

# Lexikos 3

---

*Redakteur*

*Editor*

P. Harteveld

AFRILEX-REEKS 3:1993

AFRILEX SERIES 3:1993

BURO VAN DIE WAT

STELLENBOSCH

Die publikasie van hierdie boek is moontlik gemaak  
deur 'n ruim subsidie van die L.W. Hiemstra-Trust.

The publication of this book has been made possible  
by a generous subsidy from the L.W. Hiemstra Trust.

Uitgewer      Publisher

BURO VAN DIE WAT  
Universiteit van Stellenbosch  
Privaatsak X5018 Private Bag  
7599 STELLENBOSCH

Kopiereg © 1993 deur die uitgewer  
Alle regte streng voorbehou  
Eerste uitgawe 1993

Tipografie en uitleg  
deur Pieter Harteveld en  
Hermien van der Westhuizen  
Bandontwerp deur Piet Grobler

Geset in 10 op 12 pt Palatino

Gedruk en gebind deur Rotapress  
Stewartstraat 59 Goodwood

ISBN 0 9583222 3 6

Geen gedeelte van hierdie boek mag sonder skrif-  
telike verloop van die uitgewer gereproduseer of in  
enige vorm of deur enige elektroniese of mega-  
niese middel weergegee word nie, hetsy deur foto-  
kopiëring, plaat- of bandopname, vermikrofilming  
of enige ander stelsel van inligtingsbewaring

---

# Een gegevensbank van 14de-eeuwse Middelnederlandse dialecten op computer

Pieter van Reenen en Maaïke Mulder, *Vakgroep Taalkunde,  
Vrije Universiteit, Amsterdam, Nederland*

---

## **Abstract: A Database of 14th Century Middle Dutch Dialects on Computer.**

This study describes the organization of a database, created at the Free University Amsterdam, consisting of about 2,500 charters in Middle Dutch, about 750,000 tokens. The charters were written in the Netherlands and Flanders between 1300 and 1400. Only original texts were chosen, which are dated and located. Most of them have not been published before and are to be found in the archives of The Hague, Leiden, Zwolle, Antwerp, Brussels, etc. It is shown how a charter is treated on its way from the archives to its computer format, how the data are structured and how new information — morphological coding — is added. Special attention has been paid to the way how information has been stored which is not easy to be handled on computer directly. The enriched text may be used for different purposes, not only lexicographical but even dialectal. The second part of the study serves to illustrate how the data can be exploited in view of dialect research. As an example, the verb "zullen" has been chosen. It is shown that the vowels of this verb are not the same in the different Dutch speaking areas. The results are represented on dialect maps.

**Keywords:** CHARTERS, COMPUTER FORMATS, COMPUTERIZATION, DUTCH, ELECTRONIC LANGUAGE DATA BASE, ETYMOLOGY, GRAMMATICAL ANALYSIS, LEXICOGRAPHY, MANUSCRIPTS, MIDDLE DUTCH DIALECTS, PRONUNCIATION, SEMI-AUTOMATIC CODING OF TEXT

**Samenvatting:** Deze studie beschrijft de organisatie van een databank, aangelegd aan de Vrije Universiteit Amsterdam, van ongeveer 2.500 Middelnederlandse oorkonden, bij elkaar zo 'n 750.000 tokens. De oorkonden zijn geschreven in Nederland en Vlaanderen tussen 1300 en 1400. Alleen originele teksten zijn gekozen, die gedateerd zijn en waarvan de plaats van herkomst bekend is. De meeste zijn nooit eerder gepubliceerd, en zijn afkomstig uit de archieven van Den Haag, Leiden, Zwolle, Antwerpen, Brussel, etc. Getoond wordt hoe een oorkonde wordt behandeld van archief tot computerformaat, hoe de gegevens gestructureerd zijn en hoe nieuwe informatie — morfologische codering — wordt toegevoegd. Speciale aandacht wordt besteed aan de manier waarop informatie is opgeslagen die niet makkelijk direct door de computer kan worden verwerkt. De verrijkte tekst kan voor verschillende doeleinden worden gebruikt, niet alleen lexicografisch maar ook dialectaal. Het tweede deel van de studie laat zien hoe de gegevens kunnen worden aangewend voor dialectonderzoek. Bij wijze van voorbeeld is het werkwoord "zullen"

