32.8%) and also among the most frequent variants in Zeeland (29.3%), both of which are the southernmost regions of the investigated language area. This possibly suggests a north–south difference with regard to the choice of alternative relative pronouns, viz. predominant *hetwelk* in the north versus relatively high-frequent *hetgeen* alongside *hetwelk* in the south.

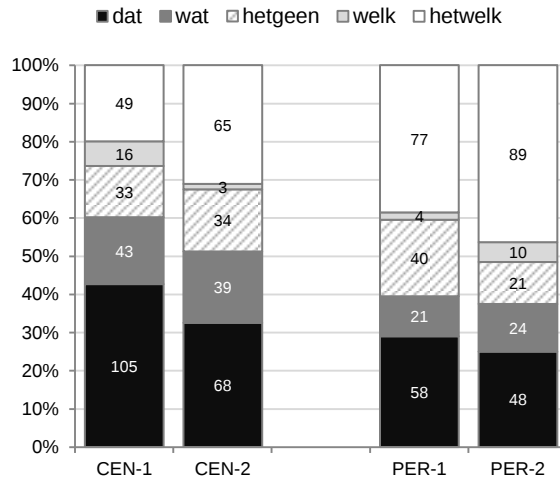
In the early nineteenth century, the high share of *dat* in Utrecht decreases from 54.3% to 31.4%. It also drops in Groningen (from 25.0% to 13.8%), whereas it remains rather stable in Friesland, North Holland and South Holland. In North Brabant and Zeeland, *dat* slightly gains ground. The *w*-form *wat* increases in some regions, particularly in South Holland (from 10.3% to 23.3%), Utrecht (from 12.4% to 21.4%) and Zeeland (from 17.3% to 26.2%). More generally, a diachronic stability can be attested in many regions, most notably in Friesland and South Holland, but also in North Brabant and Zeeland.

With regard to the alternative forms *hetwelk* and *hetgeen*, the suggested north–south difference is still visible in the nineteenth-century data. Particularly in the region of Groningen, the use of *hetwelk* increases considerably from 48.7 to 68.8%, consolidating its position as the predominant neuter relative pronoun. In Friesland, the preference for *hetwelk* (36.1%) over *hetgeen* (11.5%) remains remarkable, too. In the southern regions of Zeeland and North Brabant, on the other hand, *hetgeen* continues to be a comparatively strong variant in usage with 23.1% and 22.2%, respectively.

Variation across centre and periphery

The distribution of neuter relative pronouns across the second spatial dimension, viz. variation across the centre (CEN) and the periphery (PER), is presented in Figure 3.

Figure 3. Distribution of neuter relative pronouns across centre–periphery and time.



In the late eighteenth century, the prevalent variant used in the centre is the *d*-form *dat* (41.0%). The remaining variants are considerably less frequent: *hetwelk* (21.5%), *wat* (16.8%), *hetgeen* (13.6%) and *welk* (6.3%). In the periphery, however, *hetwelk* is the main variant in usage (39.1%), outnumbering *dat* (28.0%), *hetgeen* (20.7%) and the comparatively low-frequent *n*-form *wat* (10.1%).

In the early nineteenth century, the prevalence of *dat* diminishes in the centre (from 41.0% to 32.2%), whereas *hetwelk* gains ground and becomes an almost equally frequent second variant in usage (30.8%). The use of *wat* increases from 10.1% to 18.5%. In the periphery, the share of *hetwelk* grows even further, increasing from an already strong 39.1% to 45.6%. At the same time, *dat* slightly decreases from 28.0% to 24.6%.

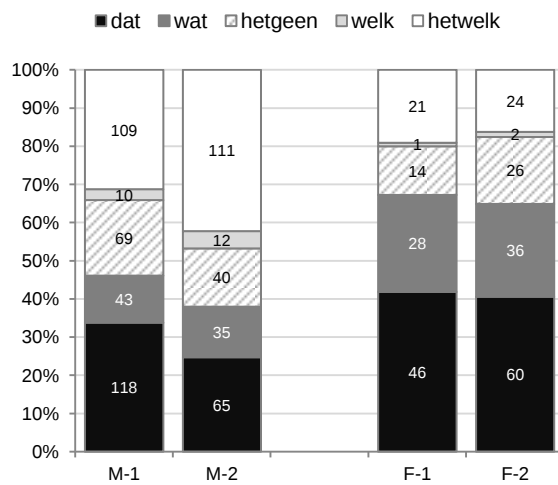
In sum, the general tendencies, i.e. less *dat*, more *hetwelk*, are similar in both the centre and the periphery. However, the prevalence and increase of *hetwelk* turns out to be more pronounced in the periphery, which may be largely due to the frequent use of this variant in the two northern regions of Friesland and Groningen. In contrast, the distribution of variants in the centre is rather balanced.

Gender variation

Presenting the distribution of variants across gender, Figure 4 reveals remarkable differences in the use of neuter relative pronouns between male (M) and female (F) writers.

In the eighteenth-century results, the most frequently occurring variants among male writers are *dat* (33.8%) and *hetwelke* (31.2%). In the ego-documents by their female contemporaries, the prevalence of *dat* as the main variant is considerably stronger (41.8%), whereas *hetwelke* occurs in only 19.1%. The use of the *w*-form turns out to be another gender difference in the first period. In fact, *wat* is only the fourth most frequent variant used by men (12.3%), but the second most frequent variant used by women (25.5%).

Figure 4. Distribution of neuter relative pronouns across gender and time.



The gender differences increase further in the nineteenth-century period. The distribution of variants in the texts written by women remains extremely stable. The *d*-/*w*-forms *dat* and *wat* only minimally decrease from 41.8% to 40.5% and 25.5% to 24.3%, respectively. In the texts written by men, however, the developments are more visible. The relative frequency of *hetwelke* increases from 31.2% to 42.2% and takes the position as the main variant in usage at the cost of *dat*, which drops from 33.8% to 24.7%. Interestingly, the (varying) shares of *wat* do not change considerably in male and female texts.

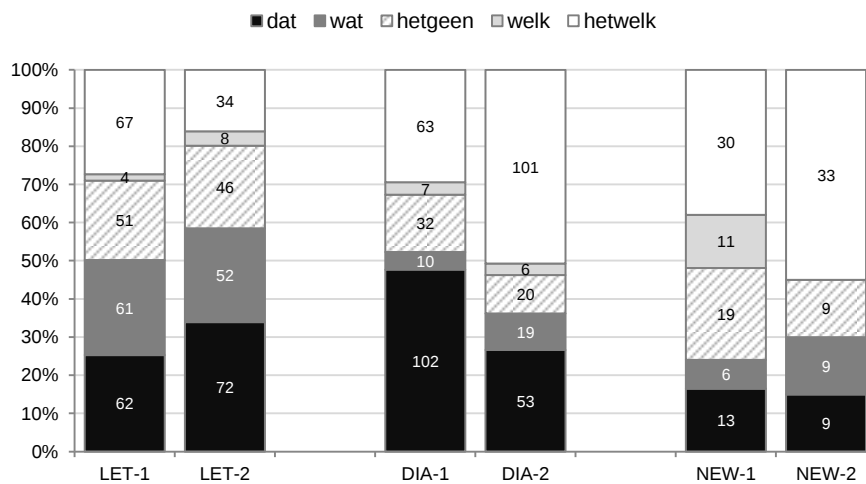
Genre variation

Investigating the last external variable, Figure 5 presents the distribution of neuter relative pronouns across the three genres in the *Going Dutch Corpus*, viz. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW).

The results for the eighteenth-century period reveal notable genre differences. In private letters, no fewer than four similarly frequent variants are used: both the *d*-/*w*-forms *dat* (25.3%) and *wat* (24.9%) as well as the additional pronominal forms *hetwelke* (27.4%) and, to a slightly lesser extent, *hetgeen* (20.8%) occur in more than twenty per cent each. In contrast, diaries and travelogues from

the first period clearly have a prevalent variant. The *d*-form *dat* occurs in almost half of all instances (47.7%), whereas the *v*-form *wat* is used in only 4.7%. The alternative *betwelk*, however, also has a rather high share of 29.4% in this genre.

Figure 5. Distribution of neuter relative pronouns across genre and time.



The newspaper genre shows yet another distribution of variants in the late eighteenth century. In these printed and published texts, two of the additional forms, viz. *betwelk* and *betgeen*, are the predominant variants, occurring in 38.0% and 24.1%, respectively. They both outnumber *dat* (16.5%) and particularly *wat* (7.6%). Interestingly, the older alternative form *welk*, which is a marginal variant in both types of ego-documents, has the highest share in newspapers (13.9%). Furthermore, the comparison of the three genres suggests that the use of *wat* in the first period is restricted to private letters, whereas it rarely occurs in diary and newspaper texts.

The distribution of variants in the private letters seems to evolve from a range of similarly frequent variants in the late eighteenth century towards one slightly more dominant variant in the early nineteenth century. In fact, the *d*-form *dat* increases from 25.3% to 34.0%. At the same time, the use of *betwelk* drops from 27.4% to 16.0%, whereas *wat* (24.5%) and *betgeen* (21.7%) generally remain stable.

The developments in diaries and travelogues are in sharp contrast to those in the private letters. The use of *dat* considerably decreases from 47.7% to 26.6%, whereas *betwelk* gains ground in period 2 and even becomes the main variant in this genre with 50.8%. A similar development in the use of *betwelk* can also be attested for early nineteenth-century newspapers, where this variant further consolidates its dominance, increasing from 38.0% to 55.0%. With regard to the *v*-form, *wat* remains a comparatively low-frequent variant in diaries and travelogues (9.6%) as well as in newspapers (15.0%), although its share increases in both genres.

Diachronically, this means that diaries and travelogues develop towards a distribution similar to newspapers. In order to trace the remarkable rise of *betwelk* in diaries and travelogues on a more detailed level, Table 3 zooms in on the distribution of variants across gender in this genre. Although it is important to take into account the overrepresentation of male writers (40 texts by 40 individuals) and thus a less representative number of female writers (10 texts by 10 individuals), the results indicate interesting tendencies.

Table 3. Distribution of neuter relative pronouns across gender and time (diaries and travelogues).

	<i>dat</i>		<i>wat</i>		<i>hetgeen</i>		<i>welk</i>		<i>hetwelk</i>		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
M-1	77	41.6	9	4.9	29	15.7	7	3.8	63	34.1	185	100
M-2	35	22.0	12	7.6	17	10.7	6	3.8	89	56.0	159	100
F-1	25	86.2	1	3.5	3	10.3	0	0.0	0	0.0	29	100
F-2	18	45.0	7	17.5	3	7.5	0	0.0	12	30.0	40	100

In the eighteenth-century data, male diarists prefer *dat* (41.6%) and *betwelk* (34.1%). In texts by female diarists, on the other hand, *dat* (86.2%) is predominant, whereas *hetgeen* and *wat* are rare, and *welk* and *betwelk* even absent. In the nineteenth century, male writers increasingly use *betwelk* (from 34.1% to 56.0%), at the cost of *dat* (from 41.6% to 22.0%). The *d*-form *dat* may remain the main variant used by female writers (45.0%), but the rise of *betwelk* can also be attested here, increasing from no attestations in the first period to 30.0% in second period. However, it seems that particularly men were establishing *betwelk* as the main variant in nineteenth-century diaries and travelogues.

6 Discussion

The case study in this chapter focused on the use of neuter relative pronouns in subject and object position. Like in other West Germanic languages, one of the major developments in the Dutch relativisation system is the change from originally demonstrative *d*-forms to originally interrogative *w*-forms. In the case of the neuter relative pronoun, this change has resulted in the gradual replacement of *dat* by *wat*. However, in the late eighteenth- and early nineteenth-century period under investigation, the change appeared to be still in its incipient stage in most contexts. Apart from the fact that it may simply be a more recent change, another important factor of the comparatively slow rise of *wat* may be the presence of many more competing forms. In fact, the contemporary set of variants for the neuter relative pronoun comprised no fewer than five options. In addition to *dat* and *wat* as the traditional *d*- and *w*-forms, the alternative forms *welk*, *betwelk* and *hetgeen* occurred in

language practice, all of which already date back to the sixteenth century or even to the Middle Dutch period. Nevertheless, both *welke* and the extended variant *betwelke* are based on the interrogative pronoun *welke*, which means that even if the change from *d*- to *w*-forms of the neuter relative pronoun was not as advanced as in the case of adverbial relativisers, there was still a strong competition from originally interrogative forms.

One aim of this case study was to test the proposed generalisation with respect to definiteness on corpus data, in order to see whether *wat* enters the language from indefinite to definite contexts, as claimed by van der Horst (1988). It is clear that the differentiation between various types of antecedents is helpful to the extent that the *w*-form *wat* was indeed the dominant form in two contexts. One of these is the most indefinite contexts of free relatives (I), whereas the other one is context IIIa, comprising the idiomatic combination *al(les) wat*. This means that the corpus-based results did not confirm a simple left-to-right movement, as suggested on van der Horst's (1988) definiteness cline. At the same time, *wat* was a marginal form in almost all other contexts, again indicating that the change did not strictly follow the cline. For the period under investigation here, it can be concluded that this cline is of little relevance for the neuter relative pronoun⁷³. What is more remarkable about the suggested internal factor is the diachronic stability of the results.

Given the fact that *wat* was the main variant in context IIIa, one would have expected a similar pattern in context II, referring to a clause or sentence. It is notable, though, that *dat* is less frequent in context II (as it is in I and IIIa), compared to contexts IV–VI. However, the *d*-form was not only replaced by *wat*, but also by the additional forms *betgeen* and *betwelke*, particularly in context II.

The diachronic stability is not only striking with respect to the internal factor, but also in terms of the external factors of region and gender. The general distribution for the nineteenth-century period was, in fact, very similar to that for the eighteenth-century period. The most obvious difference was the increased frequency of *betwelke*, largely at the expense of *dat*. Similarly, the distribution in regions such as Friesland, Groningen, North Brabant (notably all regions of the periphery) and Zeeland was fairly stable across time. One surprising outcome indicating a possible regional pattern was the relative prominence of *betwelke* in the north, and the relative prominence of *betgeen* in southern regions. This finding calls for further investigation as it is interesting to see regional differences in variants which are considered formal or typical of written language. If confirmed in future research, it would also imply that the rise of *betwelke* is a change from the periphery to the centre, as suggested by the centre–periphery distribution across time.

The gender distribution was also relatively stable across time, particularly in the case of female writers. Male writers, in contrast, showed an increase in the use of *betwelke* at the cost of *dat*. However, the results for gender turned out to be slightly different when cross-tabulated with genre.

⁷³ The irrelevance of the type of antecedent as an internal factor has also been argued by Romaine (2009: 143–144) with respect to Middle Scots.

In general, genre was undoubtedly the most important external variable affecting the distribution of neuter relative pronouns. Nineteenth-century newspapers as well as diaries and travelogues showed a remarkable increase in the use of *betwelke*. Furthermore, diaries in particular also showed a decrease in the use of *dat*. Interestingly, diaries and travelogues were more similar to private letters than to newspapers in period 1, at least with respect to the share of *betwelke*, while diaries and travelogues from period 2 align with newspapers. Thus, from a diachronic point of view, the two ego-document genres diverged. Taking into account possible gender differences, the rise of *betwelke* actually occurred in diaries and travelogues written by both men and women.

Against the background of a diachronically relatively stable distribution of neuter relative pronouns, the increase of *betwelke* in diaries and travelogues is notable. Is it possible to relate this development to contemporary language norms? Recall that Weiland's (1805) national grammar was much more elaborate with regard to relativisation than any of its predecessors. Comparing the norms and rules for relativisation found in Weiland (1805) and the preceding eighteenth-century normative tradition to the usage patterns found in the *Going Dutch Corpus*, it can be concluded that the influence of the metalinguistic discourse on language practice must have been fairly limited. Weiland rejected *d*-forms (context I), which was also confirmed by the corpus results. Eighteenth-century grammarians and Weiland (1799) wanted to restrict the use of *betgeen* to free relatives. In practice, however, it occurred in almost every context in both periods. Furthermore, Weiland (1805) proposed a distribution of forms in accordance with types of antecedents, for example *dat* and *betwelke* for clauses and sentences as antecedent (context II), *wat* in combination with *al(les)* (IIIa), *dat*, *betwelke* (and dispreferred *wat*) with nominalised adjectives (IV), and so on. In all these cases, actual language usage proved to be more variable with almost all possible variants occurring in almost all contexts. At the same time, there were considerable similarities between Weiland (1805) and usage patterns to the extent that frequently rejected forms were only marginally used in certain contexts. However, as already mentioned, the distribution across time remained relatively stable, which implies that the overlap with Weiland (1805) is already found in the late eighteenth century, when the influence of Weiland's grammar of the early nineteenth century can obviously be excluded.

At one point, the influence of Weiland (1805), either direct or indirect, can be assumed. He combined the traditional forms with the stylistic or register differences already proposed by ten Kate (1723). Weiland (1805) assigned forms such as the masculine and feminine pronouns *welke* and *denwelke* to the so-called 'solemn' style, whereas *die* and also neuter *dat* were described as forms of the so-called 'plain' style. Extending these observations to the neuter paradigms, the corpus results revealed that the 'solemn' form *betwelke* gained ground in newspapers as well as in diaries and travelogues, at the cost of the 'plain' form *dat*. Although it is difficult to prove that this was a direct result of Weiland's (1805) intervention, it does signal a situation in which *betwelke* was primarily associated with formality or

‘solemnity’ – in any case more strongly than other variants such as *dat.* Both Weiland and the corpus data testified to this situation.

In this respect, the diachronic results for diaries and travelogues are certainly remarkable. The developments in the distribution of neuter relative pronouns highlight the special position of these sources as a genre on the oral–literate continuum (cf. Section 3.1.2 in Chapter 4). Although diaries and travelogues have been typically categorised and treated as ego-documents like private letters, it has to be kept in mind that we are dealing with distinct subgenres of ego-documents (Elspaß 2012: 162). Moreover, they are usually less ‘oral’ and more ‘standard’-like (Schneider 2013: 66, cf. also Rutten 2012b for a Dutch example). Writers of ego-documents, both men and women, created a divergence of private letters on the one hand, and diaries and travelogues on the other, by adopting *betwelle* considerably stronger only in the latter type of ego-documents.

To sum up, this case study on neuter relative pronouns revealed a considerable amount of regional, gender and particularly genre variation, whereas the diachronic distribution appeared to be relatively stable across time. In the second case study on relativisation, presented in Chapter 11, I will investigate the masculine and feminine singular and plural relative pronouns. A more general conclusion of the results drawn from both case studies will be provided at the end of the following chapter.

Morphosyntactic variables (2)

Masculine and feminine singular and plural relative pronouns

1 Corpus analysis

Building on the analysis of neuter relative pronouns presented in Chapter 10, this second morphosyntactic case study focuses on masculine and feminine singular and plural relative pronouns. A historical overview of the Dutch relativisation system was provided in Section 1 of the previous chapter. For a comprehensive outline of norms and preferences related to the use of relative pronouns (i.e. including masculine and feminine forms), see Sections 2 and 3 of Chapter 10, discussing Weiland (1805) and eighteenth-century metalinguistic discourse, respectively. In addition to the corpus analysis of masculine and feminine singular and plural relative pronouns, the present chapter also contains a general conclusion drawn on the basis of the findings from two individual case studies on relative pronouns (Section 3).

1.1 Method

For the corpus analysis of masculine and feminine singular (common gender in present-day Dutch) and plural relative pronouns in subject and object position, the occurrences of the following variants were extracted:

- Nominative: *die*, *wie*, *welke*, *develke*;
- Accusative: *dien* (masc. sg.), *die* (fem. sg.; pl.), *wien* (masc. sg.), *wie* (fem. sg.; pl.), *welken* (masc. sg.), *welke* (fem. sg.; pl.) *denwelken* (masc. sg.), *develke* (fem. sg.; pl.).

For the sake of clarity, these variants will be referred to as *die*, *wie*, *welke* and *develke* (comprising both nominative and accusative forms) throughout this chapter.

Possible spelling variation, for instance the *k/ck/c* variation in *(de)welke*, was also taken into consideration. All-non relative occurrences were filtered out by hand, including *die* functioning as demonstrative pronoun, *wie* as interrogative pronoun as well as in the formulaic expression *wie weet* ‘who knows’, and *welke* as interrogative pronoun. Ultimately, 2,473 occurrences of masculine and feminine singular and plural relative pronouns in subject and object position were extracted from the *Going Dutch Corpus* and used for the analysis. Before the quantitative

results will be presented in Section 1.2, I briefly discuss the use of the possible variants (*die, wie, welke, dewelke*) in different contexts, illustrated by examples taken from the *Going Dutch Corpus*. Two internal factors will be taken into account, viz. the definiteness as well as the animacy of the antecedent.

The definiteness of the antecedent was empirically tested on neuter relative pronouns in Chapter 10 (Section 5.2). Referring to the different types of antecedents in van der Horst's (1988) definiteness cline, van der Wal (2002b: 32) points out that masculine and feminine relative pronouns can have four possible antecedents:

In the case of DIE/WIE there are four possible antecedents (categories 1, 3; 5 and 6 above). In Modern Dutch, DIE is still the common gender and plural relative pronoun which occurs with an indefinite or definite antecedent (categories 3, 5 and 6) and functions as subject, indirect or direct object. It is the more remarkable as at the end of the sixteenth century, both DIE and WIE occurred as free relatives. [...] In Modern Dutch WIE is the current free relative form, but the pronoun has not made its way into the other categories yet, in any case not in accepted Standard Dutch.

In other words, the antecedents of masculine and feminine relative pronouns may be free or headless relatives (context I), indefinite pronouns (context III), indefinite noun phrases (context V), and definite noun phrases (context VI). However, the corpus analysis of neuter relative pronouns in Chapter 10 has shown that the relevance of the antecedent's definiteness is fairly limited. Moreover, in the case of masculine and feminine relative pronouns, definiteness is generally a less interesting factor due to the relatively restricted spectrum of possible contexts, in which masculine and feminine forms can occur. Whereas neuter forms can be used in all contexts, the vast majority of masculine and feminine forms refers to either indefinite or definite noun phrases as antecedents (i.e. contexts V and VI). In other words, there are too few contexts to investigate possible diachronic developments on the cline. Although I will occasionally refer to these contexts, the definiteness of the antecedent will not be studied as an internal variable again.

Another possible internal factor that has been argued to affect the choice of particular relative pronouns is the semantic property of animacy. Assuming that the animacy of the antecedent conditions the distribution of variants, all occurrences of masculine and feminine relative pronouns in subject and object position referring to noun phrases in the definiteness contexts V (indefinite noun phrases) and VI (definite noun phrases)⁷⁴ were manually coded as either animate or inanimate. Here, 'animate' refers to humans, animals, collective nouns, but also God and other deities⁷⁵ (cf. also Bergs 2005: 146), whereas 'inanimate' comprises all

⁷⁴ Context I, i.e. free or headless relatives (antecedents embedded), and context III, i.e. indefinite pronouns (mostly referring to human entities), are excluded from this analysis.

⁷⁵ In his case study on relative clauses in Middle English family letters, Bergs (2005: 146) applies a tripartite classification of antecedents into animate (AN), inanimate (INA) and deity (DE), the latter of which is "an umbrella term that subsumes entities such as 'God', 'Jesus', 'The Holy Trinity', saints, etc.". According to the literature, "these referents may

non-human/non-animal referents (including dead animals as food, e.g. *nieuwe haring* ‘new herring’). Particularly in modern linguistics, the concept of animacy has often been treated as a hierarchy, theoretically allowing a fine-grained classification of antecedents. Working with similar historical data as in Hundt & Szmrecsanyi’s (2012) corpus-based study on animacy in Early New Zealand English, I follow their approach by applying the binary animate–inanimate distinction here.

In the following, the variability of masculine and feminine relative pronouns in late eighteenth- and early nineteenth-century usage will be illustrated by examples taken from the *Going Dutch Corpus*. The originally demonstrative *d-*form *die*, which was the default form in the earliest periods of Dutch, occurs as free relative in (1). While *wie* is the free relative form in present-day Standard Dutch, there are no occurrences of nominative *wie* or accusative *wien* as free relatives in the eighteenth- and nineteenth-century data of the *Going Dutch Corpus*. However, the case of dative *wien* in (2) shows that the *w-*form does occur in this context. In contrast, the alternative pronominal forms *welke* and the extended alternative *denwelke* do not seem to be used as free relatives at all:

- (1) **die** een vrouw vraagt, moet weten, van waar hij dezelve zal maintainereen
 ‘who(ever) asks a woman, must know from which he should maintain her’
- (2) **wien** het ook ten deele moge vallen
 ‘whom(ever) it may be allocated to’

In examples (3–5), the antecedent is an indefinite pronoun such as *elke*, *ieder*, *iemand*, *niemand*, *veel*, *weinig* and so forth, commonly referred to as context III⁷⁶. In the corpus, there are only attestations of the *d-*form in this context:

- (3) *iemand* [...] **die** morgen vroeg voor mij naar Harderwyk gaat
 ‘someone [...] who goes to Harderwijk for me tomorrow morning’
- (4) *elk*, **die** belang in de algemeene welvaart stelt
 ‘anyone who is interested in [the] general prosperity’
- (5) *alle* **die** na mij vragen
 ‘all who ask for me’

have had an influence on the choice of the relativizer”. With regard to the investigated period and the selected sources of the present study, comparatively rarely referring to God and other divine entities (especially in newspapers, diaries and travelogues), such a distinction seems less relevant.

⁷⁶ Unlike the idiomatic expression *al(les) wat* (as opposed to other indefinite pronouns like *iets*, *niets*, *veel*) in the discussion of neuter relative pronouns (Chapter 10), there is no comparable (grammaticalised) case with regard to masculine and feminine relative pronouns. Therefore, a split of context III does not seem necessary here.

In contexts V and VI, the relative pronouns have indefinite or definite noun phrases as antecedents. It is in examples (6–12) where the full spectrum of variation becomes apparent, as practically all forms occur in reference to noun phrases. Also considering the factor of animacy, the (a) examples refer to animate antecedents, whereas the (b) examples have inanimate antecedents (if applicable).

In context V, i.e. indefinite noun phrases, *die* (6), *welke* (7) and *dewelke* (8) each occur with animate and inanimate antecedents:

- (6) a. *eene goedaardige kleine oude vrouw, **die** een weinig Hollandsch en Fransch spreekt*
 ‘a good-natured small old woman, who speaks a little Dutch and French’
- b. *een sterke Donderbui, **die** met eenen harden wind en sterken regen begon*
 ‘a heavy thunderstorm, which started with a strong wind and heavy rain’
- (7) a. *een elendige kok, **welke** nog geen eens aardappelen kan kookken*
 ‘a miserable cook, who cannot even cook potatoes’
- b. *eene ervaring, **welke** ik nimmer zal vergeten*
 ‘an experience, which I will never forget’
- (8) a. *25 Jagers [...] **de welke** alle weeken een stuk wild aan ‘t hof moeten besorgen*
 ‘25 hunters [...] which have to deliver a piece of game to the court every week’
- b. *een zwaare hoofdpyn **dewelke** wel haast met braken ge verzeeld ging*
 ‘a bad headache which was almost accompanied by vomiting’

The only form that does not occur in this context is *wie*, although it should be noted again that there are several attestations of the *w*-form declined for genitive and dative case referring to indefinite noun phrases in the corpus, for instance in *een persoon, **wien** ‘t acces geweigerd is* ‘a person, who was denied access’ (animate antecedent) and *een groote tafel 12 voeten lang 8 breed **wiens** byzonderheid was dat zy uit een stuk gemaakt was* ‘a large table, twelve feet long, eight wide, whose special quality it was that it was made out of one piece’ (inanimate antecedent).

Finally, all four possible forms occur in context VI (see examples (9–12)), referring to definite noun phrases. In the corpus data, *die* (9), *welke* (11) and *dewelke* (12) occur with both animate and inanimate antecedents, whereas *wie* (10) only occurs with an animate referent.

- (9) a. *de Postmeester **die** ny dubbelt geld boden*
 ‘the postmaster whom we offered twice the money’
- b. *de heerlyke Koffij **die** de abdis heden ochtend in haar Klooster had doen bereiden*
 ‘the delicious coffee that the abbess had prepared in her convent this morning’

- (10) [eene nieuwe bierbrouwerij van] de Heer Jan van Cleef **wie** lekker bier brouwt
 ‘[a new brewery of] Mister Jan van Cleef who brews delicious beer’
- (11) a. *die wedúwe, welke reeds hoog bejaard was*
 ‘the widow, who was already very old’
 b. *de gezonde lucht welke men hier inademt*
 ‘the healthy air which one breathes here’
- (12) a. *de voornaamste actrice van Parys Mad^e du Gazon [...] dewelke hier voor 20 Representaties 1200 Livres trekt*
 ‘the most prominent actress of Paris, Mademoiselle du Gazon [...] who gets 1200 livres for 20 shows here’
 b. *de Water machiene [...] de welke aan alle de brouwers Water uit de Schelde verschaft*
 ‘the water machine [...] which supplies all brewers with water from the Scheldt’

While the main purpose of these examples is to provide a first overview of the possible forms in different contexts, as they occur in the *Going Dutch Corpus*, it is crucial to conduct a quantitative analysis in order to shed light on variation and change in the distribution of relative pronouns. The corpus results are presented and discussed in Section 1.2 and 1.3, respectively.

1.2 Results

The overall distribution of Dutch masculine and feminine relative singular and plural pronouns in subject and object position (i.e. nominative and accusative forms) is shown in Table 1 (P1 = 1770–1790, P2 = 1820–1840).

In the entire *Going Dutch Corpus*, the *d*-form *die* clearly emerges as the prevalent relative pronoun for masculine, feminine and plural referents. In fact, the only competing form in language usage is the additional form *welke* with a share of 27.7%. Both the *n*-form *wie* (0.3%) and the extended variant *dewelke* (0.9%) occur very marginally.

Table 1. Distribution of masculine/feminine relative pronouns across time.

	<i>die</i>		<i>wie</i>		<i>welke</i>		<i>dewelke</i>		Total	
	N	%	N	%	N	%	N	%	N	%
P1	888	72.3	3	0.2	315	25.6	23	1.9	1,229	100
P2	871	70.0	4	0.3	369	29.7	0	0.0	1,244	100
Total	1,759	71.1	7	0.3	684	27.7	23	0.9	2,473	100

Compared to the neuter relative pronouns investigated in Chapter 10, there is a considerably more limited set of variants in actual language use. In fact, only two forms (*die*, *welke*) are commonly used in the late eighteenth and early nineteenth century, as opposed to at least four competing forms of the neuter pronoun (*dat*, *wat*, *hetgeen*, *hetwelk*, to a lesser extent *welk*).

Diachronically, the distribution of variants presented in Table 1 turns out to be remarkably stable across periods, which is in line with the results for the neuter relative pronouns. Some small-scale developments are interesting, though. First of all, *welke* generally seems to gain ground in the nineteenth-century period, increasing from 25.6% to 29.7%. The increase of *welke* is mainly at the cost of *die*, which minimally decreases from 72.3% to 70.0%. It has to be further investigated, though, whether this is a general tendency, or whether there are genre-specific developments involved, similar to the case of neuter relative pronouns. Second, the extended form *develke* only occurs (marginally) in the eighteenth-century period (1.9%), but is completely absent in the nineteenth-century period. With only a few attestations in each period, the *n*-form *wie*⁷⁷ remains very marginal across time.

Genre variation

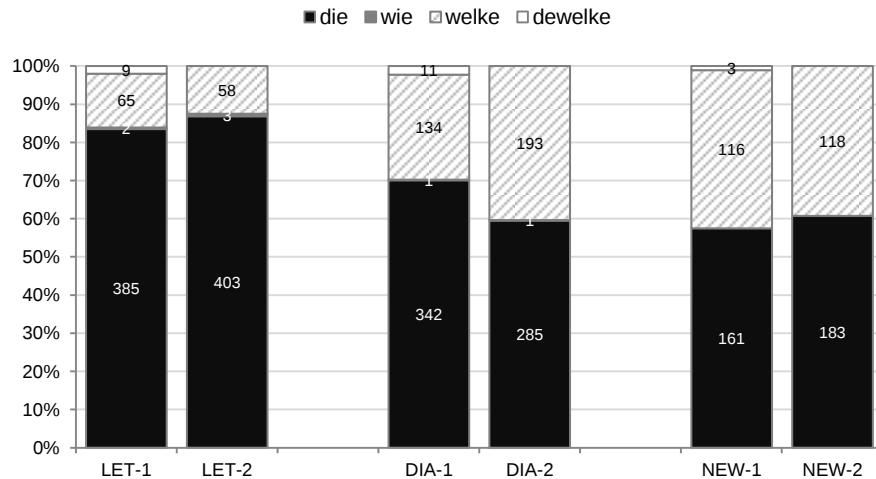
In the previous analysis of neuter relative pronouns (Chapter 10), genre proved to be the most relevant (external) variable affecting the distribution of variants. Figure 1 shows the distribution of variants across the genres of the *Going Dutch Corpus*, i.e. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW).

Again, the results reveal considerable genre differences. In private letters from the eighteenth-century period, the *d*-form *die* is by far the most frequently used relative pronoun (83.5%). The additional pronominal form *welke* has a comparatively low share of 14.1%. The remaining variants *develke* (2.0%) and particularly *wie* (0.4%) are extremely marginal in usage.

In the second ego-document genre, viz. diaries and travelogues, the prevalence of *die* is less pronounced than in private letters. Occurring in 70.1%, *die* has to be considered as the main variant, but the considerable share of *welke* (27.5%) gives evidence of two competing forms in these sources. Very similar to private letters, *develke* (2.3%) and *wie* (0.2%) are low in frequency.

In newspapers, i.e. the printed, published genre of the corpus, *die* and *welke* are most evenly distributed, occurring in 57.7% and 41.4%, respectively. In other words, they coexist as main variants in newspaper writing. There are three attestations of *develke* (1.1%), whereas *wie* is completely absent in these sources.

⁷⁷ When we also consider genitive and dative forms (including occurrences after prepositions), the originally interrogative *n*-forms like *wiens* and *wier* are somewhat more frequent. In fact, they have been used in the historical genitive since the Middle Dutch period. In the entire *Going Dutch Corpus*, one can find 41 attestations of *wiens* (gen. masc. sg.; also gen. neut. sg. of *wat*), 24 of *wier* (gen. fem. sg., pl.), 44 of *wien* (dat. masc. sg.) (cf. Weiland 1805: 113-114 for the paradigm), and 15 of *wie* after prepositions.

Figure 1. Distribution of masculine/feminine relative pronouns across genre and time.

The genre-specific distribution of variants remains relatively stable in the early nineteenth-century results. In the private letters, *die* slightly increases from 83.5% to 86.9%, whereas *welke* decreases from 14.1% to 12.5%. Generally speaking, the distribution in the letter data has to be regarded as stable, though. A similar stability in the distribution of variants can be attested in nineteenth-century newspapers. *Die* (60.8%) is still the most frequently used variant, but coexists alongside the second variant *welke* (39.2%).

In contrast, *die* considerably loses ground in diaries and travelogues, dropping from 70.1% to 59.5%, mainly in favour of *welke*, which increases from 27.5% to 40.3%. In fact, the only considerable changes are attested in these sources. In line with the remarkable diachronic developments in the use of the neuter relative pronouns (Chapter 10), diaries and travelogues seem to undergo genre-specific developments from the late eighteenth to the early nineteenth century. They appear to diverge from the more ‘oral’ ego-documents (private letters) in the first period towards the more ‘written’ formal language of printed, published texts (newspapers) in the second period.

On the whole, there are no genre-specific differences with regard to the use of *wie*. In both periods, there are only a few occurrences in the two ego-document genres and no attestations at all in newspapers. The previously at least marginally used *dewelke* disappears in all three genres in the nineteenth-century period.

The genre differences regarding the use of masculine and feminine relative pronouns – i.e. primarily *die* in private letters, but both *die* and *welke* coexisting in newspapers (and increasingly in diaries and travelogues) – suggest a link to Weiland’s (1805) awareness of stylistic or register variation. As outlined in Section 2 of the previous chapter, the relativisers *welke* and (less frequent) *dewelke* were assigned to the so-called ‘solemn’, more formal style, whereas *die* was more typical of the plain style. In line with Weiland’s (1805) comments, the corpus findings

presented here reveal that *welke* is considerably more frequent in printed, published texts than in the ‘common’ language of handwritten private letters, where *die* prevails. The question arises whether and to what extent this can be interpreted as a reflection of Weiland’s (1805) observations. Did his comments on stylistic or register variation set the pattern for early nineteenth-century practices? In the case of newspapers and private letters, the differences are already visible in the late eighteenth-century data, suggesting a situation in which *welke* must have been perceived as more formal than *die* – before Weiland (1805). Given the diachronic stability in these two genres, his direct influence on the nineteenth-century period was probably fairly limited. However, the increase of the ‘solemn’ form *welke* (at the expense of *die*) in nineteenth-century diaries and travelogues is remarkable and once again highlights the special position of this genre, certainly from a diachronic point of view. While it is not possible to prove Weiland’s intervention on this development either, it might be argued that his observations discussed in the national grammar promoted the general awareness and status of *welke* as a more formal variant as compared to *die*.

Animacy of the antecedent

The internal factor of animacy is investigated with a selected data set from the *Going Dutch Corpus*. In the previous section on genre variation, diaries and travelogues as well as newspapers turned out to be particularly interesting with regard to variation, with both *die* and *welke* being used as coexisting forms. Therefore, these two sub-corpora were selected to investigate the influence of the antecedent’s animacy. The third genre, i.e. private letters, seemed less suitable for this analysis, as *die* appeared to be by far the most frequent form in usage, thus leaving little room for variation to study. The results are presented in Table 2.

Table 2. Distribution of masculine/feminine relative pronouns across animate and inanimate antecedents.

	<i>die</i>		<i>wie</i>		<i>welke</i>		<i>dewelke</i>		Total	
	N	%	N	%	N	%	N	%	N	%
Animate-1	245	69.4	1	0.4	102	28.9	5	2.1	353	100
Inanimate-1	246	61.0	0	0.0	148	36.7	9	4.0	403	100
Animate-2	235	67.0	1	0.3	115	32.8	0	0.0	351	100
Inanimate-2	226	53.6	0	0.0	196	46.4	0	0.0	422	100

In the late eighteenth century, both *die* and *welke*, i.e. the two main variants in usage, refer back to animate and inanimate antecedents. The generally more frequent *die* has a share of 69.4% with animate antecedents, and a share of 61.0% with inanimate antecedents. *Welke* occurs in 28.9% of all instances with animate antecedents, and in 36.7% with inanimate antecedents. In other words, the

distribution across animate and inanimate antecedents is fairly well-balanced in the eighteenth-century period, although one might argue that *welke* is slightly more likely to refer to inanimate antecedents than to animate antecedents.

In the early nineteenth century, the differences are somewhat more pronounced. While the distribution of forms is diachronically fairly stable in the case of animate antecedents, one can witness a relative increase of *welke* with inanimate antecedents (from 36.7% to 46.4%) at the expense of *die* (from 61.0% to 53.6%). While this is generally in line with Weiland's (1805) preference for the more 'solemn' form *welke* in the higher registers, the increase of *welke* is, at least to some extent, conditioned internally by animacy, which is a factor that had not been considered by Weiland.

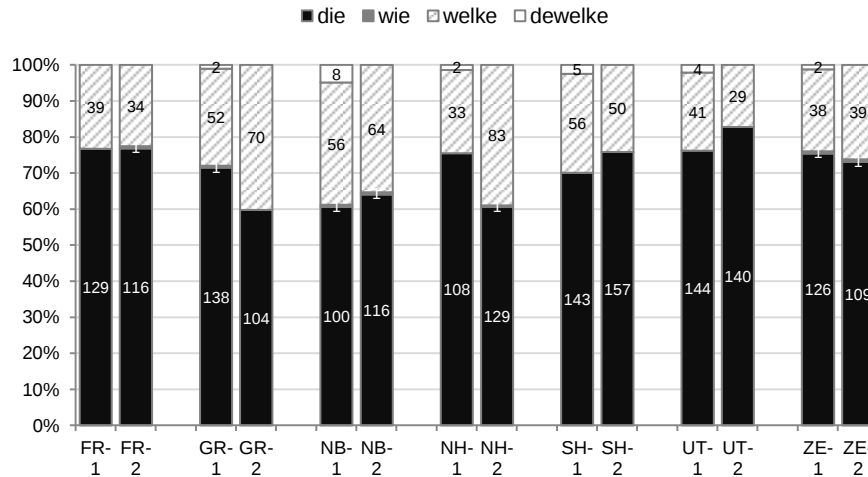
Taking into account the low number of tokens, the results for *wie* and *develke* can only be assessed tentatively. The extended form *develke*, exclusively used in the eighteenth-century period, seems to be slightly more common with inanimate antecedents (4.0% as opposed to 2.1% with animate antecedents), although this might very well be due to the small sample. The *n*-form *wie* occurs twice in the selected data set (one attestation for each period), both of which refer to animate antecedents. While this might not be representative enough to draw any solid conclusions, the remaining (five) instances in the entire *Going Dutch Corpus* were also considered in order to support these tentative findings. In fact, all seven occurrences of nominative and accusative *wie/wien* refer back to animate antecedents. Here, the relative pronoun appears to be restricted to animate antecedents indeed.

