



Universiteit
Leiden
The Netherlands

Pluricentriciteit in de taalgeschiedenis: bouwstenen voor een geïntegreerde geschiedenis van het Nederlands (16de - 19de eeuw)

Van de Voorde, I.

Citation

Van de Voorde, I. (2022, April 19). *Pluricentriciteit in de taalgeschiedenis: bouwstenen voor een geïntegreerde geschiedenis van het Nederlands (16de - 19de eeuw)*. LOT dissertation series. Amsterdam, LOT. Retrieved from <https://hdl.handle.net/1887/3284980>

Version: Publisher's Version

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

Downloaded from: <https://hdl.handle.net/1887/3284980>

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

Hoofdstuk 4 – Corpus en methode

1 Inleiding

Om pluricentriciteit in het Nederlandse taalgebied historisch te onderzoeken is er nood aan een corpus dat verschillende eeuwen en genres bestrijkt, en waarin specifieke regio's betrokken worden. Voorgaand onderzoek leverde al verschillende historisch-taalkundige corpora op voor het Nederlands. We denken hierbij o.m. aan het VKN-corpus, dat bestaat uit administratieve en gerechtelijke teksten uit hoofdzakelijk 1823 en 1829. De naam van dit corpus (VKN) verwijst dan ook naar de periode van het Verenigd Koninkrijk der Nederlanden. De teksten zijn afkomstig uit verschillende regio's in de Zuidelijke Nederlanden (Vosters, 2011, pp. 187-217). Het *Brieven als Buit* (BaB)/*Letters as Loot* (LaL) corpus, dat aan de Universiteit Leiden tot stand kwam, bevat buitgemaakte privébrieven uit de periode 1664-1674 en uit de periode 1776-1784. Deze rijke collectie brieven liet toe om sociale en regionale variatie binnen de Noordelijke Nederlanden te onderzoeken (Rutten & Van der Wal, 2014, pp. 1-17). Nog recenter is het *Going Dutch Corpus*. Dit multigenre corpus omvat egodocumenten en kranten uit de periode 1770-1790 en uit de periode 1820-1840. Het corpus heeft eveneens een regionale dimensie, met materiaal uit verschillende regio's in de Noordelijke Nederlanden (Krogull, 2018, pp. 37-69).

De doeleinden voor het onderhavige onderzoek zijn relatief omvattend. Allereerst vraagt een studie van pluricentriciteit om specifieke regio's. Zo moeten we putten uit Noordelijk én Zuidelijk materiaal, afkomstig uit centrale en perifere gebieden. Bovendien willen we met pluricentriciteit een historisch perspectief innemen, waardoor we een langere periode in de geschiedenis moeten beschouwen. Tot slot willen we genreverschillen onderzoeken, waardoor een multigenre corpus ook gewenst is. Een corpus dat aan deze voorwaarden voldoet bestond tot nog toe niet voor het Nederlands. Voor andere Germaanse talen werden in het verleden wel al diachrone, multigenre corpora ontworpen. Als voorbeeld noemen we hier het *Helsinki Corpus of English Texts* (o.a. Kytö, 1996) en *ARCHER* (o.a. Biber, Finegan & Atkinson, 1994) voor het Engels. Het Duits heeft dan weer het *GerManC* corpus (o.a. Durrell et al., 2012). In deze reeks volgt nu het

Historical Corpus of Dutch (HCD), een diachroon multigenre corpus dat ontworpen en samengesteld werd aan de Vrije Universiteit Brussel en de Universiteit Leiden (zie voor details Paragraaf 4).

Het HCD strekt zich uit van de 16^{de} tot en met de 19^{de} eeuw, met een dwarsdoorsnede rond het midden van elke eeuw (i.e. 1550, 1650, 1750 en 1850). Deze jaartallen vormen het uitgangspunt, maar voor de gekozen teksten werd een marge voorzien van twintig jaar voor en twintig jaar na elk jaartal.¹³ Het gaat dus om de volgende periodes: 1530-1570, 1630-1670, 1730-1770 en 1830-1870. Naast de indeling in periodes, bevat het corpus ook de vereiste regionale dimensie: voor elke periode werden bronnen uit de Noordelijke en Zuidelijke Nederlanden verzameld, met aandacht voor centrale en perifere gebieden. De centra situeren zich in Holland (Noorden) en Brabant (Zuiden), terwijl Zeeland (Noorden) en Vlaanderen (Zuiden) de perifere gebieden vertegenwoordigen. Een derde en laatste dimensie van het corpus bestaat in het genre. We onderscheiden administratieve teksten, egodocumenten en pamfletten. Dat betekent dat het corpus zowel handgeschreven teksten (administratieve teksten en egodocumenten) als gedrukte teksten (pamfletten) bevat. Met administratieve teksten betrekken we bovendien formele teksten in het onderzoek, terwijl egodocumenten net beschouwd worden als minder formele teksten. Zoals we zullen zien bestaan pamfletten uit verschillende subgenres, waardoor ze kunnen verschuiven op het continuüm van formeel naar minder formeel. Paragraaf 2 vangt aan met een bespreking van de drie dimensies in het HCD, namelijk de genredimensie, de diachrone dimensie en de regionale dimensie. We besteden in deze paragraaf ook aandacht aan de opbouw en omvang van het corpus. In Paragraaf 3 gaan we in op elk van de drie genres afzonderlijk. De afbakening van de genres staat daarbij centraal. Paragraaf 4 en 5 belichten respectievelijk het proces van selectie en digitalisering van de teksten in het corpus. In Paragraaf 6 specificeren we tot slot de methode die we bij de corpusanalyses gehanteerd hebben.

¹³ Zeer uitzonderlijk werd nog enkele jaren (maximaal vijf) afgeweken van deze marge, omdat de corpusopbouw anders in het gedrang zou komen. We hebben deze uitzondering enkel voor de egodocumenten moeten invoeren.

2 Historical Corpus of Dutch: opbouw en omvang

2.1 Selectiecriteria

Het HCD werd opgebouwd rond drie dimensies: genre, periode en regio.

Allereerst is het HCD een *multigenre* corpus, bestaande uit drie genres:

- Administratieve teksten
- Egodocumenten
- Pamfletten

Deze genres moeten voorkomen in vier tijdsvensters (*diachroon*):

- 1530-1570
- 1630-1670
- 1730-1770
- 1830-1870

Tot slot bevat het corpus een *regionale* dimensie, met teksten uit vier regio's:

- Brabant
- Holland
- Vlaanderen
- Zeeland

Voor de corpusopbouw streefden we voor elk genre naar 10.000 woorden per regio per periode. Die 10.000 woorden werden bij voorkeur gespreid over meerdere documenten. Concreet richtten we ons voor de administratieve teksten op twee teksten van elk 5.000 woorden, en voor de egodocumenten en pamfletten op vijf teksten van elk 2.000 woorden. Dat impliceert dat we in de meeste gevallen te maken hebben met fragmenten, en dus geen volledige teksten. In de mate van het mogelijke lieten we het begin van de transcriptie wel overeenkomen met het begin van de tekst. Voor de bepaling van het aantal teksten (2 x 5.000 voor ADM, 5 x 2.000 voor EGO en PAM) namen we het genre zelf als uitgangspunt. Van de drie genres zijn de administratieve teksten het meest homogeen. Het was mogelijk om in dit deelcorpus heel vergelijkbare documenten op te nemen (cf. Sectie 3.1). In de twee andere deelcorpora is enige heterogeniteit niet uitgesloten. Egodocumenten zijn eerder heterogeen op grond van individuele verschillen

tussen scribenten (o.a. Nevalainen & Raumolin-Brunberg, 2012, pp. 32-33). Er is dus meer variatie tussen de schrijvers van de egodocumenten. Pamfletten zijn ook heterogeen omdat verschillende soorten teksten tot het genre ‘pamflet’ gerekend kunnen worden (cf. Sectie 3.3). Dat betekent dat we binnen het genre verschillende subgenres kunnen onderscheiden. Daarom kozen we voor egodocumenten en pamfletten een andere spreiding dan voor de administratieve teksten. De uiteindelijke verdeling van het aantal teksten per genre, eeuw en regio wordt weergegeven in de tabellen 1-3.

ADM	16E	17E	18E	19E
Brabant	2	2	2	2
Holland	15	10	5	3
Vlaanderen	1	2	3	3
Zeeland	2	2	2	2

Tabel 1: aantal administratieve teksten per eeuw en regio

EGO	16E	17E	18E	19E
Brabant	1	3	3	3
Holland	5	5	5	5
Vlaanderen	1	3	6	4
Zeeland	1	5	5	5

Tabel 2: aantal egodocumenten per eeuw en regio

PAM	16E	17E	18E	19E
Brabant	8	7	5	5
Holland	2	7	7	5
Vlaanderen	8	4	6	5
Zeeland	0	9	8	5

Tabel 3: aantal pamfletten per eeuw en regio

Uit Tabel 1 leiden we af dat we voor Brabant en Zeeland konden voldoen aan het criterium van twee teksten. Voor Vlaanderen noteren we een enkele keer één tekst in de 16^{de} eeuw, en drie teksten in de 18^{de} en 19^{de} eeuw. Voor Holland vinden we echter aanzienlijk meer teksten, vooral in de 16^{de} en 17^{de} eeuw. Het betreft hier excerpten uit tekstedities die we grondig vergeleken met de originele handschriften. Deze werden als afzonderlijke tekstbestanden opgeslagen, waardoor ze in het corpus als aparte teksten voorgesteld worden. Ook voor Holland in de 18^{de} eeuw maakten we gebruik van drie dergelijke excerpten, aangevuld met twee teksten die specifiek voor

dit project getranscribeerd werden. Voor de overige teksten in dit deelcorpus geldt dat we meer dan twee teksten opnamen indien de twee geselecteerde teksten geen 10.000 woorden opleverden.

Voor de egodocumenten en pamfletten namen we vijf teksten van ongeveer 2.000 woorden als uitgangspunt. In Tabel 2 en Tabel 3 zien we de verdeling van het aantal teksten per genre. Bij de egodocumenten kon het criterium van vijf teksten goed gehandhaafd worden voor Holland en Zeeland. In het pamflettencorpus zien we voor elke regio steeds vijf pamfletten in de 19^{de} eeuw. Voor beide deelcorpora geldt dat we afweken van de vijf teksten in functie van het aantal woorden. Indien er geen vijf geschikte teksten gevonden werden, namen we meer woorden op van de teksten die we wel konden verzamelen. Anderzijds werd het aantal verhoogd indien de vijf teksten samen geen 10.000 woorden opleverden.

2.2 Genredimensie

In dit proefschrift opteren we systematisch voor de term ‘genre’ en niet voor de alternatieve term ‘tekstgenre’, aangezien de selectie van de teksten op basis van niet-talige criteria gebeurd is. Talige en niet-talige criteria stemmen overeen met de interne en externe criteria die Lee (2001, p. 38)¹⁴ aanhaalt om de term ‘genre’ te definiëren:

*A genre [...] is defined as a category assigned on the basis of **external** criteria such as intended audience, purpose, and activity type, that is, it refers to a conventional, culturally recognised grouping of texts based on properties other than lexical or grammatical (co-)occurrence features, which are, instead, the **internal** (linguistic) criteria forming the basis of *text type* categories (Lee, 2001, p. 38; vet en cursief in origineel)*

In Hoofdstuk 2 (Sectie 3.1) haalden we aan dat de traditionele taalgeschiedschrijving vaak het verhaal van standaardisatie presenteert (zie ook Rutten, Vosters & Vandenbussche, 2014, pp. 1-2). Dit hangt nauw samen met de bronnen die daartoe bestudeerd werden: overwegend gedrukte teksten, literaire en formele teksten, vaak afkomstig van de sociale

¹⁴ Lee (2001) baseert zich voor dit onderscheid op Biber (1988).

bovenlaag. De *language history from below* bood tegenwicht door de focus te verschuiven naar minder formele, handgeschreven bronnen van taalgebruikers buiten de elite (Elspaß, 2012b, p. 156; pp. 160-161; Van der Wal, 2006, p. 7). Rutten, Vosters en Vandenbussche (2014, p. 2) pleiten voor een geïntegreerd perspectief, waarbij zowel gestandaardiseerd als minder gestandaardiseerd taalgebruik bestudeerd wordt. Dit perspectief werd gehanteerd bij de opbouw van het multigenre *Going Dutch Corpus* door Krogull (2018, pp. 49-50). Hij stelt dat “[...] 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)” (Krogull, 2018, p. 50). Op dezelfde manier kiezen we met het HCD voor een geïntegreerd perspectief met drie genres: handgeschreven administratieve teksten, handgeschreven egodocumenten en gedrukte pamfletten. Deze genres worden in Paragraaf 3 uitvoeriger besproken. Tabel 4 geeft de woordenaantallen per genre weer.