gekozen. Getoond wordt dat de klinkers van dit werkwoord niet hetzelfde zijn in de verschillende Nederlands sprekende gebieden. De resultaten worden weergegeven op dialectkaarten.

**Sleutelwoorden:** COMPUTER FORMATEN, COMPUTERISERING, ELECTRONISCHE TAALGEGEVENS BANK, ETYMOLOGIE, GRAMMATIKALE ANALYSE, LEKSIKOGRAFIE, MIDDELNEDERLANDSE DIALECTEN, NEDERLANDS, OORKONDEN, SEMI-AUTOMATISCHE CODERING VAN TEKST, TEKSTEN, UITSpraak

## 1. Inleiding

Wie de oudste fase van de dialecten in het hele Nederlandse taalgebied toegankelijk wil maken voor onderzoek kan niet verder teruggaan dan de 14de eeuw.<sup>1</sup> Pas in die eeuw zijn van Groningen tot Zuid-Limburg en West-Vlaanderen voldoende teksten voor dat onderzoek bewaard gebleven: originele oorkonden waarvan de datum en plaats van redactie bekend is. Figuur 1 laat zo 'n oorkonde zien.<sup>2</sup>

De oorkonden zijn in groten getale te vinden in de Nederlandse en Belgische archieven. Wij sporen ze daar op, maken een selectie en fotograferen ze. Op het ogenblik (eind maart 1993) hebben we ruim 2.500 oorkonden bij elkaar, ongeveer 750.000 tokens.

Deze teksten lenen zich uitstekend voor taalkundig en lexicografisch onderzoek. Vanuit het oogpunt van de computer maakt het in wezen niet uit om welk type onderzoek het gaat, al zijn er wel graduele verschillen.

De processen en methoden die hieronder worden beschreven hebben dus niet alleen zuiver linguïstische toepassingen maar ook lexicografische. In de taalkunde kan daarbij worden gedacht aan etymologisch onderzoek, in de lexicografie aan het ontwerpen van woordenboeken voor oudere taalfasen.

Wij willen aan dat corpus veelvuldig en systematisch taalkundige informatie onttrekken met het oog op een onderzoek naar uitspraakverschillen tussen de dialecten binnen een precies in de tijd afgebakende periode. Om dat te bereiken hebben wij het corpus op computer gebracht, van extra informatie voorzien, en in verband met de bevraging en verrijking van het corpus pro-

<sup>1</sup> Het onderzoek maakt deel uit van het onderzoeksprogramma Corpusgebaseerde Woordanalyse Vuleit 88/9. Het corpus is mede tot stand gekomen dankzij NWO/ST 300-166-011. Jan Berns, Ton Goeman en Henk Meijering zijn wij dankbaar voor suggesties ter verbetering van deze studie.

<sup>2</sup> In de 13de eeuw vinden we ook al heel wat oorkonden, maar die zijn voornamelijk uit Vlaanderen afkomstig, zie Gysseling (1977). Verder zijn uit de 13de en 14de eeuw literaire teksten bewaard gebleven, maar die vermelden — in tegenstelling tot oorkonden — meestal hun plaats en datum van ontstaan niet. Bovendien zijn het vrijwel nooit originelen, en weten we niet hoe letterlijk ze zijn gecopieerd. Daarom kunnen ze, in tegenstelling tot originele oorkonden, niet dienen als primaire getuigen voor het dialectonderzoek.