Given the limited relevance of the animacy of the antecedent, particularly on the distribution of the two main variants *die* and *welke*, this internal factor will not be taken into consideration in the following analyses of masculine and feminine relative pronouns.

Regional variation

The relative distribution of variants across the seven selected regions of the Northern Netherlands is displayed in Figure 2 (FR = Friesland, GR = Groningen, NB = North Brabant, NH = North Holland, SH = South Holland, UT = Utrecht, ZE = Zeeland).

Clear regional patterns in the distribution of masculine and feminine relative pronouns do not emerge, although there is also variation to some extent. Generally, the *d*-form *die* is the most frequent variant in usage across all seven regions. In the eighteenth-century period, its share varies from 60.6% in North Brabant to more than 75% in Zeeland, North Holland, Utrecht and Friesland. The share of *welke* ranges between 21.7% in Utrecht and 33.9% in North Brabant. Furthermore, it is notable that the infrequent form *develke* is not restricted to particular regions, but is practically spread across the entire language area.

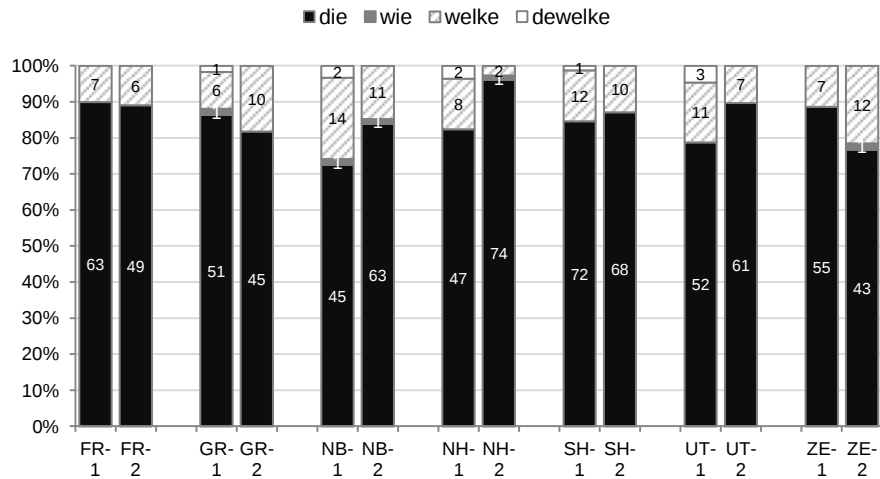
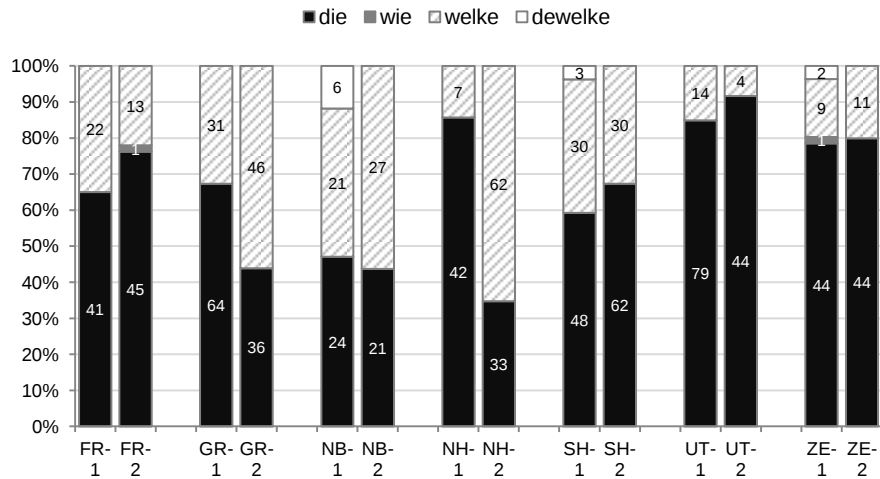
Figure 2. Distribution of masculine/feminine relative pronouns across region and time.

As pointed out before, *dewelke* disappears in nineteenth-century usage (in all regions). Overall, the distribution of variants remains rather stable in the second period, particularly in Friesland and Zeeland. In Groningen and North Holland, the share of *welke* increases from 26.9% to 40.2%, and from 23.1% to 39.0%, respectively. Recall that the neuter counterpart *betwelke* also appeared to be somewhat more common in the northern regions (cf. Section 5.2 of Chapter 10). In Utrecht, on the other hand, *die* clearly consolidates its status as dominant variant, increasing from 76.2% to 82.8%. As a result, the regional differences increase in the second period, with Groningen and North Holland being the regions with the highest shares of *welke* on the one hand, and Utrecht being the region with the strongest prevalence of *die* on the other.

In addition to the general distribution of variants across regions in the entire corpus, a closer look is taken at regional variation in two diverging types of ego-documents, i.e. private letters (Figure 3), and diaries and travelogues (Figure 4).

Starting with private letters, Figure 3 shows that *die* is the prevalent variant in all regions in both periods. In the eighteenth century, the difference ranges from 72.6% (North Brabant) to 90.0% (Friesland), and from 76.8% (Zeeland) to 96.1% (North Holland) in the nineteenth century.

There is more variation and change in the diaries and travelogues, as can be seen in Figure 4. In the eighteenth-century period, *die* is the most frequently used form in practically all regions except for North Brabant. Here, *die* occurs in less than half of all instances (47.1%), whereas *welke* and *dewelke* together have a share of 52.9%. It has to be kept in mind, though, that the data for North Brabant derive from only two different diarists in the first period and only one diarist in the second period. Idiosyncratic preferences might influence the distribution in texts from this region.

Figure 3. Distribution of masculine/feminine relative pronouns across region and time (private letters).**Figure 4.** Distribution of masculine/feminine relative pronouns across region and time (diaries and travelogues).

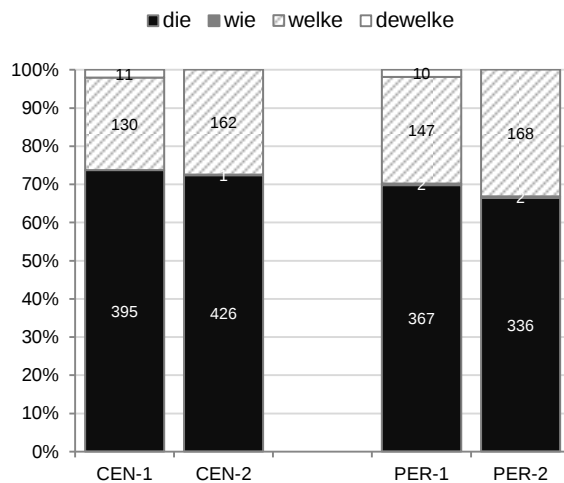
Diachronically, the most striking change can be attested in the North Holland data. The *d*-form drops from 85.1% to 34.7%, while *welke* considerably gains ground from 14.3% to 65.3%. This is in sharp contrast to the opposite developments attested in private letters from the same region, where *die* becomes practically the only variant in usage in the nineteenth-century period. This shift of main variants used in diaries and travelogues from North Holland (and in a slightly

less extreme way also Groningen), i.e. from mainly *die* in the first period to two-thirds of *welke* in the second period, clearly affects the overall distribution across regions presented in Figure 2. The distribution in texts from other regions is much more stable across time, particularly in Zeeland, but also in Utrecht, South Holland and North Brabant, the latter of which still maintains its high share of *welke* (56.3%).

Variation across centre and periphery

Furthermore, spatial variation in the distribution of variants was investigated across the centre (CEN) and the periphery (PER), as presented in Figure 5.

Figure 5. Distribution of masculine/feminine relative pronouns across centre–periphery and time.



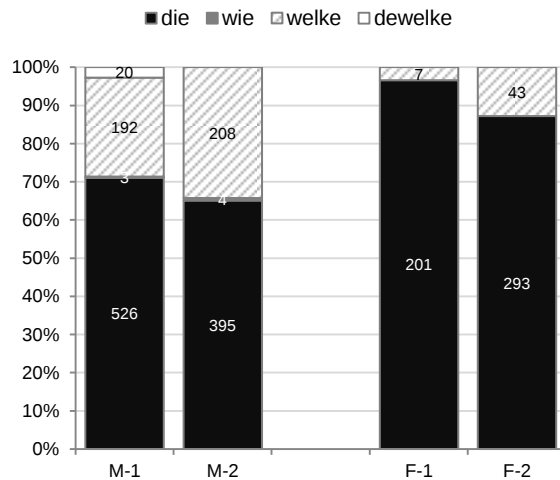
Both synchronically and diachronically, differences between the centre and the periphery appear to be limited. As could be expected from the results of the seven investigated regions, *die* is the most frequently used form (around 70%) in both the centre and the periphery as well as in both periods. *Welke* is the second most frequent form in usage (around 30%). At least in the eighteenth-century data, the extended form *dewelke* still occurs equally marginal in the centre and the periphery with eleven and ten instances, respectively.

The increase of *welke* in the nineteenth-century results is slightly more pronounced in the periphery (increase from 28.0% to 33.2%) than in the centre (from 24.3% to 27.5%). Recall that the increase of neuter *hetwelke* was also stronger in the data from the periphery. Overall, both the distribution of variants and the diachronic tendencies are very similar, though.

Gender variation

Based on data from the two sub-corpora of handwritten ego-documents, Figure 6 presents the results of the distribution of variants across gender (M = male writers, F = female writers).

Figure 6. Distribution of masculine/feminine relative pronouns across gender and time.



The gender differences are certainly striking. In the late eighteenth century, male writers most frequently use the *d*-form *die*, which occurs in 71.0%. However, with a share of 25.9%, the additional pronominal form *welke* is also quite common in ego-documents by men. Moreover, two other (minor) variants can be attested: the extended form *dewelke* (2.7%) as well as three attestations of the *w*-form *wie* (0.4%). There is much less variation in the eighteenth-century data of female writers. In fact, *die* is used in 96.6% of all instances, leaving hardly any room for other variants. The only other form in the texts written by women is *welke* (3.4%), whereas there are no attestations of *wie* or *dewelke* at all.

In the early nineteenth century, the distribution of variants turns out to be rather stable in the data by male writers. The most frequent relative pronoun is *die*, occurring in 65.1%, which is a slight decrease as compared to the late eighteenth century. The share of *welke*, however, increases from 25.9% to 34.3%. In other words, male writers use the additional pronominal form in more than two-thirds of all instances. The use of the *w*-form *wie* is still very marginal (0.7%; four tokens), whereas *dewelke* no longer occurs. In ego-documents written by women, the *d*-form *die* maintains its dominant position with a still high share of 87.2%. However, like in the data by men, *welke* gains ground (from 3.4% to 12.8%), at the expense of *die*. Again, neither *wie* nor *dewelke* occur in the data by women.

Given the relevance of genre and the growing divergence between the two types of ego-documents in particular, I additionally cross-tabulated gender and genre, presented in Tables 3a (private letters) and 3b (diaries and travelogues).

Table 3a. Distribution of masculine/feminine relative pronouns across gender and time (private letters).

	<i>die</i>		<i>wie</i>		<i>welke</i>		<i>dewelke</i>		Total	
	N	%	N	%	N	%	N	%	N	%
Male-1	235	76.3	2	0.7	62	20.1	9	2.9	308	100
Male-2	184	81.8	3	1.3	38	16.9	0	0.0	225	100
Female-1	150	98.0	0	0.00	3	2.0	0	0.0	153	100
Female-2	219	91.6	0	0.0	20	8.4	0	0.0	239	100

Table 3b. Distribution of masculine/feminine relative pronouns across gender and time (diaries and travelogues).

	<i>die</i>		<i>wie</i>		<i>welke</i>		<i>dewelke</i>		Total	
	N	%	N	%	N	%	N	%	N	%
Male-1	291	67.2	1	0.2	130	30.0	11	2.5	433	100
Male-2	211	55.2	1	0.3	170	44.5	0	0.0	382	100
Female-1	51	92.7	0	0.0	4	7.3	0	0.0	55	100
Female-2	74	76.3	0	0.0	23	23.7	0	0.0	97	100

On closer inspection, it turns out that in private letters written by men, *welke* actually loses ground in period 2, slightly dropping from 20.1% to 16.9%, whereas the share of *die* slightly increases from 76.3% to 81.8%. This is surprising as it is against the general development of *welke* among male writers displayed in Figure 6. In the data contributed by male diarists, however, *welke* gains ground from 30.0% to 44.5%. Evidently, the relative stability in the overall distribution of variants used by men is, to a certain extent, based on the genre-specific developments in diaries and travelogues. In other words, the results from diaries and travelogues are crucial for the attested gender variation.

Another interesting finding is the development of *welke* in the private letters written by women, increasing from a very marginal 2.0% to 8.4%. While this is generally in line with the diachronic rise of *welke*, it also shows that there is gender variation on a small scale, as the same variant declines in the equivalent data from their male contemporaries. Even more pronounced is the increase of *welke* in diaries and travelogues written by women, from 7.3% to 23.7%.

2 Discussion

The case study in this chapter focused on the masculine and feminine singular (present-day common gender) and plural relative pronoun in subject and object

position. Four forms were investigated: (1) the originally demonstrative *d*-form *die*, (2) the originally interrogative *w*-form *wie*, (3) the additional pronominal form *welke*, and (4) the extended alternative *dewelke*. All variants were mentioned and discussed in Weiland's (1805: 119-120) grammar and also occurred in the *Going Dutch Corpus*.

After giving a general overview of the possible forms in Section 1.1, illustrating their usage in different contexts of definiteness as well as with animate and/or inanimate antecedents, variation and change in the distribution of relative pronouns was investigated quantitatively. In the corpus analysis in Section 1.2, it was shown that out of the four possible variants, only two were actually frequent in usage, viz. *die* and, to a lesser extent, *welke*. Both the *w*-form *wie* and the extended form *dewelke* turned out to be very marginal variants in the corpus data.

Diachronically, the distribution of masculine and feminine relative pronouns was remarkably stable, which is in line with the findings of the neuter forms in Chapter 10. Two tendencies were noticeable, though. First, *dewelke* was only attested as a minor variant in the late eighteenth century, but no longer occurred in the early nineteenth-century data. Secondly, *welke* appeared to gain ground in the second period. A similar increase was attested for the neuter form *hetwelke*.

Focusing on genre variation, the results in this chapter revealed great differences between the investigated sources. Most notably, private letters showed a strong preference for the *d*-form *die*, whereas both *die* and *welke* co-occurred as two competing main variants in newspapers. As already seen in the analysis of the neuter relative pronoun, diaries and travelogues take a special intermediate position in the multi-genre corpus design between private letters on the one hand, and newspapers on the other. While *die* lost ground in nineteenth-century diaries and travelogues, the use of *welke* considerably increased, converging towards the proportion also found in newspapers. This indicates a genre-specific change with regard to the use of relative pronouns, diverging from the more oral private letters in period 1, towards the more written style of newspapers in period 2.

Geographical variation in the distribution of forms was comparatively limited. Although there was some degree of variation across the seven investigated regions (e.g. the surprising decrease of *die* in nineteenth-century diaries from North Holland), clear patterns could not be detected. Similarly, variation between the centre and the periphery was marginal.

On the contrary, gender proved to be a robust social variable once more. Particularly in the eighteenth-century period, the distribution of variants across genders showed evidence of striking differences. Whereas male writers used both *die* and *welke*, female writers used *die* in practically all instances. Furthermore, the extended form *dewelke* only occurred in ego-documents written by men. In the nineteenth-century period, the gender differences somewhat converged, but still remained visible in the results. One can argue that women preferably used the more common form *die*, whereas men also used the more formal options *welke* and (at least in period 1) *dewelke* along *die*.

Finally, in addition to the external variables, one internal variable was taken into consideration. Since the animacy of the antecedent has often been

considered as a relevant factor affecting the distribution of relative pronouns, this general assumption was tested on data from the sub-corpora of diaries and travelogues, and newspapers. However, the influence of this semantic property appeared to be fairly limited. In fact, the two competing main variants *die* and *welke* equally referred to both animate and inanimate antecedents, even though it was suggested that *welke* in the nineteenth-century period was slightly more likely to refer to an inanimate antecedents than *die*. Despite the very few number of attestations of *wie* in subject and *wien* in object position, a tentative conclusion could be drawn as all (seven) occurrences of the *w*-form had an animate antecedent.

In sum, the two strongest factors affecting the choice and distribution of variants were of external nature. First and foremost, genre proved to have the most crucial effect again. Second, gender should not be underestimated either, as the corpus analysis revealed clearly distinct patterns, particularly in the eighteenth-century period.

The effectiveness of Weiland's (1805) norms on actual language usage is more difficult to assess. Although Weiland provided elaborate information on the use of the different forms, particularly compared to his eighteenth-century predecessors, his direct influence seems to be limited. Either directly or indirectly reflected in the corpus results was Weiland's awareness of stylistic or register variation with regard to the use of (plain) *die* as opposed to (solemn) (*de*)*welke*. Evidently, the different text sources in the corpus had genre-specific preferences for particular forms – already visible in the late eighteenth century. In particular *welke* and *denwelke* appeared to signal a higher formality than the more common *die*. This was confirmed by the corpus results (for both diachronic cross-sections) as well as by Weiland's (1805) remarks on stylistic variation.

Another aspect possibly indicating Weiland's influence concerns the use of *wie* as free relatives (i.e. context I of the definiteness cline). According to Weiland (1805: 120), only *w*-forms (masc./fem. *wie*, also neuter *wat*) can function as free relatives. Although the definiteness of the antecedent was not investigated quantitatively here, instances of nominative *wie* or accusative *wien* as free relatives could not be attested. In contrast, examples of the *d*-form *die* functioning as free relatives were easily found, both in the eighteenth- and the nineteenth-century period. It should be noted, though, that Weiland's (1805: 120) example *wien ik mijn word geef* illustrates the use of dative *wien*, whereas the analysis presented here focused on subject and object position only. This supports earlier assumptions that "the relative pronoun *wie* only occurs in *casus obliqui* and after prepositions" (van der Wal 2002b: 31). Indeed, the *Going Dutch Corpus* data gives sporadic evidence of dative *wien* as a free relative. With regard to the generally low frequency one can probably conclude that the *w*-form *wie* as a free relative must have been in its incipient stage in the investigated time period, and that Weiland's influence did not considerably enhance this development.

3 Relative pronouns: General conclusion

Chapters 10 and 11 presented two closely related case studies of variation and change the Dutch relativisation system, investigating both neuter relative pronouns (Chapter 10) as well as the masculine and feminine singular and plural relative pronouns (this chapter) in subject and object position. With regard to the focus on the possible effectiveness of language norms on actual language usage, the discussions in Weiland's (1805) grammar and the preceding eighteenth-century normative tradition were also taken into consideration, as outlined in Sections 2 and 3 of the previous chapter.

Combining the corpus results of two comparable case studies on Dutch relative pronouns, some more general conclusions can be drawn. First of all, the diachronic stability of the results was one of the most remarkable findings in both cases. On the one hand, it might be argued that this is due to the relatively limited time frame under investigation, certainly with regard to the slow and gradual changes in the relativisation system. On the other hand, the elaborateness of Weiland's grammar with respect to relative pronouns, providing more explicit norms and rules, might also be regarded as a turning point in the (normative) history of Dutch relativisation, possibly affecting the distribution of variants in the second period.

Apart from the diachronic stability attested in both case studies, another similarity concerns the relevance of external factors conditioning language usage patterns. In both case studies, genre proved to be the most crucial variable. The additional pronominal forms *betwelke* (for neuter) and *welke* (for masculine/feminine) were particularly strong in newspapers, whereas they were considerably less frequent in private letters. In these sources, the traditional *d*-forms *dat* (for neuter, alongside *wat*) and *die* (for masculine/feminine) were relatively frequent. Strikingly, the third genre, i.e. diaries and travelogues, took a special intermediate position in both case studies, especially from a diachronic point of view. For both neuter and masculine/feminine relative pronouns, a genre-specific evolution in diaries and travelogues could be testified. With respect to the distribution of relative pronouns and the additional forms *betwelke* and *welke* in particular, diaries and travelogues were more similar to private letters in the late eighteenth-century period. As a result of the remarkable increase of *betwelke* and *welke* (at the expense of the *d*-forms *dat* and *die*), respectively, diaries and travelogues from the early nineteenth century diverged from the other type of handwritten ego-documents, converging towards a more 'written', formal style of the printed and published texts in the corpus.

Moreover, gender variation was attested in both case studies. For neuter as well as for masculine and feminine relative pronouns, the results revealed the tendency that men used the more formal alternative forms *betwelke* and *welke* more frequently than women, who, in turn, used the common forms *dat* (and also *wat*) and *die* relatively more often.

When compared to genre and gender, geographical variation was less pronounced. Although some interesting tendencies were shown, for instance the north–south differences with respect to the additional neuter forms *betwelke* and

betgeen, clear patterns did not emerge. It appears that space did not play a major role in the investigated time period, at least with respect to relativisation.

For each of the two case studies, one internal factor was taken into account and tested on the *Going Dutch Corpus* data. While both the definiteness of the antecedent (neuter relative pronouns) and the animacy of the antecedent (masculine/feminine relative pronouns) have traditionally been considered as decisive factors conditioning variation in the use of relative pronouns, the corpus-based results presented here did not support their supposed relevance. In the case of the definiteness of the antecedent, it is true that specific forms were more dominant in some contexts (e.g. *wat* as free relative and in combination with *al(les)*). At the same time, it was also shown that practically all variants occurred in (almost) all contexts. Given this highly variable situation, van der Horst's (1988) generalisation that originally interrogative *w*-relativisers replace the originally demonstrative *d*-forms successively from the most indefinite to the most definite contexts, could not be confirmed without certain limitations. Similarly, the second internal factor, the animacy of the antecedent, did not overly affect the choice of forms, as both main variants of the masculine/feminine relative pronouns (*die*, *welke*) occurred frequently with both animate and inanimate antecedents. However, one could notice that *welke* gained ground with inanimate antecedents in the nineteenth-century data, which might indicate that the general increase of *welke* was, to some extent, internally conditioned. Despite the limited number of occurrences in the corpus, the *w*-form seemed to be restricted to animate referents. In general, the investigated external variables, particularly genre, turned to be considerably more relevant than the internal factors, though.

One of the striking differences between the two case studies on relative pronouns concerns the number of co-occurring or competing forms in usage. In the case of the neuter relative pronoun, at least four forms (*dat*, *wat*, *betgeen*, *betwelk*) were frequently used, as opposed to only two main variants of the masculine/feminine relative pronoun (*die*, *welke*). At the same time, both case studies highlighted the presence and relevance of the so-called alternative or additional pronominal forms like *betgeen*, *betwelk* and *welke*. These were strong competitors of the traditional *d*-forms and gradually rising *w*-forms, possibly even delaying the change from *d*- to *w*-forms. Therefore, the history of relativisation has to be studied by taking into account its full inventory of contemporary forms. In fact, variation and change in the Dutch relativisation system can hardly be reduced to the major shift from *d*- to *w*-relativisation. In line with previous findings on the chronology of *d*- to *w*-changes, it is evident that developments *dat* > *wat* and *die* > *wie* were still in their incipient stages in the investigated time period, in any case in subject and object position.

Competition between synonymous forms could also be witnessed with regard to the alternative pronominal forms derived from the interrogative pronoun *welk*, i.e. *welk* and the extended variant *betwelk* for neuter, as well as *welke* and the extended variant *dewelke*. In each pair, only one form emerged as the 'winning' form. In the case of the neuter relative pronoun, the extended form *betwelk* was considerably more common than *welk*, which was only a marginal variant in the

corpus data. The opposite was true in the case of the masculine/feminine pronoun. Here, the shorter form *welke* appeared to be much more frequent, whereas the extended option *denwelke* was rare in the late eighteenth century and even disappeared in nineteenth-century usage.

Finally, the normative influence on actual language practice needs to be assessed. For the eighteenth-century period, any normative influence on the distribution of neuter and masculine/feminine relative pronouns can probably be excluded. For the most part, metalinguistic comments and normative publications from this period did not provide explicit rules or guidance for the use of relative pronouns. A frequently recurring exception concerned the neuter pronoun *hetgeen*, commonly restricted to its use as free relative and rejected in other contexts, for instance with nominal antecedents. In eighteenth-century usage, however, *hetgeen* occurred in almost all contexts, indicating that normative influence must have been limited at the most.

The situation of only sporadic norms on relativisation changed with the introduction of Weiland's (1805) national grammar, which provided relatively elaborate information and norms, either explicitly through rules, or more implicitly deduced from his numerous examples. However, even in the post-Weiland period, the normative influence on actual language usage is difficult to assess, particularly in comparison with the obvious influence of Siegenbeek's (1804) prescriptions attested in the analyses of orthographic issues. First of all, the corpus results revealed that most of Weiland's (1805) preferences for specific forms in specific contexts were already found in late eighteenth-century usage, when his influence was ontologically impossible. With respect to the diachronic stability of the results in the early nineteenth-century, one might argue that Weiland's (1805) discussion on the use of relative pronouns primarily contained accurate observations of contemporary usage patterns rather than ground-breaking prescriptions which affected subsequent language usage. Moreover, although Weiland's (1805) preferred (and dispreferred) specific forms of relative pronouns in different contexts, much more options occurred alongside these forms in both periods under investigation. This also suggests that the normative influence probably must have been fairly limited.

However, with regard to Weiland's (1805) remarks on stylistic or register variation, assigning forms like *welke* and *denwelke* to the 'solemn' style as opposed to *die* and *dat* typical of the 'plain' style, either direct or indirect influence can be assumed. In fact, the corpus results revealed an increase of the 'solemn' forms in diaries and travelogues and (at least in the case of *hetwelke*) newspapers, whereas the 'plain' forms gained ground in private letters. Although it cannot be proved that it directly reflects Weiland's (1805) intervention, he might have been a factor that consolidated the status and general perception of *hetwelke* and *(de)welke* as more formal variants on the one hand, and *dat* and *die* as forms of more common writing on the other hand.

In future research on relative pronouns, other positions than the subject and object positions investigated here, need to be taken into account. It has been argued that *w*-forms probably enter the language via more specific positions such as

genitival constructions, and subsequently spread to more frequently relativised positions such as subjects and objects (Romaine 2009: 151-152, cf. also van der Wal 2002b). A historical-sociolinguistic and corpus-based approach would shed more light on these general assumptions.

Furthermore, while the case studies in Chapters 10 and 11 focused on relative pronouns, it might also be interesting to reassess the use of relative adverbs and relative pronominal adverbs with data from the *Going Dutch Corpus*, compared to earlier findings in seventeenth- and eighteenth-century private letters (Rutten & van der Wal 2014: ch. 8). Moreover, Weiland (1805) also commented on these adverbial relativisers, which means that his possible normative influence on language practice could also be evaluated with respect to this aspect of relativisation.

Morphosyntactic variables (3)

Genitive case

1 The genitive case in Dutch

Dutch, like other Germanic languages, originally had a fully-fledged morphological case system with different case inflections for nominals and adnominals, depending on their function. For the Old and Middle Dutch periods, four cases are commonly distinguished, viz. the nominative, the genitive, the dative, and the accusative (e.g. van der Wal & van Bree 2008: 132-135). These cases were further distinguished by gender (i.e. masculine, feminine, neuter) and number (i.e. singular, plural). The genitive performed a number of functions, for instance, indicating possession in a broad sense, as well as partitive and temporal relations (van der Horst 2008: 148-150). Moreover, the inflected genitive case occurred both prenominally (*des konings zoon* ‘the king’s son’) and postnominally (*de zoon des konings* ‘the son of the king’).

Nevertheless, the case system had already started to decline as early as the Middle Dutch period. Similar to the developments in other Germanic languages, the increasing loss of inflection, commonly referred to as *deflection*, led to a situation in which historical synthetic forms competed with alternative periphrastic constructions. In the case of the genitive case, the inflected forms were gradually replaced by an analytical construction with the preposition *van* ‘of’ (*de zoon van de(n) koning*). According to Scott (2014: 107), “the *van*-construction had become constructionalised as an alternative to the adnominal genitive” by the Middle Dutch period. In present-day Dutch, remnants of the genitive case are for the most part restricted to fixed and fossilised expressions (e.g. *de tand des tijds* ‘the ravages of time’), or to archaic and formal language (e.g. *het pad des levens* ‘the path of life’) (cf. ANS 3.4.1).

In the seventeenth, eighteenth and nineteenth centuries, the synthetic genitive still occurred in written texts, first and foremost in the higher registers. It is generally assumed that the fully-fledged case system had no longer been used in spoken Dutch by that time (van der Horst 2008: 1074). However, in the context of language cultivation, aiming to remodel and elaborate the Dutch language based on the Latin ideal, grammarians of the sixteenth and seventeenth centuries continued to promote the use of four or even six cases, adding the vocative and the ablative to the four traditional cases. The latter was presented, for example, in the first printed grammar of Dutch, the *Twe-spraack vande Nederduitsche letterkunst* of 1584 (cf. e.g. van der Bree & van der Wal 2008: 191-195). These attempts to revive the old case markers created a growing discrepancy between language norms and language usage. Nominal inflection remained a widely discussed topic, though, both in

metalinguistic discourse in the eighteenth century (cf. Section 3) as well as in Weiland's (1805) national grammar of Dutch (cf. Section 2).

2 Discussion in Weiland (1805)

Concerning the decline of the Dutch case system and the rise of alternative (prepositional) constructions replacing the historical synthetic forms, the treatment of the genitive case in Weiland's (1805) national grammar exemplifies the gap between language norms and language practice.

As outlined in Section 1, the synthetic genitive case was primarily used in the written language and especially in higher registers, while it had practically disappeared from spoken language. Given the fact that the national grammar was aimed at the whole Dutch population, including schoolchildren, and generally had a strong educational focus (Rutten 2016: 124), one might expect an increasing attention given to the analytical *van*-construction, which had largely replaced the synthetic genitive case in colloquial Dutch. Remarkably, Weiland's (1805) national grammar strongly promoted the synthetic forms of the case system, including the genitive case.

To begin with, Weiland (1805: 75) provided a concise definition of the genitive case in his general discussion of the traditional Dutch case inflections:

De *tweede* naamval is die, welke de betrekkingen der zelfstandige naamwoorden op elkander aanwijst, en de zelfstandige naamwoorden zamen voegt. Zoo menigvuldig nu de gesteldheden der dingen en derzelver betrekkingen op elkander zijn, zoo menigvuldig zijn ook de gevallen, waarin een zelfstandig naamwoord den *tweeden* naamval moet aannemen.

'The *second* case is that case which indicates the relations of nouns with each other, and which joins the nouns. As manifold as the conditions of the things and their relations with each other are, as manifold are also the cases, in which a noun must take the *second* case.'

Additionally, Weiland (1805: 75) illustrated the various roles the genitive case could perform, for instance in *het werk mijner handen* 'the work of my hands' and *Gods geboden* 'God's commandments' (agent), *de heer des huizes* 'the master of the house' and *de bezitter eens grooten vermogens* 'the owner of a great fortune' (possession), *de zeden onzer eeuw* 'the morals of our century' and *de aangenaamheid dezer landboeve* 'the comfort of this farm' (time and place), and so on. Although these examples already indicate Weiland's preference for the synthetic genitive case, it was in the paradigms where his officialised choices were explicitly presented.

Table 1 provides an outline of the prescribed forms of various genitive markers offered in the paradigms in Weiland (1805).

Table 1. Prescribed forms of various genitive markers (singular/plural) in the paradigms by Weiland (1805).

	Genitive singular		
	Masculine	Feminine	Neuter
Definite art.	<i>des</i>	<i>der</i>	<i>des</i>
Indefinite art.	<i>eens</i>	<i>eener</i>	<i>eens</i>
Demonstrative pr.	<i>dezes</i>	<i>dezer</i>	<i>dezes</i>
Possessive pr.	<i>mijns</i>	<i>mijner</i>	<i>mijns</i>
	Genitive plural		
	Masculine	Feminine	Neuter
Definite art.	<i>der</i>	<i>der</i>	<i>der</i>
Indefinite art.	—	—	—
Demonstrative pr.	<i>dezer</i>	<i>dezer</i>	<i>dezer</i>
Possessive pr.	<i>mijner</i>	<i>mijner</i>	<i>mijner</i>

Systematically, Weiland (1805) only mentioned the synthetic forms of the genitive case in his paradigms. These included both the definite and indefinite article, forms of the demonstrative pronouns (as illustrated by *deze*) and the possessive pronouns (as illustrated by *mijn*). At least in the running text, Weiland (1805: 76) mentioned the analytical construction with the preposition *van* as a frequently used alternative for the genitive case:

In plaats van den *tweeden* naamval, bedient men zich dikwerf van het voorzetsel *van*, en zegt: *eene teekening van Rubbens, een lierzang van Klopstock, de keizerin van Rusland, de psalmen van David* enz., terwijl *een lierzang Klopstocks, de keizerin Ruslands* enz.; in den gewonen schrijfstijl, buiten gebruik, doch in poezij, misschien, te dulden is. In den verheven stijl, zegt men *Davids psalmen* enz.⁷⁸

In place of the *second* case, one often uses the preposition *van*, and says: *eene teekening van Rubbens, een lierzang van Klopstock, de keizerin van Rusland, de psalmen van David* etc., while *een lierzang Klopstocks, de keizerin Ruslands* etc. are obsolete in the common writing style, but perhaps tolerable in poetry. In the elevated style, one says *Davids psalmen* etc.⁷

Notably, Weiland's examples to illustrate the use of the *van*-construction only have proper names (*David, Rusland*) as possessors, which typically lack an adnominal such as articles or pronouns.

⁷⁸ Here, Weiland also commented on stylistic variation in the case of yet another alternative genitival construction, viz. the possessive *s*-construction (cf. also Rutten 2016: 130-131).

Furthermore, Weiland (1805: 76) identified three cases in which the *van*-construction is the only option. These contexts refer to (1) a specific origin of a person, (2) the material of an object, or (3) the age, size, weight or value:

In sommige gevallen wordt het voorzetsel *van* altijd gebruikt, bij voorbeeld, 1. wanneer geslacht, afkomst en Vaderland aangewezen worden: *een mensch van geringe afkomst, een Amsterdammer van geboorte* enz.; 2. wanneer de stof genoemd wordt, waaruit iets gemaakt is: *eene doos van zilver, een ring van goud* enz.; 3. wanneer ouderdom, grootte, gewigt en waarde bepaald worden: *een kind van twee jaren, een ton van twintig emmeren, een man van groote verdiensten* enz.

In some cases the preposition *van* is always used, for example, 1. when gender, descent and native country are indicated: *een mensch van geringe afkomst, een Amsterdammer van geboorte* etc.; 2. when the material, of which something is made, is mentioned: *eene doos van zilver, een ring van goud* etc.; 3. when age, size, weight and value are defined: *een kind van twee jaren, een ton van twintig emmeren, een man van groote verdiensten* etc.⁷

It is striking that in all three contexts illustrated here, an article is missing. Rutten (2016e: 131) explains that “since inflectional case in Dutch is primarily expressed on adnominals, typically on the article or a pronoun, they therefore disfavor the synthetic genitive”. It seems that Weiland did not take into account the crucial role of the missing article. The way in which Weiland’s acknowledged the existence of the alternative construction with *van* is thus very limited and highlights the preference for the synthetic genitive case even more.

Weiland’s (1805) distinct choice in favour of the historical genitive suggests that the Dutch population at large was supposed to learn the synthetic forms, although they had largely disappeared from colloquial language practice. One might assume that this choice further intensified the discrepancy between language norms and actual language practice, which will be investigated empirically with data from the *Going Dutch Corpus* in Section 5.2. But first, Section 3 will shed light on the developments in the normative discussion of the eighteenth century, in order to assess the position of Weiland (1805) against the background of the preceding metalinguistic discourse.

3 Eighteenth-century normative discussion

Throughout the eighteenth century, morphology and nominal inflection in particular were among the core topics, if not the most important, in metalinguistic discourse (Simons & Rutten 2014: 69). Rutten (2009a, 2016e) focuses on the broad developments in the eighteenth-century normative tradition, suggesting a shift from elitist (1700–1740) to ‘civil’ (1740–1770) to national grammar (from 1770 onward; cf. also Chapter 2) – changes which are also reflected in the treatment of the genitive case.

In the beginning of the eighteenth century, i.e. the period of elitist grammar, metalinguistic texts more or less exclusively presented the synthetic

forms of the genitive case. In van Hoogstraten's (1700) preface, we only find inflectional case forms, for instance for the definite article *des* (m.), *der* (f.), *des* (n.), *der* (all gender plur.), and for the indefinite article *eens/eenes* (m.), *eener* (f.), *eens* (n.).

Similarly, Moonen (1706: 46-47) only mentioned the synthetic forms for the definite article, i.e. *des* (m.), *der* (f.), *des* (n.), *der* (all gender plur.), as well as for the indefinite article, i.e. *eens* (m.), *eener* (f.), *eens* (n.). On closer inspection, however, one can find a few occasional exceptions with analytical *van*-forms as options for possessives, i.e. *van onzen* or *onzes* (m.) and *van ons* or *onzes* (n.), but only *onzer* (f., all gender plur.). Furthermore, Moonen (1706: 84-85) also mentioned the *van*-construction in the running text as a frequently used alternative:

De Teeler is de tweede Naemval [...] Dees draegt tot Merktekens in het Eenvouwige Getal *Eens, Eener, Des, Der*; als in *Eens Mans, Des Mans, Eens Kints, Eener Vrouwe, Der Vrouwe*; en in het Meervouwige *Der*; als in *Der Mannen, Vrouwen, Kinderen*.

Waer voor veeltyts *Van Eenens, van Den, van Het* gebruikt wordt; als, *De sleutel van den hof*, voor *De sleutel des hofs*, *het hooft van eenen walvisch*, voor *het hooft eens walvisches*, *het welvaeren van de stadt*, voor *het welvaeren der stadt*, *de rug van het paert*, voor *de rug des paerts*.

Voornaemelyk als de Noemer van het manlyke en Onzydige Geslachte in E en S eindigt; als *Van den wille, van den vrede, van het einde, van het huis, van het vlas, van den kroes, van het mes, van den glans*; die alle de S zeer noode achter zich willen lyden

'The genitive is the second case [...] In the singular, its markers are *Eens, Eener, Des, Der*, as in *Eens Mans, Des Mans, Eens Kints, Eener Vrouwe, Der Vrouwe*, and in the plural *Der*, as in *Der Mannen, Vrouwen, Kinderen*.

Oftentimes, *Van Eenens, van Den, van Het* are used for this, as *De sleutel van den hof*, for *De sleutel des hofs*, *het hooft van eenen walvisch*, for *het hooft eens walvisches*, *het welvaeren van de stadt*, for *het welvaeren der stadt*, *de rug van het paert*, for *de rug des paerts*.

Particularly if the nominative of the masculine and neuter gender ends in E and S, as *Van den wille, van den vrede, van het einde, van het huis, van het vlas, van den kroes, van het mes, van den glans*, all of which can hardly bear the S behind them'

Moonen argued that the use of the *van*-construction was phonologically motivated, mainly when the possessor is a masculine or neuter noun ending in *e* or *s*.

Verwer (1707) generally listed synthetic forms only, except for some variation attested in the neuter paradigms, viz. *des* or *van den* in the singular, and *der* or *van de* in the plural.

In contrast to his contemporary grammarians, Sewel (1708: 38-39) presented the historical genitive and the *van*-construction as two options. At least in the case of the definite article, Sewel listed both synthetic and analytical forms, e.g. *des* or *van den* (m.), *der* or *van de* (f.), *des* or *van het* (n.), but only *der* for the plural. In contrast, for the indefinite article we only find the synthetic genitive forms, i.e. *eens* (m./n.) and *eener* (f.). In the paradigms for possessive pronouns, Sewel (1708: 122-123) also showed some variation, presenting, for example, the synthetic genitive forms *myns* (m.) and *myner* (f.), but only analytical *van myn* (n.), and both *myner* and *van myne* (all gender plur.).

A different eighteenth-century approach, with a strong focus on stylistic variation, can be found in ten Kate (1723; cf. also Noordegraaf & van der Wal 2002). As summarised by Rutten (2016: 127), ten Kate interpreted “the difference between the usage of case endings in higher registers such as formal and literary language, and the lack of case endings in other registers as a matter of style”. In the case of the genitive, he assigned the synthetic and analytical genitive forms to three stylistic levels, viz. the sublime (*hoogdravend*), the solemn (*deftig*) and the plain (*gemeenzæm*) style. As can be seen in ten Kate’s (1723 I: 337) illustration of the indefinite article, exemplified by the feminine singular forms, the synthetic genitive was assigned to the sublime style, whereas the analytical *van*-construction occurred in the solemn and plain styles:

Hoogdr⁷⁹: EENER GROOTEN of GROOTER VROUWE.
 Deft: VAN EENE GROOTE VROUWE (of ingetr: VROUW’).
 Gem: VAN EEN’ GROOTE VROUW.

The same stylistic variation was also presented in other paradigms, for instance for the definite article and the possessive pronoun, again in the feminine singular (ten Kate 1723 I: 339):

Hoogdr: DER (en MYNER) GROOTEN of GROOTER VROUWE, of VAN DE (en MYNE) GROOTE VROUWE.
 Deft: VAN DE (en MYNE of MYN’) GROOTE VROUW’.
 Gem: VAN DE (en MYN) GROOTE VROUW.