GENRE

Administratief	154.879
Egodocumenten	150.674
Pamfletten	157.695

Tabel 4: woordenaantallen per genre

Met 10.000 woorden per regio voor elk van de vier periodes, zouden we per genre aan 160.000 woorden komen. De aantallen in Tabel 4 wijken hiervan af. Voor de administratieve component verzamelden we enkel voor Vlaanderen in de 16^{de} eeuw iets minder dan 5.000 woorden. Het lagere aantal voor dit genre wordt hierdoor verklaard. Het aantal woorden dat we voor de egodocumenten aflezen wijkt tot 10.000 woorden af van de beoogde 160.000 woorden. Voor Zeeland in de 16^{de} eeuw konden we slechts een tekst verzamelen, die iets meer dan 3.000 woorden opleverde. Daarnaast konden we het woordenaantal voor Brabant in de 19^{de} eeuw slechts optrekken tot ongeveer 7.000 woorden. Het aantal voor de pamfletten sluit het meeste aan bij het beoogde woordenaantal. Er werden echter geen pamfletten gevonden voor Zeeland in de 16^{de} eeuw, en slechts twee pamfletten voor Holland in diezelfde eeuw. Het aantal in de tabel is desondanks hoog omdat het criterium van 10.000 woorden voor dit deelcorpus soms overschreden werd.

2.3 Diachrone dimensie

Het idee voor het HCD was een corpus post-middeleeuws Nederlands van de 16^{de} tot en met de 19^{de} eeuw. We namen telkens het midden van de eeuw als referentiepunt: 1550, 1650, 1750 en 1850. Om voldoende teksten te kunnen verzamelen werd een marge van twintig jaar voor en twintig jaar na ieder jaartal ingebouwd. Dat resulteert in vier periodes: 1530-1570, 1630-1670, 1730-1770, 1830-1870. Het HCD strekt zich dus uit over vier eeuwen, een langere periode die ook vereist is indien men pluricentriciteit historisch wil onderzoeken. Bovendien corresponderen de jaartallen 1550, 1650, 1750 en 1850 met belangrijke fases in de geschiedenis van het Nederlands, die deels berusten op de traditionele visie en deels recentere inzichten weerspiegelen. Zo representeert het midden van de 16^{de} eeuw de periode voor de staatkundige scheiding tussen Noord en Zuid, toen Brabant nog beschouwd werd als het centrum van de Lage Landen. Het jaartal 1650 symboliseert het einde van de eerste fase van het standaardisatieproces, en 1750 het mogelijke begin van de aparte Noordelijke en Zuidelijke normatieve traditie. Het midden van de 19^{de} eeuw correspondeert tot slot met de discussie tussen integrationisten en particularisten in België (cf. Hoofdstuk 2, Paragraaf 2; Hoofdstuk 3, Sectie 4.2). Tabel 5 vermeldt het aantal woorden per eeuw.

EEUW	16E	17E	18E	19E
	92.824	128.048	123.111	119.265

Tabel 5: woordenaantallen per eeuw

Indien we de woordenaantallen per eeuw berekenen, zouden we met drie genres en vier regio's idealiter uitkomen op 120.000 woorden per eeuw. Voor de 16^{de} eeuw stellen we een opmerkelijk lager aantal vast. De grootste lacunes in ons corpus zijn dan ook te situeren in de 16^{de} eeuw, zoals in Sectie 2.2 al duidelijk werd. Voor de 17^{de} en 18^{de} eeuw wijken de aantallen dan weer naar boven af. Deze aantallen verkregen we door soms meer dan 10.000 woorden per regio per genre op te nemen. Voor de 19^{de} eeuw lezen we tot slot bijna 120.000 woorden af.

2.4 Regionale dimensie

De teksten in het corpus zijn afkomstig uit twee centrale en twee perifere gebieden in de Nederlanden. De centra worden in dit corpus

vertegenwoordigd door Holland (Noorden) en Brabant (Zuiden). Zeeland (Noorden) en Vlaanderen (Zuiden) representeren vervolgens de perifere gebieden. Figuur 6 geeft deze gebieden weer op de kaart. Het corpus bevat teksten uit grote steden als Amsterdam en Antwerpen, evenals materiaal uit kleinere gemeenten en dorpen. De regio's zijn zo gekozen dat het centrum in het Noorden en in het Zuiden grenst aan de respectievelijke periferie, en de perifere gebieden onderling tevens aanpalende gebieden zijn. Zoals reeds vermeld vereist de studie van pluricentriciteit specifieke regio's om patronen van diffusie te kunnen waarnemen. We kunnen onderzoeken in welke mate nieuwe varianten uitgaan van de twee centra Holland en Brabant. Daarnaast worden we in staat gesteld om Noord-Zuiddynamieken te bestuderen, alsook centrum-periferiedynamieken binnen het Noorden en het Zuiden van het taalgebied. Voor de centrum-periferiedynamieken gaan we ervan uit dat nieuwe vormen over het algemeen sneller hun weg vinden van het centrum naar een aangrenzend gebied. De varianten verspreiden zich dan van het centrum naar de periferie zonder de staatsgrenzen te overschrijden. Een verspreiding van periferie naar periferie is evenzeer mogelijk. In dat geval zijn varianten niet gebonden aan staatsgrenzen, maar maken ze deel uit van een dialectcontinuüm (cf. Hoofdstuk 3). Daarom is het belangrijk dat de periferieën in het HCD aangrenzend zijn. Dergelijke distributie wordt ook een areale distributie genoemd. De combinatie Zeeland-Vlaanderen maakt het dan ook mogelijk om het concept pluriarealiteit te beoordelen voor het Nederlands.



Figuur 6: centrum en periferie in de Noordelijke en Zuidelijke Nederlanden (VL = Vlaanderen, BR = Brabant, ZE = Zeeland, HO = Holland)

De indeling in centra en periferieën kan problematisch zijn. De begrippen 'centrum' en 'periferie' zijn namelijk niet stabiel door de eeuwen heen. Zo was Brabant in de 16^{de} eeuw duidelijk een politiek, economisch en cultureel centrum, maar dit veranderde vanaf de 17^{de} eeuw (cf. Hoofdstuk 2, Sectie 2.1). De periferie in het Noorden wordt dan weer vertegenwoordigd door Zeeland, terwijl Zeeland demografisch en economisch gezien een belangrijke regio was in de 16^{de} en 17^{de} eeuw (cf. Krogull, 2018, p. 65). Geografisch gezien is Zeeland natuurlijk wel een perifeer gebied ten opzichte van Holland, en Vlaanderen ten opzichte van Brabant. Er zijn echter nog argumenten aan te voeren om Holland en Brabant als centra te blijven beschouwen tegenover respectievelijk Zeeland en Vlaanderen. In de 17^{de}

eeuw was Brabant dan wel niet meer het centrum van de Lage Landen, maar Antwerpen in Brabant bleef wel het belangrijkste handelscentrum in het Zuiden (Willemyns, 2003b, p. 138; zie ook Janssens, 2014, pp. 247-248). Bovendien bleef Antwerpen in de 17^{de} eeuw nog even een cultureel centrum dankzij de bloei van de barokschilderkunst (Rubens, Van Dyck) (Willemyns, 2003b, p. 138). In het Noorden behield Holland tijdens de Gouden Eeuw dan weer het machtsoverwicht (Frijhoff & Spies, 2000, p. 158). Met ons corpusonderzoek willen we nagaan in welke mate de centra daadwerkelijk (taalkundige) centra waren. De corpusopbouw rechtvaardigt dus het gebruik van de begrippen ‘centrum’ en ‘periferie’ in dit onderzoek.

REGIO

Brabant	117.694
Holland	117.578
Vlaanderen	121.039
Zeeland	106.937

Tabel 6: woordenaantallen per regio

Tabel 6 toont de woordenaantallen voor de regionale dimensie. Per regio onderzoeken we vier periodes en drie genres, die elk 10.000 woorden voortbrengen. Dit zou resulteren in 120.000 woorden per regio. Afwijkingen van dit aantal verklaren we ook hier door een tekort aan teksten (naar onder) of door een lichte verhoging van de 10.000 woorden (naar boven).

2.5 Algemeen

Wanneer we de drie dimensies in ons corpus combineren, krijgen we het volgende overzicht:

	16E	17E	18E	19E	Totaal
Administratief	34.356	39.811	40.470	40.242	154.879
Brabant	9.936	9.902	9.943	10.945	40.726
Holland	9.710	9.928	10.007	9.817	39.462
Vlaanderen	4.889	9.938	10.170	9.857	34.854
Zeeland	9.821	10.043	10.350	9.623	39.837
Egodocumenten	32.699	39.455	40.279	38.241	150.674
Brabant	9.177	9.446	10.711	7.129	36.463
Holland	10.134	10.031	9.932	10.024	40.121
Vlaanderen	10.232	9.935	9.763	11.140	41.070
Zeeland	3.156	10.043	9.873	9.948	33.020
Pamfletten	25.769	48.782	42.362	40.782	157.695
Brabant	10.145	9.911	9.878	10.571	40.505
Holland	4.946	12.199	10.638	10.212	37.995
Vlaanderen	10.678	14.710	9.871	9.856	45.115
Zeeland	0	11.962	11.975	10.143	34.080
Eindtotaal	92.824	128.048	123.111	119.265	463.248

Tabel 7: woordenaantallen per genre, eeuw en regio

Uit Tabel 7 kunnen we afleiden dat het over het algemeen gelukt is om 10.000 woorden per combinatie van genre, eeuw en regio te verzamelen. Tegelijk worden ook de lacunes nogmaals duidelijk. Het deelcorpus administratieve teksten is het meest compleet. Enkel voor Vlaanderen in de 16^{de} eeuw lezen we 4.889 woorden af. Bij de egodocumenten zien we twee lacunes: voor Zeeland in de 16^{de} eeuw en in mindere mate voor Brabant in de 19^{de} eeuw. In het pamflettencorpus ontbreken pamfletten in de 16^{de} eeuw, met name voor Holland en Zeeland. We kunnen hieruit besluiten dat de meeste lacunes in de vroegste periode van ons corpus te situeren zijn, wat niet zo verwonderlijk is. Vooral in het pamflettencorpus werden soms meer dan 10.000 woorden per regio en per eeuw verzameld, waardoor dit genredeelcorpus bijna 160.000 woorden telt. De twee andere deelcorpora tellen elk meer dan 150.000 woorden. Dit resulteert in een totale corpusomvang van 463.248 woorden, verdeeld over 209 teksten. Het aantal teksten per genre staat vermeld in Tabel 8. Een overzicht van alle teksten in het corpus kan men raadplegen in Bijlage 1 van dit proefschrift.