grammatuur ontwikkeld. Alle oorkonden zijn nu op computer beschikbaar. Dat is veel werk geweest, ook veel extra denkwerk over bits en bytes, maar een andere keus hadden we niet. Zonder computer zou ons onderzoek niet haalbaar zijn. Hadden we volstaan met alleen maar transcriberen en collationeren van de oorkonden, dan zouden ze alle 2.500 nog steeds zo goed als onbruikbaar zijn geweest voor dialect- en diachroon onderzoek. We hadden een onderzoeksinspanning moeten leveren die het menselijk vermogen eenvoudig te boven gaat. Daarom is het onderzoek dat wij beogen nooit eerder van de grond gekomen. En wat wij nu tot stand brengen moet ook voor het nageslacht zorgvuldig bewaard blijven als uitgangsbasis voor uitbreiding en verder onderzoek op het gebied van de taalkunde en de lexicografie. Bij de aanleg van het corpus op computer en de ontwikkeling van de programmatuur is daar zoveel mogelijk rekening mee gehouden.

Ten eerste worden zoveel mogelijk algemene normen gehanteerd. Al zijn die er nog steeds niet veel, waar mogelijk wordt bij de aanleg van het corpus en de ontwikkeling van de programmatuur gewerkt volgens richtlijnen van het Text Encoding Initiative (TEI) en de Stichting Tekstcorpora en Databestanden in de Humaniora (STDH). Standaardisering van procedures werkt op termijn drempelverlagend en werkbeparend voor de onderzoekers.

Een tweede uitgangspunt is dat alles wat de linguïst aan informatie uit de oorkonden wil bewaren in een of andere vorm ook op de computer terecht komt. Het komt nog steeds veel voor — tot schade van het onderzoek — dat gegevens worden opgeofferd, omdat dat makkelijker werken op de computer zou zijn. Veel zogenaamde data-base benaderingen lijden aan dit euvel. Zij hebben veel weg van het beruchte bed van Procrustus. Bij de opslag op computer hoeft helemaal geen informatie verloren te gaan. Onze teksten worden daarom in hun geheel opgeslagen.

Ten derde is de ontwikkelde programmatuur modulair: taalgebonden componenten (bijvoorbeeld de Franse of Latijnse component) kunnen worden vervangen door andere taalgebonden componenten (de Nederlandse component) zonder dat de programmatuur op andere punten hoeft te worden aangepast. Op het ogenblik werkt de programmatuur met Middelnederlands, Oudfrans, Modern Frans, Arabisch, Latijn of Middelnederduits. Ook de opslag en verwerking der gegevens op computer is modulair. Een zoekcomponent kan onafhankelijk van een opslagcomponent worden aangepast of vervangen. Bovendien is de programmatuur zo veel mogelijk machine-onafhankelijk. Zo kan het overzetten van corpus en van programmatuur naar andere computersystemen met een minimum aan problemen tot stand komen.

Tenslotte moet de aanleg en verrijking van het corpus, gezien het vele handwerk, zo efficiënt mogelijk verlopen. Door alle handelingen zo eenvoudig mogelijk te houden kunnen we tijd en moeite sparen danwel onze productie verhogen. Met het oog hierop hebben wij een algemeen tekstformaat gedefinieerd: het *invoerformaat*. Het aanleveren van teksten in invoerformaat geeft de linguïst of lexicoloog het minste werk, maar maakt dit formaat daarmee

nog niet geschikt voor ondervraging van het corpus met computerprogramma-tuur. Daartoe moet het corpus eerst worden omgezet in het *computerformaat*. Invoer- en computerformaat bevatten nog niet alle informatie die de linguïst of lexicoloog in de tekst zou willen aantreffen. Wij zouden onze teksten op zijn minst morfologisch gekodeerd willen zien, te zijner tijd ook syntactisch. De morfologische codering wordt in ons corpus semi-automatisch tot stand gebracht. Het resultaat noemen we een tekst in *onderzoeksformaat*. Pas dan kan het onderzoek naar middeleeuwse dialecten goed beginnen.