Notably, ten Kate also mentioned the analytical *van*-construction in the elevated style, as an alternative to the synthetic genitive. Generally, though, it can be summarised that the lower the style or register, the less case marking we find in ten Kate’s (1723) paradigms.

From van Belle’s (1748: 42ff.) paradigms illustrating the inflection of nouns, one can deduce a clear preference for synthetic forms, which he mentioned with regard to the genitive: *des Heeren (Wet)*, *der kwaade Vrouwe (Ontteering)*, *des eersten Kinds (Bloedstorting)*. Interestingly, only examples of the prenominal genitive were used here. Van Belle (1755: 23ff.), referring to the genitive case as the *Afdaaler*, only presented synthetic forms, while the *van*-construction remained unmentioned:

De *Afdaaler* dient, om eene Zaak voor de stellen, zo als dezelve tót iets óf iemant, als afdaalende, voorkomt; en wordt betekend in de Lédekens *Eenes, Eener, Des*, en *Der*, en in de Naam- óf Voornaamwoorden en Eigene Naamen, door agteraanvoeginge van *e, en, r*, en *s*. Als:
De Sterke EENES ManS
’S MensEN Zoon.
De Zwakheid EENER VrouwE.
De Barmhertigheid DES KoningS.

⁷⁹ These abbreviations refer to the three stylistic levels distinguished by ten Kate (1723), i.e. sublime (*Hoogdr* = *hoogdravend*), solemn (*Deft* = *deftig*) and plain (*Gem* = *gemeenzæm*).

WillemS Stoutheid.
De Gunst DER KoninginNE.
MynER VrouwE Moeder.

‘The genitive serves to introduce a thing (object), as it occurs in relation to, or descends from, something or somebody. It is denoted by the articles *Eenes*, *Eener*, *Des*, and *Der*, and in the nouns or pronouns and proper names through suffixation of *e*, *en*, *r* and *s*. [...]’

From around the mid-1800s onward, i.e. the period of so-called ‘civil’ grammar, metalinguistic texts were characterised by “the plain efforts to rephrase the relatively complex grammatical descriptions of the earlier period of elitist grammar in simpler terms, and thus to make knowledge of the grammar of Dutch accessible to a larger audience” (Rutten 2016e: 125). With regard to the genitive case, the analytical *van*-construction was consistently presented as an alternative of the synthetic case forms in the paradigms of normative grammars.

A good example of this widened approach is the grammar by Elzevier (1761: 52-53), in which both analytical and synthetic forms were mentioned and illustrated:

een persoon of zaek die een’ wezentlyken eigendom heeft aen personen of zaken die by hem gestelt worden, dus zegt men: *de lof* DES *mans*, of *van den man*, *is groot*, *de lof* DER *vrouw*, of *van de vrouw*, *is groot*, *de lof* DES *konstgenootschaps*, of *van het konstgenootschap*, *is groot*. Hier blykt nu klaer dat *man*, *vrouw* en *konstgenootschap* hier in de tweeden Naemvall’ staen, of de *Eigenaers* zyn van den lof.

‘a person or thing which has an essential possession of persons or things, thus one says *de lof* DES *mans*, or *van den man*, *is groot*, *de lof* DER *vrouw*, or *van de vrouw*, *is groot*, *de lof* DES *konstgenootschaps*, or *van het konstgenootschap*, *is groot*. Here it becomes obvious that *man*, *vrouw* and *konstgenootschap* are in the second case here, or are the possessors of the praise (*lof*)’.

In Elzevier’s (1761: 60-61) paradigm for the definite article, the options with *van* were listed first, followed by the synthetic forms, i.e. *van den* or *des* (m.), *van de* or *der* (f.), *van het* or *den* (n.), *van de* or *der* (all gender plur.). The same is true for the indefinite pronoun, i.e. *van eenen* or *eens* (m.), *van eene* or *eener* (f.), *van een* or *eens* (n.). The order of mention might suggest a preference for the analytical *van*-construction. Nevertheless, by taking into account other possible genitive markers, one can only find inflected forms in Elzevier’s paradigms for possessive pronouns. Still, there were some exceptions, e.g. only *van onzen*, *van ons* (no synthetic option), signalling variation in the paradigms.

Similar to Elzevier’s approach, van der Palm (1769: 10-12) considered both synthetic and analytical options:

Vr. Welke zyn de merkteekens van dezen Naemval?

Antw. In het enkelvoudige, in het mannelyke en onzydige geslacht, heeft hy tot merkteekens de lidwoordjes *Eens* en *Des*, als *eens mans*, *des mans*, *eens kinds*, en in het

vrouwelyke *eener* en *der*, als *eener vrouwe*, *der dochter*, in het meervouwige in alle geslachten *der*, als *der mannen*, *der vrouwen*, *der kinderen*.

Vr. Gebruikt men somtyds ook niet *van eenen*, *van den*, en *van het*?

Ant. Somtyds bedient men zich van de woorden, *van eenen*, *van den*, *van het*; hebbende dit voornamelyk plaets, wanneer de Naemwoorden van het mannelyke en onzydige geslagt in *E*, *S* of *Sch* eindigen; als *van den wille*, *van het huis*, *van het vleesch*; want dan geschiedt dit, om het wangeluidt, dat de byvoeging van *eene S* in de woorden geeft, voortekomen.

Q: Which are the markers of this case?

A: In the singular, in the masculine and neuter gender, its markers are the articles *Eens* and *Des*, as in *eens mans*, *des mans*, *eens kinds*, and in the feminine gender *eener* and *der*, as in *eener vrouwe*, *der dochter*, in the plural of all gender *der*, as *der mannen*, *der vrouwen*, *der kinderen*.

Q: Doesn't one use *van eenen*, *van den*, and *van het* sometimes?

A: Sometimes one uses the words *van eenen*, *van den*, *van het*. This primarily occurs when the nouns of the masculine and neuter gender end in *E*, *S* or *Sch*, as *van den wille*, *van het huis*, *van het vleesch*; because then is occurs to prevent the cacophony, that the addition of an *S* in these words brings.'

Moreover, in van der Palm's paradigms, for instance for the definite article (1769: 22-23), both synthetic and analytical forms were presented: *des* or *van den* (m.), *der* or *van de* (f.), *des* or *van het* (n.), *der* or *van de* (all gender plur.). Likewise, van der Palm (1769: 24) listed *eens* or *van eenen* (m.), *eener* or *van eene* (f.), *eens* or *van een* (n.) as options for the indefinite article in the genitive case. The order possibly indicates a preference for the historical synthetic forms, though. The use of *van* was phonologically motivated, mainly assigned to occurrences in which masculine or neuter nouns end in *e*, *s* or *sch*, which is probably a reference to Moonen (1706).

An extensive set of paradigms is included in de Haes (1764: 21ff.), illustrating the inflections of definite and indefinite articles, demonstrative pronouns, possessive determiners and many more. He was rather consistent in presenting both the synthetic (mentioned first throughout) and analytical forms, e.g. *des* or *van den* (m.), *der* or *van de* (f.), *des* or *van het* (n.), *der* or *van de* (all gender plur.), and *eens* or *van eenen* (m.), *eener* or *van eene* (f.), *eens* or *van een* (n.).

Tollius (1776: 46-47) consistently mentioned both synthetic and analytical forms of the definite article, i.e. *des* or *van den* (m.), *der* or *van de* (f.), *des* or *van het/van den* (n.), *der* or *van de* (all gender plur.), and the indefinite article, i.e. *enes* or *van enen* (m.), *ener* or *van ene* (f.), *enes* or *van een* (n.). Stijl & van Bolhuis (1776: 79-80) also presented both synthetic and analytical options, for example of the definite article, i.e. *des* or *van den* (m.), *der* or *van de* (f.), *van het* or *des* (n.; analytical first!), all gender plural *der* or *van de*, and the indefinite article, i.e. *eenes* (*eens*) or *van eenen* (m.), *eener* or *van een* (*eene*) (f.), *eens* or *van een* (n.).

A somewhat more conservative approach can be found in Zeydelaar (1781: 140), who mainly presented synthetic forms only, for instance of the definite article, i.e. *des* (m.), *der* (f.), *des* (n.), *der* (pl.), and the indefinite article, i.e. *eens* (m.), *eener* (f.) and *eens* (n.). We do find the analytical forms presented alongside the synthetic forms in a few neuter paradigms, e.g. *van een klein Kind* or *eens kleinen Kinds*,

and *van de* or *der kleine Kinderen*, in the case of possessives *van mijn* (n.), but *mijns* (m.), *mijner* (f.), *mijner* or *van mijne* (all gender plur.).

The final decades of the eighteenth century witnessed another development in metalinguistic discourse. In the context of the debate about a national grammar of Dutch, grammatical descriptions usually preferred the analytical *van*-construction, and in some cases even no longer took into account the synthetic forms in the paradigms (Rutten 2016e: 125–126).

Van Bolhuis (1793: 24–28) illustrated the genitive case by an example with the preposition *van*, viz. *de brand van dat huis*. The restriction to analytical forms is also reflected in the paradigms for the definite and indefinite articles: *van den* (m.), *van de* (f.), *van het* (n.), *van de* (all gender plur.), and *van eenen* (m.), *van ene* (f.), *van een* (n.), respectively. Interestingly, the inflected genitive forms occur alongside the *van*-construction in the paradigms of possessive determiners (van Bolhuis 1793: 43), i.e. *mijns* or *van mijnen* (m.), *mijner* or *van mijne* (f.), *mijns* or *van mijn* (n.), *mijner* or *van mijne* (all gender plur.), as well as in the paradigms of demonstrative pronouns (van Bolhuis 1793: 46): *dezes* or *van dezen* (m.), *dezer* or *van deze* (f.), and so forth. The complete restriction to analytical forms was thus not extended to all possible genitive markers.

Just before the turn of the century, the *Rudimenta* (1799: 6–7) presented both synthetic and analytical forms of the definite and indefinite article, i.e. *des* or *van den* (m.), *der* or *van de* (f.), *van het* or *des* (n.), *der* or *van de* (all gender plur.), and *eens* or *van eenen* (m./n.), *eener* or *van eene* (f.). The varying order of mentioning the forms is noticeable, generally listing the synthetic forms before the option with *van*, but *van* before the inflected genitive in the case of the neuter singular. There is some more subtle variation with regard to definite and indefinite nouns in the paradigms illustrating the inflection of nouns: *des Broeders* or *van den Broeder* (m.), but *van de* or *der Zusters* (f.), *van het deksel* or *des deksels* (n.) and. It is arguable whether and to what extent the (inconsistent) order of mention had any meaning, i.e. whether or not we can deduce the grammarians' preferences. In any case, the *Rudimenta* (1799) presented synthetic and analytical forms next to each other.

Rutten (2016e: 126) interprets the shift from synthetic forms in the beginning of the eighteenth-century to analytical forms towards the end of the eighteenth-century as “an ongoing effort to appropriate the language of ‘the population’ instead of solely the language of ‘the elite’”. With regard to the genitive case, this is reflected in the metalinguistic discussions by a shift from synthetic to analytical forms prescribed in eighteenth-century normative grammars.

In the light of this generalisation, i.e. that eighteenth-century normative works changed from a preference for synthetic forms, to both synthetic and analytical forms, and finally towards primarily analytical forms, Weiland's (1805) national grammar has to be assessed as conservative. Recall that he presented only the synthetic genitive forms, consistently neglecting the analytical option with *van* in his paradigms (cf. Section 2). According to Rutten (2016e: 129), Weiland's official 1805 grammar is “a return to the prescriptions of the period of elitist grammar” and “a move away from the radical choice to focus only on analytical forms in the late eighteenth century”, such as van Bolhuis (1793).

4 Previous research

The genitive case in Dutch has gained a considerable amount of attention in the research literature, also from a historical-sociolinguistic perspective. Not least because of the opposite developments in language practice on the one hand, and language norms on the other, there have been several studies focusing on the genitive case, mainly in seventeenth-, and to a lesser extent, eighteenth-century Dutch. In this respect, the present case study also builds upon previous findings by extending it to the late eighteenth and the early nineteenth century.

Certainly serving as a point of departure, Scott (2014) provides an in-depth analysis of the genitive case in Dutch (as well as in German), covering a wide time span between the sixteenth century to the situation in present-day Dutch. Taking into account a variety of text sources, from published prose texts to more informal ego-documents⁸⁰, Scott demonstrates the influence of standardisation and codification on the use of the genitive case and alternative constructions. He argues that prescriptive grammars were influential in preserving the declining genitive case and in promoting its use, which was not only reflected in more formal published texts but also in personal ego-documents, particularly from the nineteenth century. It is one of the aims of this case study to examine whether and to what extent these findings can be confirmed with data from the *Going Dutch Corpus*. Furthermore, Scott (2014) highlights the internal factor of forms (i.e. masculine/neuter versus feminine/all gender plural), which will also be tested in this case study.

Another approach that establishes a link between the history of the genitive case and the effects of standardisation is Vezzosi (2000). Investigating a period from the twelfth to the nineteenth century, she claims that the ‘anomalous’ development of genitival constructions back towards syntheticity is probably the outcome of the standardisation process in general, and the “interference between the spontaneous drift and language standardization” (Vezzosi 2000: 115).

From a historical-sociolinguistic perspective and based on the *Letters as Loot* corpus of Dutch private letters, Nobels (2013) and Nobels & Rutten (2014) as well as Simons (2013) and Simons & Rutten (2014) investigate variation and change in genitival constructions in the late seventeenth- and late eighteenth-century, respectively. Whereas the PhD dissertations by Nobels (2013) and Simons (2013) primarily explore and map variation in genitival constructions, Nobels & Rutten (2014) and Simons & Rutten (2014) more specifically focus on the relation between norms and usage. With regard to the possible influence of codified language norms, they find only limited evidence that language users actually adhered to these prescriptions. For the late seventeenth century, Nobels & Rutten (2014) confirm the widespread assumption that the genitive case was probably rarely used in spoken language, but was rather part of fixed formulae memorised and used by

⁸⁰ Scott (2014) comprises both private letters and diaries as ‘informal ego-documents’. However, it is important to note that these sources differ considerably in terms of formality and conceptual orality (Schneider 2004: 78), and are therefore distinguished as two genres in the *Going Dutch Corpus* (cf. Chapter 4).

letter writers from all social backgrounds. For the late eighteenth century, Simons & Rutten (2014) reveal that the use of the genitive case remained fairly stable – with some social variation, though. Whereas the synthetic forms continued to decrease among writers from the lower and middle ranks, the results show a rise of historical genitives in the upper (middle) ranks, possibly due to the influence of normative publications from the eighteenth century. Given the temporal overlap with the present case study, these previous findings with letter data are reassessed from a multi-genre perspective, i.e. based on the *Going Dutch Corpus*, in Section 5.

Another study on the Dutch genitive case is Weerman et al. (2013), focusing on the loss of case marking in seventeenth-century texts from Amsterdam. Based on a corpus representing varying levels of formality, such as formal prose, official documents (*Justitieboek*) and personal letters (*Sailing Letters*), they also argue that genitive case marking no longer existed in informal writing. This is more or less confirmed by a comparison with earlier findings by Weerman & de Wit (1999), based on texts from the City of Bruges from the thirteenth to the fifteenth centuries. At the same time, the genitive case in the seventeenth century had also become a feature of formal norms, indicating a change from above. The quantitative results presented in Weerman et al. (2013) suggest a clear pattern that the higher the formality of texts, the more historical (case-marked) genitives occur.

Shifting the focus to eighteenth- and early nineteenth-century metalinguistic discourse, Rutten (2016e) portrays the changes in the normative tradition of the Northern Netherlands, illustrated by the treatment of genitival constructions. As already discussed in Section 3, he suggests a development from the period of ‘elitist’ grammar, favouring the synthetic genitive, to ‘civil’ grammar, offering both synthetic and analytical options, and to ‘national’ grammar, preferring the analytical forms. Rutten (2016e: 135) further remarks that it will be one of the main tasks for future research to investigate to what extent the normative prescriptions had an influence on actual language usage. This question is, in fact, addressed in Section 5.

5 Corpus analysis

5.1 Method

This case study investigates variation and change in the use of (1) the historical synthetic (adnominal) genitive case, both in prenominal and post-nominal position, and (2) the analytical construction with the preposition *van*, which had been the strongest competitor of the synthetic genitive forms ever since the Middle Dutch period.

In addition, two other alternative prenominal constructions have been attested as alternatives of the genitive case in the history of Dutch, viz. the

possessive *s*-construction (see example (1)), and the periphrastic possessive *z'n*-construction (see example (2))⁸¹:

- (1) *onse dierbaare Moeders ziekte*
'our dear mother's illness'
- (2) *de kaptyn syn dogter*
'the captain's daughter'

However, it is often claimed that the *s*-construction and the *z'n*-construction primarily occur with animate or, even more specifically, human possessors (Scott 2014: 103), although counterexamples are easily found. In contrast, the *van*-construction does not have those restrictions (Scott 2014: 103), making it the only alternative genitival construction that can be regarded as fully synonymous with the historical genitive case.

The following examples, all of which have inanimate possessors, indicate that many instances of the synthetic genitive (see (3)a and (4)a, as taken from the *Going Dutch Corpus* data) cannot be replaced by the *s*-construction (see (3)c and (4)c) and/or the *z'n*-construction (see (3)d and (4)d), whereas an alternative construction with the preposition *van* is possible (see (3)b and (4)b).

- (3) a. *de regen des vorigen daags*
'the rain of the previous day'
- (3) b. *de regen van den vorigen dag*
- (3) c. *?de vorigen daags regen*
- (3) d. **de vorigen dag z'n/zijn regen*
- (4) a. *de bezorging der trommel met kaneelkoekjes*
'the delivery of the tin with cinnamon biscuits'
- (4) b. *de bezorging van de trommel met kaneelkoekjes*
- (4) c. **de trommels [met kaneelkoekjes] bezorging*
- (4) d. **de trommel [met kaneelkoekjes] d'r/haar [z'n/zijn] bezorging*

With regard to their semantical and functional restrictions, I will not consider the *s*-construction and the *z'n*-construction as fully-fledged and entirely interchangeable variants of the (historical) genitive case in the sense of a sociolinguistic variable.

For the corpus analysis, I extracted the occurrences of several adnominal genitive markers and their analytical counterparts with the preposition *van* in the *Going Dutch Corpus*. Largely based on the selection presented in Scott (2014: 122), these markers cover articles (both definite and indefinite), demonstrative pronouns and possessive pronouns, as summarised in Table 2.

⁸¹ These two alternative genitival constructions are discussed in more detail in, e.g., Weerman & de Wit (1999), Nobels (2013), Simons (2013), and Scott (2014).

This selection of markers implies that constructions without an adnominal word were categorically excluded, as they can only occur with the *van*-construction, but not with the synthetic genitive (Rutten 2016e: 132). Two examples taken from the *Going Dutch Corpus* are *de aankomst van oorlogschepen* ‘the arrival of warships’ and *de ontwikkeling van welvaart* ‘the development of prosperity’, both of which have no synthetic equivalent.

Table 2. List of investigated genitive markers and their analytical counterpart with *van* (normalised spelling).

	Genitive case	<i>van</i> -construction
Articles (definite, indefinite)	<i>des</i> (‘s) (m./n.), <i>der</i> (f./all plur.)	<i>van den</i> (m.), <i>van het</i> (‘t) (n.), <i>van de</i> (f./all plur.)
	<i>eens</i> (m./n.), <i>eener</i> (f.)	<i>van eenen</i> (m.), <i>van een</i> (n.), <i>van eene</i> (f.)
Demonstrative pronouns	<i>dezes</i> (m./n.), <i>dezer</i> (f./all plur.)	<i>van dezen</i> (m.), <i>van dit</i> (n.), <i>van deze</i> (f./all plur.)
	<i>diens</i> (m./n.), <i>dier</i> (f./all plur.)	<i>van dien</i> (m.), <i>van dat</i> (n.), <i>van die</i> (f./all plur.)
Possessive pronouns	<i>mijns</i> (m./n.), <i>mijner</i> (f./all plur.)	<i>van mijnen</i> (m.), <i>van mijn</i> (n.), <i>van mijne</i> (f./all plur.)
	<i>ons</i> (m./n.), <i>onzer</i> (f./all plur.)	<i>van onzen</i> (m.), <i>van ons</i> (n.), <i>van onze</i> (f./all plur.)
	<i>zijns</i> (m./n.), <i>zijner</i> (f./all plur.)	<i>van zijnen</i> (m.), <i>van zijn</i> (n.), <i>van zijne</i> (f./all plur.)
	<i>haars</i> (m./n.), <i>harer</i> (f./all plur.)	<i>van haren</i> (m.), <i>van haar</i> (n.), <i>van hare</i> (f./all plur.)
	<i>huns</i> (m./n.), <i>hunner</i> (f./all plur.)	<i>van hunnen</i> (m.), <i>van hun</i> (n.), <i>van hunne</i> (f./all plur.)
	<i>uws</i> (m./n.), <i>uwer</i> (f./all plur.)	<i>van uwen</i> (m.), <i>van uw</i> (n.), <i>van uwe</i> (f./all plur.)

All undesired occurrences were filtered out by hand. In the case of the synthetic genitive forms, I excluded the absolute genitive (mostly temporal adverbials such as *des winters* or *‘s ochtends*), the partitive genitive (e.g. *de meeste hunner*), and fixed expressions with a genitive (e.g. *des noods*), all of which lack an analytical counterpart with the preposition *van*. In the case of prepositional constructions with *van*, I filtered out proper names with *van de(n)* (e.g. *de beer van de Capelle*, *Mejufvr van den Berg*), specific verbs or phrasal verbs with the preposition *van* (e.g. *spreken van*, *afscheid nemen van*), the so-called *schat van een kind*-construction⁸² (Simons 2013: 260; cf. also Paardekooper 1956), and temporal markers of the type *van de week* and *van de zomer*. Furthermore, possible spelling variation was taken into account, for instance *e/ee* and *s/z* variation in *dezes/deezes/deses/deeses*, *a/aa* in *haren/haaren*, *ij/j/y* in *mijn/mijn/myn*, and so forth.

⁸² The *schat van een kind*-construction only occurs marginally in the *Going Dutch Corpus*, e.g. in *dat lieve schat van een kient*.

The selection of occurrences for the corpus analysis does include various types of more specific, fixed contexts, in which the synthetic genitive and/or the alternative *van*-construction occur, such as dates and formulae. In fact, previous historical-sociolinguistic research on the Dutch genitive case, mainly focusing on seventeenth- and eighteenth-century private letters (e.g. Nobels & Rutten 2014; Simons & Rutten 2014), has attested that “context is a major factor of influence in the distribution of the genitive case and alternative constructions” (Nobels & Rutten 2014: 40). Largely following the categorisation introduced by Nobels (2013) and Simons (2013), I thus distinguish neutral contexts from more specific contexts, viz. dates, religious formulae, and other (non-religious) formulae, such as epistolary formulae and fixed expressions. Finally, prepositional expressions like *uit hoofde* (+ genitive/*van*) will also be considered as a separate context.

To begin with, the neutral context covers practically all occurrences of the synthetic genitive and the analytical *van*-construction which are not (overly) dominated by formulae, fixed expressions, or dates. Examples (5–8) illustrate the neutral, more creative use of these constructions:

- (5) *het gegons **der** muggen om onze ooren*
‘the buzzing of the mosquitos around our ears’
- (6) *tot dat de klok **onzer** maag zo hard begon te luiden*
‘until the bell of our stomach began to toll so loudly’
- (7) *de deur **van de** kelder*
‘the door of the cellar’
- (8) *het geklots **van eenen** zwaren waterval*
‘the splashing of a heavy waterfall’

Examples (9–12) illustrate the context of dates and other temporal references in general. Previous research has shown that this context is one of the major factors affecting the distribution of genitival constructions (e.g. Nobels & Rutten 2014: 39–40). In a very typical type of the genitive case used in dates, the noun (usually *maand* ‘month’ or *jaar* ‘year’) is omitted (see example (12)). There are 126 occurrences of this type in the entire *Going Dutch Corpus*.

- (9) *den 28sten **der** vorige maand*
‘the 28th of the previous month’
- (10) *de 26ste **van die** maand*
‘the 26th of that month’
- (11) *In den nacht **van den** 1 Sept.*
‘In the night of the 1st September’
- (12) *[de brief van] den 19 **deeses***
‘[the letter of] the 19th of this [month]’

The context of religious formulae covers any kind of formulaic or fixed expressions related to religion, including biblical references, as shown in examples (13–16):

- (13) *de byzondere gunst en goedheid **des** Allerhoogsten*
‘the extraordinary mercy and goodness of the Almighty’
- (14) *de werken **zýner** schepping*
‘the works of his creation’
- (15) *de zalige voorproeven **van den** Hemel*
‘the blissful foretastes of heaven’
- (16) *den dood **van onzen** zaligmaker*
‘the death of our Saviour’

The fourth context comprises all other formulae (i.e. non-religious), which can be either epistolary formulae, typically occurring in letter writing (see examples (17–19), or other formulaic and/or fixed expressions (see example (20)). It should be noted that the occurrences categorised as epistolary formulae do not necessarily have to be entirely fixed, but typically recur in comparable forms in the introduction and/or the ending of a letter, i.e. those parts of the common structure that tend to be largely formulaic (cf. also Rutten & van der Wal 2014: ch. 3). Similarly as in example (12) (*den 19 deeses*), a specific epistolary formula omits the noun *brief* ‘letter’ (or semantically related nouns, e.g. *bericht*, *missive* ‘message’; see example (17)), occurring 13 times in the entire corpus.

- (17) *onder het schrijven **deses***
‘while writing this [letter]’
- (18) *de betuiging **mijner** achting*
‘the expression of my respect’
- (19) *de beste verzekeringen **van zyn** volmaakte welstand*
‘the best assurances of his complete well-being’
- (20) *eene **dezer** daagen*
‘one of these days’

Given the focus on letter writing in Nobels (2013) and Simons (2013), as opposed to the multi-genre approach of this dissertation, the suggested (genre-specific) context of addresses will not be considered here. Instead, another context appeared to be quite prominent in the *Going Dutch Corpus* across all genres, viz. prepositional expressions (*voorzetseluitdrukkingen*) such as *uit hoofde*, *ten aanzien*, *bij gelegenheid*, and so forth, which are varyingly followed by an inflected genitive form or by the preposition *van*. These occurrences are illustrated by examples (21–24):

- (21) *uit hoofde **der** grote warmte*
‘in consideration of the (great) heat’
- (22) *uit hoofde **van de** drukte der straten*
‘in consideration of the busyness of the streets’
- (23) *ten aanzien **der** burgerlijke en kerkelijke huwelijken*
‘with regard to the civil and church weddings’
- (24) *ten aanzien **van het** beloop der wallen*
‘with regard to the slope of the walls’

The role of contexts, i.e. neutral versus specific/formulaic, will also be taken into consideration throughout the corpus analysis in Section 5.2.

5.2 Results

To begin with, Table 3 presents the general distribution of the synthetic genitive case and the analytical *van*-construction in the entire *Going Dutch Corpus*.

Table 3. Distribution of the genitive case and *van*-construction across time.

	Period 1: 1770–1790				Period 2: 1820–1840			
	Genitive case		<i>van</i> -construction		Genitive case		<i>van</i> -construction	
	N	%	N	%	N	%	N	%
Definite art.	601	38.1	978	61.9	697	43.2	915	56.8
Indefinite art.	9	6.1	138	93.9	28	22.1	99	78.0
Demonstrative	168	54.6	140	45.5	126	51.9	117	48.2
Possessive	93	24.2	292	75.8	118	32.7	243	67.3
Total	871	36.0	1,548	64.0	969	41.4	1,374	58.6

The overall results, comprising all genitive markers (summarised in Table 2), reveal that the analytical *van*-construction (64.0%) clearly outweighs the historical genitive case (36.0%) in the late eighteenth-century period. In the early nineteenth-century period, the *van*-construction is still the most frequently used option with a share of 58.6%, but the use of the synthetic genitive slightly increases from 36.0% to 41.4%.

When we look at the various groups of genitive markers individually, considerable differences become apparent. First of all, forms of the definite article are by far the most frequent genitive markers (roughly two-thirds of all occurrences), which is why the distribution in this group, i.e. 61.9% (*van*-construction) versus 38.1% (synthetic forms), is fairly similar to the overall results.

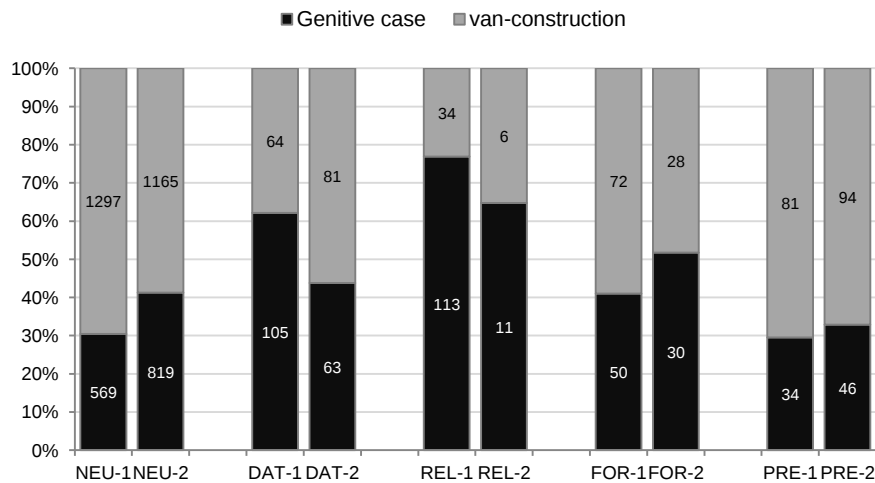
For the indefinite article, the *van*-construction is even more markedly the preferred option in the eighteenth-century data with a high share of 93.9%. The *van*-construction is also the most frequently used option with possessive determiners, occurring in 75.8%. The synthetic genitive, on the other hand, is comparatively strong in the group of demonstratives with a share of 54.6%, which can be explained by the frequent usage of inflected genitive forms of *deze*, i.e. *dezes* and *dezer*, as often preserved in dates.

In the nineteenth-century data, the distribution of constructions occurring with the definite article, demonstrative and possessive pronouns is somewhat fluctuating, but in general, synthetic and analytical constructions in these three groups of markers are fairly stable across time. In the case of the indefinite article, however, there is a notable increase of the synthetic genitive from 6.1% to 22.1%.

Context

As mentioned before, the context in which the synthetic genitive case or the alternative with the preposition *van* occurs, is one of the major factors affecting the distribution of constructions (Nobels & Rutten 2014: 40). Therefore, Figure 1 presents the general results across the various contexts distinguished in Section 5.1, i.e. the neutral context (NEU), dates (DAT), religious formulae (REL), other formulae (FOR) and prepositional expressions (PRE).

Figure 1. Distribution of the genitive case and *van*-construction across context and time.



The results for the context-specific distribution of constructions indeed reveal a considerable amount of variation. In neutral contexts, the *van*-construction has a relatively high share of 69.5%, outnumbering the synthetic genitive (30.5%) in the eighteenth-century period. Interestingly, one can witness an increase of the synthetic genitive from 30.5% to 41.3% in the nineteenth-century period. On the

assumption that neutral contexts do not typically preserve historical case inflections, the increase of the genitive suggests an effect of the prescriptions in Weiland's (1805) national grammar. As outlined in Section 2, Weiland presented only synthetic forms in his officialised paradigms.

Remarkably, the rise of synthetic genitive forms is not reflected in some of the more specific (i.e. non-neutral) contexts. Particularly in the context of dates, the results show a diachronic decrease of the synthetic forms. In the eighteenth-century data, dates most frequently take the synthetic genitive (62.1%). In the nineteenth-century data, however, the use of the historical genitive in the context of dates drops to 43.8%, whereas the *van*-construction becomes prevalent with a share of 56.3%. These tendencies certainly need a more detailed investigation when zooming in on genre variation.

Religious formulae, especially in the eighteenth century, are mostly used with the synthetic genitive (76.9%). Diachronically, the alternative *van*-construction seems to gain ground in this context (from 23.1% to 35.3%), but it has to be noticed that the nineteenth-century data set comprises no more than seventeen instances in total. In fact, the absolute decrease of religious formulae is more striking than the developments in the relative distribution. It can be concluded that, at least in the genres of the *Going Dutch Corpus*, religious formulae no longer play a major role in the nineteenth century.

The context of other (non-religious) formulae, possibly the most heterogeneous category, shows the most balanced distribution. Whereas the analytical *van*-construction is slightly dominant in the eighteenth century with a share of 59.0%, the synthetic genitive is still a frequently occurring option in 41.0%. In the nineteenth century, this context practically reaches a well-balanced 50/50 distribution of constructions.

In the case of prepositional expressions, the distribution of constructions is remarkably similar to that of neutral contexts. In the eighteenth century, they mainly occur with *van* (70.4%) rather than with the historical genitive (29.6%). Diachronically, the distribution is rather stable, with only a minor increase of the synthetic genitive in the nineteenth century.

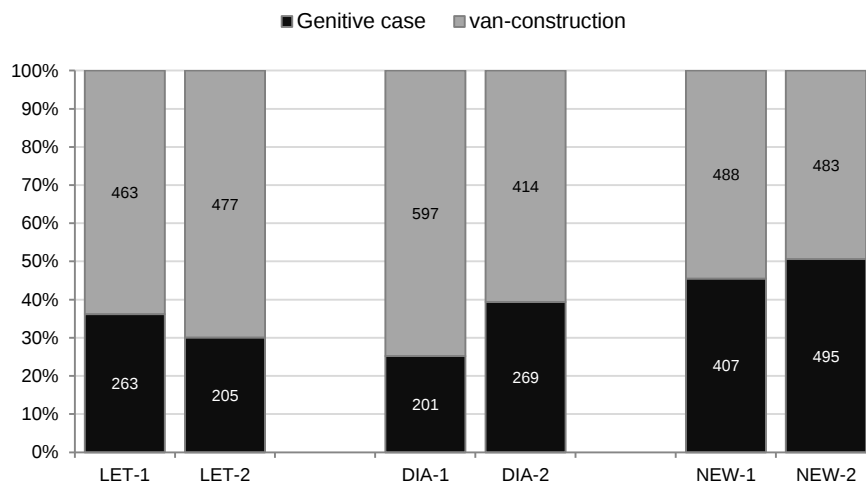
In sum, the results from the *Going Dutch Corpus* confirm earlier findings in that the role of context proves to be a relevant factor of influence on the distribution of the genitive case and the alternative *van*-construction, particularly in the late eighteenth-century period. Whereas dates and particularly religious formulae frequently occur with the synthetic genitive (diachronically decreasing, though), the *van*-construction is the preferred construction in neutral contexts as well as with prepositional expressions. At the same time, the increase of synthetic forms in neutral contexts probably indicates the normative influence of Weiland (1805).

Genre variation

Whereas previous research on the Dutch genitive case has often focused on private letters (e.g. Nobels & Rutten 2014, Simons & Rutten 2014; cf. also Weerman et al.

2013 for a mixed corpus), the design of the *Going Dutch Corpus* allows a multi-genre approach to the distribution in actual language practice as well as to the possible influence of Weiland's (1805) national grammar prescribing the synthetic forms. The results across the three investigated genres, i.e. private letters (LET), diaries and travelogues (DIA), and newspapers (NEW), are presented in Figure 2.

Figure 2. Distribution of the genitive case and *van*-construction across genre and time.



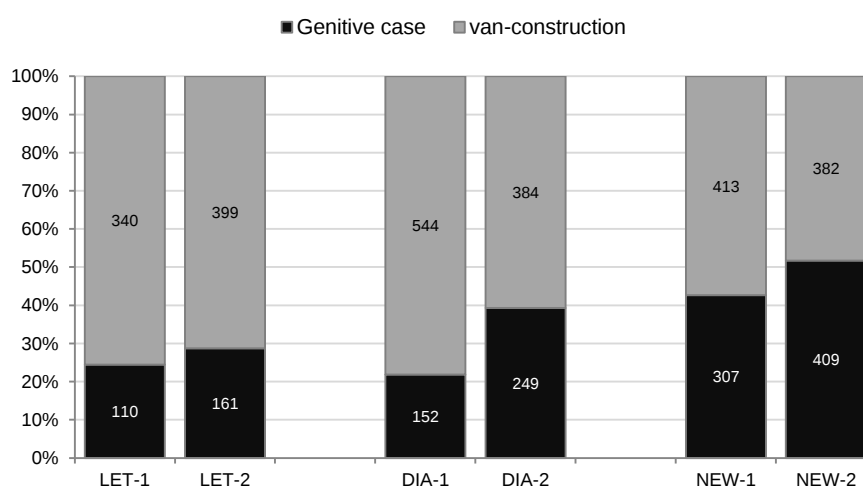
In the late eighteenth-century period, the analytical *van*-construction is prevalent in the two types of ego-documents, i.e. in private letters, occurring in 63.8%, and particularly in diaries and travelogues with a share of 74.8%. In the newspaper data, the distribution of constructions is more balanced, with the analytical construction (54.5%) occurring alongside the synthetic forms (45.5%).

In the early nineteenth-century period, the synthetic genitive case further loses ground in private letters, decreasing from 36.2% to 30.1%, while the *van*-construction consolidates its dominant position with a share of 69.9%. In diaries and travelogues, however, the use of the synthetic genitive increases from a relatively low share in the first period (25.2%) to almost 40% in the second period. A slight increase of the genitive from 45.5% to 50.6% is also witnessed in the newspaper data.

Does the increase of the synthetic forms in newspapers and especially in diaries and travelogues reflect the influence of Weiland's (1805) national grammar, clearly promoting the declining case system in his paradigms? In contrast to the increase of genitive forms in the two conceptually more 'written' genres, Figure 2 also suggests that the use of the historical genitive does not gain ground in the most 'oral' genre of the *Going Dutch Corpus*, i.e. private letters. Here, the genitive case slightly drops in favour of the analytical *van*-construction.

However, in order to assess the possible effect of Weiland (1805) on early nineteenth-century language usage, a more fine-grained investigation of the results is needed. As the various contexts have previously been argued to be a relevant factor, Figure 3 shows the distribution of the genitive case and the *van*-construction in neutral contexts exclusively.

Figure 3. Distribution of the genitive case and *van*-construction across genre and time (neutral contexts only).



The selection of neutral contexts demonstrates that the synthetic genitive actually increases across all three genres of the *Going Dutch Corpus*. In contrast to the results presented in Figure 2, the distribution in private letters even reveals a slight increase of the synthetic forms from 24.4% in the eighteenth century to 28.8% in the nineteenth century. However, the *van*-construction still outweighs the historical genitive case in these neutral contexts.

The increase of genitive forms is more pronounced in the other two genres. In diaries and travelogues, the share of the synthetic genitive rises from a relatively low 21.8% in period 1 to 39.3% in period 2. In newspapers, the share of the synthetic genitive is already relatively high in period 1 (42.6%), but it gains even more ground in period 2, increasing to 51.7%, alongside the more or less equally frequent *van*-construction (48.3%).

When compared to Figure 2 (all contexts included), one can see that the relative distribution of constructions in the two sub-corpora of diaries and travelogues as well as newspapers is, in fact, fairly similar to the results shown in Figure 3 (neutral context only). In the sub-corpus of private letters, however, both the distribution of constructions, particularly in the late eighteenth-century period, and the diachronic tendencies in this genre show considerable differences between the occurrences across all contexts on the one hand, and the separated neutral

context on the other. It can be assumed that these differences signal the influence of (non-neutral) contexts, especially in private letters.

In order to assess to what extent the specific and/or formulaic contexts influence the results, Table 4 displays the distribution of neutral and non-neutral contexts in the three sub-corpora.

Table 4. Distribution of neutral and non-neutral contexts.

	Period 1: 1770–1790				Period 2: 1820–1840			
	Neutral		Non-neutral		Neutral		Non-neutral	
	N	%	N	%	N	%	N	%
Private letters	450	62.0	276	38.0	560	82.1	122	17.9
Diaries/travelogues	696	87.2	102	12.8	633	92.7	50	7.3
Newspapers	720	80.4	175	19.6	791	80.9	187	19.1
Total	1,866	77.1	553	22.9	1,984	84.7	359	15.3

In the sub-corpus of private letters, at least from the eighteenth-century period, non-neutral contexts have a relatively high share of 38.0% and, therefore, have to be taken into account as a relevant factor of influence on the overall results. The special role of formulae in letter writing comes as no surprise as previous research based on the seventeenth- and eighteenth-century *Letters as Loot* corpus (Rutten & van der Wal 2014) has shown that private letters characteristically contain epistolary and religious formulae as well as formulaic references to dates. Simons & Rutten (2014: 65) point out that “historical letters offer a combination of formulaic language and more creative parts [...], and that formulaic contexts are more likely to preserve historic forms such as the genitive case”. With regard to the observation that the share of synthetic genitive forms in eighteenth-century private letters is notably higher across all contexts (Fig. 2) than in neutral contexts only (Fig. 3), these tendencies can probably also be attested in the letter data of the *Going Dutch Corpus*.

In contrast to the prevalence of formulae in (eighteenth-century) private letters, the relevance of non-neutral contexts as a factor of influence is considerably lower in the two remaining sub-corpora. In diaries and travelogues, the overall share of occurrences in neutral contexts is, in fact, remarkably high with 87.2% in period 1 and even 92.7% in period 2. In the newspaper data, the share of neutral contexts is around 80% in both periods.