GENRE	# TEKSTEN
Administratief	58
Egodocumenten	60
Pamfletten	91
Totaal	209

Tabel 8: verdeling aantal teksten per genre

3 Afbakening genres

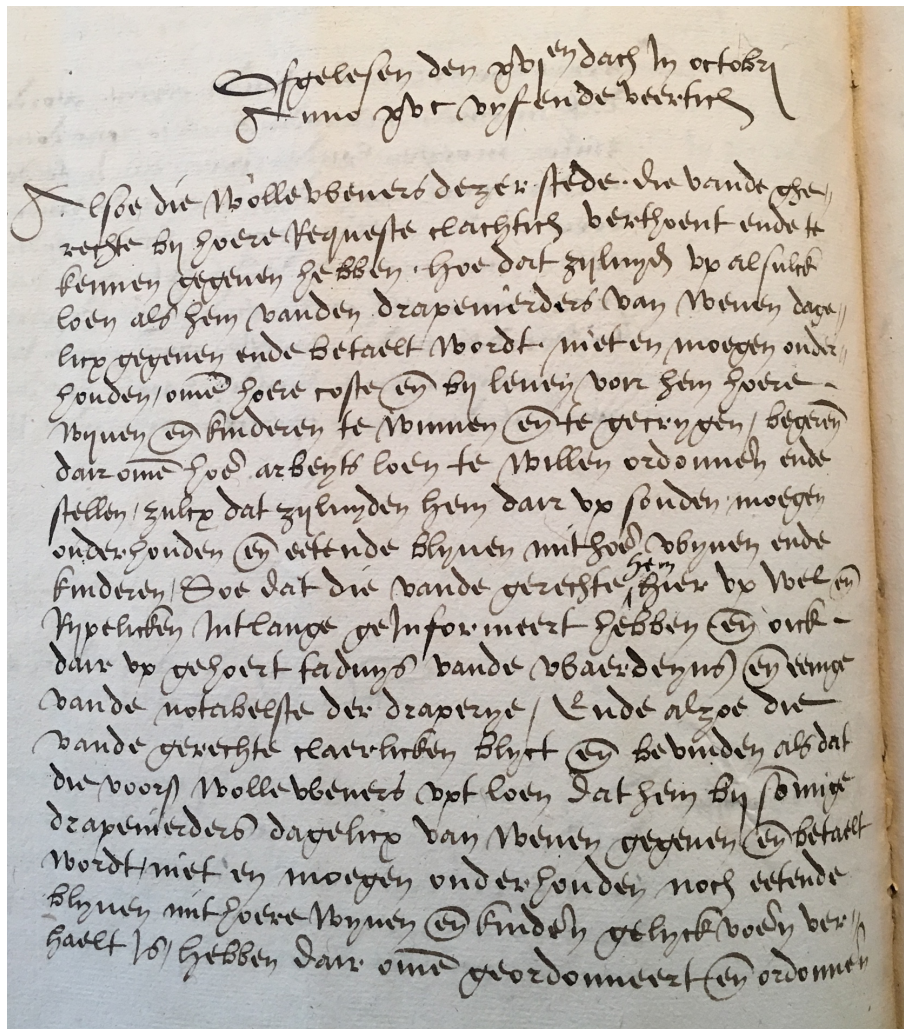
De combinatie van verschillende periodes, regio's en genres vraagt om een bedachtzame selectie van de teksten in het corpus. Zo moet een genre voorkomen van de 16^{de} tot in de 19^{de} eeuw en bovendien vergelijkbaar zijn over de verschillende eeuwen heen. Waar we bij de egodocumenten soms beperkt zijn in de hoeveelheid beschikbaar materiaal, kan een uiteenlopende reeks bronnen onder de noemer 'administratieve teksten' geplaatst worden. Om een evenwichtig corpus te bekomen is het dus noodzakelijk om elk deelcorpus zo consistent mogelijk te houden. In deze paragraaf volgt een afbakening van de drie genres.

3.1 Administratieve teksten

Een veelheid aan tekstvormen kan tot de categorie 'administratieve teksten' gerekend worden. Hierdoor was een heldere afbakening cruciaal. Na verkennend onderzoek werd beslist om de administratieve teksten te centreren rond enerzijds gilden en nijverheid, en anderzijds rond het algemeen bestuur. Onder documenten over het gildewezen vallen o.m. aflezingboeken (zie Figuur 7) en privilegeboeken. Dit zijn omvangrijke boeken waarin de wetgeving voor de gilden en nijverheid binnen de stad genoteerd werd. Voorbeelden uit het corpus zijn *Keur betreffende den verkoop van vleesch* (Van Dillen, Stadsarchief Amsterdam, 1545) en *Verbod van het gerecht om de warp met het wiel te spinnen* (Posthumus, Erfgoed Leiden en Omstreken, 1537). Documenten afkomstig van het bestuur zijn bijvoorbeeld registers van de schepenbank, resolutieboeken en notulen van de gemeenteraad. Deze documenten verschillen niet zozeer van de eerste categorie, aangezien het ook beslissingen van hogerhand betreft. De schepenbank besliste o.m. over de verdeling van goederen bij een overlijden en over de aflossing van schulden. Daarnaast bewaarde de schepenbank ook resolutieboeken (zie Figuur 8). Resolutieboeken bevatten verschillende

resoluties, die erg vergelijkbaar zijn met de latere notulen van de gemeenteraad (zie Figuur 9).

In tegenstelling tot de egodocumenten (cf. Sectie 3.2) hebben we bij de ambtelijke teksten naar alle waarschijnlijkheid steeds te maken met geoefende scribenten. Deze auteurs hanteerden beroepshalve de pen, wat doorgaans een impact heeft op de kwaliteit van het handschrift. De ruime keuze aan teksten liet ons bovendien toe om de 'nette' notulen op te nemen in plaats van minuten van notulen, die soms onoverzichtelijk neergepend waren.



Figuur 7: excerpt uit een aflezingboek (Erfgoed Leiden en Omstreken, Archief der Secretarie van de stad Leiden [SA I] – nr. 388, 1545)

Transcriptie^{15,16}:

Alsoe die wolleweuers dezer stede die vande gherechte
by hoere Requeste clachtich verthoent ende te
kennen gegeuen hebben hoe dat zyluyden vp alsulck
loen als hem vanden drapenierders van weuen dagelicx
gegeuen ende betaelt wordt niet en moegen onderhouden,
om[m]e hoere coste en[de] by leuen voir hem hoere
wyuen en[de] kinderen te winnen en[de] te gecrygen, begeren[de]
dair om[m]e hoe[re] arbeyts loen te willen ordonne[re]n ende
stellen, zulcx dat zyluyden hem dair vp souden moegen
onderhouden en[de] eetende blyuen mit hoe[re] wyuen ende
kinderen, Soe dat die vande gerechte hem hier vp wel en[de]
Rypelicken Jnt lange geInformeert hebben en[de] oick
dair vp gehoert taduys vande waerdeyns en[de] eenige
vande notabelste der draperye, Ende alzo die
vande gerechte claerlicken blyct en[de] bevinden als dat
die voors wolleweuers vpt loen dat hem by so[m]mige
drapenierders dagelicx van weuen gegeuen en[de] betaelt
wordt, niet en moegen onderhouden noch eetende
blyuen mit hoere wyuen en[de] kinde[re]n gelyck voe[re]n verhaelt
Js, hebben dair om[m]e geordonneert en[de] ordonne[re]n

¹⁵ De transcripties worden hier in functie van de leesbaarheid weergegeven zonder tags. Een onzekere lezing werd gecursiveerd in onze voorbeelden. Voor de analyses werden de transcripties wel voorzien van tags (cf. Paragraaf 5, Sectie 5.3).

¹⁶ De transcriptie bij Figuur 7 werd overgenomen uit *Bronnen tot de geschiedenis van de Leidsche textielnijverheid 1333-1795* (Posthumus, 1910-1922), en door de auteur van dit proefschrift aangepast aan de transcriptierichtlijnen die voor dit project gehanteerd worden. Hierbij volgden we de bestaande transcriptie; extra tekst in de marge, titels e.d. werden niet toegevoegd aan de transcriptie.

n^o 100
 Ter vergaderinge van
 den heere Drotsaert ende
 schepenen der vrijheid van
 Walhem gehanden ten huise
 van p^{re}ster de Drost den
 13 october 1745/

Den heere Drotsaert deser heere^{lyck} h^ojd^e
 doende syne propositie, seght dat den bester
 dinge der contributien van over ettelijcke
 weken is gebeurt ende dat daer over tot
 als nu alnogh geenen boeck en is geraemt
 door dien dat de schepenen deser vrijheid
 tot noch toe daer over geenen vasten voet
 en hebben gerecolvert ende mits de raeminge
 ende repartitie geenen uijtstel meer en sijdt
 soo versoect den gemelden heere Drotsaert
 dat als nu in promptis ende sonder voordere
 uijtstel sal worden geprocedert tot de
 gerequiseerde raeminge ende repartitie
 want hier soo protestert den selven van
 alle costen, schade ende intresten die
 daer uijt souden komen te resulteren
 justinerende voor soo vele hem dit aengaet
 te worden gehanden voor diligent.

De schepenen deser vrijheid van walhem
 gesien hebbende de bovenslaende propositie

Figuur 8: begin van een resolutie van de schepenbank (Stadsarchief Mechelen,
 Schepenbank van de Vrijheid Walem – nr. 100, 1745)

Transcriptie:

Ter vergaederinge van
den heere Drossaert ende
schepenen der vrijhijdt van
Walhem gehauden ten huijse
van peeter de Prost desen
13 october 1745

Den heere Drossaert, deser heerelijckhijdt
doende sijne *propositie*, seght dat de aenbestedinge
der contributien van over ettelijcke
weken is gebeurt ende dat daer over tot
als nu alnogh geen en is geraemt
door dien dat de schepenen deser vrijhijdt
tot nogh toe daer over geen en vasten voet
en hebben geresolveert ende mits de raeming
ende repartitie geen uijtstel meer en lijdt
soo versoeckt den gemelden heere Drossaert
dat als nu in promptis ende sonder voorderen
uijtstel sal worden geprocedeert tot de
gerequireerde raeming ende repartitie
faute dier soo protesteert den selven van
alle costen, schaede ende intresten die
daer uijt sauden komen te resulteren
sustinerende voor soo vele hem dit aengaet
te worden gehauden voor diligent.

De schepenen deser vrijhijdt van Walhem
gesien hebbende de bovenstaende propositie

Openbare Vergadering
 Woensdag den 1^{en} October
 1851
 — beginnend als de Raadsleden
 bijzondert het Raads-
 A. volvers. —

na gedaan gebed zijn de notulen
 van den 17 September te voorgesprek
 en goedgekeurd.

Door den Heer voorzitter worden over
 geleide en voorgesprek de volgende provin-
 ciale bladen.

N^o 128 Besluit van Heeren Gedeputeerde
 Staten van 22 September 1851 N^o 40, houdende
 voorschrijven hoe in het vervolg met
 aangewezene worden bij van den gemeente
 Raad uitgaande stukken, of dergelike in
 een openbare dan wel in een besloten
 vergadering zijn behandeld.
 Welk stuk is aangenomen voor infor-
 matie om daarop te attenderen. —

N^o 129 Besluit van als voer van 25^{en}
 September 1851 Aⁿ 6016 2afde opgave
 van het bedrag der personele belas-
 tingen door de Contrangers der Directie
 belastingen. —
 Welk stuk almeê is aangenomen
 voor kennisgeving om daarop te ligner-
 len te attenderen. —

N^o 130 Besluit als boven de dato 29^{en}
 September 1851 N^o 1 eerste bijeenkomst
 van eenige nieuwe Gemeente Raadsleden
 waan omtrent 15 besloten: het voor-
 — Schreef

Provinciale
 bladen. —

Figuur 9: Notulen van een vergadering van de gemeenteraad van Veere (Zeeuws Archief,
 Archief van de Gemeente Veere – nr. NGR9, 1851)

Transcriptie:

Openbare Vergadering
Woensdag den 1o October
1851
tegenwoordig alle de Raadsleden
uitgezonderd het Raadslid
A. Volkers. -

na gedaan gebed zijn de notulen
van den 17 September II voorgelezen
en goedgekeurd.
Provinciale bladen. -
Door den Heer voorzitter worden overgelegd
en voorgelezen de volgende provinciale
bladen te weten:
no 128 Besluit van Heeren Gedeputeerde
staten van 22 september JL no 40, houdende
voorschriften hoe in het vervolg moet
aangewezen worden bij van den gemeente
Raad uitgaande stukken of dezelve in
eene openbare dan wel in een besloten
vergadering zijn behandeld
welk stuk is aangenomen voor informatie
om daarop te attenderen. -
no 129 Besluit van als voren van 25e
september 1851 A no 6016 *zafda* opgave
van het bedrag der personele belastingen
door de ontvangers der Directe
belastingen. -
welk stuk almede is aangenomen
voor kennisgeving om daarop te zyner
tyd te attenderen. -
no 130 Besluit als boven de dato 29e
september 1851 no i eerste bijeenkomst
van eenige nieuwe gemeente Raadsleden
waaromtrent is besloten: het voorschreve

3.2 Egodocumenten

De term ‘egodocumenten’ ontleen we aan de historicus Jacques Presser. Hij bedacht de term rond het jaar 1950 als verzamelnaam voor o.m. dagboeken, memoires, private brieven en autobiografieën (Dekker, 2002a, p. 7; Dekker, 2002b, pp. 13-14). Presser zelf omschreef egodocumenten als teksten waarin de auteur als ik-figuur voortdurend aanwezig is (Presser, 1958; geciteerd in Dekker, 2002b, p. 14). Dekker (2002b, p. 14) spreekt kortweg over “[...] a text in which an author writes about his or her own acts, thoughts and feelings”.