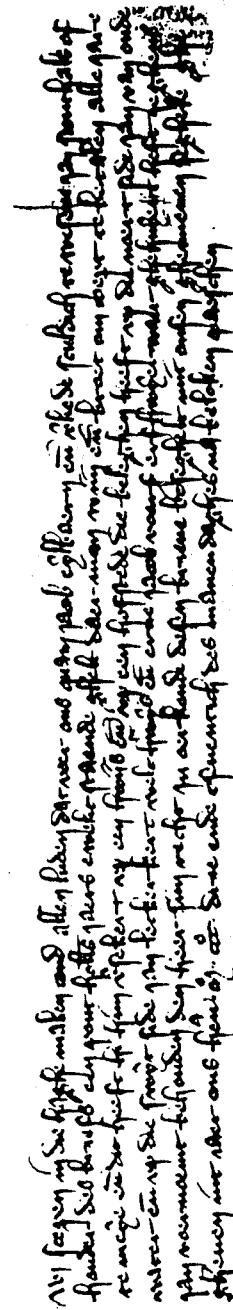
In het vervolg zullen wij in sectie 2 nader aandacht schenken aan de opbouw van ons corpus: wat is precies de informatie die wordt toegevoegd en hoe wordt het ene formaat in het andere omgezet. Voor een technischer en uitvoeriger behandeling van dit aspect wordt verwezen naar Huber (1989). Alle inspanningen hebben tot doel te komen tot resultaten die we anders niet hadden kunnen bereiken. Het is in dit stadium al duidelijk dat een van de resultaten die al heel eenvoudig verkregen kunnen worden een woordenlijst is die, met verwijzing naar kontekst, een mooie aanzet vormt voor het maken of aanvullen van een Middelnederlands woordenboek. Maar er kan veel meer. In sectie 3 laten wij zien met welk gemak dialectgegevens kunnen worden aange-maakt, mede dankzij het kaartontwerp van E. Wattel, zie Wattel (1990). Bij wijze van voorbeeld geven wij een paar kaarten van de klinkers van het werk-woord *zullen*. Ze zijn tot stand gekomen met behulp van frekwentietabellen, die (vrijwel) automatisch uit het corpus kunnen worden gedestilleerd. Een paar samenvattende tabellen staan ook in de tekst. Bij wijze van contrast bezien we onze resultaten kort in het licht van eerdere publikaties.

## 2. Het oorkondenbestand op computer: van archiefstuk tot onderzoeks-formaat

Zoals een rups een pop wordt, en een pop een vlinder, zo verandert ook een oorkonde van vorm. Van archiefstuk tot object van taalkundig of lexicogra-fisch onderzoek op computer onderscheiden we vijf stadia:

Archiefstuk > transcriptie > invoerformaat > computerformaat > onderzoeks-formaat

Tot en met invoerformaat gaat het om puur handwerk. Het wordt besproken in 2.1. Het omzetten van invoerformaat in computerformaat gaat volautoma-tisch. Hoe wordt besproken in 2.2. Het onderzoeksformaat komt interactief tot stand, dat wil zeggen semi-automatisch, zie 2.3. In deze sectie zal ook de dis-ambiguering van homografen worden behandeld.



- [1] Wij scepen in die haghe maken cond allen luden dat voer ons quam jacob eylbaern en(de) v(er)liede sculdich te wesene jan stouthals of
- [2] houder des briefs een pont holl(an)s jaers ewelic staende ghelt daer men wiin en(de) broet om coept te betalen alle jare
- [3] te meye en(de) dit heeft hi hem versekert vp een hwijs en(de) vp een hofstede die belegghen heeft vp die noert side jan van oude+
- [4] water en(de) vp die swyt side jan leckerkerk welc hwijs en(de) erue jacob voerseit erfhuywerwaer ghehuyert heft jegghens
- [5] jan voernoemt behouden den heer siin recht jn oerkonde desen brieue beseghelt mit onsen ghemienen seghele ghe+
- [6] gheuen int iaer ons h(er)en M ccc drie en(de) tseuentich des manendaghes na beloken paeschen

Figuur 1: Een oorkonde met transcriptie uit Den Haag van 1373.  
Archief: Tafel van de Heilige Geest. Inv. 590.