Zooming in on the detailed distribution across the various contexts in the sub-corpus of private letters, Table 5a gives more insights into the role of contexts.

To begin with, it can be seen that the synthetic genitive is particularly dominant in religious formulae, first and foremost in period 1 with a share of 81.5%. Synthetic forms also occur in 47.8% in the context of dates, and in 36.0% in (non-religious) formulae. These comparatively high shares of the historical genitive case clearly affect the overall distribution in eighteenth-century private letters.

Table 5a. Distribution of the genitive case and *van*-construction across context and time (private letters).

	Period 1: 1770–1790				Period 2: 1820–1840			
	Genitive case		<i>van</i> -construction		Genitive case		<i>van</i> -construction	
	N	%	N	%	N	%	N	%
Neutral	110	24.4	340	75.6	161	28.8	399	71.3
Dates	33	47.8	36	52.2	20	46.5	23	53.5
Religious	88	81.5	20	18.5	11	68.8	5	31.3
Formulae	27	36.0	48	64.0	8	22.2	28	77.8
Prepositional	5	20.8	19	79.2	5	18.5	22	81.5
Total	263	36.2	463	63.8	205	30.1	477	69.9

Diachronically, however, the overall use of religious and non-religious formulae found in the private letter data (both synthetic and analytical) decreases considerably in period 2. Moreover, the analytical *van*-construction increasingly replaces the historical genitive in the nineteenth-century remnants of these formulaic contexts. Interestingly, the rather well-balanced distribution in the context of dates, slightly in favour of the *van*-construction, remains stable across both periods. Prepositional expressions predominantly occur with the *van*-construction in both periods, with a stable share of around 80%. The neutral context is, in fact, the only context in which the synthetic forms gain ground in period 2. The (apparent) decrease of the genitive case in private letters attested in Figure 2 has to be explained mainly by the shrinking numbers of religious and other formulae.

Taking into account previous findings, the early nineteenth-century increase of the synthetic genitive in neutral contexts might be interpreted as a continuation of the diachronic developments attested for the seventeenth and eighteenth centuries, viz. the steep increase of the historical synthetic genitive in private letters by upper middle class and particularly upper class writers (Simons & Rutten 2014: 67). Recall that texts written by the upper middle and the upper ranks form the majority of private letters included in the *Going Dutch Corpus*, which allows a comparison with these results. Simons & Rutten (2014: 67) argue that “[t]he rise of the genitive case in neutral contexts in letters allocated to the U(M)C [...] suggests that the emphasis on case in metalinguistic discourse and its increasing normativity did influence the language use of those groups of letter writers”. Furthermore, they emphasise that the rise of the genitive case in letters linked to the upper (middle) ranks “is not accounted for by an increase of formulaic language” (Simons & Rutten 2014: 69), given that the use of formulaic language decreased diachronically, especially among the upper middle and upper class (cf. also Rutten & van der Wal 2014). With respect to Weiland’s (1805) unambiguous

preference for traditional case inflections, it can be assumed that the (slight) increase in the use of the genitive case found in neutral contexts in private letters reflects the influence of normative grammars, in this case Weiland (1805).

Table 5b zooms in on the distribution across contexts in the sub-corpus of diaries and travelogues. As shown in Table 4, the influence of non-neutral contexts on the general distribution is less relevant than in the case of private letters.

Table 5b. Distribution of the genitive case and *van*-construction across context and time (diaries and travelogues).

	Period 1: 1770–1790				Period 2: 1820–1840			
	Genitive case		<i>van</i> -construction		Genitive case		<i>van</i> -construction	
	N	%	N	%	N	%	N	%
Neutral	152	21.8	544	78.2	249	39.3	384	60.7
Dates	12	66.7	6	33.3	3	37.5	5	62.5
Religious	25	67.6	12	32.4	0	0.0	1	100
Formulae	3	50.0	3	50.0	8	100	0	0.0
Prepositional	9	22.0	32	78.1	9	27.3	24	72.7
Total	201	25.2	597	74.8	269	39.4	414	60.6

Particularly in the eighteenth century, the analytical option is most markedly used in the neutral context (78.2%) and with prepositional expressions (78.1%). Religious formulae and dates, like in private letters, tend to take the historical genitive case.

Finally, Table 5c shows the distribution across contexts in newspapers.

Table 5c. Distribution of the genitive case and *van*-construction across context and time (newspapers).

	Period 1: 1770–1790				Period 2: 1820–1840			
	Genitive case		<i>van</i> -construction		Genitive case		<i>van</i> -construction	
	N	%	N	%	N	%	N	%
Neutral	307	42.6	413	57.4	409	51.7	382	48.3
Dates	60	73.2	22	26.8	40	43.0	53	57.0
Religious	0	0.0	2	100	0	0.0	0	0.0
Formulae	20	48.8	21	51.2	14	100	0	0.0
Prepositional	20	40.0	30	60.0	32	40.0	48	60.0
Total	407	45.5	488	54.5	495	50.6	483	49.4

In contrast to the ego-documents, religious formulae are practically absent from newspapers, even in the eighteenth century. Interestingly, dates in newspapers seem to be predominantly used with the genitive case in period 1 (73.2%), but are increasingly replaced by the *van*-construction in period 2 (from 26.8% to 57.0%).

This remarkable shift feels somewhat counterintuitive given Weiland's (1805) prescriptions in favour of the synthetic genitive. On closer inspection, however, it is necessary to distinguish two different types of constructions falling under the category of dates. The first type, illustrated by examples (25–26), usually refers to the bigger unit of time within the date itself, most frequently *maand*, which is often omitted, highlighting the formulaic nature of this construction. The second type of construction, illustrated by examples (27–28), typically refers to the date of particular events, documents (e.g. *courant*, *brieven*, *berichten*), and so forth.

- (25) *den 22 **dezer***
 'the 22nd of this [month]'
- (26) *den 7 **dezer** Maand*
 'the 7th of this month'
- (27) *de aardbeving **van den** 26 Nov.*
 'the earthquake of 26th November'
- (28) *de Nieuw-Yorkse Courant **van den** 17 Nov.*
 'the New Yorker newspaper of 17th November'

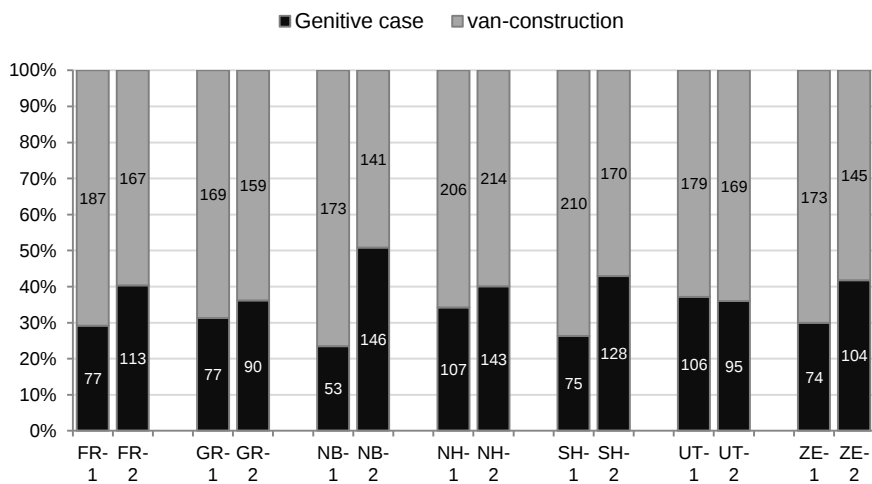
Taking a closer look at the occurrences in the newspaper data, a division of labour between the two types of constructions becomes apparent. The first type almost exclusively occurs with the synthetic genitive case, usually with formulaic *dezer*, whereas the second type favours the analytical *van*-construction, both of which is true for both the eighteenth- and the nineteenth-century period. Thus, reassessing the distribution across contexts in the newspaper data, the apparent shift in the context of dates from synthetic (period 1) to analytical (period 2) cannot be explained by the general rise of the *van*-construction in all temporal references. Instead, it might be interpreted by a slight (absolute) decrease of the *dezer*-formula and a parallel (absolute) increase of the second type of dates, clearly preferring the analytical option. Against the tendencies in Table 5c, there are no indications that the *van*-construction actually replaced the genitive in the first type of dates.

To sum up, the genre-related results presented in this section suggest that Weiland's (1805) conservative choice in favour of the historical genitive probably influenced actual language usage, as the synthetic forms increased in usage across all three genres of the *Going Dutch Corpus*. This increase is particularly visible in nineteenth-century diaries and travelogues as well as in newspapers, but to some extent also in the most 'oral' genre, i.e. the private letters. This tendency is also noticed by Scott (2014: 128), who describes the eighteenth and nineteenth centuries as "a turning point, after which we find interference from the standard written norm in the shape of productively formed genitive phrases occurring even in personal egodocuments".

Regional variation

Figure 4 displays the distribution of the historical genitive and the alternative *van*-construction across the seven regions of the *Going Dutch Corpus*, i.e. Friesland (FR), Groningen (GR), North Brabant (NB), North Holland (NH), South Holland (SH), Utrecht (UT) and Zeeland (ZE). With respect to the possible influence on formulae, only the occurrences in neutral contexts were selected.

Figure 4. Distribution of the genitive case and *van*-construction across region and time (neutral contexts only).

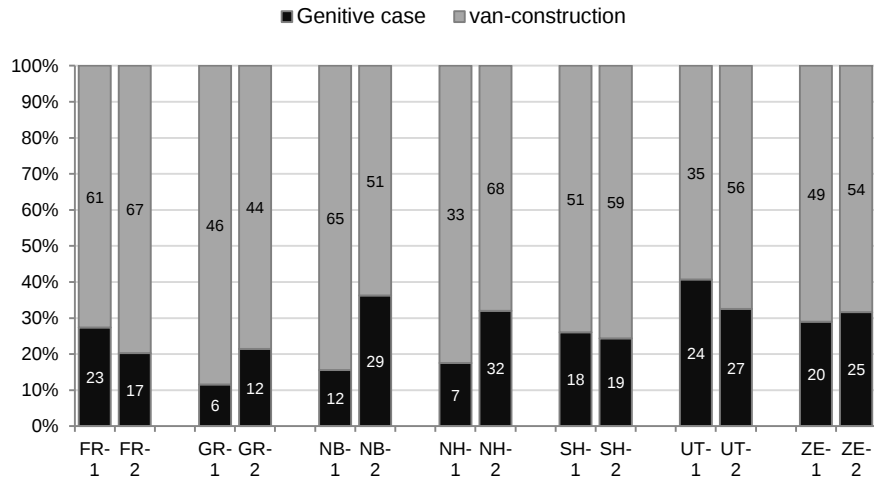


In the eighteenth-century period, the analytical *van*-construction outweighs the historical genitive in all seven regions. While the synthetic forms are particularly rare in North Brabant (23.5%), they are considerably more common in Utrecht with a share of 37.2%.

In the nineteenth-century period, the rise of the synthetic genitive case can be witnessed in practically all regions, except for Utrecht, where the distribution is more or less stable. The most notable developments are in Zeeland (from 30.0% to 41.8%), South Holland (26.3% to 43.0%) and especially North Brabant. This region has the lowest share of synthetic genitives in the first period, but at the same time shows the strongest increase of the genitive from 23.5% to 50.9% in the second period, thus co-occurring with the equally frequent *van*-construction.

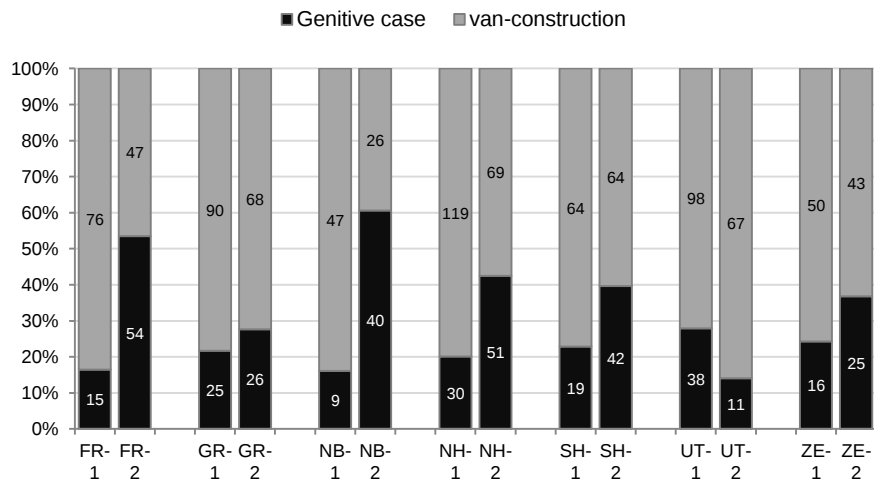
Can we expect to find more marked regional patterns by looking at the distribution across regions for each genre? Figure 5a shows the results drawn from the sub-corpus of private letters.

The eighteenth-century letter data does reveal a fairly high amount of variation. Although the analytical construction is the most common option in all regions, Utrecht stands out with a considerably higher share of historical genitives, occurring in 40.7%. In contrast, the synthetic forms rarely occur in North Brabant (15.6%) and Groningen (11.5%).

Figure 5a. Distribution across region and time (neutral contexts only; private letters).

In the nineteenth century, the regional differences are generally less extreme, although the developments vary. Groningen still has the lowest share of synthetic genitives with 21.4%, whereas there is a steep increase in North Brabant from 15.6% to 36.3%. The analytical option, on the other hand, gains some ground in Friesland and Utrecht.

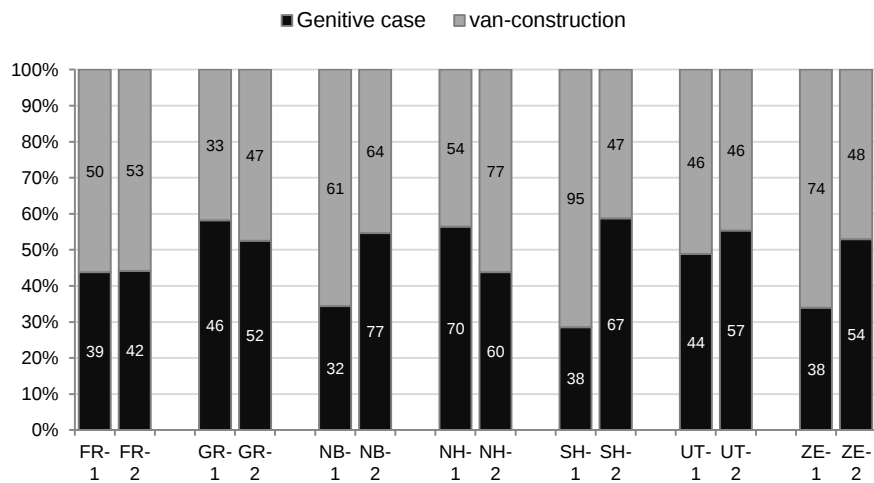
Figure 5b presents the results in the sub-corpus of diaries and travelogues.

Figure 5b. Distribution across region and time (neutral contexts only; diaries and travelogues).

In the eighteenth-century period, the variants are more or less similarly distributed across all seven regions. The analytical *van*-construction is the most frequent option in all seven regions, ranging from 72.1% in Utrecht to 83.9% in North Brabant. Like in letters from the same period, Utrecht has the highest share of synthetic genitives (27.9%), while they are least frequent in North Brabant (16.1%). There is more variation in the nineteenth-century period. The *van*-construction becomes by far the most frequently used option in Utrecht (from 72.1% to 85.9%), whereas the rise of the historical genitive can be witnessed in the remaining regions. The synthetic forms increase in both regions of the Holland area, but most notably in Friesland (from 16.5% to 53.5%) and North Brabant (from 16.1% to 60.6%). However, we have to be careful with the results from North Brabant, as the nineteenth-century data for this region is based on only one text. The distribution presented here actually represents intra-individual variation in the text produced by one single (male) diarist⁸³.

Figure 5c presents the results in the newspaper data, showing that both constructions occur in a more or less balanced way in most regions.

Figure 5c. Distribution across region and time (neutral contexts only; newspapers).



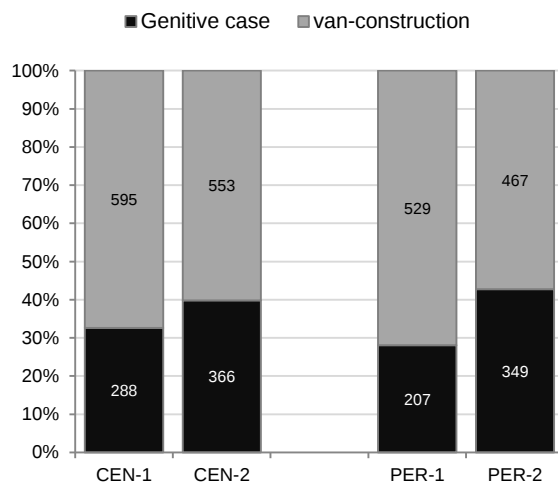
⁸³ This methodological drawback, however, allows a closer look at intra-individual variation in the writing of the male diarist from Breda (North Brabant). Except for three prepositional expressions, the majority of occurrences (66) is in the neutral context. While the diarist uses both synthetic and analytical constructions, it turns out that the different forms (i.e. masculine/neuter vs. feminine/all gender plural) play a crucial role on the choice of constructions. In fact, 95% of all synthetic genitives are feminine or plural forms. Some more variation can be found in the *van*-constructions, but these tend to be mostly with masculine or neuter nouns. Occasionally, the diarist uses both options next to each other, even within the same sentence, e.g. in *de helling eens beuvels aan den ingang van een aangenaam dal*.

The results across regions converge in newspapers from the nineteenth century, approaching the general 50/50 distribution. On the one hand, there is an increase of synthetic forms in the three southern regions, viz. Zeeland (from 33.9% to 52.9%), North Brabant (from 34.4% to 54.5%) and particularly South Holland (from 28.6% to 58.8%). On the other hand, the analytical construction gains ground in Groningen (from 41.8% to 47.5%) and North Holland (from 43.6% to 56.2%).

Variation across centre and periphery

Figure 6 shows the distribution of constructions across the centre (CEN) and the periphery (PER) in neutral contexts.

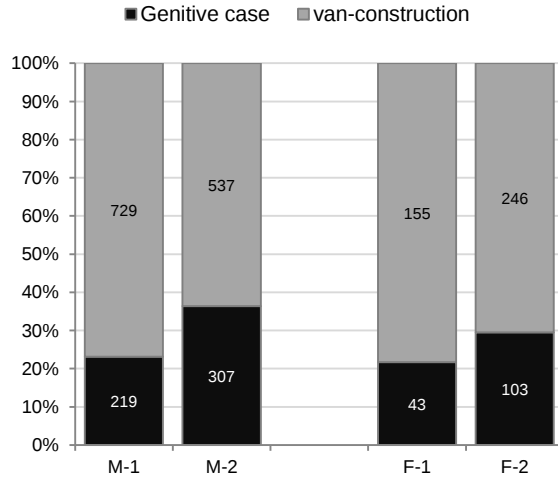
Figure 6. Distribution across centre–periphery and time (neutral contexts only).



In the eighteenth-century period, the differences between the centre (32.6% synthetic vs. 67.4% analytical) and the periphery (28.1% synthetic vs. 71.9% analytical) are fairly marginal. Similarly, no major distributional differences between centre and periphery can be attested in the nineteenth-century period, although the increase of the synthetic genitive is more pronounced in the periphery (from 28.1% to 42.8%) than in the centre (from 32.6% to 39.8%).

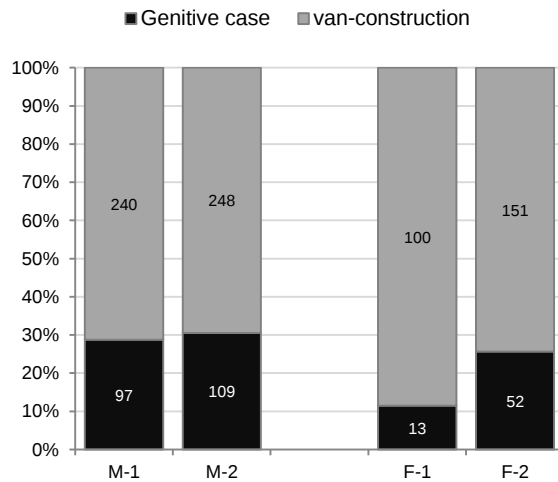
Gender variation

Figure 7 shows the distribution of constructions across gender, i.e. across male (M) and female (F) writers of ego-documents in the *Going Dutch Corpus*. Only the occurrences in neutral contexts were selected.

Figure 7. Distribution across gender and time (neutral context only).

Hardly any gender variation can be attested in the eighteenth-century data. In fact, the analytical *van*-construction is the most frequently used option among male and female writers, occurring in 76.9% and 78.3%, respectively. In the nineteenth-century data, the synthetic genitive gains ground in ego-documents written by both men (from 23.1% to 36.4%) and, to a somewhat lesser extent, women (from 21.7% to 29.5%).

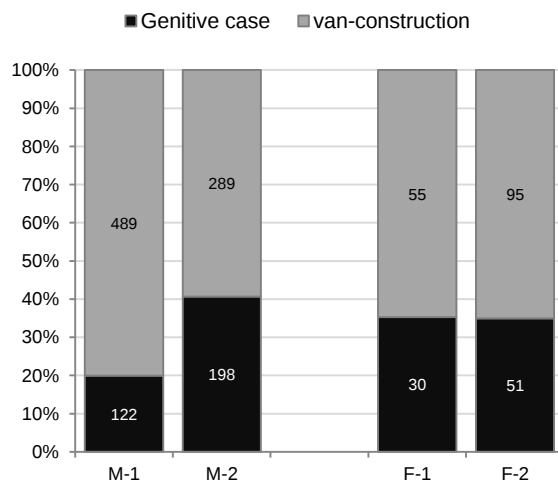
Allowing an even more fine-grained analysis and taking into consideration the genre differences between the two types of ego-documents, Figure 8a also displays the gender-related distribution of constructions in private letters.

Figure 8a. Distribution across gender and time (neutral context only; private letters).

As can be seen, eighteenth-century male letter writers predominantly use the *van*-construction (71.2%), more or less maintaining this share in the nineteenth-century period (69.5%). For female letter writers from the first period, the analytical *van*-construction is by far the most frequently option with (88.5%). However, the use of the synthetic genitives by women increases considerably in the nineteenth century from 11.5% to 25.6%. In other words, whereas eighteenth-century male letter writers were much more likely to use a synthetic genitive than their female contemporaries, the gender-related differences largely level out in the nineteenth-century data.

Zooming in on the second type of ego-documents, Figure 8b presents the distribution of constructions across genders in the sub-corpus of diaries and travelogues (in neutral contexts).

Figure 8b. Distribution across gender and time (neutral context only; diaries and travelogues).



In these texts, the gender-related distribution in neutral contexts shows a different picture than in the letter data. The *van*-construction is strikingly prevalent among male diarists in the eighteenth-century period (80.3%). In the nineteenth-century period, however, there is a steep increase of the synthetic genitive used by men, doubling its share from 20.0% to 40.7%.

In contrast, female diarists from the eighteenth century use the synthetic forms relatively frequently in 35.3%. The preference for the analytical option (64.7%) is thus less pronounced than in the case of male diarists of the same period. This distribution remains stable in period 2. Similarly to the findings in private letters, gender variation largely levels out in the nineteenth century, with a strong share of the genitive case found in diaries and travelogues by both men and women.

To sum up, gender-related differences in neutral contexts can only be attested in the eighteenth-century period, while they approximate in the nineteenth-century period in both types of ego-documents. Diachronically, one can observe a stable distribution in the case of male letter writers and female diarists, and a noticeable increase of synthetic genitive forms in the case of female letter writers and male diarists.

Internal variable: Forms

Finally, the internal variable of forms will be taken into account as a possible factor of influence, as suggested by Scott (2014). Figure 9 presents the distribution of synthetic and analytical constructions across masculine/neuter singular (M/N) and feminine singular/all gender plural forms (F/Plur).

Figure 9. Distribution of the genitive and the *van*-construction across forms and time.

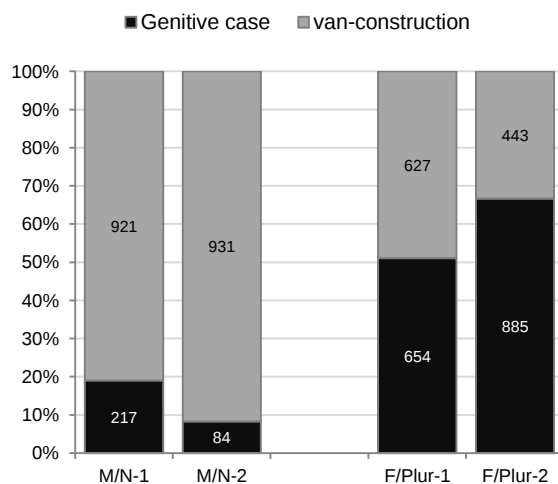
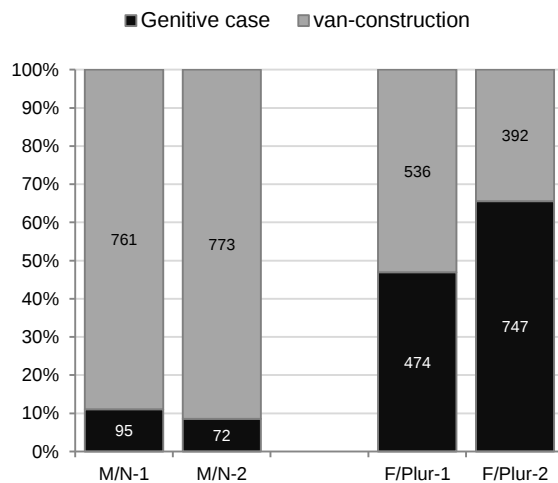


Figure 9 clearly reveals major differences between the masculine and neuter forms on the one hand, and feminine and all gender plural forms on the other hand. In the eighteenth-century data, the *van*-construction of the masculine and neuter forms (80.9%) clearly outweighs the synthetic genitive (19.1%), which further decreases in the nineteenth century, when it becomes a marginal option (8.3%).

In contrast to their masculine/neuter equivalents, the distribution of constructions with feminine/all gender plural forms is well-balanced in the eighteenth century (roughly 50/50). In the nineteenth century, the female/plural forms see a considerable increase of the synthetic genitive (from 51.1% to 66.6%) at the expense of the *van*-construction (from 49.0% to 33.4%).

Taking into account the role of contexts attested in this case study several times before, Figure 10 displays the distribution of the genitive case and *van*-construction across forms in neutral contexts only.

Figure 10. Distribution of the genitive and the *van*-construction across forms and time (neutral context only).



Generally, Figure 10 shows that the diachronic developments in Figure 9 are confirmed when only occurrences in neutral contexts are considered. In the case of masculine and neuter forms, we do see some influence of context reflected in the results, as the eighteenth-century share of the historical genitive is lower (11.1% versus 19.1%) when the non-neutral contexts are excluded. For the nineteenth-century data, i.e. when religious and other formulae no longer play a relevant role, the distribution in neutral context only is almost identical with the distribution across all contexts.

The same is true for the female and all plural gender forms. In the first period, we can notice a minor effect of the non-neutral contexts, viz. particularly dates and religious formulae, in that the synthetic genitive is somewhat lower in neutral contexts only (46.9%) as opposed to all contexts (51.1%). In the second period, however, the share of the prevalent genitive case (65.6%) is practically as high as across all contexts (66.6%). The preliminary conclusion drawn from the results in Figure 9, i.e. that the synthetic genitive becomes a marginal option in nineteenth-century usage in the case of masculine/neuter forms, whereas it considerably gains ground in the case of female/all gender plural forms, is confirmed by these findings. In fact, the variable of context does not crucially affect the opposite developments across forms.

The question arises whether the effectiveness of Weiland's (1805) paradigms prescribing the historical case system was largely dependent on the forms of the markers. On the one hand, Figure 10 gives evidence that the synthetic

genitive in masculine and neuter forms almost disappears from early nineteenth-century language practice – despite the officialised norms in favour of case inflections. On the other hand, the increase of the synthetic genitive in feminine and all gender plural forms suggests the influence of Weiland (1805) on actual language usage.

Scott (2014) also observes and discusses these form-related differences, taking into account the higher token frequency of feminine and plural genitive markers in the genitive as a relevant internal factor. He points out that “by the 19th century, most nouns occurring in the genitive were feminine singulars and plurals of all gender” (Scott 2014: 121), viz. 90 (M/N) versus 544 (F/Plur) occurrences in his nineteenth-century data. The results drawn from the *Going Dutch Corpus* and presented in Figure 9 confirm this major difference in the token frequency of genitive markers, viz. 84 (M/N) versus 885 (F/Plur) occurrences. Scott (2014: 121) argues that “[t]he high token frequency of feminine singular and all genders plural nouns in the genitive in the 19th century may well have aided the preservation of the *x der y* structure, but not a masculine/neuter equivalent”.

Also diachronically, the results from the *Going Dutch Corpus* are generally in line with Scott’s (2014: 121–122) observations for the sixteenth/seventeenth, eighteenth and nineteenth centuries:

across the three periods, the masculine/neuter singular markers tend to decrease in use while the feminine singular/all gender plural markers tend to increase in use. The genitive structure that consistently had the highest token frequency, and which therefore would have been the genitive structure most familiar to language users, was *x der y*. In addition to having the highest token frequency of any genitive determiner, *der*, as a determiner used with any plural noun as well as feminine singular nouns, had a particularly high frequency; that is to say, it was used with a large group of nouns.

With regard to the possible Weiland effect on early nineteenth-century language practice, it can be suggested that the national 1805 grammar could boost the use of feminine/all gender plural markers. At the same time, the prescriptions in Weiland (1805) apparently failed to revive the synthetic genitive in masculine and neuter markers, which had been too low in frequency by the end of the nineteenth century, particularly in neutral contexts, and were possibly no longer familiar enough to language users. In sharp contrast, the data from the post-Weiland generation sees the rise of the synthetic genitive in feminine and plural markers. Notably, this increase is not limited to the *x der y* structure mentioned by Scott (2014), as shown in Table 6a.

The increase of the historical genitive forms can not only be attested for the definite article *der*, but also for the feminine and plural forms of the indefinite article as well as possessive pronouns. While this does not categorically rule out the special role of the *x der y* structure and its preserving effect (i.e. by far the most frequent structure with the historical genitive case), the findings in Table 6a give evidence that the early nineteenth-century rise of the synthetic case forms affects

more genitive markers than just *der*. Therefore, the influence of Weiland (1805) can probably be assumed here.

Table 6a. Distribution across forms and time (female/all gender plural forms).

	Period 1: 1770–1790				Period 2: 1820–1840			
	Genitive case		<i>van</i> -construction		Genitive case		<i>van</i> -construction	
	N	%	N	%	N	%	N	%
Definite art.	419	47.0	472	53.0	621	67.2	303	32.8
Indefinite art.	8	44.4	10	55.6	26	59.1	18	40.9
Demonstrative	144	72.7	54	27.3	125	72.3	48	27.8
Possessive	83	47.7	91	52.3	113	60.4	74	39.6
Total	654	51.1	627	49.0	885	66.6	443	33.4

Table 6b confirms that no such effect can be attested for the masculine/neuter markers, where the genitive in the definite article – the equivalent *x des y* structure – as well as the synthetic forms of the indefinite article, demonstrative and possessive pronouns are for the most part replaced by the *van*-construction.

Table 6b. Distribution across forms and time (masculine/neuter forms).

	Period 1: 1770–1790				Period 2: 1820–1840			
	Genitive case		<i>van</i> -construction		Genitive case		<i>van</i> -construction	
	N	%	N	%	N	%	N	%
Definite art.	182	26.5	506	73.6	76	11.1	612	89.0
Indefinite art.	1	0.8	128	99.2	2	2.4	81	97.6
Demonstrative	24	21.8	86	78.2	1	1.4	69	98.6
Possessive	10	4.7	201	95.3	5	2.9	169	97.1
Total	217	19.1	921	80.9	84	8.3	931	91.7

6 Discussion

Investigating another crucial morphosyntactic variable in the context of the Dutch *schrijftaalregeling*, this chapter focused on variation and change in the use of the (adnominal) genitive case and the alternative prepositional *van*-construction in late eighteenth- and early nineteenth-century Dutch. Building upon a vivid research

tradition on genitival constructions in Dutch, primarily in the seventeenth (Nobels & Rutten 2014, Weerman et al. 2013) and eighteenth century (Simons & Rutten 2014), this case study aimed to examine the effectiveness of metalinguistic discourse and particularly Weiland's (1805) national grammar on actual language usage.

As outlined in Section 1, Dutch originally had a fully-fledged case system, including the genitive case, which had been in decline since the Middle Dutch period. As a result of the increasing loss of inflections in general, and the parallel rise of the alternative constructions, particularly with the preposition *van*, a situation of competition between synthetic and analytical genitival constructions had emerged. By the eighteenth century, inflected genitive forms still occurred in written language and especially in higher registers, while they had (presumably) disappeared from spoken and colloquial language.

Nevertheless, eighteenth-century grammarians still had a strong focus on nominal inflection. In Section 3, it was pointed out that the historical genitive forms were widely preferred in the paradigms of early eighteenth-century grammars (e.g. Moonen 1706, Verwer 1707), although the analytical *van*-construction was an increasingly accepted alternative towards the middle of the century (e.g. Elzevier 1761, van der Palm 1769), and even became the preferred option in the final decades (e.g. van Bolhuis 1793). Against the background of these developments in metalinguistic discourse, it is striking that Weiland's (1805) officialised grammar of Dutch returned to the prescription of synthetic genitive forms only. One of the central questions was whether and to what extent this conservative choice in favour of the historical case inflections influenced early nineteenth-century language practice. Did the gap between norms and usage grow even further?

In Section 5, the possible effectiveness of Weiland (1805) was investigated through a corpus-based analysis of the synthetic genitive case and its analytical alternative with the preposition *van*. The general results revealed that the *van*-construction was prevalent in the late eighteenth-century data with a share of 64.0%. However, in the early nineteenth-century (i.e. post-Weiland) period, one could see that the synthetic forms gained some ground in usage, increasing from 36.0% to 41.4%. While these tendencies suggested a 'Weiland effect', a more fine-grained analysis appeared to be necessary in order to assess the normative influence.

Previous research (e.g. Nobels & Rutten 2014, Simons & Rutten 2014) has demonstrated that the role of contexts is a crucial factor of influence, in that the genitive case is more likely to be preserved in formulaic contexts (such as dates, religious or epistolary formulae, etc.) than in the more creative, neutral contexts. Indeed, a considerable amount of variation across contexts was also attested in the present case study, especially for the eighteenth-century period. The *van*-construction was the most frequently used construction in neutral contexts, whereas dates and especially religious formulae mostly occurred with the genitive case. In the nineteenth-century period, however, the share of the synthetic genitive forms increased in the neutral context, which supports the assumption of a

normative influence of Weiland's (1805) grammar prescribing these historical case inflections.

With respect to genre variation, the general results (across all contexts) indicated that the synthetic forms gained ground in newspapers and particularly in diaries and travelogues, whereas the *van*-construction seemed to consolidate its dominant position in private letters, at the expense of the synthetic forms. By taking into account the influence of contexts and selecting only occurrences in the neutral context, the results were somewhat different, though. To begin with, the distribution of constructions in eighteenth-century private letters revealed that the *van*-construction was considerably more frequent in the neutral context than across all contexts, signalling a fairly strong influence of formulaic contexts. On closer inspection, religious formulae in letters (especially from period 1) mostly occurred with synthetic genitive forms, affecting the distribution of constructions. In the other two genres, the role of contexts turned out to be a less relevant factor of influence, though.

Furthermore, the nineteenth-century data (neutral context only) nicely illustrated the genre-related gradation also attested in previous case studies, in that the highest share of the prescribed synthetic forms was found in newspapers (around 50%), followed by diaries and travelogues (around 40%) and, with the lowest share in the most 'oral' genre, viz. private letters (around 30%). Diachronically, the synthetic genitive gained ground in all three genres of the *Going Dutch Corpus*. This general development across all genres, including private letters, suggests the effectiveness of Weiland's (1805) national grammar on language usage. In fact, these results also confirmed previous findings (e.g. Simons & Rutten 2014, Scott 2014), viz. that ego-documents written by members of the upper middle and upper ranks – which are also the socio-economic groups of writers primarily represented in the *Going Dutch Corpus* – saw the increase of the historical genitive case, most probably due to normative influence.

The analysis of possible geographical variation did not reveal marked patterns, even though there was some variation across the seven selected regions. For instance, the diachronic increase of the synthetic genitive in North Brabant was observed in all three genres. With respect to the centre and the periphery, hardly any variation could be attested. In this case study, it can probably be concluded that space was no longer a decisive external factor in the period under investigation.

The sociolinguistic variable of gender revealed that, at least when only neutral contexts were considered, the synthetic genitive case increased in texts written by both men and women. It was slightly more frequent among male users, though. Some further differences became apparent when zooming in on the gender-related results for each of the two ego-document genres individually. In the first period, male letter writers were more likely to use the genitive case than female letter writers, whereas the opposite was true in diaries and travelogues. In second period, those gender differences more or less levelled out.

Finally, the internal variable of forms was investigated, suggested to be a major factor conditioning the distribution of constructions. Indeed, very marked differences could be revealed between the masculine and neuter forms on the one

hand, and feminine and all gender plural forms on the other hand. In line with previous observations, mainly by Scott (2014), the *van*-construction was overwhelmingly used with masculine/neuter nouns, whereas the genitive case had become a marginal option by the late eighteenth century. In sharp contrast, the feminine/all gender plural equivalents were still frequently used with the historical genitive case, and even increased considerably in the early nineteenth century. Moreover, Scott (2014) refers to the conserving effect of the *x der y* structure, which was not only the most frequently used genitive structure but also the most familiar one to language users. Nevertheless, the results drawn from the *Going Dutch Corpus* demonstrated that the rise of the synthetic genitive forms was not limited to the definite article, but could also be attested for other groups of genitive markers. In sum, it is likely that a combination of the officially prescribed and promoted genitive forms in Weiland's (1805) grammar on the one hand, and a generally high familiarity among language users (especially *x der y*) on the other hand, helped to increase the use of feminine/all gender plural genitive markers in the early nineteenth century. In fact, no such effect could be attested for the masculine/neuter genitive markers. It might be assumed that they had become too low in frequency and had no longer been familiar to most language users, which is why they could not be 'revived' in actual language use – despite the official prescriptions.

Coming to the general question whether and to which extent Weiland's (1805) national grammar could influence the use of the historical genitive case in actual language practice, it may be concluded that normative influence was, at least to a certain degree, reflected in the corpus results. Not only did the use of synthetic forms (in neutral contexts) considerably gain ground in nineteenth-century diaries and travelogues as well as in newspapers, but also in the most 'oral' genre, viz. private letters. Moreover, both male and female writers increasingly used the inflected case forms in the post-Weiland period.

Nonetheless, it must not be forgotten and trivialised that the analytical *van*-construction had actually been established as the prevalent construction in late eighteenth-century language practice, certainly in handwritten ego-documents. The increasing relevance of the *van*-construction in usage is generally in line with the developments in the eighteenth-century normative tradition. Despite losing some ground in the early nineteenth century in favour of the synthetic genitive case, the *van*-construction remained the most frequently used option in handwritten ego-documents and particularly in private letters. In newspapers, i.e. the printed and most 'written' genre investigated, synthetic and analytical constructions co-occurred as equally common options.

Still, Weiland's (1805) conservative choice and his effort to officially revive the historical genitive case on a national level can be assessed as partly successful – certainly when we consider the fact that the synthetic forms had largely disappeared from spoken/colloquial language, and had primarily been preserved in the higher registers.

Inter- and intra-individual variation and change

The Martini Buys family correspondence

1 Introduction and research objectives

While the analyses of orthographic and morphosyntactic phenomena presented in Chapters 5–12 mainly focused on variation and change in the community at large, it is also important to take into account the individual as an additional factor possibly affecting distributional patterns. In their study on language and the individual based on data from the *Corpus of Early English Correspondence*, Nevalainen & Raumolin-Brunberg (2016: 202f.) zoom in on the linguistic behaviour of individuals, which “complements the aggregate picture that we have drawn [...] on the language use of people grouped together on the basis of sociolinguistic variables such as gender, age, region and socio-economic status”.

Following previous historical-sociolinguistic research on individual variation in (primarily English) private letters, for example Raumolin-Brunberg’s (2005, 2009) work of lifespan changes, Austin (1991, 1994) on the Clift family correspondence, as well as various studies on the Paston Letters (e.g. Bergs 2005; Hernández-Campoy 2016) and the Bluestocking Letters (e.g. Sairio 2008), this chapter examines inter- and intra-individual variation and change in Dutch family correspondence. The data, spanning the second half of the eighteenth and the first half of the nineteenth century, is drawn from a corpus of private letters from the Martini Buys family archives, which are, for the most part, stored in *Het Utrechts Archief* (Utrecht), with additional material collected from the *Brabants Historisch Informatie Centrum* (‘s-Hertogenbosch). The *Martini Buys Correspondence Corpus* was specifically designed and compiled for this case study and will be presented in Section 2.