Egodocumenten hebben gedurende de laatste decennia steeds meer terrein gewonnen in de historische sociolinguïstiek. Ze vervullen dan ook een bijzondere rol binnen het onderzoeksgebied. Niet in het minst omdat ze beschouwd worden als teksten die het alledaags gesproken taalgebruik zo dicht mogelijk benaderen (Elspaß, 2007, p. 5; Elspaß, 2012b, p. 156). Daarnaast kunnen ze de traditionele taalgeschiedenis, hoofdzakelijk gebaseerd op literaire en formele teksten, aanvullen (Elspaß, 2012b, p. 156; Van der Wal, 2006, p. 7). Van der Wal en Rutten (2013, p. 12) merken terecht op dat hoewel egodocumenten kenmerken van de gesproken taal bevatten, het wel geschreven communicatie blijft. Het mag dus wel duidelijk zijn dat het taalgebruik in egodocumenten geen rechtstreekse reflectie van de gesproken taal is. Egodocumenten vertonen wel kenmerken van de gesproken taal, maar bevatten eveneens kenmerken typisch voor het geschreven taalgebruik (zie ook Rutten & Van der Wal, 2011). Het onderzoek naar die typisch geschreven kenmerken is eveneens waardevol.

De spanning tussen het geschreven en gesproken taalgebruik wordt o.m. aangehaald door Koch en Oesterreicher (1985). Zij stellen een conceptueel continuüm van gesproken en geschreven taal voor. De pool ‘gesproken’ komt overeen met *Sprache der Nähe*, de pool ‘geschreven’ met *Sprache der Distanz*. Deze begrippen worden door Elspaß (2012b, p. 157) als *language of immediacy* en *language of distance* gekenmerkt. Afhankelijk van een aantal parameters kan een gesproken of geschreven uiting op het continuüm geplaatst worden. Het gaat dan om parameters als formaliteit, familiariteit tussen de gesprekspartners en spontaniteit. Zo gaat *immediacy* gepaard met informeel en spontaan taalgebruik, terwijl *distance* formeel en gepland taalgebruik impliceert (Elspaß, 2012b, p. 157; Koch en Oesterreicher, 1985). Private brieven worden vanwege de interactie tussen de gesprekspartners

(dialoog) in de richting van de *language of immediacy* geplaatst, nog voor dagboeken en reisverslagen (monoloog) (Elspaß, 2012b, p. 162; Van der Wal & Rutten, 2013, p. 2). In het HCD werd evenwel geen privécorrespondentie opgenomen. Aan het begin van de corpusopbouw werd al duidelijk dat het met name voor de 16^{de} eeuw een uitdaging zou worden om voldoende brieven voor alle regio's te verzamelen. Simons (2013, p. 24) merkt op dat egodocumenten, en dan vooral brieven, niet altijd "op grote schaal" beschikbaar geweest zijn.¹⁷ De keuze viel dus op dagboeken (Figuur 10) en reisverslagen (Figuur 11). Wanneer het gaat over egodocumenten worden beide types over het algemeen na elkaar genoemd. Krogull (2018, pp. 55-57) bespreekt het onderscheid tussen beide, maar stelt vast dat de lijn tussen dagboek en reisverslag soms dun is. Hij maakt om die reden geen onderscheid tussen dagboeken en reisverslagen in het *Going Dutch Corpus*. Reisverslagen zijn in wezen "yet another type of diaries, written in travel settings" (Krogull, 2018, p. 56). In het HCD worden ze eveneens als één genre beschouwd. Met de keuze voor dagboeken en reisverslagen zouden we in principe egodocumenten voor elk van de onderzochte eeuwen kunnen vinden.

Elspaß (2012b, p. 159) argumenteert dat er vanaf de vroegmoderne tijd ook dagboeken geproduceerd werden door leden van de lagere sociale klassen. We probeerden dan ook in de mate van het mogelijke verschillende sociale klassen samen te brengen in het egodocumentencorpus. Overeenkomstig met het *Brieven als Buit corpus* (Rutten & Van der Wal, 2014, pp. 9-10) maakten we een sociale stratificatie op grond van het beroep van de scribenten. Bij vrouwelijke scribenten keken we naar het beroep van de echtgenoot, indien er geen beroep bekend was. We baseerden ons op Frijhoff en Spies (2000) om de verschillende beroepen in sociale groepen in te delen. Hoewel deze indeling zich toespitst op de Noordelijke Nederlanden in de 17^{de} eeuw, achten wij ze geschikt om een algemeen beeld van de sociale groepen in ons egodocumentencorpus te schetsen. Frijhoff en Spies (2000, pp. 189-190) onderscheidden zes groepen:

¹⁷ De waarde van het *Brieven als Buit corpus* (Nobels, 2013; Rutten & Van der Wal, 2014; Simons, 2013) wordt hierdoor des te meer benadrukt.

SOCIALE GROEPEN¹⁸

1. Adel en patriciaat
2. Rijke kooplieden, academici, hoge ambtenaren
3. Leraren, dorpspredikanten, notarissen, klerken
4. Lagere ambtenaren, schoolmeesters, kleine schippers
5. Zeelieden, soldaten, knechts, handarbeiders
6. Bezitslozen, zwervers, bedelaars, invaliden

Tabel 9: indeling in sociale groepen (gebaseerd op Frijhoff & Spies, 2000, pp. 189-190)

Ze duiden groep 2 en 3 aan met 'brede burgerij' (Frijhoff & Spies, p. 190). Voor onze stratificatie namen we deze aanduiding over, en beschouwden we 2 en 3 ook als één sociale groep. Ons deelcorpus egodocumenten bevat 60 teksten, geschreven door 59 verschillende auteurs.¹⁹ Van 6 scribenten bleef het beroep onbekend. De overige scribenten konden als volgt ingedeeld worden:

GROEP	n =
1	6
2+3	43
4	2
5	2
6	0

Tabel 10: sociale stratificatie in het egodocumentencorpus

Uit Tabel 10 leiden we af dat de verdeling over sociale groepen in ons corpus allesbehalve gelijk is. Het leeuwendeel van de scribenten staat relatief hoog gerangschikt op de sociale ladder. Slechts 4 scribenten bekleden een lagere positie.²⁰ Het bleek voor dit onderzoek niet mogelijk om verschillende sociale klassen te integreren in het corpus. De sociale klasse van de scribenten is echter geen variabele in het onderzoek en zal in het vervolg van dit proefschrift geen rol meer spelen. Voor nu volstaat het om te concluderen dat we de geschriften van een redelijk homogene groep scribenten kunnen analyseren. Net als de sociale klasse maakt ook gender geen deel uit van het

¹⁸ De voorbeelden van beroepen die in deze tabel gegeven worden zijn niet exhaustief.

¹⁹ Balthasar Moretus II is verantwoordelijk voor twee teksten in het subcorpus.

²⁰ Van 6 auteurs is de functie niet gekend. Dit aantal kan dus eventueel iets hoger liggen.

onderzoek. Met 51 mannen en 8 vrouwen in het corpus is het niet aangewezen om verschillen tussen mannelijke en vrouwelijke scribenten te onderzoeken.

Bij de selectie van de teksten hanteerden we enkele criteria. In eerste instantie opteerden we voor teksten die niet geschreven werden met een groot publiek voor ogen. Bij voorkeur schreven de auteurs van egodocumenten voor zichzelf of voor de naaste familie. Het persoonlijke karakter van de teksten was evenzeer een voorwaarde. Naast de genoemde dagboeken en reisverslagen, vallen hieronder dan ook zogeheten familiekronieken. Deze familiekronieken bevatten (summiere) aantekeningen bij gebeurtenissen binnen de familie, zoals huwelijken, geboortes en overlijdens. Soms maken ze deel uit van een omvangrijker handboek of manuaal met inkomsten en uitgaven. Aan het begin of einde van zo'n handboek treffen we dan enkele pagina's met persoonlijke opmerkingen aan. Daarnaast kregen ook gewone kronieken een plaats binnen het deelcorpus egodocumenten. Omdat er voor het Zuiden onvoldoende teksten aan de oppervlakte kwamen door enkel naar dagboeken en reisverslagen te zoeken, werd beslist om de zoekopdracht uit te breiden met 'kroniek'. Het type kroniek dat we opnamen in dit deelcorpus is vergelijkbaar met de eerdergenoemde familiekronieken. Alleen worden in de gewone kronieken meestal geen familiale gebeurtenissen beschreven, maar wel lokale gebeurtenissen die de auteur als ooggetuige waarnam. Een goed voorbeeld is de 19^{de}-eeuwse Mechelse kroniek Jonghmans, waarin de plaatselijke kapelaan Jan Jonghmans o.m. beschrijft wat er allemaal leefde op politiek vlak. Deze auteur bericht eveneens over zijn familie (Stadsarchief Mechelen, Kronieken en jaarboeken – Serie xxviii nr. 2, 1824-1838). Dergelijke kronieken hebben dan veel weg van een dagboek. Kronieken met een louter geschiedkundig karakter, die meer neigen naar de *language of distance* (Elspaß, 2012b, p. 157; Koch & Oesterreicher, 1985), werden niet opgenomen in dit deelcorpus. Tot slot was het gegeven van ooggetuige bij de selectie van kronieken heel belangrijk. Indien de auteur de – soms historische – gebeurtenissen niet zelf had beleefd, kwam de tekst niet in aanmerking.

1631 & 1632: 1631 21 maart syn wi verhuyst wt londen
 naar Bachnaai. naar dat wi 5 iare gewoont
 hadden te Londen.
 1632 den 26 maart dat is de 5 april merwon
 tijt op een maendag des avonds te 10 ure
 is geboren onse dochter Hosina. in dit ons
 woonhuys te Bachnaai bynton Londen.
 Ende is gedooft van cosyn ambrosius in de
 duytse kerche te Londen daer onse andere
 kinderen gedooft syn. dese 8 ste april op
 een sondaeg. des doop acht dagen wtgestelt
 synde om dat myne woerde may siet was
 de gesünge waren cosyn abram bot
 en moynen susamond vadermachers.
 moynen. absent synde heeft loffrou carpen
 hier des ter tijt in engeland synde haer
 plaotse bewaert. de gesünge hebben haer
 tot gedachtenisse gegere dit volgende
 een groote vergulden cop heeft gegere
 cosyn bot. moynen susamond heeft haer
 gegere een vergulde schale daer in staet
 : als d'arch begiolt acht zielen van
 swatere vloot.
 : soo gedoncht Hosina dijn doopte in griffi
 bloot:

des cop die
 woest onse 21
 de schale. 19

Figuur 10: bladzijde uit het dagboek van Maria de Fraeije (Zeeuws Archief, Verzameling
 Aanwinsten Zeeuws Archief – nr. 181, 1631-1632)

Transcriptie:

1631 en 1632: 1631 21 meert zyn wij verhuyst wt Londen
naer hacknaij. naer dat wij 5 iaren gewoont
hadden te Londen.

1632 den 26 meert dat is den 5 april nieuwen
stijl op eenen maendach des avonts ten 10 uren
is geboren onse dochter Josijna. in dit ons
woonhuijs te hacknaij buijten londen.

ende is gedoopt van cosyn ambrosius in de
duytse kercke te londen daer onse andere
kinderen gedoopt zijn. desen 8 sten april op
eenen sondach. des doop acht dagen wtgesteld
synde om dat mijnen weerden man sieck was
de getuijgen waren cosyn abram bec
en moijken susanneken radermachers.
moijken absent synde heeft Joffrou carpentier
doen ter tyt in engelant synde haer
plaetse bewaert. de getuijgen hebben haer
tot gedachtenisse gegeven dit volgende
eenen grooten vergulden cop heeft gegeven
cosyn bec. moyken susanneken heeft haer
gegeven een vergulde schale. daer in staet
: als d'arck behielt acht zielen van
swaters vloet.

: soo gedenckt Josina dijns doopts in christi
bloet:

den cop die weecht oncen. 21 de schale. 19

Wat aangaat de Kleeding der Bulwers,
Picadors, Banderilleros en Matadors, dit
kan ik u niet beschrijven, het opera van Savys
alleen kan u daar een gedacht van geven.