## 2.1 Van archiefstuk tot invoerformaat

De weg van archiefstuk tot oorkonde in invoerformaat moet zo snel mogelijk worden afgelegd. Het is veel handwerk dat niet kan worden geautomatiseerd. Wij bespreken kort hoe de selectie, transcriptie en verrijking van de oorkonden in dit stadium tot stand komt.

### 2.1.1 Selectie van archiefstukken

De grote meerderheid van onze 14de-eeuwse oorkonden is niet uitgegeven en moet van (foto's van) de perkamenten originelen in de archieven worden overgeschreven en gecollationeerd (papieren dokumenten zijn nog heel zeldzaam). In de archieven kiezen wij zo mogelijk zo 'n 100 oorkonden per plaats, gelijkelijk over de eeuw verdeeld. Wij selecteren bij voorkeur goed leesbare teksten, wat in de tweede helft van de eeuw meestal geen probleem geeft: er is overvloed aan keuze. We streven geen volledigheid na, zoals Gysseling (1977) heeft gedaan voor de 13de eeuw, daarvoor is er te veel materiaal en te weinig tijd beschikbaar (we zijn al 10 jaar bezig!). Soms kunnen we gebruik maken van bestaande tekstitgaven, bijvoorbeeld Erens (1950, 1952), Moors (1952), Vangassen (1964). We kiezen alleen teksten die plaatselijk of regionaal taalgebruik bevatten, dat wil zeggen ze moeten zelf vermelden uit welke plaats of streek zij afkomstig zijn. Alleen in Drente hebben we geen plaatselijke oorkonden kunnen vinden en hebben we ons tevreden gesteld met streekgebonden oorkonden, en de onderscheiden streken zijn vrij groot. Ook elders — bijvoorbeeld in Noord-Brabant — onderscheiden wij naast plaatsen wel regio's, maar die zijn dan vaak beperkt in omvang.

### 2.1.2 Transcriptie van de oorkonden

Wij slaan de documenten niet als foto's op computer op, maar wij transcriberen ze diplomatiek: *u* blijft *u* (bijvoorbeeld in *ghegheuen*), *v* blijft *v* (bijvoorbeeld in *vs*), hoofdletters en kleine letters uit het origineel blijven gehandhaafd, leestekens eveneens, aaneengescreven woorden blijven aaneengescreven (*vanden* is heel frekwent). Enzovoort. Wij laten codicologische en paleografische aspecten buiten beschouwing en doen ook geen onderzoek naar zegels of handen. Wel kiezen wij zo veel mogelijk *verschillende* handen, voorzover daarvoor in een archief voldoende materiaal beschikbaar is. Na transcriptie van het perkament, kan een oorkonde eruit zien als in figuur 2.

Amsterdam 1395

WJ Ricout ianssoen en(de) ian bru"ns  
zoen / scepene in aemstelre damme .  
orconden dat voir ons quam clais  
godevairdssone en(de) geliede dat hi  
vircoft heuet pieter minnen enen  
scepenbrief die sijns vader  
plach te wesen daer de[se] scepen  
brief doergesteken is / en(de)  
droech hem daer alle rechten van  
op. En(de) clays [godevairdsso]ne  
gelieder of betaelt te wesen den  
lesten pe(n)ni(n)c metten eersten /  
Jn oirkonde desen brieue beseghelt  
mit onsen seghelen Geg(heuen) int  
jaer 1000. driehondert . viue  
en(de)tneghentich  
op alregodsheilighen dach ----