Compared to the large-scale analyses conducted with the multi-genre *Going Dutch Corpus*, the overarching research questions of this dissertation are approached from a slightly different perspective. While the effectiveness of the early nineteenth-century language policy on actual language usage will also play a central role here, this chapter generally aims at more exploratory and smaller-scale investigations on inter- and intra-individual variation. In fact, this case study attempts to (re)assess the possible impact of language policy, and Siegenbeek’s (1804) orthography in particular⁸⁴, from a micro-level perspective by zooming in on

⁸⁴ For the purpose of the present case study, it was deliberately chosen to focus on orthographic variables, given the complexity of morphosyntactic issues analysed in Chapters

individual writers. Several informants actually have contributed private letters written both before and after the *schrijftaalregeling* had been introduced in 1804/1805. The present chapter addresses a number of (partly interrelated) research questions which the macro-level analyses with the *Going Dutch Corpus* have left unanswered.

First of all, the aspect of intra-individual (in)consistency will be examined. Generally speaking, how consistent or inconsistent were individual writers in the use of particular spelling variants? Is variation in the overall distribution of variants (i.e. detected in the community at large) based on individual writers who used one specific variant invariably, or did writers use coexisting variants side by side? Moreover, did this change after the spelling had been regulated in 1804, assuming that these official norms also increased the awareness of spelling and spelling norms or rules?

Secondly, this case study will also take into account the possible influence of age and individual lifespan changes. While the design of the multi-genre *Going Dutch Corpus* does not consider age as a sociolinguistic variable, this case study allows to investigate whether and to what extent age-related changes can be detected in private letter writing. Table 1 schematically presents the patterns of change in the individual and the community, as introduced by Labov (1994) and further developed by Sankoff & Blondeau (2007).

Table 1. Patterns of change in the individual and the community*.

	Individual	Community
(1) Stability	Stable	Stable
(2) Age-grading	Unstable	Stable
(3) Lifespan change	Unstable	Unstable
(4) Generational change	Stable	Unstable
(5) Communal change	Unstable	Unstable

* after Sankoff & Blondeau (2007: 563, originally adapted from Labov 1994: 83)

The first pattern (1) describes diachronically stable linguistic variation overall, i.e. with regard to individual behaviour as well as to the community at large. The second pattern (2) represents stability on the level of the community, but instability in the linguistic behaviour of the individual, also referred to as age-grading⁸⁵. The pattern of generational change (4) concerns a situation with stability

10-12, conditioned by both internal and external factors. In fact, a substantial amount of occurrences is necessary for a reliable analysis of these features, whereas individual spelling preferences are more likely to be identified on the basis of a lower number of tokens.

⁸⁵ Raumolin-Brunberg (2005: 38) points out that the concept of age-grading refers to “changes in which the use of a variant or variants recurs or increases at a particular age in

on the idiolectal level, despite ongoing language change in the community at large. This means that linguistic forms acquired during the formative years remain unchanged in adulthood, irrespective of changes on the communal level. In contrast, the pattern of lifespan change (3) refers to individual changes in the direction of ongoing change in the community. Finally, the fifth pattern of communal change (5) sees people from the same community altering their language in the same direction. As Raumolin-Brunberg (2009: 171) rightly adds, “[i]t is noteworthy that both lifespan change and communal change involve instability in the individual and the community”. The difference between the two concepts lies in perspective, viz. that “lifespan change looks at the individual and communal change at the community” (ibid.)

Investigating changes in the national community of Dutch language users, the pattern of communal change was, in fact, investigated in Chapter 5–12. This chapter shifts the focus to the patterns of generational and lifespan change, testing whether and to what extent individual language users altered their linguistic behaviour in the direction of official prescriptions or not. More specifically, the case study in this chapter investigates the effectiveness of Siegenbeek’s (1804) orthography on the language use of adult individuals, who had not been exposed to the *schrijftaalregeling* directly through school education. Did they change their spelling practices in the direction of official norms, possibly acquiring ‘new’ variants in adulthood and thus testifying to the pattern of lifespan change, i.e. changes over the lifespan of individuals in the direction of changes in the rest of the community (Nevalainen & Raumolin-Brunberg 2016: 203; cf. also Sankoff 2007: 563)? Or does the Martini Buys family correspondence confirm the pattern of generational change, i.e. the stability of the individual across the lifespan but change in the wider community? Certainly in the context of top-down language policy, this seems a particularly interesting research question.

Finally, this case study also attempts to reassess variation and change in the data drawn from the *Going Dutch Corpus* (and the sub-corpus of private letters) on a micro level. For instance, the role of adult language users in the communal developments of the post-*schrijftaalregeling* period needs to be discussed. Furthermore, it should be interesting to see whether the micro-level findings reveal any differences between the investigated spelling features in terms of general awareness.

In Section 2, I first outline the compilation and design of the *Martini Buys Correspondence Corpus*, followed by a concise genealogy of the Martini Buys family. The analyses of five orthographic variables will be presented in Section 3. Section 4 contains a general discussion of the findings.

successive generations” (cf. also Labov 2001: 76). However, with historical data, age-grading is particularly difficult to trace, which is why it is not considered in this chapter.

2 The Martini Buys family correspondence

2.1 Compiling the *Martini Buys Correspondence Corpus*

The corpus of private letters from the Martini Buys family archives is best regarded as a separate, stand-alone supplement to the multi-genre *Going Dutch Corpus*. In fact, it was designed and compiled for the purpose and research questions of this case study on individual variation and change. Sixteen letters from the Martini Buys family correspondence are also part of the sub-corpus of private letters in the *Going Dutch Corpus*, comprising ten texts from the late eighteenth-century period and six texts from the early nineteenth-century period. These texts are also included in this supplement corpus. The additional Martini Buys private letters had partly been collected to be used in the *Going Dutch Corpus*, but were ultimately excluded due to the limit of 2,000 words per writer, i.e. in order to avoid the overrepresentation of particular informants. The bulk of letters, however, was collected specifically for the compilation of this supplement corpus.

For the *Martini Buys Correspondence Corpus*, the same transcription conventions were applied as for the transcriptions of handwritten ego-documents included in the *Going Dutch Corpus* (cf. Chapter 4). However, as we are dealing with a separate case study, a few methodological adjustments could be made, primarily in order to ‘loosen’ the selection criteria of material. Firstly, unlike the *Going Dutch Corpus*, there is no maximum number of words per writer in the supplement corpus. Some more prolific writers naturally left more letters (and thus more words for the corpus) than other members of the family. The focus on individual variation as well as the micro-level approach allow these differences. Secondly, the periods are less strictly defined. Whereas the *Going Dutch Corpus* delimits the two diachronic cross-sections to the periods between 1770–1790 (period 1) and 1820–1840 (period 2), the corpus of family correspondence has slightly more flexible periods, also comprising data from the years 1791–1802 (for period 1), 1806–1819 and 1841–1848 (for period 2). In contrast to the texts selected for the *Going Dutch Corpus*, even some private letters without the exact date of writing were included, at least when the approximate date could be reconstructed and deduced from the context (e.g. dates of birth and death of family members mentioned in the texts). Nonetheless, in order to match the general research questions and design of the multi-genre *Going Dutch Corpus*, the two diachronic cross-sections defined as ‘before 1804/1805’ and ‘after 1804/1805’, i.e. before and after the introduction of the *schrijftaalregeling*, will still be applied here. This allows me to assess the possible normative influence of (in this case) Siegenbeek’s (1804) orthography on the use of particular spelling features in the Martini Buys family correspondence.

2.2 Size and structure of the *Martini Buys Correspondence Corpus*

The supplement corpus comprises almost 64,000 words, consisting of 102 private letters written by and sent to members of the Martini Buys family, spanning three

generations of eleven informants (four males, seven females) from the second half of the eighteenth and the first half of the nineteenth century (cf. Table 2).

Table 2. The *Martini Buys Correspondence Corpus*.

Informant (M/F)	Lifespan	Regional origin and mobility	Date of letters (Age of writer)	Texts	Words
GENERATION I					
Antoni Martini (M)	1728–1800	Den Bosch (NB), Leiden (SH)	1787–1799 (59–71)	7	4,678
Eva Maria Adriana Buys (F)	1735–1811	Den Bosch (NB), Helvoirt (NB)	1787–1815 (52–80)	7	4,963
GENERATION II					
Anna Maria Emelia (F)	1763–1848	Den Bosch (NB), Helvoirt (NB)	c. 1780s–1825 (c. 20s–62)	8	5,218
Paulus Hubert (M)	1765–1836	Den Bosch (NB), Leiden (SH), Amsterdam (NH)	1787–1799 (22–34)	12	11,975
Geertruid Johanna Antonia Strick van Linschoten (F)	1767–1843	Utrecht, Loenersloot (UT)	1788–1834 (21–67)	16	8,632
Sibilla (F)	1765–1828	Den Bosch (NB), Utrecht	1788–1825 (23–60)	10	5,139
Hendrik Bernard (M)	1768–1848	Den Bosch (NB), Leiden (SH), Vught (NB)	1789–1835 (21–67)	16	7,926
GENERATION III					
Catharina Andrea Geertruid (F)	1796–1861	Amsterdam (NH), Utrecht	1825–1843 (29–47)	12	6,653
Antoni Adriaan (M)	1798–1873	Amsterdam (NH), North Brabant	1831–1833 (33–35)	7	6,519
Eva Maria Adriana (F)	1801–1869	Amsterdam (NH)	c. 1810s–1848 (c. 17–47)	6	2,009
Anna Maria Emilia (F)	1806–1875	Amsterdam (NH)	1823 (17)	1	259
Total				102	63,971

This corpus design makes it possible to investigate inter- and intra-individual variation and change in private letter writing of closely related individuals from a similar and thus comparable background. Nevertheless, there might also be differences in external constraints such as gender, age, educational background, status, mobility and so on (Raumolin-Brunberg 2009: 173), which have to be taken into consideration.

It should be noted that the choice of this particular family is arbitrary and may not necessarily be representative of the late eighteenth- and early nineteenth-century society of the Northern Netherlands. However, the availability of a considerable amount of private letters written by various family members of three generations makes this collection of family correspondence particularly interesting and suitable with regard to the research questions outlined in Section 1. Furthermore, the genealogy of the Martini Buys family is relatively well-documented (cf. de Meij 2011), which allows for a micro-level interpretation of the results.

Table 2 shows the design of the corpus with some basic information about the family members (i.e. gender, lifespan and age, regional origin and mobility across lifespan) and the data (i.e. span of letter writing, number of texts and words). A more detailed genealogical background of the Martini Buys family is provided in Section 2.3.

2.3 Family background

According to de Meij (2011), the origins of the Martini Buys genealogy go back to the late sixteenth century. Tonnis Martens (1580–1661) was born in the German village of Buer (part of present-day Gelsenkirchen in North Rhine-Westphalia). Martens was a cooper, who moved to Wesel, near the Dutch border, together with his wife Enneken in the early 1630s. It was their grandson Antoni (1657–1730) who brought the family and the Latinised family name Martini to the Northern Netherlands. He came to Utrecht in 1677 to study theology, later became ordinand in Leiden and ultimately moved to ‘s-Hertogenbosch with his wife Geertruid Buys. Their son Hendrik Bernard (1693–1776) held various positions in the city administration of ‘s-Hertogenbosch. He and his wife Anna Maria Emelia van Schagen were the parents of the oldest member included in the *Martini Buys Correspondence Corpus*, viz. Antoni Martini (1728–1800). He was born in ‘s-Hertogenbosch, studied law in Leiden and became Pensionary (*raad en pensionaris*) of his hometown in 1756. One year later he married Eva Maria Adriana Buys (1735–1811), also born in ‘s-Hertogenbosch as the daughter of lieutenant-general Paulus Hubert Buys. She is the second informant of the first generation in the corpus.

Antoni and Eva Maria Adriana had four children, all of which were born in ‘s-Hertogenbosch and contributed private letters to the second generation of informants in the corpus: Anna Maria Emelia (*Mietje*) (1763–1848), Paulus Hubert (*Pau*) (1765–1836), his twin sister Sibilla (1765–1828), and Hendrik Bernard (*Hein*) (1768–1848). The latter went to Leiden for his law studies, but returned to Brabant

and 's-Hertogenbosch, where he worked as a lawyer and held several administrative positions. Hendrik Bernard, who was married four times and was the father of twelve children, came into the possession of the manor of Geffen, hence his later name Martini van Geffen. Paulus Hubert, his elder brother, also studied law in Leiden and moved to Amsterdam in the 1790s, working as a lawyer and starting a stockbroking company. In 1791, he married Geertruid Johanna Antonia (*Truitje*) Strick van Linschoten (1767–1843), a daughter of general Andries Jan Strick van Linschoten from Utrecht. Anna Maria Emilia, the eldest sister, first married Jacob Frederik Roosendaal and, after his death, François André de Jonge. Sibilla, at the age of 47, married Daniel Gerard van den Burgh, a lawyer from Utrecht.

The informants of the third generation are the children of Paulus Hubert and Geertruid Johanna Antonia: Catharina Andrea Geertruid (*Cato*) (1796–1861), Antoni Adriaan (*Toon*) (1798–1873), the twin sisters Eva Maria Adriana (*Mimi*) (1801–1869) and Sibilla Paulina Elisabeth⁸⁶ (*Paulien*) (1801–1870), and Anna Maria Emilia⁸⁷ (*Emè*) (1806–1872), who were all born in Amsterdam. Antoni Adriaan, like his father Paulus Hubert, studied law in Leiden and later worked as a lawyer and stockbroker in Amsterdam. He married Cornelia Henriëtte Constance van Eijs and was the father of four children. Catharina Andrea Geertruid married Jan van den Bergh, and her sister Anna Maria Emilia married Gulian Tutein Nolthenius. The twin sisters remained unmarried.

In social terms, the three generations of the Martini Buys family are best characterised as a well-to-do family. With all male family members being lawyers, and most of them holding respectable administrative positions, they can probably be associated with the upper layer of late eighteenth- and early-nineteenth century society, especially in their home region of North Brabant.

3 Corpus analysis

3.1 Method

Building on the findings from the analyses of orthographic variables with the *Going Dutch Corpus* presented in Chapters 5–9, three consonantal and two vocalic spelling

⁸⁶ Sibilla Paulina Elisabeth is part of the family correspondence, but only indirectly. Although some letters included in the corpus were signed by *Mimi* & *Paulien*, they were actually written in the hand of her twin sister Eva Maria Adriana (*Mimi*), which can be deduced from the letters Eva Maria Adriana wrote on her own. Therefore, Sibilla Paulina Elisabeth is not listed as an informant in Table 2.

⁸⁷ Surprisingly, de Meij's (2011: 10) otherwise very detailed inventory of the Martini Buys family archives does not mention Anna Maria Emilia: "Uit het huwelijk van Paulus Hubert Martini Buys en Geertruid Johanna Antonia Strick van Linschoten worden drie dochters en een zoon geboren" 'Of the marriage [...] three daughters and one son were born' (also missing in the genealogy, cf. de Meij 2011: 108). However, the baptismal registers in the Amsterdam City Archives clearly attest the birth of Anna Maria Emilia in 1806.

features are investigated with the *Martini Buys Correspondence Corpus*. The five orthographic variables are briefly outlined below:

- (1) **Syllable-final /xt/:** The orthographic representation of the consonant cluster /xt/ in syllable-final position as <cht> or <gt>.⁸⁸ Both variants were officially prescribed by Siegenbeek (1804), but for etymologically distinct groups of words (e.g. <gt> for *bragt* < *brenge* as opposed to <cht> for *kocht* < *kepen*) (cf. Chapter 5).
- (2) **Final /t/ in d-stem verbs:** The orthographic representation of final /t/ in second and third person singular and second person plural present tense indicative forms of verbs with a *d*-stem as either <dt>, <d> or <t>. Siegenbeek (1804) prescribed <dt> as the official spelling variant (cf. Chapter 6).
- (3) **Word-medial and word-final /s/:** The orthographic representation of word-medial and word-final /s/ (< Wgm. *sk>) as <sch> or <s>. The spelling with <sch> was prescribed by Siegenbeek (1804) (cf. Chapter 7).
- (4) **Long e's in open syllable:** The orthographic representation of etymologically distinct long e's in open syllable, traditionally referred to as sharp-long *ê* and soft-long *ē*. Siegenbeek (1804) officialised the phonology-based system with <ee> for sharp-long *ê* in open syllable, and <e> for soft-long *ē* in open syllable (cf. Chapter 8).
- (5) **West Germanic *ī:** The orthographic representation of Wgm. *ī as <ij>, alternatively realised as <ÿ> in handwriting, or <y>. Siegenbeek (1804) officially prescribed the double-dotted spelling <ij>, rejecting the Greek-derived and thus 'foreign' <y> (cf. Chapter 9).

For the corpus analyses of orthographic variables in the *Martini Buys Correspondence Corpus*, the same methods and search queries were used as in the previous analyses with the *Going Dutch Corpus*. For more methodological details, see the corresponding chapters.

3.2 Results

Variable (1): Syllable-final /xt/

Table 3 shows the corpus results for the first variable, i.e. the orthographic representation of syllable-final /xt/ in two categories of words, referred to as *cht*-words and *gt*-words, respectively. The overall results drawn from the *Going Dutch Corpus*, as well as in the sub-corpus of private letters, indicate that Siegenbeek's (1804) division into *cht*-words (with <cht>, due to final devoicing) and *gt*-words

⁸⁸ As the third variant <ght> turned out to be marginal even in the entire *Going Dutch Corpus*, it will not be taken into consideration in this case study.

(with <gt>, due to etymology) was adopted in nineteenth-century language practice. Whereas <gt> was clearly the predominant eighteenth-century spelling for all words with syllable-final /xt/, the newly promoted variant <cht> was successfully established as the prime variant for *cht*-words.

Table 3. Distribution of variants for syllable-final /xt/ in *cht*-words and *gt*-words.

	/xt/ in <i>cht</i> -words				/xt/ in <i>gt</i> -words			
	<cht>		<gt>		<gt>		<cht>	
	N	%	N	%	N	%	N	%
<i>GD Corpus</i>								
P1 (Total)	241	17.7	1,117	82.3	838	79.4	217	20.6
P2 (Total)	987	75.7	316	24.3	992	82.3	213	17.7
P1 (Letters)	99	12.2	711	87.8	338	89.4	40	10.6
P2 (Letters)	470	67.1	230	32.9	363	80.5	88	19.5
GEN. I	23	29.9	54	70.1	30	73.2	11	26.8
Antoni	21	55.3	17	44.7	7	38.9	11	61.1
Eva M. A.	2	5.1	37	94.9	23	100	–	–
GEN. II	95	35.6	172	64.4	117	63.2	68	36.8
Anna M. E. <i>pre-/post-Siegenbeek</i>	–	–	50 31/19	100 100/100	13 5/8	100 100/100	–	–
Paulus H.	57	71.3	23	28.8	32	33.3	64	66.7
Geertruid J. A. <i>pre-/post-Siegenbeek</i>	10 –/10	16.4 –/27.8	51 25/26	83.6 100/72.2	28 11/17	96.6 100/94.4	1 –/1	3.4 –/5.6
Sibilla <i>pre-/post-Siegenbeek</i>	3 –/3	15.8 –/60.0	16 14/2	84.2 100/40.0	21 16/5	91.3 100/71.4	2 –/2	8.7 –/28.6
Hendrik B. <i>pre-/post-Siegenbeek</i>	25 –/25	43.9 –/54.3	32 11/21	56.1 100/45.7	23 5/18	95.8 100/94.7	1 –/1	4.2 –/5.3
GEN. III	91	88.3	12	11.7	48	88.9	6	11.1
Catharina A. G.	32	91.4	3	8.6	19	86.4	3	13.6
Antoni A.	42	85.7	7	14.3	22	100	–	–
Eva M. A.	17	89.5	2	10.5	5	62.5	3	37.5
Anna M. E.	–	–	–	–	2	100	–	–

Generation I: The private letters from the first generation of informants reveal inter-individual differences. Antoni used both <cht> and <gt> in what would become Siegenbeek's *cht*-words and *gt*-words, with a modest preference for <cht> in both groups. A lexically or etymologically conditioned pattern cannot be

attested, though, as *wacht*, *dacht*, *zicht* occur alongside *wagt*, *dagt*, *zigt*. On the other hand, Antoni spelled the rather formulaic *verzogt* consistently with <gt>. The letters by his wife Eva Maria Adriana show an invariable use of <gt> for all words with syllable-final /xt/, except for two tokens with <cht> (*dochter*, *versocht*). However, Nevalainen & Raumolin-Brunberg (2016: 97) argue that “if the individual score falls between 0 and 10, or 90 and 100 [per cent], the person has been considered to have an invariable grammar”, according to which “an occasional occurrence of the minority variant does not change the overall impression”. Following this methodology, Eva Maria Adriana was thus a consistent user of <gt>, both before (1787–1797) and after (1810–1815) the *schrijftaalregeling*.

Generation II: The second generation turns out to be particularly intriguing, revealing considerable variation in the use of variants between the five informants. The differences already become apparent in the letters produced before Siegenbeek (1804). Like her mother Eva Maria Adriana, the eldest daughter Anna Maria Emelia consistently used <gt>, with no single attestation of <cht>. The same invariable usage of <gt> as the only variant is also found in the letters by her siblings Sibilla and Hendrik Bernard as well as her sister-in-law Geertruid Johanna Antonia. Interestingly, Paulus Hubert, the eldest son, already used <cht> alongside the less frequent <gt>, unlike his siblings and his wife, but like his father Antoni. To some extent, his choice of variants seems to be lexically conditioned. Of all 64 attestations of <cht> in the group of *gt*-words, no fewer than 55 tokens are found in words with *REGT*/*RECHT*, and particularly in his fairly formulaic usage of *oprecht* ‘sincere’, which he systematically spelled with <cht>. On the other hand, we find a less consistent spelling of <cht> and <gt> in *opzichte*, *voorzichten* alongside *opzigte*, *voorzigt*, and in *vervachten*, *onverwachte* alongside *verwagten*, *onverwagte*.

How did Siegenbeek’s intervention ‘from above’, i.e. the etymologically motivated split into *cht*-words and *gt*-words, affect this generation of letter writers, who did not acquire the feature during their formative years of childhood and youth? Unfortunately, for Paulus Hubert no private letters seem to be preserved or available from the post-Siegenbeek period. For the remaining family members, however, interesting patterns can be attested. Like her mother Eva Maria Adriana, Anna Maria Emelia continued to use <gt> as the only variant in her nineteenth-century texts from 1806 and 1825. In other words, the official division into *cht/gt* did not affect her spelling practices, as <cht> is completely absent from her texts. Her sister Sibilla, on the other hand, seemed to be aware of the ‘new’ <cht> spelling. Despite an admittedly low number of tokens, Sibilla’s letters from 1810 onward contain both <gt> and <cht>, the latter of which was completely absent in her 1788–1790 texts. Occasionally, she also used the newly acquired <cht> for *gt*-words (*verricht*, *doorzicht*), which might be interpreted as hypercorrect forms. With a higher number of tokens, the same tendency can be confirmed in the post-Siegenbeek letters by sister-in-law Geertruid Johanna Antonia. Only using <gt> in her letters from 1788–1789, her texts written between 1817–1834 see the emergence of <cht> in the category of *cht*-words, co-occurring with the still more frequent <gt>. Similarly, Hendrik Bernard’s letters from the period 1821–1835 also show a considerable increase of <cht>, particularly for *cht*-words (from complete

absence to more than fifty per cent). Nevertheless, in this group of words <cht> (e.g. 13x *dochter*, 8x *echter*) still co-occurred with <gt> (e.g. 8x *wagt*, 7x *kogt*), though not interchangeably with the same lexical items.

While Siegenbeek's rules related to the *cht/gt* issue might not have been applied entirely successfully, i.e. in conformity with the prescribed spelling, the lifespan changes are remarkable. Apart from the eldest sister Anna Maria Emelia, who turned out to be conservative across her lifespan, her siblings Sibilla and Hendrik Bernard as well as sister-in-law Geertruid Johanna Antonia must have acquired <cht> as a new spelling variant even in their late thirties or forties. These findings suggest a high awareness of the officially promoted coexistence of <gt> and <cht>.

Generation III: In the third generation of letter writers, Siegenbeek's division into *cht*-words and *gt*-words was applied most unproblematically. In the case of *cht*-words, <cht> was used in more than 85% by the three siblings Catharina Andrea Geertruid, Antoni Adriaan and Eva Maria Adriana. Furthermore, *gt*-words were also predominantly spelled in line with Siegenbeek, certainly in the letters by Catharina Andrea Geertruid (86.4% <gt>) and Antoni Adriaan (100% <gt>).

Variables (2)–(3): Final /t/ in *d*-stem verbs; word-medial and word-final /s/

Table 4 shows the corpus results for the other two consonantal variables, i.e. the orthographic representation of final /t/ in second and third person singular and second person plural present indicative forms of *d*-stem verbs on the one hand, and the orthographic representation of word-medial and word-final /s/ (<Wgm. *sk>) on the other. In both cases, the results from the *Going Dutch Corpus* indicated a considerable increase of the officially prescribed variants in early nineteenth-century language usage, i.e. <dt> and <sch>, respectively. However, while <sch> had already been established as the predominant variant by the late eighteenth century, there was a striking shift from <d> in the eighteenth century to <dt> in the nineteenth century.

The orthographic representation of final /t/ in particular forms of *d*-stem verbs is one of the relatively low-frequent phenomena in this corpus study, and does not allow a fine-grained interpretation. Despite the limited amount of tokens in the *Martini Buys Correspondence Corpus*, some tendencies can be discussed, though.

Generation I: Although the total number of tokens is hardly representative, all three historical variants occur in the private letters from the first generation. Antoni used <dt> (2x *wordt*) and <d> (*goedvind*), whereas <d> (*soud*, *zend*) and the phonetic spelling with <t> (2x *wort*, *bout*) occur in the letters written by his wife Eva Maria Adriana.

Generation II: The <d> spelling is practically the only variant which was consistently used by all family members from the second generation, both before and after Siegenbeek (1804). The only two instances of <dt> (*wordt*) and <t> (*ondervint*) are found in the letters by Geertruid Johanna Antonia from 1825. Generally, the officialisation of <dt> did not seem to have affected the second

Martini Buys generation at all, since they did not alter their spelling in the nineteenth century. To some extent, this dominance of <d> might be related to the regional origin of the informants. The members of the first and second generations were all born in 's-Hertogenbosch in the region of North Brabant, where, according to the general findings based on the *Going Dutch Corpus*, <d> was relatively dominant in the late eighteenth century and remained a strong second variant in the early nineteenth century (cf. Chapter 6).

Table 4. Distribution of variants for final /t/ in *d*-stem verbs and word-medial/final /s/.

	Final /t/ in <i>d</i> -stem verbs						Word-medial and word-final /s/			
	<dt>		<d>		<t>		<sch>		<s>	
	N	%	N	%	N	%	N	%	N	%
<i>GD Corpus</i>										
P1 (Total)	62	17.1	258	71.1	43	11.8	607	72.7	228	27.3
P2 (Total)	293	66.3	138	31.2	11	2.5	690	95.0	34	5.0
P1 (Letters)	34	18.8	108	59.7	39	21.5	237	56.6	182	43.4
P2 (Letters)	133	56.8	91	38.9	10	4.3	299	92.9	23	7.1
GEN. I	2	25.0	3	37.5	3	37.5	31	88.6	4	11.4
Antoni	2	66.7	1	33.3	–	–	21	100	–	–
Eva M. A.	–	–	2	40.0	3	60.0	10	71.4	4	28.6
GEN. II	1	1.4	72	97.3	1	1.4	110	96.5	4	3.5
Anna M. E.	–	–	9	100	–	–	14	87.5	2	12.5
Paulus H.	–	–	28	100	–	–	28	100	–	–
Geertruid J. A.	1	5.9	15	88.2	1	5.9	30	100	–	–
Sibilla	–	–	14	100	–	–	19	95.0	1	5.0
Hendrik B.	–	–	6	100	–	–	19	95.0	1	5.0
GEN. III	13	48.1	14	51.9	–	–	45	97.8	1	2.2
Catharina A. G.	7	41.2	10	58.8	–	–	21	100	–	–
Antoni A.	6	60.0	4	40.0	–	–	17	100	–	–
Eva M. A.	–	–	–	–	–	–	7	87.5	1	12.5
Anna M. E.	–	–	–	–	–	–	–	–	–	–

Generation III: For the third generation, again, the number of tokens is too low for a detailed interpretation. What we can see, though, is that both

Catharina Andrea Geertruid and her brother Antoni Adriaan used <d> and <dt> as more or less equally frequent variants. The inconsistent spelling of the high-frequency verb *worden* as *wordt* (6 tokens) and *word* (11 tokens) in both Catharina Andrea Geertruid's and Antoni Adriaan's letters possibly indicates a limited awareness (or acceptance) of <dt> as the prescribed norm.

Coming to the orthographic representation of /s/ (derived from West Germanic *ʃk) in word-medial and word-final position, the results from the *Martini Buys Correspondence Corpus* reveal <sch> as the prevalent variant across all three generations.

Generation I: Antoni consistently spelled <sch> in all instances, whereas there is some more variation in the letters written by his wife Eva Maria Adriana, with <sch> (71.4%) coexisting alongside <s> (four tokens), all of which are attestations of *tussen*.

Generation II: Similar to the case of <d> in the previous variable, there was a pronounced preference for one variant, i.e. <sch>, across all members of the second generation. Only a few tokens with <s> are occasionally found in texts by Anna Maria Emelia (2x *tussen*), Sibilla (*vis*) and Hendrik Bernard (*gevenste*). Whereas the general results drawn from the sub-corpus of private letters indicate a much higher share of <s>, the Martini Buys family had developed a clear preference for <sch>.

Generation III: In line with Siegenbeek's prescription, but at the same time also continuing the practices of the previous generation, the third generation invariably used <sch> as the only variant. Only one token of <s> (*tussen*) can be attested in a letter written by Eva Maria Adriana.

Variable (3): Long *e*'s in open syllable

Table 5 shows the distribution of spelling variants representing sharp-long *ê* and soft-long *ē* in open syllables in the *Martini Buys Correspondence Corpus*. According to the findings from the *Going Dutch Corpus*, <ee> had been established as the main variant for sharp-long *ê* in open syllable by the late eighteenth century, whereas <e> and <ee> were co-occurring variants for soft-long *ē* in open syllable. Siegenbeek (1804) officialised the phonology-based system, i.e. sharp-long <ee> and soft-long <e>, which was successfully adopted in the nineteenth-century community at large.

Generation I: For words with sharp-long *ê* in open syllable, both Antoni and Eva Maria Adriana almost exclusively used the digraph <ee>, with the occasional token of <e> (three and one, respectively). For words with soft-long *ē* in open syllable, the single grapheme <e> was predominantly used. It appears that the historical-phonological distinction of sharp-long *ê* and soft-long *ē* was reflected in the spelling practices of the two informants from the first generation. Interestingly, the distinction was most accurately applied in Eva Maria Adriana's letters (86.4% <e> for soft-long *ē*), whereas her husband shows somewhat more variation (67.1% <e>, 32.9% <ee>).

Table 5. Distribution of variants for sharp-long *ê* and soft-long *ē* in open syllable.

	Sharp-long <i>ê</i>				Soft-long <i>ē</i>			
	<ee>		<e>		<ee>		<e>	
	N	%	N	%	N	%	N	%
<i>GD Corpus</i>								
P1 (Total)	1,364	90.1	150	9.9	1,664	59.5	1,133	40.4
P2 (Total)	1,806	91.5	167	8.5	3,043	92.5	246	7.5
P1 (Letters)	528	88.3	70	11.7	728	50.7	707	49.3
P2 (Letters)	654	86.4	103	13.6	1,419	88.6	182	11.4
GEN. I	58	93.5	4	6.5	125	76.7	38	23.3
Antoni	34	91.9	3	8.1	55	67.1	27	32.9
Eva M. A.	24	96.0	1	4.0	70	86.4	11	13.6
GEN. II	268	94.7	15	5.3	475	63.0	279	37.0
Anna M. E. <i>pre-/post-Siegenbeek</i>	22 11/11	100 100/100	–	–	13 11/2	14.6 20.4/5.7	76 43/33	85.4 79.6/94.3
Paulus H.	117	97.5	3	2.5	224	80.3	55	19.7
Geertruid J. A. <i>pre-/post-Siegenbeek</i>	29 11/18	100 100/100	–	–	13 6/7	11.1 10.0/12.3	104 54/50	88.9 90.0/87.7
Sibilla <i>pre-/post-Siegenbeek</i>	35 18/17	94.6 94.7/94.4	2 1/1	5.4 5.3/5.6	45 13/32	54.2 32.5/74.4	38 27/11	45.8 67.5/25.6
Hendrik B. <i>pre-/post-Siegenbeek</i>	65 13/52	86.7 76.5/89.7	10 4/6	13.3 23.5/10.3	180 33/147	96.8 91.7/98.0	6 3/3	3.2 8.3/2.0
GEN. III	91	95.8	4	4.2	235	93.6	16	6.4
Catharina A. G.	34	97.1	1	2.9	94	94.0	6	6.0
Antoni A.	51	96.2	2	3.8	90	91.8	8	8.2
Eva M. A.	5	83.3	1	16.7	41	95.3	2	4.7
Anna M. E.	1	100	–	–	10	100	–	–

Generation II: Like in the previous generation, and also very much in line with the general situation in the wider community, words with sharp-long *ê* were primarily spelled with <ee> across all family members. Compared to the practically invariable usage of <ee> in the letters by Anna Maria Emelia (100%), Paulus Hubert (97.5%), Geertruid Johanna Antonia (100%) and Sibilla (94.6%), there are slightly more attestations of <e> for sharp-long *ê* in Hendrik Bernard's letters (89.7%). The relative distribution of variants, with a strong prevalence of <ee>, is generally stable across the pre- and post-Siegenbeek periods.

For words with soft-long $\bar{e}$, considerably more variation can be attested, involving <e> and <ee> as co-occurring variants, which is also in line with the general findings drawn from the *Going Dutch Corpus*. In the letters written by Anna Maria Emelia and Geertruid Johanna Antonia, <ee> is the preferred variant, occurring in 85.4% and 88.9%, respectively. This means that these two female writers did not distinguish sharp-long and soft-long $\bar{e}$'s by spelling and apparently applied the alternative syllabic system (i.e. <ee> in open syllable). In contrast to Anna Maria Emelia and Geertruid Johanna Antonia, the two male informants of the second generation used the <e> spelling for soft-long $\bar{e}$. Paulus Hubert (80.3%) and his younger brother Hendrik Bernard (96.8%) thus applied the phonology-based system. Their sister Sibilla appears to take an intermediate position, using <e> (54.2%) alongside <ee> (45.8%) for soft-long $\bar{e}$.

Zooming in on the diachronic changes in the letter data from before and after Siegenbeek (1804), more interesting patterns become visible. While the brothers Paulus Hubert and Hendrik Bernard already made a distinction between <ee> for sharp-long $\acute{e}$ and (primarily) <e> for soft-long $\bar{e}$ in the late eighteenth century, different developments can be witnessed in the letters by their female contemporaries of the second generation. Both Anna Maria Emelia and Geertruid Johanna Antonia continued to apply the syllabic system with <ee> in open syllables. Evidently, the officialised phonology-based distinction into <ee> and <e> did not affect their spelling practices after 1804.

Remarkably, we do see changes in the direction of Siegenbeek's prescription in the letters written by their sister(-in-law) Sibilla. In line with the official spelling norms, she shifted from <ee> (67.5%) as the main variant for soft-long $\bar{e}$ in her letters from 1788–1790 to <e> (increasing from 32.5% to 74.4%) in her letters from 1810–1825. Although <e> was part of her pre-Siegenbeek practices, she must have acquired the phonology-based system even beyond her formative years. Inconsistent spellings, for instance *deze/deeze*, *mede/meede*, co-occurred across her lifespan, though.

Generation III: The phonology-based system as prescribed by Siegenbeek (1804) was consistently applied in the letters from the third generation. Catharina Andrea Geertruid and Antoni Adriaan use <ee> for sharp-long $\acute{e}$ in 97.1% and 96.2%, respectively, and <e> for soft-long $\bar{e}$ in 94.0% and 91.8%, respectively. Even though the number of tokens is lower in the case of their younger sisters Eva Maria Adriana and especially Anna Maria Emilia, it can be seen that they also spelled according to the phonology-based system. It should be taken into account that all members of the fourth generation were born and raised in Amsterdam, where the phonological distinction between sharp-long $\acute{e}$ and soft-long $\bar{e}$ had already merged centuries earlier. This clearly supports the conclusion that the orthographic distinction grounded on phonology must have been acquired as a direct or indirect result of Siegenbeek's prescription.

Variable (5): West Germanic *ī

Table 6 shows the relative distribution of spelling variants representing Wgm. *ī, comprising the double-dotted forms <ij> and <ÿ>, the undotted <y> and alternative forms. In the *Going Dutch Corpus*, the distribution of dotted and undotted variants remained surprisingly stable after Siegenbeek (1804) had prescribed <ij> as the national variant. In fact, only newspapers shifted from <y> in the eighteenth century to <ij> in the early nineteenth century, whereas there was clearly more resistance in adopting the double-dotted spelling among writers of ego-documents.

Table 6. Distribution of variants for Wgm. *ī.

	West Germanic *ī							
	<ij>		<ÿ>		<y>		Other	
	N	%	N	%	N	%	N	%
GD Corpus								
P1 (Total)	322	3.6	3,156	35.3	4,668	52.2	800	8.9
P2 (Total)	989	10.1	3,369	34.4	4,934	50.4	497	5.1
P1 (Letters)	178	3.3	2,621	48.7	2,259	42.0	325	6.0
P2 (Letters)	127	2.1	2,458	40.4	3,172	52.2	320	5.3
GEN. I	–	–	203	50.0	189	46.6	14	3.4
Antoni	–	–	203	97.1	5	2.4	1	0.5
Eva M. A.	–	–	–	–	184	98.4	3	1.6
GEN. II	66	3.2	703	33.6	1,005	48.0	318	15.2
Anna M. E. <i>pre-/post-Siegenbeek</i>	–	–	–	–	211 118/93	100 100/100	–	–
Paulus H.	–	–	624	91.2	4	0.6	56	8.2
Geertruid J. A. <i>pre-/post-Siegenbeek</i>	–	–	–	–	532 269/263	99.8 100/99.6	1 0/1	0.2 –/0.4
Sibilla <i>pre-/post-Siegenbeek</i>	66 –/66	23.1 –/57.9	45 –/45	15.7 –/39.5	174 172/2	60.8 100/1.8	1 –/1	0.3 –/0.9
Hendrik B. <i>pre-/post-Siegenbeek</i>	–	–	34 27/7	9.0 24.1/2.6	84 21/63	22.2 188/23.7	260 64/196	68.8 57.1/73.7
GEN. III	1	0.1	431	53.3	219	27.1	158	19.5
Catharina A. G.	–	–	210	58.0	3	0.8	149	41.2
Antoni A.	1	0.3	198	61.3	115	35.6	9	2.8
Eva M. A.	–	–	5	4.8	100	95.2	–	–
Anna M. E.	–	–	18	94.7	1	5.3	–	–

Generation I: The two informants of the first generation had distinct individual spelling preferences with regard to this feature. Antoni preferred the double-dotted <ÿ> (97.1%), whereas his wife Eva Maria Adriana almost exclusively used the undotted <y> (98.4%). In other words, there is practically no intra-individual variation, but a consistently applied choice for one variant.

Generation II: The idiosyncratic nature of this spelling variable becomes even more apparent in the letters written by the second generation. Anna Maria Emelia and Geertruid Johanna Antonia invariably used <y> in their letters, both before and after Siegenbeek (1804). On the other hand, Paulus Hubert, like his father Antoni, primarily used <ÿ> (91.2%) already in the late eighteenth century. His younger brother Hendrik Bernard was far less consistent in the use of the double-dotted spelling. His letters from 1789–1797, written during his twenties, contain <ÿ> (24.1%) alongside <y> (18.8%) but most frequently the idiosyncratic <ÿ̇> with some kind of acute accent (57.1%). In his post-Siegenbeek letters from 1821–1835, written in his fifties and sixties, the accented <ÿ̇> had become his prevalent variant (73.7%), whereas the undotted <y> is found in 23.7%. Interestingly, the officially prescribed spelling decreased to a marginal 2.6%.

In terms of normative influence, the most remarkable results are found in the private letters written by Sibilla. Examining her entire data set, one could easily get the impression that she used the double-dotted variants <ij> and <ÿ> alongside the undotted <y>. On closer inspection, though, it appears that Sibilla systematically shifted from the undotted to the double-dotted spelling across her lifespan. In her letters from 1788–1790, written in her early to mid-twenties, <y> invariably occurs in 100% of all instances. In the letters from 1810–1825, between her mid-forties to the age of sixty, Sibilla's choice of variants radically changes in the direction of Siegenbeek's prescriptions, using either <ij> or <ÿ> in 97.4% of all instances. Except for two occasional attestations, her previous <y> spelling no longer occurs, which underscores her awareness of the officialised norm.

At least within this particular family correspondence, the results for the late eighteenth century signal gender as a factor that conditioned the choice of variants. While all four female informants from the first and second generation used <y> in their private letters, father Antoni and his eldest son Paulus Hubert were users of <ÿ>. Hendrik Bernard further developed an idiosyncratic preference for the accented <ÿ̇>.