Dinsdag 14 Mei 1850

Sedert ik in deze hoofdstad aangekomen ben,
heb ik niet opgehouden te zoeken, naar een
huis waar ik my in den kost wil stellen, want
anders indien ik in het hotel bleef, kon ik
met een pensioen van 4000 franken niet
toekomen, eenen inwooner van Madrid aan
wie ik aanbevolen ben, door zynen broeder
van Barcelona, helpt my in myne opkoe-
ringen, en by gelooft een huis ondekt te
hebben, waar ik voor 16 a 20 realen dagt
(4 a 5 frank) zal eene seboone kamer hebben

Figuur 11: bladzijde uit het 19de-eeuwse reisverslag van François-Auguste Gevaert (Universiteitsbibliotheek Gent, Manuscripten, 1850)

Transcriptie:

Wat aangaat de kleeding der chulors,
Picadors, Banderillors en Matadors, dit
kan ik u niet beschryven, het opera van Parys
alleen kan u daar een gedacht van geven.

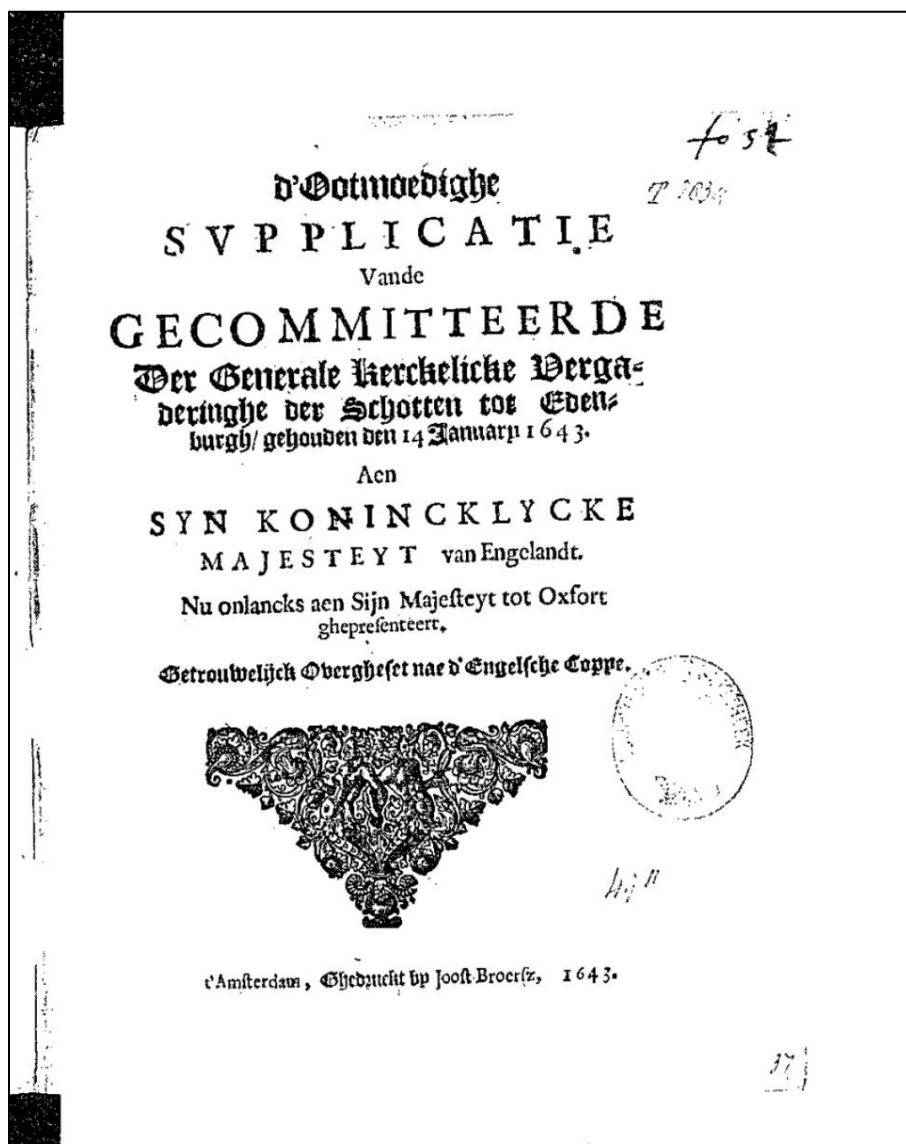
Dynsdag 14 Mei 11 uur 'S morg

Sedert ik in deze hoofdstad aangekomen ben,
heb ik niet opgehouden te zoeken, naar een
huis waar ik my in den kost wil stellen, want
anders indien ik in het hotel bleef, zou ik
met een pensioen van 4000 franken niet
toekomen, eenen inwoner van Madrid aan
wie ik aanbevolen ben, door zynen broeder
van Barcelone, helpt my in myne opzoekingen,
en hy gelooft een huis ondeckt te
hebben, waar ik voor 16 a 20 realen dags
(4 a 5 franks) zal eene schoone kamer hebben

3.3 Pamfletten

Het deelcorpus pamfletten bevat als enige component in het HCD gedrukte teksten. Hoewel de term 'pamflet' nog niet echt gebruikt werd om de oudste drukwerken over politieke en religieuze aangelegenheden aan te duiden (Harline, 1987, p. 2), kunnen we dit type drukwerk aantreffen in elk van de onderzochte tijdsvensters. De pamfletten in ons corpus komen voor een groot deel voort uit de collectie van Knuttel, aangevuld met pamfletten uit diverse Vlaamse archieven (cf. Paragraaf 4). Knuttel, van opleiding theoloog, kreeg als tijdelijk ambtenaar de opdracht om de pamfletten in de Koninklijke Bibliotheek van Den Haag (KB) te catalogiseren. Dat leverde een zeer uitgebreide catalogus op met titels tot 1853. Wel dient hierbij te worden opgemerkt dat niet alle pamfletten in de KB beschreven zijn in Knuttel. Met name de pamfletten die na zijn werk in de bibliotheek terechtgekomen zijn, ontbreken (Van der Hoeven, 1978, pp. xix-xx). Onze catalogisering bevat o.m. tijdingen, ordonnanties en supplicaties, evenals samenspraken en plakkatens. Het type pamflet wordt dan al duidelijk in de titel (zie Figuur 12 voor een voorbeeld van een supplicatie). Dit geldt vooral voor de 16^{de}-, 17^{de}- en 18^{de}-

eeuwse pamfletten. Zowel in de eerste drie eeuwen als in de 19^{de} eeuw, zijn er ook pamfletten zonder karakterisering als tijding, samenspraak e.d. in de titel. Voorbeelden zijn *De invloed van de godsdienst op het openbaar en staatkundig leven* (KB, KW Pfl 29058, 1849) of *Eenige weldenkende vlaemsche Burgers aen hunne zoogezegde onafhankelyke of conservatieve broeders* (FelixArchief, Gelegenhheidsdocumenten – nr. 203#312, 1860).



Figuur 12: 17^{de}-eeuws pamflet uit de Knuttel-collectie (KB, KW Pfl 4911, 1643)

Transcriptie:

d'Ootmoedighe
SVPPPLICATIE

Vande

GECOMMITTEERDE

Der Generale Kerckelicke Vergaderinghe
der Schotten tot Edenburgh,
gehouden den 14 January 1643.

Aen

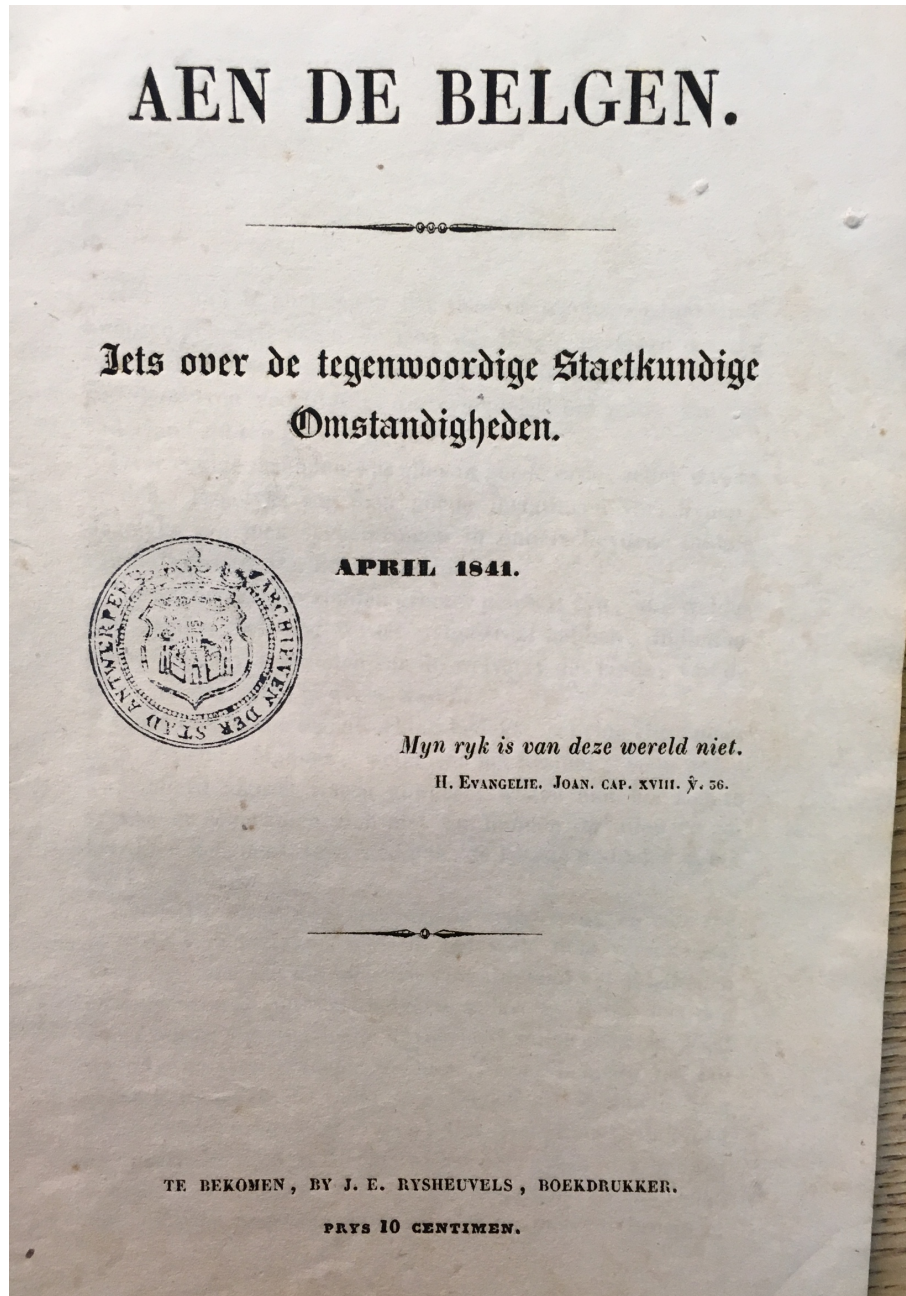
SYN KONINCKLYCKE

MAJESTEYT van Engelandt.

Nu onlancks aen Sijn Majesteyt tot Oxfort
ghepresenteert.

Getrouwelijck Overgheset nae d'Engelsche Coppe.

t'Amsterdam, Ghedruckt by Joost Broersz, 1643.



Figuur 13: 19^{de}-eeuws pamflet (FelixArchief, Gelegenheidsdocumenten – nr. 203#4156, 1841)

Transcriptie:

AEN DE BELGEN.

Iets over de tegenwoordige Staetkundige
Omstandigheden.

April 1841.

Myn ryk is van deze wereld niet.
H. EVANGELIE. JOAN. CAP. XVIII. V. 36.

TE BEKOMEN, BY J.E. RYSHEUVELS, BOEKDRUKKER.
PRYS 10 CENTIMEN.