Noot: Ricout (regel 1) kan zijn Ricont

**Figuur 2: Tekst van een oorkonde na zijn transcriptie waarin veel voorkomende moeilijkheden zijn verwerkt.**

### 2.1.3 Identificatie van teksten

De Middeleeuwse teksten moeten stuk voor stuk makkelijk zijn te herkennen en terug te vinden. In de meesten archieven zijn ze dat ook, maar voor ons taalkundig of leksikografisch onderzoek hebben we een andere klassifikatie nodig, waarin in elk geval plaats van herkomst en tijd van redactie zijn verwerkt. Wij hebben dat uniform voor alle dokumenten opgelost door middel van een letter-cijfercombinatie, bijvoorbeeld *\_o:E109'39509*, zie figuur 3. Die zit als volgt in elkaar.

<i>_o:</i>	Wat volgt is een oorkonde
<i>E109'</i>	De letter-cijfercombinatie voor Amsterdam
<i>395</i>	Het jaartal 1395
<i>09</i>	De negende oorkonde (uit 1395) uit Amsterdam.

De plaats-tijdcode hoeft maar eenmaal boven de oorkonde vermeld te worden en geldt tot de volgende oorkonde, die weer met *\_o:* begint.

Later kan op elk gewenst moment de *\_o:code* automatisch aan alle woorden van de betrokken oorkonde gekoppeld worden. Hoe, daar gaan wij hier nu niet verder op in.

Letter-cijfercombinaties als E109 geven aan om welk dialect het gaat volgens de klassifikatie van Kloeke. Zo kunnen de middeleeuwse gegevens op dialectkaarten worden geplot en is vergelijking mogelijk met dialectgegevens van andere herkomst, bijvoorbeeld uit het hedendaagse Nederlands.<sup>3</sup> Een voorbeeld hiervan wordt in sectie 2 gegeven.

Tussen het Kloekennummer en het jaartal staat een "", een dummy. Er had ook een *a*, een *b* etc. kunnen staan, want het Kloekennummer heeft wel eens een extensie (het gaat dan om een later aan de lijst toegevoegde plaats). Zelf hebben wij vaak een *r* toegevoegd ter aanduiding van *regio*. De regio krijgt het nummer van de centrale plaats binnen de regio, waarbij wij het Kloekennummer met 500 hebben verhoogd. Dat geeft nooit problemen met het aantal cijfers — dat blijft altijd drie —, want het Kloekennummer is altijd lager dan 500. E109 is een plaats (Amsterdam), E609 is diezelfde plaats gelegen in de regio (regio Amsterdam). De dummy "" zorgt ervoor dat alle nummers even lang zijn, ook als er geen extensie nodig is: 10 karakters. Die uniformiteit is heel belangrijk voor latere computerbewerkingen.

Met letter-cijfercombinaties als E109`39506 identificeren wij ieder document ondubbelzinnig. Bovendien geven wij tegelijk plaats of regio van herkomst en het jaar binnen het millennium (van 1000 tot 1999) aan.

```
_o:E109`39509
WJ Ricout3_n,u-ofn? ians3$soen en(de) ian3 bru"ns3
zoen / scepene in aemstelre+damme3
orconden dat voir ons quam clais3
godevairds$sone en(de) geliede dat hi
vircoft heuet pieter3 minnen3 enen
scepenbrief die sijns vader
plach te wesen daer> de[se] scepen+
brief <doer$gesteken is / en(de)
>ldroech hem daer>2 alle rechten <2van
op<1. En(de) clays3 [godevairds3$so]ne
geliede$r+of betaelt te wesen den
lesten pe(n)ni(n)c metten eersten /
Jn oirkonde desen brieue beseghelt
mit onsen seghelen Geg(heuen) in$t
jaer 1000. driehondert . viue
en(de)tneghentich
op alre$gods$heilighen dach ----
```

**Figuur 3: De oorkonde uit figuur 2 in invoerformaat**

<sup>3</sup> Dit corpus is eveneens beschikbaar aan de Vrije Universiteit, die voor dit project samenwerkt met het P.-J. Meertens-Instituut.