Generation III: The spelling of Wgm. **i* continued to be largely dependent on idiosyncratic preferences even in the post-Siegenbeek generation, with <ij> being the officially prescribed variant. Catharina Andrea Geertruid varied between the double-dotted <ÿ> (58.0%) and an alternative form with diacritics somewhere between <ÿ̇> and <ÿ̈> (41.2%). The latter, however, can probably be interpreted as a more 'sloppy' version of the 'neat' double-dotted <ÿ>, by connecting the two dots. Apparently the undotted <y> was no option for Catharina Andrea Geertruid, as there are only three attestations of the rejected variant in her letters. Her brother Antoni Adriaan inconsistently used both <ÿ> (61.3%) and <y> (35.6%), alternating between the two variants even within the same texts. His awareness of the prescribed double-dotted spelling must have been

limited, which might be somewhat surprising considering his academic background and his work as a lawyer, both suggesting familiarity and proficiency with standard language norms. His sister Eva Maria Adriana, in contrast, spelled <y> in 95.2% of all cases, against the prescribed norm. Despite the considerable lower number of tokens, it appears that the youngest sister Anna Maria Emilia primarily used <ÿ> (94.7%). In sum, the findings for the *ij/y* issue testify to a high degree of inter-individual variation between the siblings of the third generation. Although they probably acquired <ij> during their formative years, it seems that the awareness of this double-dotted spelling as the official variant was not overly prominent, or at least not consistently adopted in private letter writing practices.

As already discussed, the comparatively strong persistence of the rejected <y> in the nineteenth century could be related to the very nature of this orthographic variable, involving a diacritic rather than a ‘proper’ graphemic representation. It was also discussed by Vosters et al. (2010: 99), who consider the minimal (diacritic) difference between the variants in handwriting as a possible explanation for the lack of change. In Chapter 9, I argued that the presence or absence of the two dots might not have been as salient to language users as it was to Siegenbeek and many of his eighteenth-century predecessors. What the results from this case study mainly indicate is a high degree of inter-individual variation, i.e. idiosyncratic but mostly consistent choices of a particular variant, irrespective of standard norms, and to a considerably lesser extent intra-individual variation (particularly Hendrik Bernard and Antoni Adriaan). Remarkably, even the siblings from the third generation, when <ij> had been officially prescribed, have idiosyncratic preferences. On the other hand, it should be emphasised that the case of Sibilla also testifies to a conscious shift towards the official norm, systematically replacing <y> by the prescribed double-dotted spelling.

4 Discussion

The case study presented in this chapter focused on inter- and intra-individual variation and change in the *Martini Buys Correspondence Corpus*, a specifically compiled corpus of private correspondence from the Martini Buys family archives, spanning the second half of the eighteenth and the first half of the nineteenth century. Taking a micro-level approach to individual behaviour, and investigating five orthographic variables, a number of (partly interrelated) research questions could be addressed.

To begin with, one interesting aspect that could hardly be examined with the large-scale data set of the *Going Dutch Corpus* concerns the consistency of individual spelling preferences and practices. In other words, did individual writers use coexisting variants inconsistently, or did they (more or less) consistently choose one particular variant? While the presented findings from three generations proved to be too diverse to be summed up concisely, it appears that most informants tend to have preferences for particular variants. In fact, the second generation, particularly in the pre-Siegenbeek texts, showed a fairly high degree of consistency

in spelling practices. In the period before the official spelling regulation, the three female members Anna Maria Emelia, Geertruid Johanna Antonia and Sibilla invariably used <gt>, <d>, <sch>, <y> and, except for Sibilla, <ee> in open syllables. Paulus Hubert was consistent in his use of <d>, <sch>, <ÿ> and more or less consistently applied the phonology-based system with <ee> and <e>. His brother Hendrik Bernard invariably spelled <gt>, <d>, <sch>, but was more inconsistent in the use of the two vocalic variables, especially with regard to undotted, double-dotted and accented variants of the *ij/y* variable.

The third generation, exposed to the official Siegenbeek norms during their formative years, turned out to be consistent in the use of <sch>, <cht> for *cht*-words and <gt> for *gt*-words, <ee> for sharp-long *ê* and <e> for soft-long *ē* in open syllables (according to the phonology-based system). Interestingly, in the case of final /t/ in forms of *d*-stem verbs, it appeared that the inconsistent use of variants only started after 1804/1805. Both Catharina Andrea Geertruid and Antoni Adriaan used <dt> and <d> as co-occurring variants, whereas the previous generation invariably used <d>. This pattern possibly underlines the transitional character of the first half of the nineteenth century, as both the former main variant <d> and the officialised <dt> were in use. Similarly, Antoni Adriaan interchangeably used the prescribed <ÿ> alongside the rejected <y>. His sister Catharina Andrea Geertruid did not use the undotted variant, but often varied between a neatly double-dotted <ÿ> and a more sloppy <ÿ̃>/<ÿ̂>. Their younger sister Eva Maria Adriana, on the other hand, almost invariably used <y>, against the Siegenbeek norm, underscoring the inter-individual differences in consistency with regard to the *ij/y* variable.

Furthermore, this chapter addressed the issue of individual lifespan change as a possible effect of language policy. By zooming in on those informants of the Martini Buys family from which private letters written before and after Siegenbeek's (1804) orthography have been preserved, this case study also shed light on the impact of top-down language policy on spelling practices of adult language users, whose formative years had been completed years or even decades before the *schrijftaalregeling* was introduced. The general results drawn from the *Going Dutch Corpus*, at least for the most orthographic variables (cf. Chapters 5–9), indicated a communal shift, i.e. a general change of spelling practices in the direction of the officially prescribed norms in the community at large. It raised the question whether and to what extent language users in adulthood participated in this change. Did they acquire 'new' variants irrespective of transmission through (school) education?

For the informant from the first generation, Eva Maria Adriana Buys, who was around seventy years old when the language policy was introduced, no changes across lifespan could be witnessed. In all five cases, she maintained her spelling preferences until the early nineteenth century. Her eldest daughter Anna Maria Emelia as well as her daughter-in-law Geertruid Johanna Antonia, both from the second generation, did not alter their spelling across lifespan either. These three cases of (female) writers, in fact, testify to the pattern of generational change, "in which there is idiolectal stability despite ongoing change in the community"

(Raumolin-Brunberg 2009: 171). In other words, despite the top-down language policy and official spelling norms, the use of spelling variants in the letters by Eva Maria Adriana, Anna Maria Emelia and (except for the fairly cautious adoption of <cht>) Geertruid Johanna Antonia remained stable, as “the linguistic forms acquired in childhood remain unchanged” (ibid.).

The second generation also offers two interesting counterexamples, though, giving evidence of lifespan changes in the spelling practices of Hendrik Bernard and Sibilla. Whereas the use of <d> remains stable (i.e. not changing to prescribed <dt>), the dynamics in the use of the other features shed some more light on the effectiveness and awareness of Siegenbeek’s (1804) official norms among adult language users. First of all, Hendrik Bernard acquired <cht> as a new variant, which had been absent in his eighteenth-century letters, but co-occurred with <gt> in his texts from the 1820s and 1830s. Although he worked as a lawyer and held various administrative functions in North Brabant, which presupposes familiarity with standard language norms, he did not adopt <ij> (or <ÿ>) as the main variant, but rather developed an accented form <ÿ> in his private letters – possibly his idiosyncratic representation of <ÿ>? While this is beyond the scope of this case study, it might be interesting to see whether <ÿ> (or a more properly double-dotted form) also occurs in Hendrik Bernard’s professional documents, for example business letters. Remarkably, his sister Sibilla, unlike her female contemporaries, consciously altered her spelling practices in the direction of the official Siegenbeek norms. In fact, these changes could be witnessed in three cases (with varying ‘success rates’). Firstly, Sibilla, like her brother, adopted <cht> as a new variant alongside <gt>. Secondly, she consolidated <e> for soft-long *ē* as opposed to <ee> for sharp-long *ē*. Thirdly, she radically shifted from <y> before Siegenbeek to <ij>/<ÿ> after Siegenbeek.

These lifespan changes in the private letters by Hendrik Bernard and, even more strikingly, Sibilla, must testify to a considerable awareness of spelling norms even among adult language users in the early nineteenth century – beyond the direct acquisition through education. A possible explanation for Sibilla’s adoption of official spelling variants has to be found in different means of contact with these norms. As the results in Chapters 5–9 unambiguously attested the use of Siegenbeek’s prescribed variants in nineteenth-century newspapers, it could be assumed that adult language users like Sibilla Martini acquired their knowledge of spelling norms through the reading of newspapers and other sorts of published writing. In these texts, readers were exposed to the invariable use of spelling variants in conformity with official prescriptions, which possibly raised their awareness for the newly promoted orthographic conventions regardless of formal (school) education.

Addressing the awareness of forms, Raumolin-Brunberg (2009: 173) emphasises that the direction of change (in Labovian terms) should be taken into account, arguing that “[i]t may be a different matter to adopt changes from below, i.e., shifts that emanate from below the level of social awareness, and changes from above, i.e., shifts that stem from prestigious sources, often acquired with full public awareness”. At least some level of public awareness of official spelling norms must

have been present in the case of the language policy of the Northern Netherlands, although the spelling regulations reached – or affected – by no means all informants of the transitional generation(s).

Finally, how do the micro-level findings drawn from this case study further enrich the interpretation of variation and change witnessed in the community at large, i.e. with the multi-genre *Going Dutch Corpus*? What might be perceived as a methodological drawback is the inclusion of adult language users in the nineteenth-century cross-section whose formative years had been completed long before the introduction of the *schrijftaalregeling* in 1804/1805. Although adult language users in the early nineteenth century were not directly exposed to the national language policy through education and thus during their formative years, this chapter has shown that they could still participate in the language change (from above), altering their spelling practices across lifespan. In fact, individual informants of the second generation, and the case of Sibilla Martini in particular, adopted the official spelling norms in adulthood at a later point in life. While the impact of the top-down written language regulation must have been of a more indirect kind – compared to the younger generation of writers, who were exposed directly to the language-in-education policy – the results indicate that the effectiveness of the national language policy can actually be examined even in texts produced by the adult generation. It is true that conservative individuals from the pre-*schrijftaalregeling* generation, for instance Anna Maria Emelia and Geertruid Johanna Antonia in this case study, might skew the overall results to some extent. On the other hand, this case study has shown that even individuals from the post-*schrijftaalregeling* generation sometimes preferred and used variants which clearly deviate from the official Siegenbeek norms. Summing up, I therefore argue that the ‘older’ generation of writers should be taken into account as a possible factor conditioning the results of the nineteenth-century data, but are ultimately best treated as a legitimate part of the community at large and, more concretely, of the nineteenth-century cross-section of the *Going Dutch Corpus*.

Conclusion

1 Introduction

In the present dissertation, I examined the effectiveness of the Dutch national language policy introduced in the early 1800s on language practices in the Northern Netherlands. Analysing the newly compiled *Going Dutch Corpus*, a diachronic multi-genre corpus of late eighteenth- and early nineteenth-century Dutch, it was the central aim to test and assess the impact of the so-called *schrijftaalregeling* ‘written language regulation’ on actual usage. The official regulations for spelling (Siegenbeek 1804) and grammar (Weiland 1805) were commissioned by the Dutch government, intended to be used in the educational and administrative domains. The question at the heart of this dissertation was whether and to what extent the top-down language policy measures exerted influence on patterns of language variation and change. In other words, how successful was the *schrijftaalregeling* in spreading the officialised norms for the Dutch standard variety across the community at large? Did the citizens of the young nation follow Siegenbeek’s and Weiland’s prescriptions, as envisaged by the government?

In this concluding chapter, I bring together the most relevant findings drawn from all empirical case studies in order to provide an overarching assessment of the central research objectives, discussing striking patterns and differences, and ultimately determining policy effectiveness. Section 2 briefly discusses the results from eight orthographic and morphosyntactic variables, which were analysed on the basis of the *Going Dutch Corpus*. Particularly in the case of morphosyntax, external factors such as genre appeared to play an important role. The variational dimensions of genre, space, gender and the individual will be further discussed in Section 3, highlighting the major findings for each of these external variables. Section 4 then assesses the changing relationship between language norms and language use in the period under investigation, seeking to determine the possible success of the Dutch language policy. Finally, Section 5 contains the concluding remarks and an outlook for future research.

2 Measuring policy success: Orthography and morphosyntax

In order to measure and assess the effectiveness of Siegenbeek’s (1804) orthographic and Weiland’s (1805) grammatical prescriptions, eight linguistic variables were investigated in Chapters 5–12. Combining quantitative and qualitative methods by taking into account actual language use (*Going Dutch Corpus*) and metalinguistic discourse (*Normative Corpus of the Northern Netherlands*) (Chapter

4), this dissertation presented a systematic method to measure the potential impact of top-down language policy on usage patterns. The selection of case studies comprised five orthographic variables (Chapters 5–9), followed by three morphosyntactic variables (Chapters 10–12).

Orthographic variables

Starting with orthographic issues, three consonantal and two vocalic features were examined, all of which were debated in the eighteenth-century normative tradition and officially regulated in the national orthography by Siegenbeek (1804).

The first case study focused on the orthographic representation of syllable-final /xt/ in etymologically distinct words (Chapter 5). In the eighteenth-century normative tradition, both <cht> and <gt> were already acknowledged as (co-existing) spelling variants, although consensus with respect to the spelling of individual words was still limited. Siegenbeek then officialised a purely orthographic split into two categories of words with /xt/, which I referred to as *cht*-words (to be spelled with <cht> due to final devoicing, e.g. *kocht*) and *gt*-words (to be spelled with <gt> due to etymology, e.g. *bragt* < *brenge*). The corpus analysis showed that in late eighteenth-century usage, <gt> was by far the most frequent variant, occurring in around 80% of all cases. Against this clear preference in contemporary spelling practices, Siegenbeek's fairly intricate rule was successfully adopted by nineteenth-century language users. In fact, the results from the post-Siegenbeek period displayed a neat distribution of <cht> for *cht*-words, and <gt> for *gt*-words. These changes signalled a high awareness of norms related to the *cht/gt* issue, particularly the newly promoted variant <cht>. The results even revealed an increase of hypercorrect forms with <cht> for *gt*-words, especially among women and in private letters. For the most part, however, Siegenbeek's prescribed categorisation into *cht*- and *gt*-words was applied across all genres, regions and genders.

The second case study investigated the orthographic representation of final /t/ in particular forms of verbs with a *d*-stem, such as *worden* and *vinden* (Chapter 6). Three historical spelling variants were considered, viz. <t> (phonetic), <d> (morphological) and <dt> (analogical/morphological, but also etymological < verbal ending *-det*). The latter and most complex variant was already preferred in most eighteenth-century normative works, and also prescribed by Siegenbeek. Late eighteenth-century usage, however, did not reflect the widespread metalinguistic preference in favour of <dt>. The corpus results for this period, in fact, showed that <d> was used in more than 70% of all instances, implying a discrepancy between norms and usage. This radically changed in the early nineteenth century, when a shift from <d> to <dt> as the prevalent variant took place, strongly suggesting Siegenbeek's normative influence on actual usage. While newspapers invariably adopted <dt>, <d> remained a relatively common alternative spelling in ego-documents. Nevertheless, the general increase of <dt> was witnessed across all genres, regions and genders.

In the third case study, I examined the orthographic representation of word-medial and word-final /s/ (< Wgm. *sʰ) as <sch> or <s>, for instance in *tuss(ch)en* and *wens(ch)* (Chapter 7). Throughout the eighteenth century, most linguistic commentators already had a strong preference for the spelling <sch>, which corresponded with the overall corpus results, where <sch> was attested as the main variant in more than 70%. However, private letters from this period revealed much more variation, with <sch> and <s> co-existing as two almost equally frequent variants. Furthermore, it was shown that women even preferred <s> over <sch>, testifying to a fairly diverse picture in eighteenth-century usage. Siegenbeek, following his normative predecessors, officially prescribed <sch>, which further consolidated its prevalence in language use, while <s> became a marginal variant even in private letters and among women. The unrivalled position of <sch> was somewhat surprising with regard to the changes that had taken place in spoken language. Although /sx/ was no longer pronounced in medial and final position, the spread of the prescribed ‘old’ spelling <sch> was extremely successful across all genres, regions and genders. Again, these changes could be related to normative influence.

The fourth case study tackled the heavily debated spelling of long vowels, focusing on the orthographic representation of etymologically distinct long *ē*’s in open syllable (Chapter 8). The distinction between so-called sharp-long *ē* and soft-long *ē* had disappeared in many regions, including the wider Amsterdam area, while it had been preserved in many dialects of, for instance, Zeeland and Groningen. This resulted in a highly variable situation with various competing writing systems, both in eighteenth-century norms and usage. Regardless of the merger in his native Amsterdam area, Siegenbeek prescribed the phonology-based system, according to which the etymological difference was represented in spelling, viz. by sharp-long <ee> and soft-long <e> in open syllables. While the corpus results for sharp-long *ē* showed a strong prevalence of the digraph <ee> in 90%, both before and after Siegenbeek, the results for soft-long *ē* revealed striking developments. In eighteenth-century usage, there was a modest preference for <e>, although <ee> occurred frequently as well. However, after <e> had been prescribed as the official spelling for soft-long *ē*, the use of the single grapheme increased dramatically, resulting in a neat distribution of <ee> for sharp-long and <e> for soft-long in more than 90% each – perfectly in line with the phonology-based system. Although it went against the grain of several eighteenth-century patterns and tendencies, Siegenbeek’s prescription was successfully adopted across all genres, regions and genders.

While the results from the previous four orthographic cases all signalled a strong normative effect on nineteenth-century usage patterns, the fifth case study revealed its limitations. Examining the orthographic representation of Wgm. **ī*, as in *mijn/myn* and *schrijven/schryven* (Chapter 9), it was shown that Siegenbeek’s prescription was less effective compared to the other investigated spelling issues. The choice for either <ij> or <y> was widely discussed among eighteenth-century commentators, but it was only in the final decades of the century that <ij> emerged as the preferred variant. This trend in normative discourse paved the way

for Siegenbeek, who officialised <ij> as the national variant. On the one hand, the corpus results did confirm a complete shift from <y> to <ij> in newspapers, i.e. in accordance with the prescription. On the other hand <ij>, or rather <ÿ>, did not gain ground in handwriting practices, neither in private letters nor in diaries and travelogues. In these texts, (prescribed) double-dotted and (rejected) undotted variants continued to coexist into the nineteenth century, which could possibly be explained by the close similarities between the forms, i.e. a relatively minor diacritic distinction, especially in handwriting (cf. also Vosters et al. 2010: 99). Zooming in on individual spelling practices in a corpus of family correspondence (Chapter 13), a relatively high degree of inter-individual variation of the *ij/y* issue could be attested due to idiosyncratic preferences, probably indicating that the awareness of the prescribed spelling was comparatively limited.

To sum up the results of five orthographic case studies, a considerable effect of Siegenbeek's official orthography on early nineteenth-century usage was identified. It is important to note that Siegenbeek's choice of spelling variants was, for the most part, neither innovative nor radical. In fact, he largely followed – and officialised – already existing preferences of the (late) eighteenth-century normative tradition, most notably Kluit (1763, 1777). Against the background of this more or less stable situation in metalinguistic discourse, the remarkable changes in actual language practices most likely took place under influence of the national language policy, in this case the official 1804 orthography. The striking convergence towards Siegenbeek's prescriptions (except for the case of <ij>) in the early nineteenth century was consistently observed across all three genres, in all seven regions, and in texts by both men and women.

Morphosyntactic variables

In addition to the five orthographic variables discussed in Chapters 5–9, three morphosyntactic variables were investigated in order to examine the possible normative effects of Weiland's (1805) national grammar on language practices.

Chapters 10 and 11 presented two case studies on variation and change in the Dutch relativisation system, focusing on the neuter relative pronoun in subject and object position, and the masculine and feminine singular and plural relative pronoun in subject and object position, respectively. In eighteenth-century metalinguistic discourse, relativisation was not a core topic and specific rules for the use of relativisers were generally sparse. In fact, Weiland (1805) appeared to be the first grammar providing relatively elaborate and explicit information on relativisation, from which a number of prescriptive rules could be inferred. Diachronically, however, the corpus results for both case studies on relativisation displayed a remarkably stable distribution of variants, suggesting that Weiland's direct influence on nineteenth-century usage must have been limited. On closer inspection, certain patterns and developments could be observed, though.

The investigation of neuter relative pronouns revealed a highly variable situation with no less than five variants, viz. the traditional *d-* and *w-*forms *dat* and *wat*, as well as the additional pronominal forms *betgeen*, *betwelke* and *welke*. All variants

occurred in the corpus results, both for the eighteenth- and the nineteenth-century period, although *welk* turned out to be a relatively marginal form. In the case of masculine and feminine singular and plural relative pronouns, the two most common variants in usage were the traditional *d*-form *die* and the additional pronominal form *welke*, whereas the *w*-form *wie* and the extended pronominal *develke* rarely occurred in the corpus.

To begin with, two internal factors were tested on the *Going Dutch Corpus* data, viz. the definiteness of the antecedent (neuter forms) and the animacy of the antecedent (masculine/ feminine/plural forms). Against traditional assumptions, the corpus analyses gave evidence that these factors did not crucially condition the distribution of variants. The decisive factors were external, and most notably related to genres, which are discussed further in Section 3. In fact, the results across genres revealed that the additional pronominal forms *betwelk* and *welke* were considerably more frequent in newspapers than in private letters, where, in turn, the *d*-forms *dat* and *die* were more commonly used. The genre of diaries and travelogues took a special intermediate position in the corpus design between private letters and newspapers. From a diachronic perspective, a genre-specific evolution in the use of relative pronouns could be identified. In the late eighteenth century, the distribution of variants in diaries and travelogues was fairly similar to that in private letters. In the early nineteenth century, however, the considerable increase of *betwelk* and *welke* (at the expense of the *d*-forms *dat* and *die*) led to a divergence of diaries and travelogues from private letters. At the same time, these sources converged towards the distribution found in newspapers, signalling an evolution towards a more formal, typically ‘written’ style, as opposed to the comparatively informal, more ‘oral’-like private letters. Furthermore, both case studies on relative pronouns revealed a considerable amount of gender variation. Men tended to use the ‘solemn’ forms more frequently than women, whereas the common forms *dat/wat* and *die* occurred relatively frequent in texts written by women.

Coming back to the assessment of Weiland’s normative influence, at least one aspect of his elaborate comments might have affected the use of relativisers. In fact, his awareness of and remarks on stylistic differences between different forms were, to some extent, reflected in the corpus results. Weiland assigned pronominal forms like *welke* and *develke* to the more formal or ‘solemn’ style, whereas *die* and *dat* were typical of the more informal or ‘plain’ style. In the nineteenth-century data, ‘solemn’ variants like *betwelk* and *welke* considerably gained ground in diaries and travelogues as well as in newspapers, whereas the ‘plain’ variants increased in private letter writing. This stylistic distinction was possibly an effect of normative intervention of Weiland’s grammar on the use of relativisers. Although direct influence could hardly be proven, the developments in the distribution of variants implied a sociolinguistic situation in which forms like *betwelk* and *welke* indexed formality more strongly than common *d*- and *w*-forms.

In Chapter 12, I investigated variation and change in the (adnominal) genitive case and its alternative construction with the preposition *van*. The decline of the Dutch case system had led to a competition between synthetic and analytical

genitival constructions. By the eighteenth century, inflected forms had largely disappeared from spoken and colloquial language, while they were maintained in written language and particularly in higher registers. Metalinguistic discourse still had a strong focus on nominal inflection. Towards the end of the century, however, the analytical *van*-construction was increasingly accepted and even regarded as the best option. Against these developments in language use, Weiland only prescribed the synthetic genitive forms in his official grammar. With regard to his normative influence, the corpus analysis showed that the conservative choice in favour of historical case inflections was at least to some extent effective. Zooming in on the occurrences in neutral (i.e. non-formulaic) contexts, the *van*-construction appeared to be the most frequent variant in eighteenth-century usage, occurring in almost 70%. In the nineteenth century, the share of synthetic forms surprisingly increased to more than 40%. Diachronically, the synthetic genitive gained ground in all genres, even in the most 'oral' genre of private letters, supporting the assumption of Weiland's normative influence. Furthermore, both male and female writers increasingly used the synthetic genitive forms.

However, one internal factor turned out to condition the distribution of variants considerably. The corpus results revealed major differences between masculine and neuter forms on the one hand, and feminine and plural forms on the other. In the post-Weiland data, the synthetic genitive appeared to gain ground in feminine and plural forms, which probably signalled Weiland's prescription. This tendency could not be attested for masculine and neuter forms, which had already been too infrequent by the late eighteenth century. In other words, Weiland could not 'revive' these forms, whereas the higher familiarity of more frequent structures like *x der y* probably fostered the increase of feminine/plural genitive markers. Although Weiland's effort to prescribe the synthetic genitive case was thus partly successful, it must be noted that the *van*-construction had been too established in actual language usage to be entirely replaced by historical genitive forms.

Orthography versus morphosyntax

Comparing the findings from five orthographic and three morphosyntactic variables, it can be concluded that the normative influence on spelling practices was much stronger than on grammatical issues. The drastic changes in the direction of the prescriptions in Siegenbeek's (1804) national orthography gave clear evidence of a normative effect on actual language usage within a fairly short period of time. Particularly in nineteenth-century newspapers, the prescribed spelling norms were adopted in practically all instances. More interestingly, though, Siegenbeek's influence could also be demonstrated in nineteenth-century ego-documents, as writers of private letters as well as diaries and travelogues increasingly used the prescribed variants (with the exception of <ij>).

In contrast, the analyses of morphosyntactic issues showed that the impact of Weiland's (1805) national grammar was much less obvious than in the case of orthography. On closer inspection, however, either direct or indirect influences of Weiland's grammar could be signalled for all three variables under investigation. In

the case of relative pronouns, Weiland's comments on stylistic differences seemed to have consolidated 'plain' forms as the preferred choices in private letter writing, whereas the more formal variants gained ground in newspapers and, strikingly, in diaries and travelogues. With respect to the genitive case, Weiland's conservative prescription in favour of the old synthetic forms appeared to 'revive' the historical case inflections in actual language use, at least to a certain extent. Against the ongoing chance towards analytical constructions, the synthetic forms increased across all genres, regions and genders, but were conditioned internally, only gaining ground in feminine and plural forms.

When assessing the effectiveness of concrete language policy measures like the Dutch *schrijftaalregeling*, orthography and grammar are thus best considered as two distinct levels of effectiveness. As Rutkowska & Rössler (2012: 213) rightly remark, "orthography, more than any other aspect of language, is likely to be influenced by external factors such as language planning which impose change from above the level of consciousness". The coherent findings for orthographic variables in this dissertation testify to the influence of top-down language policy on norm awareness and actual usage in the community at large. Unlike the categorical choices for spelling, visibly reflected in the corpus results, norms on morphosyntactic issues turned out to be more complex, being conditioned by both external and internal factors. In the case of relative pronouns, Weiland generally acknowledged various forms, but preferred (or dispreferred) specific variants in specific contexts and registers. Actual language practices, however, appeared to be even more variable, both before and after Weiland, making it difficult to pinpoint and assess his normative influence. In the case of the genitive (and the declining case system in general), the ongoing language change had probably progressed too far to be completely reversed by grammatical prescription. Nonetheless, an increase of the prescribed synthetic forms could be witnessed in the data.

3 Genre, space, gender and the individual: Assessing the external variables

Diachronically, the corpus results indicated the effect of early nineteenth-century policy measures on actual language use, particularly on the level of orthography and, to a lesser extent, on the level of morphosyntax. Taking into account the external variables integrated in the design of the *Going Dutch Corpus*, more fine-grained differences and patterns were revealed. This section presents the findings related to genre, space (regions and centre-periphery), gender and individual variation.

Genres

Based on the assumption that diachronic changes affect different genres to different extents, the *Going Dutch Corpus* was designed as a diachronic multi-genre corpus. The discussion of results in Section 2 already indicated that genre was

indeed a crucial factor in many case studies. The *Going Dutch Corpus* comprised two types of handwritten ego-documents, viz. private letters as well as diaries and travelogues, but also newspapers, representing printed and published texts. In line with the initial expectations, the corpus analyses more or less consistently testified to a specific genre gradation with respect to the degree of linguistic variation or, put differently, linguistic uniformity, particularly with regard to standardised and non-standard forms.

As presumed, the highest amount of linguistic variation was found in private letters, especially before the early nineteenth-century *schrijftaalregeling*, but also in the period thereafter. In line with previous historical-sociolinguistic research on letter writing (e.g. Rutten & van der Wal 2014; Nevalainen & Raumolin-Brunberg 2013; Elspaß 2005), the analyses based on the *Going Dutch Corpus* demonstrated that private letters in particular can give access to linguistic variation in the past, which had become invisible in contemporary printed texts.

In addition to private letters, another type of handwritten ego-documents was included in the *Going Dutch Corpus*, viz. diaries and travelogues. In terms of linguistic variation expected to be found, these sources were indeed more uniform than private letters. This was also in accordance with previous genre characterisations (e.g. Elspaß 2012, Schneider 2013), which emphasised the strictly monologic (rather than dialogic) nature of these texts, and their lower degree of conceptual orality, especially compared to private letters. Nevertheless, diaries and travelogues still displayed more linguistic variation than the third genre in the corpus, i.e. newspapers. The corpus results thus confirmed that diaries and travelogues constituted an intermediate position on the oral-literate continuum and also in the multi-genre corpus design.

As the only printed and published genre in the *Going Dutch Corpus*, (regional) newspapers displayed by far the strongest uniformity. The consistent and practically invariant use, particularly with regard to spelling, was already attested in the late eighteenth century, but was certainly consolidated after the written language regulations of 1804/1805. In this sense, newspapers can probably be considered representative of the written language conventions in contemporary printed and published texts.

Unlike the linguistic uniformity attested in newspapers, the two types of handwritten ego-documents call for some more attention. While private letters, diaries and travelogues have often been lumped together as ‘ego-documents’, the corpus results revealed different degrees of conformity with the official standard norms. In the late eighteenth-century, the distribution of variants found in private letters on the one hand, and in diaries and travelogues on the other, was still fairly similar. In the early nineteenth century, however, diaries and travelogues increasingly conformed to the writing conventions of newspapers and to the standard norms of the *schrijftaalregeling*, more strongly than private letters. Particularly visible in the genre-specific use of relative pronouns, diaries and travelogues diachronically diverged from the higher variability and conceptual orality still found in private letters from the same period. This means that a growing divergence between the two types of ego-documents could be witnessed in

the period around 1800. These specific developments also emphasise that an umbrella category of ‘handwritten ego-documents’ tends to be too overgeneralised for historical-(socio)linguistic research, and that diaries and travelogues clearly form a (sub-)genre in its own right, distinct from private letters.

Genre thus proved to be a robust external factor. However, at least in one case, the interrelated external factor of medium (cf. also Rutkowska & Rössler 2012: 225), rather than genre itself, appeared to be crucial. With regard to the *ij/y* spelling issue, the medium of genres had the strongest effect on the distribution of variants. First of all, printed texts only distinguished <ij> and <y>, whereas the range of representations was much more complex in handwriting. Furthermore, newspapers radically shifted from <y> to <ij>, while both handwritten genres seemed to reject the prescription in large part, as (double-)dotted and undotted forms continued to coexist, regardless of the officialised spelling.

Space (1): Regions

Taking into account possible regional variation, data from seven regions of the Northern Netherlands was included in the *Going Dutch Corpus*, viz. Friesland, Groningen, North Brabant, North Holland, South Holland, Utrecht and Zeeland. At least in the late eighteenth-century period, and particularly in handwritten ego-documents, regional variation could still be attested to a certain extent. Some of the patterns related to individual variables could be explained by (supra)regional writing practices and/or phonological differences in spoken dialects.

In the case of verb final /t/, distinct Southern writing practices played a role in the distribution of spelling variants in North Brabant and Zeeland, i.e. the two southernmost regions of the investigated language area, bordering the Southern Netherlands. Especially in metalinguistic discourse, the spelling <d> was regarded as typically Southern (versus typically Northern <dt>). In the late eighteenth-century results for North Brabant and Zeeland, <dt> appeared to be practically absent, suggesting a comparatively strong orientation towards Southern practices. Even in the nineteenth-century results, <d> continued to be a strong competitor of the officially prescribed <dt>, at least in North Brabant. A similar (southern) tendency was observed in the comparatively strong use of alternative representations of Wgm. **ī* (other than <ij>/<ÿ>/<y>) in North Brabant and Zeeland.

In the case of etymologically distinct long *e*'s in open syllables, regional patterns in the distribution of spelling variants could partly be referred to dialectal differences in merger and non-merger regions. In the eighteenth-century data, the <e> spelling for soft-long *ē* was more frequent in Zeeland and Groningen, where the historical-phonological distinction had been preserved in many dialects, than in merger regions like the Holland area. Notably, these regional differences disappeared in the nineteenth-century data.

With regard to the morphosyntactic variables under investigation, surprising north-south differences were revealed in the use of the neuter relative pronoun, indicating a relative prominence of *betwelk* in the northern regions of

Groningen and Friesland, and a relative prominence of *betgeen* in the southern regions of North Brabant and Zeeland. These geographically conditioned preferences in the relativisation system certainly deserve closer inspection in future research.

Generally, regional variation first and foremost became visible in handwritten ego-documents. Eighteenth-century newspapers, while still being published regionally, were already fairly uniform across all regions. Only in some cases, North Holland in particular seemed to be ahead of the remaining regions in the use of (spelling) variants that would become officially prescribed variants in the *schrijftaalregeling*. For instance, the consistent use of <dt> was attested in North Holland newspapers as early as the late eighteenth century, i.e. at a time when other regional newspapers still preferred <d>. Furthermore, North Holland newspapers were also more progressive in the use of <cht>, whereas newspapers from most other regions primarily spelled <gt>. Given the fact that Siegenbeek was a native of Amsterdam, it might be suggested that the printed writing practices from the (North) Holland area might have had some influence on the choice of variants officialised in his 1804 orthography – in addition to the evident influence of contemporary normative discourse.

Diachronically, and most importantly, though, the external variable of region showed that late eighteenth-century regional differences, both in ego-documents and in newspapers, were largely levelled out in the early nineteenth century. In the context of a national language, these developments imply that the envisaged aim of the government to spread a uniform variety of Dutch through the community at large was pretty much achieved.

Space (2): Centre versus periphery (and Zeeland)

The second external variable related to space, categorising the selected regions either as centre or periphery, did not reveal any remarkable patterns. Both synchronically and diachronically, general differences between these two demographic and (socio)economic groups of regions turned out to be limited, although individual case studies did display some degree of centre–periphery variation. In the analysis of neuter relative pronouns, for instance, a strong increase of *betwelke* was identified in the periphery. These results, however, were primarily due to the high frequency in the northern periphery (especially Groningen), whereas the variant was comparatively low frequent in the southern periphery (North Brabant). This example demonstrates that the centre–periphery categorisation was not without problems, as variation on this level was further conditioned geographically, for instance by north-south differences. What is more, the centre (i.e. North and South Holland as well as Utrecht) forms a contiguous area, whereas the periphery actually comprises two geographically distant areas within the Northern Netherlands (i.e. Friesland and Groningen in the far north versus North Brabant in the far south), possibly differing in terms of (supra)regional writing practices and, in more general terms, sociolinguistic space.

The seventh region of Zeeland was purposely kept apart due to its historically shifting status from centre to periphery. In some cases, the eighteenth-century results for Zeeland were similar to those for North Brabant (e.g. use of <d>, alternative representation of <y>), which might justify a categorisation of these two regions as the (southern) periphery, perhaps due to a similar orientation towards Southern writing practices. In the nineteenth century, however, Zeeland displayed a number of distinct developments, which could not be observed in North Brabant. In at least two cases, Zeeland witnessed the strongest conformity to the prescribed Siegenbeek variants, viz. <dt> (from two tokens in period 1 to more than 80% in period 2) as well as <ij>/<ÿ> (from a share of one third in period 1 to two thirds in period 2). These striking shifts in the Zeeland data indeed indicate a dynamic region, which could not easily be assigned to the centre or the periphery in the period around 1800.

Gender

The two types of ego-documents in the *Going Dutch Corpus* were written by men and women, allowing to take into account the sociolinguistic variable of gender. For the eighteenth-century period, the corpus results still displayed a fair amount of gender variation in various case studies. With respect to spelling practices, women seemed to prefer <s> for word-medial and word-final /s/ (as opposed to men primarily using <sch>), hardly spelled <cht> for syllable-final /xt/, and applied the phonetic spelling <t> for final /t/ in *d*-stem verbs (almost absent in texts by men). Moreover, women clearly preferred the common relative pronouns *dat*, *wat* and *die* over more formal pronominal forms like *hetwelke* and *welke*, both of which were fairly frequent among men. The prevalence of the alternative genitival construction with the preposition *van* was also considerably more pronounced in ego-documents by women than by men.

Diachronically, however, both men and women increasingly used the officially prescribed variants, indicating a general effect of the language policy measures on language practices of both genders. In fact, gender variation, for the most part, declined in the early nineteenth century. Nevertheless, one could still observe a minor yet strikingly consistent tendency across most case studies. The share of prescribed variants (orthography in particular) was usually higher in the use of nineteenth-century men than among their female contemporaries. In cases like <dt>, <sch>, both <cht> and <gt>, and <e> for soft-long *ē*, the corpus results revealed a 'gender gap' of ten to twenty per cent even in the nineteenth-century data. Despite similar developments in the direction of the standard norms, the remaining gender-related differences suggested that the *schrijftaalregeling* probably reached women to a somewhat lesser extent than men. In other words, men were slightly ahead of women in adopting the prescribed variants in their writing.

While it can only be speculated what might have caused these differences, it is likely that the familiarity with writing conventions and practices as well as the awareness of 'correct' forms was more strongly developed among late eighteenth- and early nineteenth-century men, certainly from the middle to the upper ranks, as

represented in the *Going Dutch Corpus*. Though still depending on education and occupation, men from these layers of society were often trained and thus more experienced in writing professionally, whereas women's writing from the same period was usually still restricted to the private sphere. As Rutten & van der Wal (2014: 396) explain, changing writing conventions were usually noticed earlier by those writers who frequently read and write. In the period under investigation, these writers were more likely to be men than women, which might explain the consistent gender-related pattern, even in the nineteenth-century results.

Individual variation

While the linguistic analyses conducted with the multi-genre *Going Dutch Corpus* in Chapter 5–12 focused on the influence of top-down language policy on the community at large, Chapter 13 examined inter- and intra-individual variation and change in a new corpus of family correspondence, spanning around 100 private letters from the second half of the eighteenth and the first half of the nineteenth century. Specifically compiled for this case study, the *Martini Buys Correspondence Corpus* allowed to zoom in on individual spelling practices across three generations of the Martini Buys family, both before and after the language policy. Some valuable micro-level insights were gained, which further enriched the general understanding of policy effectiveness in the Northern Netherlands.

With regard to intra-individual variation (i.e. inconsistent use of co-occurring variants), a certain degree of inconsistency was identified across all generations. Most informants, however, turned out to be surprisingly consistent in their use of variants. Addressing the possible influence of language policy measures on individual spelling choices, remarkable developments in the use of variants were even attested in adulthood, giving evidence of lifespan changes that could occur as an effect of language policy measures. Interestingly, certain adult informants from the transitional generation (i.e. those family members who were not directly exposed to the language policy in their formative years) appeared to adopt various prescriptions by Siegenbeek irrespective of (school) education. Not only did this underscore the general normative awareness in the early nineteenth century, it also indicated that education was not the only means of transmitting the standardised norms. On the other hand, spelling choices systematically deviating from the standardised norms were also attested in letters written by members of the youngest generation, who were exposed to the official norms in school. This case study further demonstrated that a micro-level approach to individual behaviour over time can add important nuances to a large-scale corpus study.

4 Policy versus practice

Bringing together the findings of this dissertation, it can be concluded that the concrete language policy measures introduced by the Dutch government in the early 1800s were effective in the sense that actual language usage patterns generally

displayed a striking convergence towards the prescribed norms. The case studies in Chapters 5–12 showed that these developments were particularly, though not exclusively visible in the corpus analyses of orthographic variables. Direct normative influence on grammatical issues, at least for the most part, was less obvious and thus more difficult to assess.

Tackling the intriguing interplay between language norms and language practice, the developments in early nineteenth-century usage in the Northern Netherlands clearly indicated a change in the awareness of norms and the influence these norms exerted on actual practices after the introduction of the *schrijftaalregeling* in 1804/1805. Against the background of an increasingly coherent normative tradition, with Siegenbeek (1804) in particular largely following existing and established preferences in (late) eighteenth-century metalinguistic discourse (as, e.g., in the case of <dt>, <sch>, <ij>), it must be assumed that the developments in language practice reflect an increasing awareness of linguistic norms among language users.