Pamfletten eenduidig definiëren blijkt een moeilijk punt (Harms, 2011, p. 14; Van der Hoeven, 1978, p. ix). Het *Algemeen letterkundig lexicon* vermeldt dat het begrip ‘pamflet’ als dusdanig pas gangbaar werd aan het einde van de 18^{de} eeuw en het begin van de 19^{de} eeuw. In de eeuwen die hieraan voorafgingen komen andere aanduidingen voor. ‘Vlugschrift’, ‘manifest’, ‘plakkaat’, ‘ordonnantie’, ‘octrooi’ en ‘blauwboekje’ zijn slechts enkele voorbeelden (“Pamflet”, 2012). Van der Hoeven (1978, pp. xxi-xxii) oppert dat er een verschil is tussen overheidspublicaties (bv. plakكات en ordonnanties) en de mening van particuliere personen. Hij geeft aan dat dit onderscheid niet steeds gemaakt is in pamflettencatalogi en dat ook Knuttel enkele plakكات opnam, hoewel hij dat in eerste instantie niet wilde doen. Harms (2011, p. 14) stelt eveneens dat de pamflettenverzamelingen in de 19^{de} en 20^{ste} eeuw niet samengesteld werden op basis van scherpe selectiecriteria. Het valt dus niet uit te sluiten dat ook ons deelcorpus pamfletten in zekere mate heterogeen is. Vrieler (2007, p. 18) definieert 17^{de}-eeuwse pamfletten op de volgende manier:

[E]en pamflet is een gedrukt, niet-periodiek populair medium dat zichzelf met verschillende benamingen kan afficheren, gebonden is aan de actualiteit, dikwijls anoniem verschijnt, allerlei vormen van fictie en non-fictie kan bevatten en inhoudelijk de lezer probeert te overtuigen van een bepaald standpunt.

Hij gaat dus niet uit van de benamingen, maar wel van de functies van de geschriften (Vrieler, 2007, pp. 17-18). Met deze definitie kunnen we de discussie over wat een pamflet is en wat niet enigszins omzeilen. Het staat vast dat de pamfletten verschillende vormen kunnen aannemen, maar ze hebben steeds een vergelijkbare functie. Voor de pamfletten in ons corpus hanteren we aldus de definitie van Vrieler (2007). De heterogeniteit van het pamflettencorpus impliceert ook dat bepaalde pamfletten formeler zijn dan andere. Zo troffen we in plakkatens en ordonnanties formeel taalgebruik aan, terwijl in samenspraken ook informeler taalgebruik kon voorkomen. Dat komt met name omdat in een samenspraak personages opgevoerd worden, die in dialoog gaan met elkaar. In deze fictieve dialogen zitten vaak niet-standaardtalige vormen vervat die geassocieerd worden met de spreektaal (cf. Rutten, 2008a voor de constructie van regiolecten en sociolecten in literaire werken).

3.4 Hypotheses genres

Nu de drie genres afgebakend zijn kunnen we met het oog op de latere corpusanalyses al enkele algemene voorspellingen over de genres formuleren. We hebben eerst en vooral te maken met meer formele teksten (administratieve teksten) en minder formele teksten (egodocumenten). De formelere teksten werden bovendien neergepend door geoefende scribenten, die de taal voor professionele doeleinden inzetten. De auteurs van de minder formele egodocumenten schreven daarentegen eerder incidenteel. Daarnaast bevat het HCD een gedrukt genre, namelijk de pamfletten. Deze teksten werden dus geredigeerd voor ze gedrukt werden. Op grond van de parameters formaliteit en spontaniteit kunnen we de egodocumenten dus vóór de administratieve teksten en de pamfletten plaatsen op het continuüm van gesproken en geschreven taal. Egodocumenten bevatten zoals gezegd kenmerken van de gesproken taal en de geschreven taal, maar ze staan wel dicht bij de *Sprache der Nähe* dan de andere twee genres. De administratieve teksten en de pamfletten bevatten meer kenmerken van de *Sprache der Distanz* (Elspaß, 2012b; Koch & Oesterreicher, 1985).

We argumenteren dat het taalgebruik in de gedrukte pamfletten het meest progressief zal zijn. Concreet betekent dit dat nieuwe varianten – die later de standaardvarianten zullen worden – sneller overgenomen zullen worden in

de pamfletten. Drukkers leverden immers een belangrijke bijdrage aan de standaardisatie van de taal. Met het oog op een zo groot mogelijke afzetmarkt streefden de drukkers naar een soort algemeen Nederlands. Taalvormen die te regionaal waren werden dan ook vervangen door andere taalvormen (Willemys, 2003b, pp. 110-111). De egodocumenten zijn daarentegen handgeschreven teksten die eerder in de privésfeer tot stand kwamen. Het is dan ook aannemelijk dat deze scribenten voorgeschreven standaardvarianten minder snel zullen overnemen dan de drukkers. Uit eerder historisch-sociolinguïstisch onderzoek weten we bovendien dat egodocumenten meer variatie vertonen dan gedrukte teksten uit dezelfde periode (o.a. Krogull, 2018; Nobels, 2013; Rutten & Van der Wal, 2014; Simons, 2013). Tot slot zullen we onderzoeken of het taalgebruik in de administratieve teksten eerder progressief is, of eerder conservatief. Enerzijds waren de scribenten van administratieve teksten beroepshalve met taal bezig. Men kan dan verwachten dat ze gevoelig waren voor taalnormen, en deze normen ook navolgden. Het onderzoek van Vosters (2011) ondersteunt deze aanname. Hij onderzocht administratieve en gerechtelijke teksten ten tijde van het Verenigd Koninkrijk der Nederlanden. Vosters kon aantonen dat de scribenten uit het Zuiden de opgelegde Noordelijke varianten vlot overnamen in hun geschriften. Dit gebeurde bovendien in een relatief korte periode (Vosters, 2011). Anderzijds zijn ambtelijke teksten ook statisch (Willemys, 2003b, p. 115), waardoor het taalgebruik misschien eerder conservatief zal zijn. Bovendien gaan innovaties niet per se uit van typisch geschreven genres (Nevalainen & Raumolin-Brunberg, 2017, p. 2).

4 Selectie en verzameling

Voor de inventarisatie van de bronnen werd over het algemeen dezelfde werkwijze gehanteerd. Een grondig verkennend onderzoek ging vooraf aan de selectie en raadpleging van authentieke teksten. De eerste stap bestond in het opsporen van de bronnen via online inventarissen/catalogi van archieven en bibliotheken, of papieren inventarissen ter plaatse. Concreet moest een tekst passen binnen het genre, het tijdsvenster en de afgebakende regio, zoals beschreven in Sectie 2.1. Eens de bronnen gelokaliseerd waren, bezochten we de archieven en bibliotheken in kwestie om de teksten te fotograferen. Op basis van het fotomateriaal werden transcripties gemaakt, die vervolgens gecorrigeerd werden. De finale transcripties namen we

uiteindelijk op in het HCD. In het vervolg van deze paragraaf wordt de selectie en verzameling voor elk genre afzonderlijk toegelicht. In Paragraaf 5 besteden we meer aandacht aan de digitalisering van de teksten.

Administratieve teksten

In het verleden werden reeds ambtelijke teksten ontsloten voor taalkundig onderzoek, zoals het *Digitaal Compilatiecorpus Historisch Nederlands* van Coussé (2010). Het *Compilatiecorpus* bevat een component ambtelijke teksten uit de periode 1250-1799. Dit corpus overlapt dus voor een groot deel met onze vooropgestelde tijdsspanne. Om verschillende redenen werd echter beslist om niet uit te gaan van deze bestaande transcripties. In de eerste plaats vertrekt Coussé van papieren tekstedities voor de teksten vanaf de 15^{de} eeuw (Coussé, 2010, p. 130). Ze geeft zelf aan dat er in de tekstedities weinig informatie gegeven wordt over hoe de transcripties precies tot stand gekomen zijn (ibid.). Dit is op zich geen probleem indien de transcripties grondig vergeleken worden met het handschrift. Die controle is voor de teksten in het *Compilatiecorpus* niet gebeurd.²¹ Bovendien bevat het corpus nog behoorlijk veel lacunes voor Brabant vanaf de 15^{de} eeuw en Holland vanaf de 17^{de} eeuw (Coussé, 2010, pp. 131-133). De keuze om ons deelcorpus administratieve teksten volledig zelfstandig op te bouwen komt hieruit voort. Bestaande transcripties uit tekstedities konden deel uitmaken van het corpus, op voorwaarde dat een vergelijking met het originele handschrift mogelijk was.

De samenstelling van de administratieve component gebeurde als een geheel aan de Vrije Universiteit Brussel. Zoals eerder vermeld, selecteerde ik teksten over het gildewezen of afkomstig van het algemeen bestuur. Nadat de teksten gelokaliseerd waren, werden de archieven bezocht en de handschriften gefotografeerd. Dankzij de gevorderde digitalisering van het Stadsarchief Amsterdam en het project *Itinera Nova* (Stadsarchief Leuven), waren er van de documenten uit Amsterdam en Leuven al veel scans online beschikbaar. Ook de website archieven.nl biedt veel gedigitaliseerde archiefstukken aan. De ambtelijke teksten voor Holland zijn afkomstig uit het

²¹ Van sommige tekstedities kunnen we de handschriften overigens niet meer traceren omdat ze vernietigd werden, bv. *Coutume de la salle et châtelainie d'Ypres* (L. Gilliodts-Van Severen, 1911) (correspondentie met Rik Opsommer, archivaris van het Stadsarchief Ieper).

Noord-Hollands Archief, het Stadsarchief Amsterdam en Erfgoed Leiden en Omstreken. Voor Zeeland trof ik de nodige documenten aan in het Gemeentearchief Goes, het Zeeuws Archief en het Utrechts Archief. Uit het Stadsarchief Mechelen, het Gemeentearchief Londerzeel, het Stadsarchief Leuven en het Rijksarchief Leuven haalde ik de zakelijke teksten voor Brabant. Ten slotte konden de handschriften voor Vlaanderen in het Stadsarchief Brugge, het Rijksarchief Gent en het Rijksarchief Kortrijk gevonden worden.

Egodocumenten

Het deelcorpus Noordelijke egodocumenten werd samengesteld aan de Universiteit Leiden onder leiding van Marijke van der Wal. Zo gebeurde de selectie van de teksten door laatstgenoemde. Student-assistente Emmy Stevens bezocht de archieven, fotografeerde de documenten en transcribeerde ze vervolgens aan de Leidse Universiteit. Het grootste deel van de correctiefase gebeurde ook in Leiden. Dit deelcorpus werd als een geheel aangeleverd voor het huidige onderzoek.

Een uitgebreide beschrijving van Noordelijke egodocumenten was reeds beschikbaar via de website egodocument.net, onder leiding van Arianne Baggerman en Rudolf Dekker. De inventarisaties van egodocumenten die men via de website kan raadplegen, bevatten handgeschreven egodocumenten van de 16^{de} eeuw tot 1918, waardoor ze als uitgangspunt konden gelden voor het eigen onderzoek (Baggerman, Dekker, Grabowsky & Schulte Nordholt, z.j.; Lindeman, Scherf & Dekker, 1993; 1994). Een aanzienlijk deel van de Noordelijke egodocumenten werd op basis van deze ordening samengesteld. De hiaten die na een grondige exploratie van de inventarisaties nog overbleven, konden haast overal opgevuld worden met bijkomend (archief)onderzoek. De Noordelijke egodocumenten zijn afkomstig uit de volgende archieven en bibliotheken: het Gemeentearchief Delft, de Universiteitsbibliotheken van Nijmegen, Amsterdam en Leiden, de Koninklijke Bibliotheek, het Archief van de Nederlandse Provincie der Dominikanen, het Nationaal Archief, het Noord-Hollands Archief, het Zeeuws Archief, het Koninklijk Zeeuwsch Genootschap der Wetenschappen, de bibliotheek van de Hoge Raad van Adel, het Stadsarchief Rotterdam, Erfgoed Leiden en Omstreken, het Stadsarchief Amsterdam, het Centrum voor familiegeschiedenis, het Regionaal Archief Alkmaar, het Nederlands

Scheepvaartmuseum en het Historisch Documentatiecentrum voor het Nederlands Protestantisme.

Voor de Zuidelijke egodocumenten was nog geen inventaris beschikbaar. Dit luik van het corpus werd samengesteld naar het voorbeeld van de Noordelijke egodocumenten. Hieraan ging verkennend onderzoek via online catalogi en inventarissen vooraf, waarop ik ter plaatse ging in de archieven en bibliotheken om foto's te maken van de handschriften. Voor Brabant werden de bronnen verzameld in de Koninklijke Bibliotheek van België, het Museum Plantin-Moretus, het Stadsarchief Mechelen en de Universiteitsbibliotheek Gent. De nodige egodocumenten voor Vlaanderen werden aangetroffen in de Universiteitsbibliotheek Gent, het Archief Gent, de Koninklijke Bibliotheek van België, de Rijksarchieven van Gent, Kortrijk en Brugge en de Koninklijke Oudheidkundige Kring van het Land van Waas.