Without any doubt, the top-down language policy and more concretely the *schrijftaalregeling* constituted a crucial intervention in the sociolinguistic situation of the early 1800s. By discursively constructing a national standard variety and officially laying down ‘the’ rules for ‘the’ Dutch language on behalf of the government, a clear-cut separation between standard and non-standard varieties was introduced in the Northern Netherlands. This rigid categorisation was a novelty in the long history of Dutch standardisation, officially splitting the contemporary diaglossic continuum into a hierarchical opposition between standard and non-standard, and thus creating a new sociolinguistic situation of diglossia (cf. Chapter 1; also Rutten 2016c). The almost immediate implementation of this strong standard language ideology in the national school system, and also implementing the accompanying language norms for orthography and grammar (cf. Chapter 2), proved to be forceful measures to disseminate the standard variety of Dutch and to raise the awareness of national norms for the written language in a relatively short period of time.

It should be kept in mind that these official regulations for spelling and grammar were primarily intended to be used in the administrative and educational domains, while it was not formally mandatory to follow them, certainly not in private writing. The acceptance or, in other words, the success of these norms in actual language use was thus anything but guaranteed. Nevertheless, the corpus results from this dissertation clearly indicated that the official prescriptions were not only adopted in printed, published texts like newspapers, but also in relatively informal texts from the private sphere, as the two types of ego-documents (private letters, diaries and travelogues) demonstrated. Moreover, these developments in the direction of prescribed standard norms could be witnessed in all regions under investigation, both from the centre and the periphery of the language area, and in texts produced by both men and women. This means that quite some awareness and, more importantly, a general acceptance of the national standard norms in the community at large can be assumed.

To conclude, the language policy measures were thus pretty successful in spreading the national standard variety of Dutch across the community at large, as envisaged by the government. While regional and gender differences, still present in late eighteenth-century usage, had largely levelled out in the early nineteenth century, genre continued to be the most external crucial factor conditioning the distribution of variants.

5 Concluding remarks

In this dissertation, new insights were gained into variation and change processes in late eighteenth- and early nineteenth-century Dutch in the Northern Netherlands. More specifically, the effectiveness of concrete language policy measures was investigated empirically, on the basis of the multi-genre *Going Dutch Corpus*, which was designed and compiled in order to assess policy success in the history of Dutch standardisation. Not only in the history of Dutch, but also of most other languages, the impact of top-down language planning and/or policy endeavours on actual usage patterns has hardly been investigated. Therefore, the present dissertation also contributes to the international discussion on norms and usage in historical sociolinguistics (e.g. Rutten et al. 2014a). The case of the Dutch *schrijftaalregeling* in the early 1800s, with its official regulations for orthography and grammar, served as a highly interesting and, importantly, concrete point of departure. The findings drawn from this research demonstrate that it *is* possible to ‘measure’ the effectiveness of language policy in historical contexts. Methodologically, a custom-made and balanced corpus design as well as a systematic approach, covering a wide range of linguistic variables, appeared to be essential for a comparable and sophisticated assessment of how policy affected usage patterns.

Similar historical-sociolinguistic corpus projects are currently being conducted for nineteenth-century German (Elspaß & Niehaus 2014) and nineteenth- and early twentieth-century Icelandic (van der Feest Viðarsson 2017). It should be the aim of future research to compare the results from a number of historical cases, paving the way for a more overarching perspective on policy effectiveness across various (European) languages. Such a comparative approach could shed more light on the mechanisms behind language policy and their influence on usage patterns in general, i.e. in order to determine which factors are decisive for policy endeavours to be effective and successful. It is safe to say that this dissertation on the specific case of Dutch in the Northern Netherlands around 1800 provides a substantial foundation for future (comparative) research on this intriguing topic in historical sociolinguistics.

References

I (*Normative Corpus of the Northern Netherlands*)

van Belle, Jan. 1748. *Korte wegwijzer, ter spel- spraak- en dichtkonden*. Haarlem: Izaak van der Vinne.

van Belle, Jan. 1755. *Korte schets der Nederduitse spraakkunst*. Haarlem: Christoph Henrich Bohn.

van Bolhuis, Lambertus. 1793. *Beknopte Nederduitse spraakkunst*. Leiden/Deventer/Groningen: D. du Mortier en zoon, J.H. de Lange & J. Oomkenks.

Elzevier, Kornelis. 1761. *Drie dichtproeven [...] benevens een proef van een nieuwe Nederduitse spraakkunst*. Haarlem: J. Bosch.

Francius, Petrus. 1699. *Gregorius Nazianzenus, Van de mededeelzaamheid*. Amsterdam: H. Wetsteen.

de Haes, Frans. 1764. *De nagelaten gedichten, en Nederduitse spraakkunst*. Amsterdam: Pieter Meijer.

Hakvoord, Barend. 1746. *De nieuwe Nederduitse spel-, lees- en schryf-kunst*. Deventer: Abraham van Wezel.

Heugelenburg, Martinus. 1763. *Klein woordenboek, zijnde een kort en klaar onderwijs in de Nederlandse spel, en leeskonst*. Amsterdam: J. Ratelband & J. Bouwer.

Hilarides, Johannes. 1705. *Nieuwe taalgronden der Neederduitse taal*. Franeker: Jacobus Horreus.

van Hoogstraten, David. 1700. *Aenmerkingen over de geslachten der zelfstandige naemwoorden*. Amsterdam: François Halma.

Huydecoper, Balthazar. 1730. *Proeve van taal- en dichtkunde*. Amsterdam: E. Visscher & J. Tirion.

ten Kate, Lambert. 1723. *Aenleiding tot de kennisse van het verhevene deel der Nederduitse sprake*. Two volumes. Reprint and introduction by Jan Noordegraaf & Marijke van der Wal. Alphen aan den Rijn 2001: Canaletto.

Kluit, Adriaan. 1763. Eerste verhoog over de tegenwoordige spelling der Nederduitsche taal. *Nieuwe Bydragen tot de opbouw der vaderlandsche letterkunde*, 285-352.

Kluit, Adriaan. 1777. Verhoog over de tegenwoordige spelling der Nederduitsche taal, vergeleken met de spelling der ouden, en uit dezelve ene soort van evenredigheid opgemaakt. In *Werken van de Maetschappij der Nederlandsche letterkunde te Leyden*. Leiden: Van der Eyk.

Kunst wordt door arbeid verkreegen. 1770. *Nederduitsche spraekkunst*. Leiden: Hoogeveen.

Moonen, Arnold. 1706. *Nederduitsche spraekkunst*. Amsterdam: François Halma.

Niervaart, Cornelis. 1743. *Oprecht onderwijs van de letter-kunst*. Purmerend: Pieter Jordaan.

Nylöe, Jakobus. 1703. *Aanleiding tot de Nederduitsche taal, om goetd en zuiver Nederduitsch te spreken of te schryven*. Amsterdam: Gerardus Borstius.

van der Palm, Kornelis. 1769. *Nederduitsche spraekkunst, voor de jeugdt*. Four volumes. Rotterdam: Reinier Arrenberg.

van Rhyn, Leonard. 1758. *Kort begryp der Nederduitsche spel-kunst*. Amsterdam: Jan Schuring.

Rudimenta. 1799. *Rudimenta of gronden der Nederduitsche spraake*. Uitgegeeven door de Maatschappij tot Nut van 't Algemeen. Leiden/Deventer/Utrecht: D. du Mortier en zoon, J.H. de Lange & G.T. van Paddenburg en zoon.

Sewel, Willem. 1708. *Nederduitsche spraakkunst*. Amsterdam: Erven J. Lescaijle & Assuerus Lansvelt.

Sewel, Willem. 1712. *Nederduitsche spraakkunst*. Second edition. Amsterdam: Robert Blokland.

Siegenbeek, Matthijs. 1804. *Verhandeling over de Nederduitsche spelling, ter bevordering van eenparigheid in dezelve*. Amsterdam: Johannes Allart.

Stijl, Klaas & Lambertus van Bolhuis. 1776. *Beknopte aanleiding tot de kennis der spelling, spraakdeelen, en zinteecken van de Nederduitsche taal*. Groningen: Jan Oomkens.

Tollius, Herman. 1776. *Proeve eener Aanleiding tot de Nederduitsche letterkunst*.

- Verwer, Adriaen. 1707. *Schets van de Nederlandse taal grammatica, poëtica, retorica*. Dutch translation of *Linguae Belgicae idea grammatica, poëtica, rhetorica*. Amsterdam: François Halma.
- Verwer, Adriaen. 2005. *Taalkundige geschriften (1707-1711)*. Edited by Igor van de Bilt. Amsterdam:
- Weiland, Petrus. 1799. *Nederduitsch taalkundig woordenboek*. Amsterdam: Johannes Allart.
- Weiland, Petrus. 1805. *Nederduitsche spraakkunst*. Amsterdam: Johannes Allart.
- Wester, Hendrik. 1799. *Bevatlyk onderwys in de Nederlandsche spel- en taalkunde, voor de schooljeugd*. Groningen: Jan Oomkens.
- Zeydelaar, Ernst. 1774. *Nederduitsche spelkonst*. Dordrecht: Abraham Blussé en zoon.

II (Bibliography)

- Ager, Dennis E. 2001. *Motivation in language planning and language policy*. Clevedon: Multilingual Matters.
- Algemeene denkebeelden over het nationaal onderwijs*. 1798. Amsterdam: S.n.
- Anderwald, Lieselotte. 2016. *Language between description and prescription. Verbs and verb categories in nineteenth-century grammars of English*. Oxford: Oxford University Press.
- ANS = *Algemene Nederlandse Spraakkunst*. Online edition <<http://ans.ruhosting.nl/e-ans/index.html>>.
- Auer, Anita & Victorina González-Díaz. 2005. Eighteenth-century prescriptivism in English: A re-evaluation of its effects on actual language usage. *Multilingua* 24, 317–341.
- Auer, Anita, Catharina Peersman, Simon Pickl, Gijsbert Rutten & Rik Vosters. 2015. Historical sociolinguistics: the field and its future. *Journal of Historical Sociolinguistics* 1 (1), 1–12.
- Auer, Peter. 2005. Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In Nicole Delbecque, Johan van der Auwera & Dirk Geeraerts (eds.), *Perspectives on variation: Sociolinguistic, historical, comparative*. Berlin/New York: De Gruyter, 7–42.

- Auer, Peter. 2011. Dialect vs. standard: A typology of scenarios in Europa. In Bernd Kortmann & Johan van der Auwera (eds.), *The languages and linguistics of Europe. A comprehensive guide*. Berlin/ New York: De Gruyter, 485–500.
- Austin, Frances (ed.). 1991. *The Clift Family Correspondence 1792–1846*. Sheffield: Centre for English Cultural Tradition & Language.
- Austin, Frances. 1994. The effect of exposure to standard English: The language of William Clift. In Dieter Stein & Ingrid Tiekens-Boon van Ostade (eds.), *Towards a Standard English 1600–1800*. Berlin/New York: De Gruyter, 285–313.
- Baggerman, Arianne. 2011. Lost time: Temporal discipline and historical awareness in nineteenth-century Dutch egodocuments. In Arianne Baggerman, Rudolf Dekker & Michael Mascuch (eds.), *Controlling time and shaping the self: Developments in autobiographical writing since the sixteenth century*. Leiden/Boston: Brill, 455–541.
- Bergen, Linda van & David Denison. 2007. A corpus of late eighteenth-century prose. In Joan C. Beal, Karen P. Corrigan & Hermann L. Moisl (eds.), *Creating and digitizing language corpora*. Basingstoke/New York: Palgrave Macmillan, 228–246.
- Bergs, Alexander. 2005. *Social networks and historical sociolinguistics: Studies in morphosyntactic variation in the Paston letters (1421–1503)*. Berlin: De Gruyter.
- Bergs, Alexander. 2012. The uniformitarian principle and the risk of anachronisms in language and social history. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 80–98.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- van de Bilt, Igor. 2009. *Landkaartschrijvers en landverdelers. Adriaen Verver (ca. 1655–1717), Adriaan Kluit (1735–1807) en de Nederlandse taalkunde van de achttiende eeuw*. Amsterdam: Stichting Neerlandistiek VU & Münster: Nodus.
- Blommaert, Jan & Jef Verschueren. 1998. The role of language in European nationalist ideologies. In B. Scheffelin, K. Woolard & P. Kroskity (eds.), *Language ideologies: Practice and theory*. New York: Oxford University Press, 189–210.
- Boekholt & de Booy. 1987. *Geschiedenis van de school in Nederland vanaf de middeleeuwen tot aan de huidige tijd*. Assen/Maastricht: Van Gorcum.
- de Bonth, R.J.G. et al. 1997. Nieuwnederlands, circa 1650–1880. In van den Toorn, Pijnenburg, van Leuvensteijn, van der Horst (eds.), *Geschiedenis van de Nederlandse taal*. Amsterdam: Amsterdam University Press, 361–453.

- Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Oxford University Press.
- van Bree, Cor & Arjen P. Versloot. In cooperation with Rolf H. Bremmer jr. 2008. *Oorsprongen van het Stadsfries*. Leeuwarden: Fryske Academy.
- Conde-Silvestre, J. Camilo & Juan M. Hernández-Campoy 2012. Introduction. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 1–8.
- Cooper, Robert. 1989. *Language planning and social change*. Cambridge: Cambridge University Press.
- Daems, Frans. 2002. *Zoals wij zijn wordet gij*. Is de werkwoordspelling logisch of etymologisch? In Reinhilde Haest, Luuk Van Waes & Dirk Caluwé (eds.), *Communicatief bekeken. Liber Amicorum Stijn Verrept*. Mechelen: Kluwer, 42–54.
- De Wulf, Chris, Jan Goossens & Johan Taeldeman. 2005. *Fonologische atlas van de Nederlandse dialecten. Deel IV: De consonanten*. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.
- Dekker, Rudolf. 1995. Dutch travel journal from the sixteenth to the early nineteenth centuries. *Lias. Sources and documents relating to the early modern history of ideas* 22, 277–300.
- Dekker, Rudolf. 2002. *Egodocuments and history: autobiographical writing in its social context since the Middle Ages*. Hilversum: Verloren.
- Dibbets, G.R.W. 1999. De *Nederduitsche spraekkunst* (1764) van Frans de Haes. *Tijdschrift voor Nederlandse taal- en letterkunde* 115, 42–64.
- van Driel, Lodewijk. 1992. 19th-century linguistics: The Dutch development and the German theme. In *Historiographia Linguistica* 15, 155–185.
- Durrell, Martin, Paul Bennett, Silke Scheible & Richard J. Whitt. 2012. *The GerManC corpus* (Documentation). Manchester: University of Manchester.
- Elspaß, Stephan. 2005. *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer.
- Elspaß, Stephan. 2007. A twofold view ‘from below’: New perspectives on language histories and historical grammar. In Stephan Elspaß, Nils Langer, Joachim Scharloth & Wim Vandenbussche (eds.), 2007. *Germanic Language Histories ‘from below’ (1700–2000)*. Berlin: De Gruyter, 3–9.

Elspaß, Stephan. 2012. The use of private letters and diaries in sociolinguistic investigation. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 156–159.

Elspaß, Stephan. 2014. Prescriptive norms and norms of usage in nineteenth-century German. In Gijsbert Rutten, Rik Vosters & Wim Vandenbussche (eds.), *Norms and usage in language history, 1600–1900. A sociolinguistic and comparative perspective*. Amsterdam/ Philadelphia: John Benjamins, 303–320.

Elspaß, Stephan, Nils Langer, Joachim Scharloth & Wim Vandenbussche (eds.). 2007. *Germanic Language Histories 'from below' (1700–2000)*. Berlin: De Gruyter.

Elspaß, Stephan & Konstantin Niehaus. 2014. The standardization of a modern pluriareal language. Concepts and corpus designs for German and beyond. *Orð og tunga* 16, 47–67.

EWN = *Etymologisch woordenboek van het Nederlands* (Amsterdam University Press). Online edition <www.etymologie.nl>.

Fairman, Tony. 2007. Writing and 'the standard': England, 1795–1834. In Stephan Elspaß & Wim Vandenbussche (eds.), *Lower class language use in the 19th century*. Special issue of *Multilingua*, 167–201.

van der Feest Viðarsson, Heimir. 2017. The syntax of others: 'Un-Icelandic' verb placement in 19th- and early 20th-century Icelandic. In Ingrid Tieken-Boon van Ostade & Carol Percy (eds.), *Prescription and tradition in language. Establishing standards across time and space*. Bristol: Blue Ridge Summit, 152–167.

Ferguson, Charles A. 1959. Diglossia. *WORD* 15:2, 325–340.

Gledhill, John Michael. 1973. *Aspects of the development of Dutch consonantal spelling: On the evidence of grammarians, lexicographers, and the principal variants of printed books, from Middle Dutch to the present day*. PhD dissertation. University of London.

Goossens, Jan, Johan Taeldeman & Geert Verleyen. 2000a. *Fonologische atlas van de Nederlandse dialecten. Deel II: De Westgermaanse korte vocalen in open lettergreep. Deel III: De Westgermaanse lange vocalen en diftongen*. Ghent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.

Goossens, Jan, Johan Taeldeman & Geert Verleyen. 2000b. *Fonologische atlas van de Nederlandse dialecten. Deel III: De Westgermaanse lange vocalen en diftongen*. Ghent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.

Hernández-Campoy, Juan Manuel & Juan Camilo Conde-Silvestre (eds.). 2012. *The handbook of historical sociolinguistics*. Malden, MA/Oxford: Wiley-Blackwell.

- Hernández-Campoy, Juan M. 2016. Authorship and gender in English historical sociolinguistic research: Samples from the Paston Letters. In Cinzia Russi (ed.) *Current Trends in Historical Sociolinguistics*. Berlin/NewYork: De Gruyter, 108–142.
- van der Horst, J.M. 1988. Over relatief *dat* en *wat*. *De nieuwe taalgids* 81, 194–205.
- van der Horst, J.M. 2008. *Geschiedenis van de Nederlandse syntaxis*. Two volumes. Louvain: Leuven University Press.
- Hundt, Marianne & Benedikt Szmrecsanyi. 2012. Animacy in early New Zealand English. *English World-Wide* 33 (3), 241–263.
- Instructie voor den Agent van Nationale Opvoeding*. 1798. The Hague: S.n.
- Kielkiewicz-Janowiak, Agnieszka. 2012. Class, age, and gender-based patterns. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 307–331.
- Kloek, Joost & Wijnand Mijnhardt. 2004. *1800. Blueprints for a national community*. Assen: Royal Van Gorcum & Palgrave Macmillan.
- Kloke, Gesinus Gerhardus. 1927. *De Hollandsche expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandsche dialecten*. The Hague: Martinus Nijhoff.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgebrauch. *Romantistisches Jahrbuch* 36, 15–43.
- Krogull, Andreas. 2018. De effectiviteit van de schrijftaalregeling. Historisch-sociolinguïstisch onderzoek naar taalbeleid en taalgebruik in de achttiende en negentiende eeuw. *Internationale Neerlandistiek* 56 (1), 23–42.
- Krogull, Andreas, Gijsbert Rutten & Marijke van der Wal. 2017. Relativisation in Dutch diaries, private letters and newspapers (1770–1840): A genre-specific national language? In Tanja Säily, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds.), *Exploring future paths for historical sociolinguistics*. Amsterdam/Philadelphia: John Benjamins, 157–186.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1994. *Principles of Linguistic Change*. Vol. 1, *Internal factors*. Oxford: Blackwell.

- Labov, William. 2001. *Principles of Linguistic Change*. Vol. 2, *Social factors*. Oxford: Blackwell.
- Langer, Nils. 2001. *Linguistic purism in action. How auxiliary tun was stigmatized in Early New High German*. Berlin/New York: De Gruyter.
- Langer, Nils. 2011. Historical sociolinguistics in nineteenth-century Schleswig-Holstein. *German Life and Letters* 64, 169–178.
- Lindeman, Ruud, Yvonne Scherf & Rudolf Dekker. 1993. *Egodocumenten van Noord-Nederlanders van de zestiende tot begin negentiende eeuw. Een chronologische lijst*. Rotterdam: Erasmus Universiteit.
- Lindeman, Ruud, Yvonne Scherf & Rudolf Dekker. 1994. *Reisverslagen van Noord-Nederlanders van de zestiende tot begin negentiende eeuw. Een chronologische lijst*. Rotterdam: Erasmus Universiteit.
- Lippi-Green, Rosina. 2012. *English with an accent: Language, ideology, and discrimination in the United States*. Second edition. London/New York: Routledge.
- Los, Willeke. 2005. *Opvoeding tot mens en burger. Pedagogiek als cultuurkritiek in Nederland in de 18e eeuw*. Hilversum: Verloren.
- Mannila, Heikki, Terttu Nevalainen & Helena Raumolin-Brunberg. 2013. Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*. Cambridge: Cambridge University Press, 337–360.
- Marquilhaes, Rita. 2012. A historical digital archive of Portuguese letters. In Marina Dossena & Gabrielle Del Lungo Camiciotti (eds.), *Letter writing in Late Modern Europe*. Amsterdam/Philadelphia: John Benjamins, 31–44.
- Marquilhaes, Rita. 2015. Non-anachronism in the historical sociolinguistic study of Portuguese. *Journal of Historical Sociolinguistics* 1 (2), 213–242.
- Martineau, France. 2013. Written documents. What they tell us about linguistic usage. In Marijke J. van der Wal & Gijsbert Rutten (eds.), *Touching the past: Studies in the historical sociolinguistics of ego-documents*. Amsterdam/Philadelphia: John Benjamins, 129–148.
- Marynissen, Ann. 2005. Die geographische Streuung der Familiennamentypen im niederländischen Sprachgebiet. *Niederdeutsches Wort* 45, 105–120.
- Marynissen, Ann & Damaris Nübling. 2010. Familiennamen in Flandern, den Niederlanden und Deutschland – ein diachroner und synchroner Vergleich. In

- Antje Dammel, Sebastian Kürschner & Damaris Nübling (eds.), *Konstrastive Germanistische Linguistik* (Teilband 1). Hildesheim et al.: Georg Olms, 311–362.
- Mathijsen, Marita. 1988. Spelling in de negentiende eeuw: Siegenbeek en Bilderdijk. *De negentiende eeuw* 12 (1988), 129–140.
- Mattheier, Klaus J. 1988. Nationalsprachenentwicklung, Sprachstandardisierung und historische Soziolinguistik. In Klaus J. Mattheier (ed.), *Standardisierungsentwicklungen in europäischen Nationalsprachen: Romania, Germania* (Sociolinguistica. Internationales Jahrbuch für Europäische Soziolinguistik 2. Tübingen: Niemeyer, 1–9.
- de Meij, G.M.J. 2011 [1984]. *Inventaris van de archieven van de familie Martini Buys 1559–1966 (1990) en van het Huis en de Heerlijkheid Loenersloot (1276) 1569–1954 (1986)*. Utrecht: Het Utrechts Archief.
- Meyerhoff, Miriam. 2011. *Introducing sociolinguistics*. London/New York: Routledge.
- Mihm, Arend. 1998. Arbeitersprache und gesprochene Sprache im 19. Jahrhundert. In Dieter Cherubim, Siegfried Grosse & Klaus J. Mattheier (eds.), *Sprache und bürgerliche Nation. Beiträge zur deutschen und europäischen Sprachgeschichte des 19. Jahrhunderts*. Berlin/New York: De Gruyter, 282–316.
- Mijnhardt, Wijnand. 1987. *Tot heil van 't menschdom: culturele genootschappen in Nederland, 1750–1815*. Amsterdam: Rodopi.
- Milroy, James. 1992. *Linguistic variation and change. On the historical sociolinguistics of English*. Oxford/Cambridge: Blackwell.
- Milroy, James. 2012. Sociolinguistics and ideologies in language history. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 571–584.
- Molewijk, G.C. 1992. *Spellingverandering van zín naar onzín (1200-heden)*. The Hague: SDU Uitgeverij Koninginnegracht.
- Moliner, Olivier & Evelyn Ziegler. 2017. Sprachenpolitik, Sprachenideologie und Sprachgebrauch in Luxemburg. Das Deutsche in Luxemburg im 19. Jahrhundert (1795–1920). *Sociolinguistica* 31 (1), 117–138.
- Montgomery, Michael. 1997. A tale of two Georges: The language of Irish Indian traders in colonial North America. In Jeffrey Kallen (ed.), *Focus on Ireland*. Amsterdam: John Benjamins, 227–254.
- Negele, Michaela. 2012. *Varianten der Pronominaladverbien im Neuhochdeutschen. Grammatische und soziolinguistische Untersuchungen*. Berlin: De Gruyter.

- Nevalainen, Terttu. 1996. Historical sociolinguistics. In Terttu Nevalainen & Helena Raumolin-Brunberg (eds.), *Sociolinguistics and language history. Studies based on the Corpus of Early English Correspondence*. Amsterdam: Rodopi, 11–38.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics. Language change in Tudor and Stuart England*. London: Longman.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2012. Historical sociolinguistics: Origins, motivations, and paradigms. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 22–40.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2016. *Historical sociolinguistics. Language change in Tudor and Stuart England*. Second revised edition. New York: Routledge.
- Niehaus, Konstantin. 2016. *Wortstellungsvarianten im Schriftdeutschen. Über Kontinuitäten und Diskontinuitäten in neuhochdeutscher Syntax*. Heidelberg: Universitätsverlag Winter.
- Nieuwborg, E. R. 1969. *Retrograde woordenboek van de Nederlandse taal*. Antwerpen: Plantyn.
- Nobels, Judith. 2013. *(Extra)Ordinary letters. A view from below on seventeenth-century Dutch*. PhD dissertation. Utrecht: LOT.
- Nobels, Judith & Gijsbert Rutten. 2014. Language norms and language use in seventeenth-century Dutch. Negation and the genitive. In Gijsbert Rutten, Rik Vosters & Wim Vandenbussche (eds.), *Norms and usage in language history, 1600-1900. A sociolinguistic and comparative perspective*. Amsterdam/Philadelphia: John Benjamins, 21–48.
- Nobels, Judith & Marijke van der Wal. 2012. Linking words to writers: Building a reliable corpus for historical sociolinguistic research. In Nils Langer, Steffan Davies & Wim Vandenbussche (eds.), *Language and history, linguistics and historiography. Interdisciplinary approaches*. Bern et al.: Peter Lang, 343–361.
- Noordegraaf, Jan. 2004. A matter of time. Dutch philosophy of language in the eighteenth century. In T. F. Shannon & J. P. Snapper (eds.), *Janus at the millennium: Perspectives on time in the culture of the Netherlands*. Dallas: University Press of America, 211–225.
- Noordegraaf, Jan. 2018. Pieter Weiland and his *Nederduitsche Spraakkunst*. In Rick Honings, Ton van Kalmthout & Gijsbert Rutten (eds.), *Language, literature and the construction of a Dutch national identity (1780–1830)*. Amsterdam: University Press, 145–166.

Noordegraaf, Jan & Marijke van der Wal. 2001. Lambert ten Kate (1674–1731) en de taalwetenschap / Lambert ten Kate (1674–1731) and linguistics. In Jan Noordegraaf & Marijke van der Wal (eds.), *Lambert ten Kate Hermansz, Aenleiding tot de kennis van het verhevene deel der Nederduitsche sprake (1723). Twee delen in facsimile*. Alphen a/d Rijn: Canaletto, 1–32.

Notulen = *Notulen van het Staats-Bewind der Bataafsche Republiek. September–December 1804*. 1804.

Paardekooper, P.C. 1956. Een schat van een kind. *De nieuwe taalgids* 49, 93–99.

Palander-Collin, Minna, Minna Nevala & Arja Nurmi. 2009. The language of daily life in the history of English: Studying how macro meets micro. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)*. Amsterdam/Philadelphia: John Benjamins, 1–23.

Percy, Carol. 2012. Early advertising and newspapers as sources of sociolinguistic investigation. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 191–210.

van der Ploeg, Hidde Wibius. 1800. *Het belang der waare volksverlichting*. Amsterdam: Wed. Harm. Keijzer, Corns. De Vries & Hendrik van Munster.

Poplack, Shana & Nathalie Dion. 2009. Prescription vs praxis: The evolution of future temporal reference in French. *Language* 85 (3), 557–587.

Poplack, Shana, Lidia-Gabriela Jarmasz, Nathalie Dion & Nicole Rosen. 2015. Searching for standard French: The construction and mining of the *Recueil historique des grammaires du français*. *Journal of Historical Sociolinguistics* 1 (1), 13–55.

Rademann, Tobias. 1998. Using online electronic newspapers in modern English-language press corpora: benefits and pitfalls. *ICAME Journal* 22, 49–72.

Raumolin-Brunberg, Helena. 2005. Language change in adulthood: Historical letters as evidence. In Minna Palander-Collin & Minna Nevala (eds.), *Letters and Letter Writing*, special issue of *European Journal of English Studies* 9 (1), 37–51.

Raumolin-Brunberg, Helena. 2009. Lifespan changes in the language of three early modern gentlemen. In Minna Palander-Collin, Minna Nevala & Arja Nurmi (eds.), *The Language of Daily Life in England (1400–1800)*. Amsterdam/Philadelphia: John Benjamins, 165–196.

Roberge, Paul T. 2006. Language history and historical sociolinguistics/Sprachgeschichte und historische Soziolinguistik. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics. An international*

handbook of the science of language and society, Volume 3. Berlin/New York: De Gruyter, 2307–2315.

Romaine, Suzanne. 2005. Historical sociolinguistics/historische soziolinguistik. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics. An international handbook of the science of language and society*, Volume 2. Berlin/New York: De Gruyter, 1696–1703.

Romaine, Suzanne. 2009 [1982]. *Socio-historical linguistics. Its status and methodology*. Cambridge: Cambridge University Press.

Rutkowska, Hanna & Paul Rössler. 2012. Orthographic variables. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 213–236.

Rutten, Gijsbert. 2009a. Grammar to the people. The Dutch language and the public sphere in the 18th century: With special reference to Kornelis van der Palm. *Beiträge zur Geschichte der Sprachwissenschaft* 19, 55–86.

Rutten, Gijsbert. 2009b. Uit de geschiedenis van de spelling: Over de scherp- en zachtlange [e:] en [o:]. *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 119, 85–140.

Rutten, Gijsbert. 2010. Vroegmoderne relativa: naar een diachrone constructiegrammatica. *Nederlandse Taalkunde* 15, 1–32.

Rutten, Gijsbert. 2011. In cooperation with Rik Vosters. *Een nieuwe Nederduitse spraakkunst: Taal-normen en schrijfprijken in de zuidelijke Nederlanden in de achttiende eeuw*. Brussels: VUB Press.

Rutten, Gijsbert. 2012a. ‘Lowthian’ linguistics across the North Sea. *Historiographia Linguistica* 39, 45–59.

Rutten, Gijsbert. 2012b. From adverb to conjunction and back. The (de)grammaticalization of Dutch *dan*. *Diachronica* 29 (3), 301–325.

Rutten, Gijsbert. 2016a. The ‘Golden Age Myth’: The construction of Dutch as a national language. In Gijsbert Rutten & Kristine Horner (eds.), *Metalinguistic perspectives on Germanic languages: European case studies from past to present*. Oxford: Peter Lang, 13–34.

Rutten, Gijsbert. 2016b. Standardization and the myth of neutrality in language history. *International Journal of the Sociology of Language* 242, 25–57.

- Rutten, Gijsbert. 2016c. Diaglossia, individual variation and the limits of standardization: Evidence from Dutch. In Cinzia Russi (ed.), *Current Trends in Historical Sociolinguistics*. Berlin: De Gruyter, 194–218.
- Rutten, Gijsbert. 2016d. Historicizing diaglossia. *Journal of Sociolinguistics* 20 (1), 6–30.
- Rutten, Gijsbert. 2016e. Teaching the genitive. Variation of genitival constructions in Dutch ‘national’ grammar (1800–1830). *Beiträge zur Geschichte der Sprachwissenschaft* 26, 123–138.
- Rutten, Gijsbert. 2018. Matthijs Siegenbeek in defence of Dutch. In Rick Honings, Ton van Kalmthout & Gijsbert Rutten (eds.), *Language, literature and the construction of a Dutch national identity (1780–1830)*. Amsterdam: University Press, 25–48.
- Rutten, Gijsbert & Ton van Kalmthout. 2018. Introduction. Cultural nationalism and the rise of Dutch studies. In Rick Honings, Ton van Kalmthout & Gijsbert Rutten (eds.), *Language, literature and the construction of a Dutch national identity (1780–1830)*. Amsterdam: University Press, 9–24.
- Rutten, Gijsbert, Andreas Krogull & Bob Schoemaker. Accepted. Implementation and acceptance of national language policy. The case of Dutch (1750–1850). *Language Policy*.
- Rutten, Gijsbert & Rik Vosters. 2010. Chaos and standards: Orthography in the Southern Netherlands (1720–1830). *Multilingua* 29, 417–438.
- Rutten, Gijsbert, Rik Vosters & Wim Vandenbussche (eds.). 2014a. *Norms and usage in language history, 1600–1900. A sociolinguistic and comparative perspective*. Amsterdam/Philadelphia: John Benjamins.
- Rutten, Gijsbert, Rik Vosters & Wim Vandenbussche. 2014b. The interplay of language norms and usage patterns. Comparing the history of Dutch, English, French and German. In Gijsbert Rutten, Rik Vosters & Wim Vandenbussche (eds.), *Norms and usage in language history, 1600–1900. A sociolinguistic and comparative perspective*. Amsterdam/Philadelphia: John Benjamins, 1–17.
- Rutten, Gijsbert & Marijke van der Wal. 2013. Change, contact and conventions in the history of Dutch. In *Taal & Tongval* 65 (1), 97–123.
- Rutten, Gijsbert & Marijke van der Wal. 2014. *Letters as loot. A sociolinguistic approach to seventeenth- and eighteenth-century Dutch*. Amsterdam/Philadelphia: John Benjamins.
- Rutten, Gijsbert & Marijke J. van der Wal. 2017. Discourse continuity and the written medium: Continuative relative clauses in the history of Dutch. In Daniel

- van Olmen, Hubert Cuyckens & Lobke Ghesquière (eds.), *Aspects of Grammaticalization. (Inter)Subjectification and Directionality*. Berlin: De Gruyter, 113–138.
- Rutten, Gijsbert, Marijke J. van der Wal, Judith Nobels & Tanja Simons. 2012. Negation in seventeenth- and eighteenth-century Dutch: A historical-sociolinguistic perspective. In *Neuphilologische Mitteilungen* 113, 323–342.
- Sairio, Anni. 2008. Bluestocking letters and the influence of eighteenth-century grammarians. In Marina Dossena & Ingrid Tiekens-Boon van Ostade (eds.), *Studies in Late Modern English Correspondence: Methodology and data*. Bern: Peter Lang, 137–162.
- Sankoff, Gillian & Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83 (3), 560–588.
- van Sas, Niek C.F. 2004. *De metamorphose van Nederland. Van oude orde naar moderniteit, 1750–1900*. Amsterdam: Amsterdam University Press.
- Schneider, Edgar W. 2013. Investigating historical variation and change in written documents. New perspectives. In Jack K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*. Second edition. Malden, MA/Oxford: Wiley-Blackwell, 57–81.
- Schoemaker, Bob. 2018. *Genijd der jeugd voor taal en deugd. Het onderwijs in de Nederlandse taal op de lagere school, 1750–1850*. PhD dissertation. Utrecht: LOT.
- Schoemaker, Bob & Gijsbert Rutten. 2017. Standard language ideology and Dutch school inspection reports (1801–1854). *Sociolinguistica* 31, 101–116.
- Schoonenboom, Judith. 1997. De geschiedenis van *dat*, *wat* en *hetgeen* in bijbelvertalingen. *Nederlandse Taalkunde* 2, 343–369.
- Schoonenboom, Judith. 2000. *Analyse, norm en gebruik als factoren van taalverandering. Een studie naar veranderingen in het Nederlands onzijdig relativum*. Unpublished PhD dissertation: Universiteit van Amsterdam.
- de Schutter, Georges & Hanne Kloots. 2000. Relatieve woorden in het literaire Nederlands van de 17^e eeuw. *Nederlandse Taalkunde* 5, 325–342.
- Scott, Alan K. 2014. *The genitive case in Dutch and German: A study of morphosyntactic change in codified languages*. Leiden/Boston: Brill.
- Sevic, Radmila B. 1999. Early collections of private documents: The missing link in the diachronic corpora. In Christopher Beedham (ed.), *Langue and parole in synchronic and diachronic perspective*. Amsterdam, Pergamon, 337–347.

- Sijs, Nicoline van der. 2004. *Taal als mensenwerk. Het ontstaan van het ABN*. The Hague: Sdu.
- Simons, Tanja. 2013. *Ongekend 18e-eeuws Nederlands. Taalvariatie in persoonlijke brieven*. PhD dissertation. Utrecht: LOT.
- Simons, Tanja & Gijsbert Rutten. 2014. Language norms and language use in eighteenth-century Dutch: Final n and the genitive. In Gijsbert Rutten, Rik Vosters & Wim Vandenbussche (eds.), *Norms and usage in language history. A sociolinguistic and comparative perspective*. Amsterdam/ Philadelphia: John Benjamins, 49–72.
- Taeldeman, Johan. 2013. The southwestern dialect area: Phonology. In Frans Hinskens & Johan Taeldeman (eds.), *Language and space. An international handbook of linguistic variation. Volume 3: Dutch*. Berlin: De Gruyter, 150–173.
- Tagliamonte, Sali A. & Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24 (2), 135–178.
- Takada, Hieroyuki. 1998. *Grammatik und Sprachwirklichkeit von 1640–1700. Zur Rolle deutscher Grammatiker im schriftsprachlichen Ausgleichsprozess*. Tübingen: Niemeyer.
- Tieken-Boon van Ostade, Ingrid. 1987. *The auxiliary do in eighteenth-century English: A socio-historical linguistic approach*. Dordrecht: Foris.
- Tieken-Boon van Ostade, Ingrid (ed.). 2008. *Grammars, grammarians and grammar-writing in eighteenth-century England*. Berlin/New York: De Gruyter.
- van der Velde, Isaac. 1956. *De tragedie der werkwoordsvormen. Een taalhistorische en taaldidactische studie*. Groningen/Jakarta: J. B. Wolters.
- Vezzosi, Letizia. 2000. The history of the genitive in Dutch: An evidence of the interference between language standardisation and spontaneous drift. *Studia Germanica Posnaniensia* 26, 115–147.
- Vis, George J. 2004. *Van Siegenbeek tot Lodewick: verkenningen naar de geschiedenis van de studie der Nederlandse letterkunde, speciaal in het onderwijs*. Amsterdam: Amsterdamse Historische Reeks.
- Vosters, Rik. 2011. *Taalgebruik, taalnormen en taalbeschouwing in Vlaanderen tijdens het Verenigd Koninkrijk der Nederlanden: Een historisch-sociolinguïstische verkenning van vroeg-negentiende-eens Zuidelijk Nederlands*. PhD dissertation. Vrije Universiteit Brussel.
- Vosters, Rik, Els Belsack, Jill Puttaert & Wim Vandenbussche. 2014. Norms and usage in nineteenth-century Southern Dutch. In Gijsbert Rutten, Rik Vosters &

Wim Vandenbussche (eds.), *Norms and usage in language history, 1600-1900. A sociolinguistic and comparative perspective*. Amsterdam/Philadelphia: John Benjamins, 72–100.

Vosters, Rik, Gijsbert Rutten & Marijke van der Wal. 2010. Mythes op de pijnbank. Naar een herwaardering van de taalsituatie in de Nederlanden in de achttiende en negentiende eeuw. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 120, 93–112.

Vosters, Rik, Gijsbert Rutten & Wim Vandenbussche. 2012. The sociolinguistics of spelling. A corpus-based case study of orthographical variation in nineteenth-century Dutch in Flanders. In Ans van Kemenade & Nynke de Haas (eds.), *Historical Linguistics 2009. Selected papers from the 19th International Conference on Historical Linguistics*. Amsterdam/Philadelphia: John Benjamins, 253–273.

van der Wal, Marijke. 1995. *De moedertaal centraal. Standaardisatie-aspecten in de Nederlanden omstreeks 1650*. The Hague: Sdu.

van der Wal, Marijke. 2002a. Lambert ten Kate and eighteenth-century Dutch linguistics. *Beiträge zur Geschichte der Sprachwissenschaft* 12, 49–63.

van der Wal, Marijke. 2002b. Relativisation in the history of Dutch: Major shift or lexical change? In P. Poussa (ed.), *Relativisation on the North Sea Littoral*. Munich: Lincom. 27–35.

van der Wal, Marijke. 2003. Relativiteit in de grammaticale traditie. Tussen norm en descriptie? In Els Ruijsendaal, Gijsbert Rutten & Frank Vonk (eds.), *Bonjours Neef, ghoeiden daghb Cozyn*. Münster: Nodus, 361–375.

van der Wal, Marijke. 2006. *Onvoltooid verleden tijd. Witte vlekken in de taalgeschiedenis*. Amsterdam: Koninklijke Nederlandse Academie van Wetenschappen.

van der Wal, Marijke & Cor van Bree. 2008. *Geschiedenis van het Nederlands*. Houten/Antwerpen: Spectrum.

Watts, Richard J. 2012. Language myths. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell, 585–606.

Watts, Richard J. & Peter Trudgill. 2002. *Alternative histories of English*. London: Routledge.

Weerman, Fred & Petra de Wit. 1999. The decline of the genitive in Dutch. *Linguistics* 37 (6), 1155–1192.