Hoewel de egodocumenten gedefinieerd voor deze studie zelfs in de vroegste periode voorkomen, kwam deze verzameling – in tegenstelling tot de administratieve teksten – stroever tot stand. Vooral voor de 16^{de} eeuw bleek het aanbod geschikte egodocumenten, op Holland na, eerder schaars. Bovendien worden egodocumenten niet altijd even transparant gedocumenteerd, wat een inventarisatie als die van Lindeman et al. (1993; 1994) en Baggerman et al. (z.j.) des te waardevoller maakt. Naast de bewaarcultuur bemoeilijkte de taal waarin het dagboek of reisverslag opgesteld was de collectie, met name voor Brabant in de 19^{de} eeuw. Heel vaak troffen we dagboeken in het Frans aan. Een mogelijke verklaring zou het prestige van het Frans in de hogere sociale klassen kunnen zijn. Het jonge België werd geregeerd door een Franstalige elite (cf. Hoofdstuk 2, Sectie 2.1) en het zijn net vaak de geschriften van de elite die hun weg naar de archieven vonden (Elspaß, 2012b, p. 159). Wanneer we kijken naar de auteurs van de dagboeken, reisverslagen en kronieken in ons corpus, stellen we inderdaad vast dat zij voornamelijk de hogere sociale klassen representeren (cf. Sectie 3.2). Voor Vlaanderen konden we echter wel verschillende handschriften verzamelen uit de 19^{de} eeuw. Dat kan enerzijds wijzen op een gunstige bewaarcultuur in Vlaanderen, en spreekt anderzijds tegen eventuele invloed van het Frans.

Pamfletten

Ook voor het deelcorpus pamfletten werd met vereende krachten gewerkt. Onder leiding van Gijsbert Rutten kon aan de Universiteit Leiden al een aanzienlijk deel van de pamfletten ontsloten worden. Zo werd er al een selectie van geschikte pamfletten gemaakt in de eerdergenoemde catalogus van Knuttel. Student-assistenten Hielke Vriesendorp, Rosanne van den Berg en Gillan Wijngaards hielpen bij de inventarisatie en transcripties van deze pamfletten.

Voor meer dan de helft van het deelcorpus konden we putten uit de Knuttel-collectie, die met ongeveer 32.000 pamfletten behoorlijk omvangrijk is (Van der Hoeven, 1978, p. xix). De collectie kan men digitaal raadplegen via *Dutch Pamphlets Online*. De Knuttel-collectie – die niet gecreëerd werd met het oog op een diachrone studie – kon echter niet alle periodes en regio's voor dit doctoraal onderzoek overbruggen. Met name voor Vlaanderen (16^{de}, 18^{de} en 19^{de} eeuw) en Brabant (19^{de} eeuw) waren additionele bronnen noodzakelijk. Voor Zeeland (16^{de} eeuw) ontbraken er ook nog teksten en in mindere mate vertoonde de Knuttel-collectie lacunes voor Holland (16^{de} eeuw). De lacunes voor Vlaanderen konden opgevuld worden met pamfletten uit de collecties van de Universiteitsbibliotheek Gent en het Rijksarchief Gent. Bijkomende pamfletten voor Brabant trof ik aan in het FelixArchief (Stadsarchief Antwerpen), de Erfgoedbibliotheek Hendrik Conscience en het Rijksarchief Gent. Op enkele scans van de Universiteitsbibliotheek Gent na bezocht ik de bovengenoemde archieven en bibliotheken om de originele documenten te fotograferen.

De plaats waar het pamflet gedrukt werd gold als maatstaf om de teksten geografisch te classificeren. Een uitdaging daarbij was de vermelding 'Ghegheuen in onser stadt van Bruessele' (met spellingvarianten), die we verschillende keren aan het einde van een pamflet aantroffen. Het lemma 'geven' kan echter wijzen op de dag waarop een wet afgekondigd werd. Zo vermeldt het *WNT* het voorbeeld 'Gegeven te 's-Gravenhage, 15 Januari 1886' ("Geven", 2007). Voor onze pamfletten betekent dit dat de wetten of verordeningen hoogstwaarschijnlijk voor het eerst in Brussel afgeroepen werden en vervolgens hun weg vonden naar verschillende drukkerijen. De pamfletten die in Vlaanderen gedrukt werden beschouwden we dan ook als Vlaamse pamfletten, ondanks de bovengenoemde vermelding.

5 Digitalisering

Om de teksten in het corpus digitaal te kunnen doorzoeken en analyseren, werden ze handmatig getranscribeerd en opgeslagen als tekstbestanden. Voor de transcriptierichtlijnen baseerde ik me op het transcriptieprotocol dat opgesteld en gehanteerd werd voor het *Going Dutch Corpus* aan de Universiteit Leiden (cf. Krogull, 2018). Het transcriptieprotocol van Krogull (2018) werd overigens gebaseerd op de transcriptierichtlijnen die Vosters (2011) opstelde voor het VKN-corpus. Naar het voorbeeld van het *Going Dutch Corpus* en het VKN-corpus definieerde ik specifieke richtlijnen voor de teksten in het HCD. Hoewel we te maken hebben met een deelcorpus gedrukte teksten (pamfletten) en twee deelcorpora geschreven teksten (administratieve teksten en egodocumenten), stemmen de transcriptierichtlijnen grotendeels overeen. Ik koos er dan ook voor om de transcriptierichtlijnen samen te voegen in één transcriptieprotocol voor alle teksten.²² Centraal staat het streven naar een zo diplomatisch mogelijke transcriptie. Dat houdt o.m. in dat we de regeleindes respecteren, de oorspronkelijke spelling en spatiëring behouden en geen interpunctie toevoegen indien die er niet staat. Op die manier trachten we om het originele document zo getrouw mogelijk weer te geven.

5.1 Werkwijze

Het merendeel van de teksten opgenomen in het HCD was nog niet gedigitaliseerd en diende nog getranscribeerd te worden. In sommige gevallen konden we een beroep doen op reeds bestaande transcripties. Van de 209 teksten in het corpus, zijn er 46 gebaseerd op een bestaande transcriptie. Concreet gaat het om 38 transcripties van administratieve teksten en 8 van egodocumenten. Van de pamfletten bestond tot nader order nog geen transcriptie. De overige teksten werden getranscribeerd door student-assistenten in Leiden en Brussel, vrijwilligers en studenten in het kader van Wikiscripta Neerlandica II (Vrije Universiteit Brussel & Universiteit Leiden, gefinancierd door ANV), en de auteur van dit proefschrift. Met name student-assistenten Emmy Stevens, Hielke Vriesendorp en Rosanne van den Berg verrichtten een aanzienlijk deel van het transcribeerwerk. De correctie van de transcripties gebeurde steeds

²² Het volledige transcriptieprotocol kan men raadplegen in Bijlage 2 van dit proefschrift.

door iemand anders dan de transcribent; door de auteur van dit proefschrift, Marijke van der Wal (Noordelijke egodocumenten), Charlotte Verheyden en Eline Lismont.

De bestaande transcripties (al dan niet uitgegeven) werden beschouwd als eerste transcripties die naderhand nog nagekeken dienden te worden. Zo toetsten we iedere bestaande transcriptie – hoe betrouwbaar ook – aan het oorspronkelijke manuscript om eventuele errata en inconsistenties te elimineren. De vergelijking met het originele handschrift was dan ook een voorwaarde om de transcriptie aan te wenden. Verder stemden we de transcriptie af op de transcriptierichtlijnen (zie Sectie 5.3) die voor dit onderzoek gehanteerd werden om de continuïteit van het corpus zeker te stellen.

Zoals vermeld konden we vooral voor de administratieve teksten een beroep doen op reeds beschikbare transcripties, in het bijzonder van 16^{de}- en 17^{de}-eeuwse bronnen. Voor Holland konden we transcripties uit die vroegste periodes overnemen uit *Bronnen tot de geschiedenis van de Leidsche textielnijverheid 1333-1795* (Posthumus, 1910-1922) en *Bronnen tot de geschiedenis van het bedrijfsleven en het gildewezen van Amsterdam 1510-1672* (Van Dillen, 1929-1974). Aan de werkgroep Paleografie in Zeeland (PaiZ) en de Historische Vereniging Arnemuiden ontleen we transcripties van 16^{de}- en 17^{de}-eeuwse Zeeuwse teksten. Het project *Itinera Nova* gaf toestemming om transcripties van 16^{de}-eeuwse schepenregisters uit Leuven aan te wenden voor dit corpus. Daarnaast mochten we voor Brabant gebruikmaken van de transcripties van schepenregisters die vrijwilliger Louis De Bondt maakte in samenwerking met het Rijksarchief Leuven. Tot slot konden we voor Vlaanderen deels putten uit de transcripties gepubliceerd in *Costumen van de stad en van de kasselrij Kortrijk. II. Turben afgenomen door de Kortrijkse schepenbank (1485-1581)* (Monballyu, 1989).

5.2 Metadata

Uit elke transcriptie verzamelden we een reeks metadata. Een eerste gegeven is de auteur. Bij de egodocumenten is het zo goed als steeds duidelijk wie de tekst geschreven heeft. Voor anonieme auteurs hanteerden we de aanduiding ‘Anoniem’, gevolgd door het jaartal waarin de tekst tot stand kwam (bv. Anoniem_1658). Op die manier konden anonieme auteurs ook als

aparte auteurs geanalyseerd worden. In de pamfletten wordt zelden melding gemaakt van een auteur en ook de scribenten van administratieve teksten blijven vaak onbekend. Voor deze twee genres laten we het gegeven van de auteur dus achterwege. Het jaartal van de teksten kunnen we voor elk genre achterhalen en bestaat ofwel uit vier cijfers ofwel uit twee reeksen van vier cijfers, die dan respectievelijk het begin- en eindjaartal van het document aanduiden. De laatste optie komt vooral voor bij de egodocumenten, omdat sommige dagboeken en vooral kronieken over verschillende jaren heen bijgehouden werden. Om de teksten ook in ruimte te situeren, probeerden we de oorsprong van ieder document te achterhalen. Anders dan bij de pamfletten en ambtelijke teksten – waar systematisch melding gemaakt wordt van de locatie – gingen we bij de egodocumenten in eerste instantie uit van de geboorteplaats van de auteur om deze vraag te beantwoorden. Indien de geboorteplaats van de scribent niet te achterhalen was, baseerden we ons op de vertrekplaats (reisverslagen) of de plaats waar een functie uitgeoefend werd (bv. burgemeester) om de teksten geografisch te definiëren. Soms konden we de plaatsnaam ook afleiden uit het handschrift zelf. Indien een tekst zich bijvoorbeeld volledig afspeelde in Mechelen, maar de geboorteplaats van de auteur onbekend was, noteerden we Mechelen als plaatsnaam. Plaatsnamen die nadien nog onzeker waren werden tussen vierkante haken geplaatst. In het hele corpus vormen onzekere plaatsnamen echter een absolute minderheid.

De metadata werden ook geïntegreerd in de naamgeving van elk tekstbestand. Elk bestand kreeg een unieke naamgeving, die voor alle genres volgens hetzelfde stramien opgebouwd is, bijvoorbeeld:

ADM-1550-BR-1-RAL_Schepenbank_Steenhuffel_1550_1551.txt

Uit deze naamgeving kunnen we het volgende afleiden (van links naar rechts):

- genre
- periode
- regio
- nummer van de tekst
- archief
- specificatie

- plaatsnaam
- jaartal(len) tekst

De egodocumenten en pamfletten kregen respectievelijk de aanduiding 'EGO' en 'PAM'. Verder verschilt ook de specificatie per genre. Bij de administratieve teksten werd gespecificeerd om welke soort tekst het gaat (in het voorbeeld een tekst afkomstig van de schepenbank). Bij de egodocumenten is de specificatie de auteur. Bij de pamfletten uit de catalogus van Knuttel werd het nummer van Knuttel overgenomen als specificatie; de overige pamfletten kregen geen specificatie.

5.3 Transcriptierichtlijnen

De teksten in het HCD werden getranscribeerd volgens de diplomatische methode. Dat houdt in dat we streven naar een zo authentiek mogelijke weergave van de tekst. Hoofdletters, leestekens en spatiering worden daarbij niet genormaliseerd (Nederlands Historisch Genootschap, 1988, pp. 9-10). Hieronder volgt een overzicht van de elementen uit de originele teksten die we zo getrouw mogelijk overgenomen hebben.²³

Eerst en vooral werden de regeleindes en structuur van de paragrafen gerespecteerd. Het algemene uitzicht van de tekst bleef op die manier behouden in de transcriptie. Op het niveau van de tekst zelf, verrijkten we de transcriptie met tags.

<leftmargin> en **<rightmargin>** voor aantekeningen in de marge;

<illeg/> voor volledig onleesbare woorden;

<illeg com="..."/> voor deels onleesbare woorden met een suggestie voor een (deel van een) mogelijke lezing tussen de aanhalingstekens;

<ambig> voor onzekere lezingen;

**** voor doorgehaalde woorden die nog leesbaar waren;

²³ Dit overzicht is niet exhaustief. In Sectie 5.3 worden de meest voorkomende elementen beschreven. Het volledige transcriptieprotocol werd als bijlage (cf. Bijlage 2) bij dit proefschrift gevoegd.

<cancel/> voor doorgehaalde woorden die niet meer leesbaar waren;

<ins> voor ingevoegde woorden (boven, onder of naast de regel);

<reg orig="af-|breking">afbreking</reg> voor woordafbrekingen aan het einde van een regel;

<ed com="..."> voor opmerkingen bij een woord of specifieke passage in de tekst;

<pb/> om het einde van een pagina aan te duiden.

Verder kregen eigennamen de tag **<name>**²⁴ en geografische namen de tag **<pl>**.

6 Methode

Om het concept pluricentriciteit in de Nederlandse taalgeschiedenis te beoordelen, voerden we kwantitatieve corpusanalyses uit in het HCD. In het totaal werden zes taalkundige features als afhankelijke variabelen afzonderlijk van elkaar geanalyseerd. Deze worden hieronder opgesomd. De eerste drie variabelen zijn orthografische variabelen, gevolgd door een morfonologische variabele (sjwa-apocope) en twee morfosyntactische variabelen (negatie en d-/w-relativa):

- de spelling van /a:/ in gesloten syllabe
- de spelling van /k/ in woordfinale positie
- de spelling van /t/ bij werkwoorden met een d-stam in de 2^{de} en 3^{de} persoon enkelvoud en de 2^{de} persoon meervoud van de indicatief presens
- sjwa-apocope bij werkwoordsvormen in de 1^{ste} persoon enkelvoud van de indicatief presens
- de verandering van tweeledige negatie naar moderne eenledige negatie

²⁴ Een onzekere lezing van een eigennaam werd aangeduid met een **<ambig>**-tag binnenin de **<name>**-tag.

- de verandering van d- naar w-relativa bij de betrekkelijke bijwoorden en betrekkelijke voornaamwoordelijke bijwoorden ('daar' > 'waar')

We beschouwen elke variabele vanuit de drie dimensies in het HCD, namelijk periode, regio en genre, en bekijken eveneens de interacties tussen deze onafhankelijke variabelen. De interactie tussen de onafhankelijke variabelen regio en genre enerzijds, en periode en regio anderzijds, wordt in alle analysehoofdstukken bestudeerd. In de hoofdstukken 5 en 6 visualiseren we daarnaast ook de interactie tussen de drie onafhankelijke variabelen (i.e. periode – regio – genre). Het totale aantal resultaten voor de spelling van /a:/ en de spelling van /k/ is namelijk voldoende hoog om een driedelige interactie mogelijk te maken. Voor de overige features vonden we minder resultaten, waardoor een opsplitsing over drie variabelen niet zo representatief zou zijn. In de hoofdstukken over deze variabelen (i.e. hoofdstukken 7, 8, 9 en 10) voegen we daarom de combinatie periode – genre toe.

De invloed van de bovengenoemde taalexterne factoren staat centraal in ons onderzoek. Toch rapporteren we in enkele hoofdstukken ook over de invloed van de belangrijkste taalinterne factoren, met name over variatie gebonden aan het lemma (spelling van /t/, sjwa-apocope) en de syntactische context (negatie).

6.1 Zoekmethode

De eerste stap bestond erin om de variabelen op te sporen in het corpus. Zowel het doorzoeken van het corpus als het analyseren van de resultaten gebeurde met uitgewerkte scripts in *R*. Dit is een programmeertaal die toelaat om tekstuele data te onderzoeken, analyseren en statistisch te verwerken (zie o.a. Gries, 2021; Levshina, 2015). Voor de visualisatie van de data maakten we gebruik van het *R*-pakket ggplot2 (Wickham, 2016).

Per variabele werden reguliere expressies opgesteld om het corpus te doorzoeken. Als voorbeeld van zo'n reguliere expressie nemen we hier de spelling van /a:/ in gesloten syllabe, de eerste orthografische variabele die we zullen bespreken.

`\\b\\w{0,}aa[bcdfghklmnpqrstvwxyz]\\b|\\b\\w{0,}aa[bcdfghklmnpqrstvwxyz]{2,}\\w{0,}\\b25`

Met deze zoekexpressie worden alle tokens in het corpus opgespoord waarbij /a:/ gespeld wordt als <aa>.

[6]	"bepaalt"	"aande"	"gemaakt"	"gemaakt"
[11]	"bestaan"	"genoodzaakt"	"aanschouw"	"aanwas"
[16]	"Fiscaal"	"verhaal"	"scrobbelaars"	"daar"
[21]	"schrobbelaars"	"schrobbelaarsbaas"	"haar"	"baas"
[26]	"tpaar"	"plaatse"	"haar"	"aangegeeven"
[31]	"aantemanen"	"gaan"	"aan"	"haar"

Figuur 14: uittreksel zoekresultaten met <aa> voor de spelling van /a:/

Met de reguliere expressies konden we in het corpus zoeken naar mogelijke spellingvarianten van elke variabele. Bij de spelling van /a:/ is <aa> slechts een van de mogelijke varianten, en zochten we eveneens naar <ae> en <ai>. We beschrijven in de analysehoofdstukken de zoekmethode per variabele. Een overzicht van alle reguliere expressies kan men raadplegen in Bijlage 3 van dit proefschrift.

De verkregen tokens werden vervolgens manueel gefilterd en geannoteerd in Excel. Deze werkwijze laat toe om zogenaamde vals positieve resultaten uit de dataset te verwijderen. Zo werd bij de spelling van /a:/ een woord als 'detail' verwijderd (geen /a:/), terwijl 'dair' behouden bleef (modern lemma: 'daar'). Doorgehaalde woorden of passages, ambigue lezingen, passages in een andere taal, eigennamen en geografische namen konden vooraf al uitgesloten worden van de analyses. Deze werden tijdens de transcriptiefase namelijk voorzien van tags (cf. Sectie 5.3 van dit hoofdstuk).

Voor de drie orthografische variabelen gebeurde de eerste controle door student-assistente Julie Van Ongeval. De betreffende databestanden werden nadien een tweede keer nagekeken door de auteur van dit proefschrift. De Excelbestanden met de tokens van de overige variabelen werden geannoteerd en nagekeken door de auteur van het proefschrift. In deze fase codeerden we ook eventuele taalinterne variabelen. Zo kenden we bij de spelling van /k/, /t/ en de sjwa-apocope het moderne lemma toe aan de overgebleven tokens. Bij de variabele negatie werd de syntactische context gecodeerd. Wat de orthografische variabelen betreft konden we hiervoor

²⁵ `\\b` staat voor een woordgrens; `\\w` staat voor een letterteken.

opnieuw rekenen op Julie Van Ongeval. In de analysehoofdstukken bespreken we per variabele welke contexten uitgesloten werden. Voor alle variabelen geldt weliswaar dat onduidelijke voorkomens geen deel uitmaken van de analyses.

6.2 Data-analyse

Omdat het statistisch modelleren van de gegevens niet tot de hoofddoelstellingen van het ondernomen doctoraatsonderzoek behoorde, besliste ik om uitsluitend gebruik te maken van beschrijvende statistiek. Daarnaast sluit ik me aan bij de redenering van Krogull (2018). In zijn proefschrift stelt hij dat “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” (Krogull, 2018, p. 74). In zijn onderzoek naar de effecten van het nationale taalbeleid op het werkelijke taalgebruik koos Krogull (2018, p. 74) voor een groot aantal variabelen in plaats van geavanceerdere statistische analyses. Op dezelfde manier argumenteer ik dat aan mijn studie van historische pluricentriciteit in het Nederlandse taalgebied recht gedaan kan worden met verschillende casestudy’s, waarbij taalveranderingen diachroon, geografisch en per genre in beeld gebracht worden. De benadering is eerder innovatief in de toepassing van het moderne concept pluricentriciteit op een historische context. Ik koos er dan ook voor om de voorziene tijd zo optimaal mogelijk te verdelen over verschillende variabelen.

Een correlatietest was weliswaar mogelijk in het geval van twee numerieke variabelen. Om te bepalen hoe sterk een relatie tussen twee variabelen is, wordt een correlatiecoëfficiënt berekend. De correlatiecoëfficiënt neemt een waarde aan tussen -1 en 1, waarbij -1 staat voor een perfecte negatieve correlatie en 1 voor een perfecte positieve correlatie. De waarde 0 duidt op de totale afwezigheid van een lineaire correlatie tussen de variabelen (Baayen, 2008, p. 87; Levshina, 2015, pp. 115-116). Bij normaal verdeelde data kan de onderzoeker Pearson’s r berekenen (Levshina, 2015, p. 116). Onze data zijn echter niet normaal verdeeld.²⁶ In dat geval wordt een niet-parametrische test gebruikt. Spearman’s rangcorrelatiecoëfficiënt – ook bekend als Spearman’s ρ – dient zich dan aan als mogelijkheid. Deze is

²⁶ Dit werd op voorhand voor elk feature getest.

gebaseerd op de rangorde van de data, eerder dan op de numerieke gegevens die achter deze rangorde schuilgaan (Baayen, 2008, p. 91; Levshina, 2015, p. 116; p. 132). Naast de correlatiecoëfficiënt wordt de p-waarde berekend om na te gaan of de correlatie statistisch significant is (zie o.a. Levshina, 2015, pp. 126-129). We rapporteren in de analysehoofdstukken telkens beide waardes.

6.3 Vooruitblik analysehoofdstukken

In de volgende zes hoofdstukken worden de resultaten van elke variabele afzonderlijk besproken. Elk hoofdstuk vangt aan met een afbakening van de variabele, waarbij ook de zoekmethode gespecificeerd wordt (Paragraaf 1). In die paragraaf belichten we telkens de historische achtergrond van de variabele. Bij de sjwa-apocope, negatie en relativa is de paragraaf over de afbakening iets uitgebreider dan bij de spellingvariabelen, al gaan we bij de spelling van /t/ wel in op de verschillende spellingprincipes. Morfonologische en morfosyntactische variabelen zijn over het algemeen complexer dan orthografische variabelen en behoeven dan ook iets meer uitleg. Vervolgens bekijken we hoe elke variabele zich verhoudt tot de taalnormen en het taalgebruik zelf (Paragraaf 2). De bespreking van de taalnormen werd hoofdzakelijk op secundaire literatuur gebaseerd. Een normatief corpus maakt namelijk geen deel uit van dit onderzoek, en bovendien is er voldoende literatuur voorhanden. Voor het taalgebruik zelf raadpleegden we eerdere historisch-sociolinguïstische studies. In Paragraaf 3 van elk analysehoofdstuk formuleren we enkele hypothesen. Deze hypothesen zijn gebaseerd op de corpusopbouw en op de geraadpleegde literatuur per variabele. Enerzijds kunnen we aan de hand van de genres in ons corpus algemene tendensen voorspellen (cf. Hoofdstuk 4, Sectie 3.4). Anderzijds is het mogelijk om op basis van eerder onderzoek specifieke hypothesen te formuleren. De resultaten zelf presenteren we steeds in Paragraaf 4. Elk analysehoofdstuk sluit af met een bespreking van de resultaten, die we in verband brengen met de theorie over pluricentriciteit (Paragraaf 5). De uitkomsten van de verschillende variabelen brengen we achteraf samen in de conclusie van dit proefschrift (Hoofdstuk 11).