- Weerman, Fred, Mike Olson & Robert Cloutier. 2013. Synchronic variation and loss of case: Formal and informal language in a Dutch corpus of 17th-century Amsterdam texts. *Diachronica* 30 (3), 353–381.
- Wegera, Klaus-Peter. 2013. Language data exploitation: design and analysis of historical language corpora. In Paul Bennett, Martin Durrell, Silke Scheible & Richard J. Whitt (eds.), *New methods in historical corpora*. Tübingen: Narr, 55–73.
- Weijnen, A. A. 1966. *Nederlandse dialectkunde. Tweede druk*. Assen: Van Gorcum & Dr. H.J. Prake & H.M.G. Prakke.
- Weinreich, Uriel, William Labov & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics. A symposium*. Austin/London: University of Texas Press, 95–195.
- Willemyns, Roland. 2013. *Dutch: Biography of a language*. Oxford: University Press.
- Willemyns, Roland & Wim Vandenbussche. 2006. Historical sociolinguistics: Coming of age? *Sociolinguistica. International Yearbook of European Sociolinguistics* 20, 146–165.
- Włodarczyk, Matylda. 2013. Self-reference and ego involvement in the 1820 Settler petition as a *leaking genre*. In Marijke J. van der Wal & Gijsbert Rutten (eds.), *Touching the past: Studies in the historical sociolinguistics of ego-documents*. Amsterdam/Philadelphia: John Benjamins, 201–224.
- Wright, Sue. 2012. Language policy, the nation and nationalism. In Bernard Spolsky (ed.), *The Cambridge Handbook of Language Policy*. Cambridge: University Press, 59–78.
- Yáñez-Bouza, Nuria. 2011. *ARCHER past and present (1990-2011)*. Poster presented at the 32nd Conference of the International Computer Archive of Modern and Medieval English (ICAME 32), Oslo, 1–5 June 2011 [updated to October 2013].
- Yáñez-Bouza, Nuria. 2016. Daily jottings: Preposition placement in English diaries and travel journals from 1500 to 1900. *Folia Linguistica Historica* 37, 281–314.
- Zeman, Sonja. 2013. Historische Mündlichkeit. Empirische Erörterung einer theoretischen Problemlage. *Zeitschrift für Germanistische Linguistik* 41 (3), 377–412.
- Ziegler, Evelyn. 2007. Putting standard German to the test: some notes on the linguistic competence of grammar-school students and teachers in the nineteenth century. In Stephan Elspaß, Nils Langer, Joachim Scharloth & Wim Vandenbussche (eds.), 2007. *Germanic Language Histories 'from below' (1700–2000)*. Berlin: De Gruyter, 309–330.

APPENDIX

Transcription conventions for handwritten ego-documents

Header information

```
<header>
DOCUMENT:      Rotterdam_Ledeboer_1237_207_let02 (without
                a/b/c and .jpg)
ARCHIVE:       Stadsarchief Rotterdam (full name of archives)
GENRE:         letter / diary
DATE:          1829-07-25 (i.e. YYYY-MM-DD)
PLACE:         Leuvenum (place as indicated on the source document)
TRANSCRIPTION: AK / CB / HV / GW (insert initials)
NOTES:         (Optional notes or comments)
WORD COUNT:    1048 (transcription word count without header)
</header>
```

Text file names

Save all transcriptions as Text files, e.g. *Rotterdam_Ledeboer_1237_207_let02.txt*. Transcriptions which are based on various digital images from the same source document (e.g. various pages of the same letter) are saved as one text file.

Diplomatic transcription

The aim of a diplomatic transcription is to present the source document as it actually appears. In other words, do not correct or normalise the spelling and punctuation as used in the original document.

Use of tags

Ambiguous and illegible words:

The following three tags represent three different levels between ambiguity and total illegibility:

<code><illeg/></code>	An entirely illegible word (note: for sequences of more than one illegible word, please use this tag for each word)
<code><illeg com="word..." /></code>	An illegible word, but a suggested or partial reading
<code><ambig>word</ambig></code>	A probable reading, but possibly ambiguous

Insertions and deletions

<code><ins>word</ins></code>	An inserted word or (part of a) sentence, which is added outside of the normal line of text (note: use this tag for the entire insertion)
<code>word</code>	A crossed out or deleted word which is still readable
<code><cancel/></code>	A crossed out or deleted word which is no longer readable
<code><gap/></code>	Parts of the source document which are not transcribed (e.g. long list of names, accounts of expenses)

Character layout

Given the fact that all ego-documents are handwritten (i.e. not printed), it is unlikely to find instances of bold or italicised words. Emphasis through underlining does occur regularly, though:

<code><u>word</u></code>	An underlined word
--------------------------------------	--------------------

Hyphenation

Always transcribe hyphens and alternative line break characters with – (which may be realised differently in handwriting). Hyphenated words at the end of a line are restored before the line break, but the following tag is used to indicate the original form (reg = regularisation; orig = original form):

<code><reg orig="word">word</reg></code>
--

Capitalisation

Although it might not always be clear whether a letter is capitalised or not, capitalisation in ego-documents should be transcribed as diplomatically as possible. Even when (unusual) capital letters occur in the middle of a word, please also transcribe them as capitals accordingly.

Spacing

“Spacing between words is often problematic. The general rule of thumb is that *when a white space* (not long enough to be a blank) *is clearly distinguishable, it is transcribed thus*, using the space bar once. If, however, *no space is observed between two separate and unrelated words (not compounds!)*, as is often the case, *modern spacing conventions are used*, and the words are transcribed separately. In these cases, we assume the lack of spacing to be purely typographical. *In the case of compounds or otherwise related words, a lack of spacing is interpreted as a linguistic feature, and thus transcribed diplomatically.*” (Vosters 2009)⁸⁹

Editorial comments

<ed com="comment" />

Additional comments in the transcription
(moderately used!)

Special case: ij/y

The orthographic representation of *ij/y* is particularly variable in handwritten ego-documents (cf. Chapter 9). Even within the same document, you might come across different spelling variants. The following four variants should be distinguished in the transcriptions:

<ij>	<i>Lange ij</i> with <i>i</i> and <i>j</i> as neatly separated characters
<ÿ>	<i>Griekse y</i> with two dots (i.e. <i>ÿ</i>)
<y>	<i>Griekse y</i> (no dots)
<°y>	All alternative representations, e.g. single-dotted <i>y</i> , <i>y</i> with accent marks or other diacritics (note: the exact position of dots, accents or other diacritics is not relevant)

⁸⁹ Vosters, Rik. 2009. *Linguistic aspects of law and justice in the United Kingdom of the Netherlands (1815–1830). Transcription guidelines and editorial practices*. Third edition.

Dutch summary (*Samenvatting*)

Beleid versus praktijk. Taalvariatie en taalverandering in achttiende- en negentiende-eeuws Nederlands

In dit proefschrift onderzoek ik taalvariatiepatronen en taalveranderings-processen in het Nederlands van de late achttiende en vroege negentiende eeuw. Specifieker ga ik na in hoeverre het nationale taalbeleid rond 1800 invloed heeft gehad op het daadwerkelijke taalgebruik in de Noordelijke Nederlanden, d.w.z. het gebied dat ongeveer overeenkomt met het huidige Nederland.

Vanuit een historisch-sociolinguïstisch perspectief blijken de achttiende en negentiende eeuw een bijzonder intrigerende periode voor onderzoek naar het Nederlandse taalverleden te zijn. Met name de sociaal-politieke ontwikkelingen in de decennia rond 1800, de beginfase van de natievorming en nationalistische taalplanning, leveren een interessant vertrekpunt op. In de tweede helft van de achttiende eeuw ontstond in de Noordelijke Nederlanden een sterk nationalistisch gekleurd debat, waarin de moedertaal toenemend als identiteitssymbool voor ‘de’ Nederlandse natie werd gezien. Met de opkomst van de standaardtaalideologie in Europa groeide ook de noodzaak om een uniforme standaardvariëteit van het Nederlands vast te leggen. In het laatste decennium van de achttiende eeuw werd de standaardisering een nationale kwestie en ontwikkelde de landsregering concrete plannen om een eenheidsspelling en een eenheidsgrammatica te ontwerpen. Dit resulteerde in de beginjaren van de negentiende eeuw in een officieel taalbeleid, dat de grootschalige verspreiding van de standaardtaal beoogde. In het kader van de zogenaamde *schrijftaalregeling* werd in 1804 en 1805 de Nederlandse spelling en grammatica van overheidswege gecodificeerd. De taalnormen, vastgelegd in de officiële orthografie (Matthijs Siegenbeeks *Verhandeling over de Nederduitsche spelling*, 1804) en de officiële grammatica (Petrus Weilands *Nederduitsche spraakkunst*, 1805) werden ten eerste aanbevolen voor gebruik door de overheid en in het onderwijs. Gezien het officiële karakter van de codificatie en de bemoeienis van de overheid met talige kwesties markeert de schrijftaalregeling dan ook een belangrijk keerpunt in het standaardiseringsproces van het Nederlands. Daarnaast introduceerden de officiële taalbeleidsmaatregelen een hiërarchisch onderscheid tussen het superieure Standaardnederlands enerzijds en de ‘niet-standaard’ variëteiten anderzijds. Opmerkelijk is dat de gevolgen van deze ingrijpende verandering in de sociolinguïstische situatie op het negentiende-eeuwse taalgebruik nooit eerder zijn onderzocht. Met dit proefschrift, dat tot stand is gekomen in het kader van het door NWO gefinancierde Vidi-project *Going Dutch. The Construction of Dutch in Policy, Practice and Discourse (1750–1850)* aan de Universiteit Leiden, wil ik deze lacune in de geschiedschrijving van het Nederlands opvullen.

In het inleidende eerste hoofdstuk schets ik de grote lijnen van mijn promotieonderzoek en presenteer ik de centrale onderzoeksvragen van dit proefschrift. Het doel is om na te gaan hoe effectief oftewel ‘succesvol’ de Nederlandse taalbeleidsmaatregelen zijn geweest in het verspreiden van de standaardtaalnormen onder de bevolking. Met andere woorden, volgden de burgers van de Nederlandse natie de voorschriften van Siegenbeek en Weiland? Het was in feite niet wettelijk verplicht om dat ook daadwerkelijk te doen, zeker niet in het privédomen. De effectiviteit van het taalbeleid bestudeer ik door middel van een systematische analyse van authentieke taalgebruiksdata uit de late achttiende en vroege negentiende eeuw, die bij elkaar het *Going Dutch Corpus* vormen. Dit diachrone multigenre-corpus heb ik speciaal voor dit proefschrift samengesteld om de onderzoeksvragen te kunnen beantwoorden, met name over de potentiële invloed van de schrijftaalregeling op de taalpraktijk.

Hoofdstuk 2 behandelt de historische achtergrond van mijn onderzoek, voornamelijk in de periode rond 1800. Allereerst beschrijf ik de sociaal-politieke situatie in de late achttiende en vroege negentiende eeuw, waarbij de contouren van de natievorming en het nationalisme in de Noordelijke Nederlanden worden toegelicht. In de nasleep van de Franse revolutie vonden de meest cruciale veranderingen in de zogeheten Bataafs-Franse tijd plaats. Dit leidde in 1795 tot de oprichting van de Bataafse Republiek, zoals de Noordelijke Nederlanden tussen 1795 en 1801 heetten. Ook op het gebied van taal en onderwijs trof de nieuwe regering, en de benoemde Agent van Nationale Opvoeding in het bijzonder, fundamentele maatregelen om de nationale eenheid te bevorderen. Hiertoe behoorden de invoering van een nationaal schoolsysteem en de schrijftaalregeling in 1804/1805 – het historisch vertrekpunt van dit proefschrift. Verder ga ik in dit hoofdstuk kort in op het belang van de achttiende-eeuwse beweging van *volksverlichting*. Het doel hiervan was het verspreiden van ‘verlichting’ door alle lagen van de samenleving en de vorming van een homogene natie. Ook in het metalinguïstische discours vonden in de loop van de achttiende eeuw parallelle ontwikkelingen plaats. Deze laten zich onderverdelen in drie fasen, voornamelijk bepaald op basis van de beoogde doelgroepen: van grammatica’s voor de elitaire kringen (1700–1740) naar grammatica’s voor de bourgeoisie (1740–1770), en uiteindelijk de weg naar een nationale grammatica voor de gehele bevolking (vanaf 1770). Het achttiende-eeuwse normendebat resulteerde in de schrijftaalregeling, d.w.z. de codificatie van de spelling-Siegenbeek en de grammatica van Weiland. Tenslotte wordt in dit hoofdstuk ook het vernieuwde taalonderwijsbeleid in de vroege negentiende eeuw aangesneden. De voorschriften van Siegenbeek en Weiland werden van overheidswege geïmplementeerd in het lager onderwijs, met het plan om de kennis van het Standaardnederlands zo effectief mogelijk te verspreiden. In dit proefschrift ga ik empirisch na in hoeverre dit doel in de praktijk werd bereikt.

Hoofdstuk 3 geeft een algemene inleiding in het theoretisch kader, waarin het gepresenteerde promotieonderzoek is ingebed. Centraal hierbij staat de historische sociolinguïstiek, die in de afgelopen drie decennia tot een internationaal onderzoeksveld binnen de taalkunde is uitgegroeid. Om talen in hun

sociaalhistorische context te kunnen bestuderen, maken onderzoekers uit dit interdisciplinaire veld gebruik van theorieën en bevindingen uit de moderne sociolinguïstiek, historische taalkunde, sociale geschiedenis, maar daarnaast ook de corpuslinguïstiek, dialectologie en andere verwante disciplines. Vervolgens bespreek ik de basisprincipes, waaraan de historische sociolinguïstiek ten grondslag ligt, maar ook de uitdagingen (en kansen) die het onderzoek met historische tekstbronnen met zich meebrengt. Bijzondere aandacht besteed ik hierbij aan de verschillende perspectieven op taalgeschiedenis. De traditionele geschiedschrijving van de meeste Europese talen, waaronder het Nederlands, schetst vaak een schijnbaar lineaire ontwikkeling van een diversiteit aan dialecten naar een uniforme standaardtaal. De onderliggende standaardtaalideologie heeft ertoe geleid dat het onderzoek voor een groot deel is gebaseerd op standaardtalige bronnen. Over het algemeen zijn dit gedrukte en vaak literaire werken, geschreven door mannen uit de hoogste kringen. Met andere woorden, de traditionele bronnen representeren slechts een beperkt deel van de samenleving, waardoor een substantieel deel van de taalwerkelijkheid onzichtbaar blijft. Door de nadruk te leggen op andere bronnen en door andere vragen te stellen, heeft de historische sociolinguïstiek en dan met name de *language history from below*-benadering een belangrijke bijdrage geleverd aan onze kennis over taalgeschiedenis. In plaats van gedrukte werken wordt veelal gebruik gemaakt van handgeschreven egodocumenten, zoals brieven, dagboeken en reisjournalen. Deze bronnen hebben het voordeel dat ze minder formeel zijn en dichter bij de authentieke (gesproken) taal staan dan de voorheen bestudeerde teksten. Bovendien zijn egodocumenten door een groter deel van de samenleving geschreven: mannen en vrouwen uit verschillende sociale klassen en regio's. Verder biedt dit hoofdstuk een overzicht van historisch-sociolinguïstische corpora voor verschillende Europese talen. Ook de centrale onderwerpen taalplanning, taalbeleid en prescriptivisme worden kort behandeld. Ondanks het feit dat de wisselwerking tussen taalnormen en taalgebruik in de historische sociolinguïstiek een steeds belangrijkere rol inneemt, valt op dat er tot nu weinig bekend is over de effecten van concrete taalbeleidsmaatregelen op de taalpraktijk. Deze lacune wordt in dit proefschrift verder onderzocht.

In hoofdstuk 4 introduceer ik het samengestelde corpus en de gehanteerde methode. Om na te kunnen gaan hoe effectief het Nederlandse taalbeleid op het taalgebruik daadwerkelijk is geweest, heb ik voor dit proefschrift een diachroon multigenre-corpus laat achttiende- en vroeg negentiende-eeuws Nederlands gebouwd. Het *Going Dutch Corpus* omvat ruim 420.000 woorden, gelijkmatig verdeeld over twee tijdvakken, die de generaties taalgebruikers van vóór (1770–1790) en ná (1820–1840) de schrijftaalregeling representeren. Het corpusdesign sluit dus nauw aan bij de gestelde onderzoeksvragen. Allereerst wordt in dit hoofdstuk het proces van het verzamelen, selecteren en transcriberen van de corpusdata gedetailleerd toegelicht. Vervolgens ga ik nader in op de verschillende variationele dimensies van het corpus, d.w.z. de geïntegreerde externe variabelen: de genredimensie, de diachrone dimensie, de geografische dimensie en de sociale dimensie. Ervan uitgaande dat taalveranderingsprocessen verschillende genres in verschillende mate beïnvloeden, omvat het *Going Dutch Corpus* drie soorten

bronnen: (1) privébrieven, (2) dagboeken en reisjournalen, en (3) kranten. Privébrieven gelden in de historische sociolinguïstiek als de beste teksten om de taalwerkelijkheid in het verleden te bestuderen, omdat ze van alle schriftelijke bronnen de meest informele zijn en ze het dichtst bij de authentieke gesproken taal staan. Daarnaast bevat het corpus nog een tweede soort egodocumenten: dagboeken en reisjournalen. Vergeleken met privébrieven worden deze teksten over het algemeen als schrijftaliger gekarakteriseerd en staan ze dicht bij de standaard. Naast egodocumenten bevat het corpus ook een genre gedrukte en gepubliceerde teksten, namelijk kranten. Voor historisch-sociolinguïstisch onderzoek blijken kranten bijzonder geschikt omdat ze in die tijd nog lokaal werden geproduceerd en gelezen. De drie genres in het corpus maken het mogelijk om handgeschreven en gedrukte taal met elkaar te vergelijken. De geografische dimensie van het corpus kan worden onderverdeeld in twee niveaus. Ten eerste bevat het corpus data uit zeven regio's: Friesland, Groningen, Noord-Brabant, Noord-Holland, Zuid-Holland, Utrecht en Zeeland. Ten tweede is er een onderscheid gemaakt tussen het centrum van het onderzochte taalgebied enerzijds, en de periferie anderzijds. Tot het kerngebied behoren Noord- en Zuid-Holland en Utrecht, terwijl Friesland, Groningen en Noord-Brabant de meer perifere gebieden vertegenwoordigen. In verband met sociaaleconomische verschuivingen in de loop der eeuwen, laat ik Zeeland hier buiten beschouwing. Naast de genre- en geografische dimensies implementeert het *Going Dutch Corpus* gender als een belangrijke sociale factor. Dit is echter alleen in het geval van de egodocumenten mogelijk. Hierbij heb ik gestreefd naar een zo evenwichtig mogelijke verdeling van mannelijke en vrouwelijke scribenten. Terwijl dit in het brievencorpus grotendeels is gelukt, bleek het geschikte dagboekenmateriaal in Nederlandse archieven voornamelijk door mannen te zijn geschreven. Desondanks is het uiteindelijk bereikte aandeel van zo'n een derde vrouwelijke scribenten een aanzienlijke vooruitgang ten opzichte van de traditionele taalgeschiedschrijving, waarin vrouwen nagenoeg afwezig waren. De sociale achtergrond van het merendeel brief- en dagboekschrijvers in het corpus is te karakteriseren als (welgestelde) midden- en hogere klasse, met uitzondering van de bovenste laag van de maatschappij. Gezien de relatief homogene representatie in het corpus heb ik ervoor gekozen om sociale klasse als externe factor buiten beschouwing te laten. Om de veranderingen in het taalgebruik, maar ook de keuzes van Siegenbeek en Weiland zelf tegen de achtergrond van de voorafgaande normentraditie te kunnen plaatsen, heb ik naast het *Going Dutch Corpus* een normatief corpus samengesteld met achttiende-eeuwse publicaties uit de Noordelijke Nederlanden (het *Normative Corpus of the Northern Netherlands*). Deze verzameling teksten omvat in totaal 31 spellingboekjes, grammatica's en andere taalkundige verhandelingen, gepubliceerd in de periode tussen 1699 en de schrijftaalregeling in 1804/1805. De twee referentiewerken van Siegenbeek en Weiland vormen het eindpunt van de achttiende-eeuwse normentraditie. Het *Going Dutch Corpus* en het normatieve corpus maken het mogelijk om variatie- en veranderingspatronen systematisch te analyseren. In de hoofdstukken 5–12 wordt dit volgens een vast stramien gedaan, waarbij rekening wordt gehouden met de officiële voorschriften van Siegenbeek en Weiland, de

ontwikkelingen in het achttiende-eeuwse normendiscours, en het daadwerkelijke taalgebruik op basis van het *Going Dutch Corpus*.

Wat volgt is het empirische hart van dit proefschrift. Om de effectiviteit van Siegenbeeks spelling en Weilands grammatica te kunnen meten, worden in totaal acht taalverschijnselen onderzocht. De gepresenteerde analyses omvatten ten eerste vijf orthografische variabelen (hs. 5–9), en ten tweede drie morfosyntactische variabelen (hs. 10–12). Allereerst behandel ik drie orthografische casestudy's met betrekking tot medeklinkers. In hoofdstuk 5 onderzoek ik de orthografische representatie van /xt/ aan het einde van een lettergreep in etymologisch verschillende woorden. In de normentraditie van de achttiende eeuw werden zowel <cht> als <gt> als naast elkaar bestaande varianten erkend, hoewel er nog weinig consensus was met betrekking tot de spelling van individuele woorden. Pas in het kader van de schrijftaalregeling introduceerde Siegenbeek op basis van etymologische overwegingen een categorisering van woorden met /xt/, die ik als *cht*-woorden respectievelijk *gt*-woorden omschrijf. Woorden uit de eerste categorie vereisten volgens Siegenbeek de spelling met <cht> vanwege de Auslautverhärtung in het Nederlands, bijvoorbeeld in *kocht*. In de tweede categorie schreef Siegenbeek daarentegen de spelling met <gt> voor, vanwege de etymologie met de letter <g>, bijvoorbeeld in *bragt* < *brenghen*. Uit de corpusanalyse blijkt dat <gt> in de late achttiende eeuw verreweg de meest gebruikte variant was en in rond 80% van alle gevallen voorkwam. Ondanks deze duidelijke voorkeur in de spellingspraktijk werd Siegenbeeks nogal complexe regel succesvol aangenomen. De corpusresultaten voor de vroege negentiende eeuw laten namelijk zien dat <cht> nu voornamelijk voor *cht*-woorden en <gt> voornamelijk voor *gt*-woorden werd gebruikt. Deze opvallende veranderingen signaleren een hoog bewustzijn van de normen met betrekking tot de *cht/gt*-kwestie. Dit wordt in de resultaten nog eens benadrukt door een toename in het hypercorrecte gebruik van <cht> voor *gt*-woorden, met name onder vrouwen en in privébrieven. Siegenbeeks spellingsregel werd grotendeels in alle genres, regio's en door beide genders toegepast.

De casestudy in hoofdstuk 6 focust op de orthografische representatie van /t/ in de auslaut in bepaalde vormen van werkwoorden met een stam op *d*, zoals *worden* en *vinden*. Hierbij houd ik rekening met drie historische varianten, namelijk de fonetische spelling <t>, de morfologische <d>, en <dt>, waar naast het morfologische principe ook analogie en etymologie (< historische werkwoordsuitgang -*del*) een rol spelen. De laatste en tegelijk meest complexe variant werd al gedurende de achttiende eeuw door de meeste grammatici geprefereerd en ook Siegenbeek verkoos <dt> als de officiële spellingsvariant. De corpusresultaten voor de eerste periode laten de wijdverspreide voorkeur voor <dt> echter niet zien. Hieruit blijkt namelijk niet <dt> maar <d> de meest frequente variant te zijn geweest, wat wederom op een grote discrepantie tussen taalnormen en taalgebruik wijst. Pas in de vroege negentiende eeuw kwam hier verandering in, toen een radicale verschuiving van <d> naar <dt> plaatsvond. Tegen de achtergrond van een stabiele situatie in het normendebat laat deze verandering in het gebruik de normatieve invloed van Siegenbeeks voorschriften sterk vermoeden. Op genre-niveau valt op te merken dat negentiende-eeuwse

kranten nagenoeg altijd <dt> spelden, terwijl <d> in de egodocumenten als een relatief gebruikelijk alternatief bleef bestaan. Desondanks kan de algemene toename van <dt> in alle genres, regio's en beide genders worden geconstateerd.

Hoofdstuk 7 presenteert de derde casestudy naar de orthografische representatie van /s/ (< Westgermaans *sk>) in woord-mediale en finale positie als <sch> of <s>, bijvoorbeeld in *tuss(ch)en* en *wens(ch)*. Reeds gedurende de achttiende eeuw gaven de meeste taalcommentatoren de voorkeur aan <sch>. Dit komt grotendeels overeen met de corpusresultaten voor de laat achttiende-eeuwse taalpraktijk, waar <sch> in meer dan 70% werd gebruikt. In privébrieven uit deze periode treft men daarentegen <sch> en <s> als twee nagenoeg even frequente varianten aan. Verder laten de resultaten zien dat vrouwelijke scribenten zelfs <s> boven <sch> prefereerden, wat een nogal gevarieerd beeld van het achttiende-eeuwse taalgebruik weergeeft. In navolging van zijn voorgangers schreef Siegenbeek in 1804 <sch> officieel voor. Uit de corpusresultaten blijkt dat de al dominerende positie van <sch> in de taalpraktijk hierdoor nog werd versterkt. <s> kwam hierna zelfs in brieven en bij vrouwelijke scribenten nauwelijks nog voor. De onbetwistbare positie van <sch> is enigszins verrassend gezien de veranderingen die in de gesproken taal hebben plaatsgevonden. Alhoewel /sx/ in mediale en finale positie niet meer werd uitgesproken, was de verspreiding van de 'oude' spelling <sch> voor deze posities enorm succesvol in alle genres, regio's en beide genders. Ik stel dat ook de veranderingen in het gebruik van deze orthografische variabele zijn toe te schrijven aan de normatieve invloed van Siegenbeeks spelling.

De volgende twee orthografische variabelen behandelen kwesties met betrekking tot het klinkersysteem. In hoofdstuk 8 staat de orthografische representatie van twee etymologisch verschillende lange *ē*'s in open lettergreep centraal. In het metalinguïstische debat behoorde deze kwestie tot een van de meest bediscussieerde onderwerpen. Het fonologische onderscheid tussen de zogenaamde scherplange en zachtlange *ē*'s was in vele regio's, waaronder het gebied rond Amsterdam, reeds verdwenen. In veel dialecten uit Zeeland en Groningen bijvoorbeeld bleef het onderscheid daarentegen bestaan. Hierdoor ontstond in de achttiende eeuw een uiterst variabele situatie met verschillende concurrerende spellingconventies, zowel in het normendebat als in de taalpraktijk. Ondanks de samenval van lange *ē*'s in zijn geboortestad schreef de Amsterdammer Siegenbeek het historisch-fonologische systeem voor. Het etymologische onderscheid werd hierbij bewaard en orthografisch onderscheiden, d.w.z. door een scherplange <ee> en een zachtlange <e> in open lettergreep. De corpusresultaten voor de scherplange *ē* laten zien dat de digraaf <ee> zowel vóór en ná Siegenbeek in 90% van de gevallen voorkwam. De spelling van de zachtlange *ē* onderging daarentegen een opmerkelijke verandering. In de achttiende-eeuwse data is er een lichte voorkeur voor <e> te constateren, maar ook <ee> werd vaak gebruikt. Nadat Siegenbeek de enkele <e> officieel voor zachtlange *e* in open lettergreep had voorgeschreven, nam het gebruik van <e> echter dramatisch toe. De vroeg negentiende-eeuwse resultaten laten in feite een keurige verdeling van de twee varianten zien, waarbij zowel <ee> voor scherplang als ook <e> voor zachtlang in meer dan 90% werd gebruikt, wat geheel overeenkomt met het voorgeschreven

fonologische systeem. Gezien de patronen en tendensen in de achttiende eeuw is het des te opmerkelijker dat Siegenbeeks spellingsregel omtrent de lange *e*'s zo succesvol werd toegepast in alle genres, regio's en door beide genders.

In hoofdstuk 9 onderzoek ik de orthografische representatie van de Westgermaanse **i* in woorden zoals *mijn/myn* en *schrijven/schryven*. Terwijl de vorige vier casestudy's een sterke normatieve invloed van Siegenbeeks orthografie op het negentiende-eeuwse taalgebruik signaleren, duiden de resultaten van deze laatste variabele de grenzen van de effectiviteit aan. In vergelijking met de andere spellingkwesities blijkt Siegenbeeks voorschrift hier minder invloedrijk te zijn geweest. De vraag of men beter <ij> of <y> zou moeten spellen, werd gedurende de achttiende eeuw uitvoerig behandeld. Pas in de laatste decennia tekende zich in het normendebat een voorkeur voor <ij> af. Siegenbeek sloot zich hierbij aan en legde in 1804 <ij> als nationale variant vast. De corpusresultaten bevestigen enerzijds een totale omkeer van <y> naar <ij> in de kranten, d.w.z. een verandering conform het officiële voorschrift. Anderzijds zette de spelling <ij>, of eerder de variant <ÿ>, zich niet door in handgeschreven egodocumenten - noch in privébriefen noch in dagboeken en reisjournalen. In deze bronnen bleven varianten met twee of zonder puntjes in de vroege negentiende eeuw naast elkaar bestaan, net als voor de eeuwwisseling. Als mogelijke verklaring voor de relatief stabiele verdeling van varianten in handgeschreven bronnen komt het vrij kleine diakritische onderscheid tussen de varianten in aanmerking. Mogelijk waren taalgebruikers zich hierdoor minder bewust van een officiële norm dan bij de andere spellingskenmerken. Met uitzondering van de *ij/y*-kwesitie, waarbij de officieel voorgeschreven variant <ij> zich alleen in gedrukte teksten doorzette, blijken Siegenbeeks spellingsregels opmerkelijk effectief te zijn geweest.

Naast de vijf orthografische casestudy's in de hoofdstukken 5–9 onderzoek ik in dit proefschrift ook drie morfosyntactische variabelen. In de hoofdstukken 10–12 ga ik na hoe effectief de grammatica van Weiland is geweest door wederom systematische analyses op basis van het *Going Dutch Corpus* te presenteren. Om te beginnen, behandelen de hoofdstukken 10 en 11 twee casestudy's met betrekking tot het systeem van de Nederlandse relativa. Hoofdstuk 10 focust op onzijdige betrekkelijke voornaamwoorden in subject en objectpositie. Mannelijke, vrouwelijke en meervoudige vormen van de betrekkelijke voornaamwoorden staan vervolgens centraal in hoofdstuk 11. Wat normen voor het gebruik van de relativa betreft, zijn er in achttiende-eeuwse grammatica's nauwelijks concrete voorschriften te vinden. Weiland (1805) was in feite de eerste grammaticus die dit onderwerp relatief uitvoerig behandelde en een aantal expliciete regels opstelde. Desondanks blijkt het vroeg negentiende-eeuwse taalgebruik opvallend stabiel ten opzichte van de vorige eeuw. Dit toont aan dat het normatieve effect van Weilands grammatica – in tegenstelling tot Siegenbeek spellingsregels – beduidend minder zichtbaar is. Wat de corpusresultaten eerst en vooral laten zien is een uiterst variabele situatie in het geval van de onzijdige relativa, waarvan niet minder dan vijf varianten voorkomen. Naast de traditionele *d*- en *w*-vormen *dat* en *wat* moet er dus ook rekening worden gehouden met de pronominale vormen *hetgeen*, *hetwelke* en *welke*, hoewel de laatstgenoemde relatief

weinig voorkomt. In het geval van de mannelijke, vrouwelijke en meervoudige relativa treffen we in het corpus vooral voorkomens van de *d*-vorm *die* en de alternatieve vorm *welke* aan. De *n*-vorm *wie* blijkt evenals *denwelke* nauwelijks voor te komen. Bovendien wordt in elk van de twee hoofdstukken de potentiële invloed van een interne factor onderzocht: de bepaaldheid van het antecedent bij de onzijdige relativa, en de bezieldeheid (of *animacy*) van het antecedent bij de mannelijke, vrouwelijke en meervoudige relativa. In tegenstelling tot wat enkele toonaangevende studies naar Nederlandse relativa beweren, signaleren de corpusdata in dit proefschrift geen substantiële invloed van de onderzochte interne factoren op de verdeling van de varianten. De doorslaggevende factoren blijken in beide casestudy's van externe aard te zijn. Met name genre heeft grote invloed op het gebruik van betrekkelijke voornaamwoorden. Zo blijkt uit de corpusresultaten dat de alternatieve vormen *betwelke* en *welke* in kranten aanzienlijk vaker voorkwamen dan in privébrieven, waar de *d*-vormen *dat* en *die* veel gebruikelijker waren. Het derde genre, de dagboeken en reisjournalen, neemt een speciale tussenpositie in. Vanuit een diachroon oogpunt valt vooral de toename van *betwelke* en *welke* (ten koste van *dat* en *die*) op. Hierdoor blijken dagboeken zich sterker te ontwikkelen in de 'schriftelijke' richting van de kranten, en zich dus verder te verwijderen van de andere egodocumenten. Ook zijn er in beide hoofdstukken wezenlijke genderverschillen te constateren. Mannelijke scribenten maakten vaker gebruik van de stilistisch 'deftigere' vormen dan vrouwen, terwijl de zogeheten 'gemeenzame' vormen *dat*, *wat* en *die* relatief vaak voorkomen in teksten geschreven door vrouwelijke scribenten. Een directe invloed van Weiland's grammatica op het gebruik van de betrekkelijke voornaamwoorden is in de negentiende-eeuwse data moeilijk terug te vinden. Wel valt op dat Weiland's bewustzijn van stijlverschillen tussen de formelere vormen zoals *betwelke* en *welke* enerzijds, en de informelere *d*- en *n*-vormen anderzijds genrespecifieke keuzes in het gebruik van relativa mogelijk heeft versterkt.

De laatste morfosyntactische casestudy in hoofdstuk 12 focust op de historische genitief en de alternatieve constructie met het voorzetsel *van*. Het voortschrijdende verval van het Nederlandse naamvallensysteem heeft geleid tot een competitieve situatie met zowel synthetische (*de zoon des konings*) als analytische genitiefconstructies (*de zoon van de(n) koning*). In de achttiende eeuw waren synthetische genitiefvormen uit de gesproken omgangstaal verdwenen, ook al kwamen ze nog steeds voor in de geschreven taal, voornamelijk in hogere registers. In het metalinguïstische discours bleef nominale inflectie een van de kernthema's. Gedurende de achttiende eeuw werd de analytische *van*-constructie in toenemende mate als volwaardige optie gezien, en in de laatste decennia zelfs geprefereerd. Tegen deze ontwikkelingen in, ging Weiland (1805) een stap terug en schreef hij in zijn grammatica alleen de synthetische optie voor. De corpusresultaten wijzen erop dat zijn conservatieve keuze tot op bepaalde hoogte effectief moet zijn geweest. In neutrale (d.w.z. niet-formulaire) contexten is er in de vroege negentiende eeuw een verrassende stijging van de genitief te constateren. Ondanks het feit dat de *van*-constructie in de late achttiende eeuw al in bijna 70% voorkwam, nam het gebruik van genitiefvormen vanuit een diachroon perspectief in alle genres toe.

Opmerkelijk is dat dit zelfs in de meest ‘orale’ bronnen, de privébrieven, gebeurde, wat de normatieve invloed van Weiland laat vermoeden. Bovendien is de toename terug te vinden in het taalgebruik van zowel mannen als vrouwen. Desondanks blijkt één interne factor de verdeling van de constructies aanzienlijk te conditioneren. De corpusdata laten zien dat er cruciale verschillen waren tussen mannelijke en onzijdige vormen enerzijds en vrouwelijke en meervoudsvormen anderzijds. In de negentiende eeuw steeg het gebruik van de genitief bij vrouwelijke en meervoudsvormen, mogelijk (ook) als gevolg van Weilands voorschrift. In het geval van de mannelijke en onzijdige vormen is het hem echter niet gelukt om de synthetische optie te laten herleven. Deze waren al voor de eeuwwisseling te laag in frequentie, terwijl de hogere frequentie van structuren zoals *x der y* ongetwijfeld een gunstige bijdrage heeft geleverd aan de toename van de genitief in vrouwelijke en meervoudsvormen. Weilands voorkeur voor de oude genitief was dus enigszins succesvol. De inmiddels gevestigde *van*-constructie kon hij echter niet meer verdringen. Op basis van de resultaten in de hoofdstukken 10-12 valt ten slotte te constateren dat officiële voorschriften op het gebied van orthografie veel effectiever zijn geweest dan op het gebied van grammatica.

Naast de acht casestudy's op basis van het *Going Dutch Corpus*, waarin ik de ontwikkelingen in de samenleving vanuit een multigenre-perspectief heb bestudeerd, zoom ik in hoofdstuk 13 in op de individuele taalgebruiker als mogelijke factor. Om intra- en interindividuele variatie en verandering te kunnen onderzoeken, heb ik speciaal voor dit doeleinde een brievencorpus samengesteld met persoonlijke correspondentie uit het archief van de familie Martini Buys. Het *Martini Buys Correspondence Corpus* (bijna 64.000 woorden) omvat zo'n honderd privébrieven, geschreven door elf familieleden uit drie generaties uit de tweede helft van de achttiende en de eerste helft van de negentiende eeuw. Het Martini Buys-corpus bevat dus materiaal van vóór en ná de schrijftaalregeling en concreter de spelling-Siegenbeek, waardoor het mogelijk is om de invloed op individuele spellingspraktijken onder de loep te nemen. Dezelfde vijf orthografische variabelen uit hoofdstukken 5-9 worden hier op microniveau bestudeerd, wat een vergelijking met de resultaten uit het *Going Dutch Corpus* mogelijk maakt. Ten opzichte van het succes van het taalbeleid blijkt dat enkele scribenten uit de interessante ‘tussengeneratie’, die de schrijftaalregeling op hogere leeftijd meemaakten, hun spellingskeuzes veranderden in de richting van Siegenbeeks voorschriften. Deze bevindingen werpen nieuw licht op het bewustzijn en de acceptatie van officiële spellingsnormen – onafhankelijk van taalonderwijs op school. Tevens blijkt dat er ook scribenten van de jongste generatie waren, die met hun spellingskeuzes systematisch afweken van de voorschriften. Een empirische microlevel-analyse naar individueel taalgebruik kan onze algemene kennis van beleidseffectiviteit dus aanzienlijk verrijken.

In het slothoofdstuk worden de belangrijkste bevindingen van dit proefschrift samengevat en concluderend besproken. Daarnaast sta ik in bredere zin stil bij de bijdrage die dit onderzoek levert aan het onderzoeksveld. Allereerst leveren de resultaten van de verschillende casestudy's nieuwe inzichten op over variatiepatronen en veranderingsprocessen in het Nederlands van de late achttiende

en vroege negentiende eeuw. Specifieker vult dit proefschrift een belangrijke lacune op in het onderzoek naar de geschiedenis van het Nederlands, waarin de effectiviteit van de eerste officiële taalbeleidsmaatregelen rond 1800 nooit eerder is onderzocht. Ook in de geschiedschrijving van de meeste andere (Europese) talen is er tot nu toe weinig aandacht besteed aan de invloed van taalplanning en taalbeleid op het daadwerkelijke gebruik. De hier gepresenteerde systematische benadering, voor een overgroot deel gebaseerd op het *Going Dutch Corpus*, laat zien dat het wel degelijk mogelijk is om de invloed van taalbeleid op de taalpraktijk in het verleden empirisch te onderzoeken. Hierbij blijken zowel het op maat gemaakte en gebalanceerde corpusdesign als de ruime selectie aan variabelen essentieel te zijn geweest om de beleidseffectiviteit genuanceerd te kunnen onderzoeken en beoordelen. Daarnaast levert het onderzoek een relevante bijdrage aan de levendige discussies over taalnormen en taalgebruik in de internationale historische sociolinguïstiek. De casus van de Nederlandse schrijftaalregeling in de vroege negentiende eeuw dient hierbij als een intrigerend en concreet vertrekpunt voor een systematische analyse. Daarmee vormt dit proefschrift een solide basis voor toekomstig (comparatief) onderzoek naar taalbeleid in de geschiedenissen van andere Europese talen.

Curriculum Vitae

Andreas Krogull was born on 21 September 1985 in Gelsenkirchen, Germany. In 2007, he started studying Germanic languages at the University of Duisburg-Essen, graduating with a Bachelor's degree (2-*Fach-B.A.*) in Anglophone Studies and Dutch Language and Culture. In 2013, Andreas further obtained his Master's degree (2-*Fach-M.A.*) in English Linguistics and German Language and Culture, after spending an Erasmus semester at the Radboud University Nijmegen the year before. Both during and after his MA studies, he also worked as research assistant and lecturer at the Department of Anglophone Studies in Essen.

In 2014, Andreas joined the Leiden University Centre for Linguistics (LUCL) as a PhD candidate in the NWO-funded Vidi research programme *Going Dutch. The Construction of Dutch in Policy, Practice and Discourse (1750–1850)*. In his sub-project, he conducted historical-sociolinguistic research on the effects of the Dutch language policy in the early 1800s on language practice, under the supervision of Professor Marijke van der Wal and Dr Gijsbert Rutten. The present dissertation is the outcome of his four-year period of PhD research. Moreover, Andreas worked as lecturer in the history of Dutch at Leiden University and has published several articles and book chapters.