## 2.1.4 Speciale tekens

In de oorkonden komen speciale, typisch middeleeuwse tekens voor. De "/" bijvoorbeeld is een zogenaamde Duitse komma. De Umlaut in *Bru"ns* staat in het origineel op de klinker, wij zetten hem erachter. Er zijn meer van zulke superscripten, zoals "^", dat (toevallig) staat voor een hoog geschreven "o". Andere symbolen van deze aard kunnen zonodig vrijelijk worden toegevoegd. Een lijst van zulke conventies moet per project worden bijgehouden. Uitsluitend standaard ASCII wordt gebruikt.

## 2.1.5 Aanvullingen op de transcripties: afkortingen en verminkte passages

Afgekorte passages die zijn aangevuld of verminkte tekst die is hersteld staat tussen [ ] of ( ). De ( ) worden gebruikt om aan te geven dat er een afkorting in het middeleeuwse handschrift stond die door de moderne uitgever is opgelost. De [ ] omgeven passages uit het origineel die niet leesbaar zijn, maar waarvan kon worden gereconstrueerd wat er gestaan moet hebben.

```
en~ -> en(de)
pe~ni~c -> pe(n)ni(n)c
de[se]
[godevairds]ne
```

Zulke gereconstrueerde passages en opgeloste afkortingen hebben een aparte status. Ze lijken middeleeuws maar zijn het niet. Is de middeleeuwse vorm van de woorden onderwerp van onderzoek, dan kunnen we ze dus niet gebruiken. Maar we kunnen ze ook niet weglaten. Complete woordvormen en volledige zinnen zijn wel heel belangrijk bij syntactische analyse, waar de woordvorm er als zodanig niet toe doet. (Ze maken de tekst ook makkelijker begrijpelijk voor de onderzoeker.) En ze maken het semi-automatisch coderen van het tekstmateriaal, zie 2.3, niet ingewikkelder dan nodig. Daarom verwarlozen wij zulke tekst niet, maar geven het met behulp van de haken een speciale status.

## 2.1.6 Aanvullingen op de transcripties: scheiden en samenvoegen

Heel vaak zijn in middeleeuwse teksten woorden gescheiden of aaneengescreven, waar wij dat niet zouden doen. Toch respecteren wij de middeleeuwse schrijfwijze. Tegelijk willen we kunnen beschikken over het moderne equivalent. Dat is makkelijker, want uniformer en bekender, en voorkomt verwarring. We slaan de brug tussen de twee systemen met koppel- en ontkoppelteken. De \$ in *doer\$gesteken* ("door gestoken") is zo 'n ontkoppelteken,

waar het origineel *doergesteken* heeft. Als koppelteken gebruiken we een "+": daarmee wordt *aemstelre+damme* tot een enkele eigennaam.

Bij telwoorden wordt "%" ingevoegd als ze gescheiden in het dokument staan, als ze dat niet zijn een "#". Wij kunnen op die manier het telwoord en zijn samenstellende delen zo gewenst apart onderzoeken. Romeinse cijfers in de middeleeuwse tekst worden in de computer ingevoerd gevolgd door \_g ("getal"). Anders kan men verkeerde dingen lezen. In figuur 2 staan geen voorbeelden, maar *lvi* = ongelijk "lui" en wordt *lvi\_g*. Arabische cijfers blijven (voorlopig) wat ze zijn. Bij telwoorden kan de voorcodering met % en # overbodig worden, als een aparte ontleder van telwoorden wordt ontwikkeld.

### 2.1.7 Aanvullingen op de transcripties: eigennamen en fictieve woorden

Eigennamen kunnen vaak niet automatisch als zodanig worden herkend. Is *Jan de Bakker* bakker of heet hij alleen maar zo? Ze worden daarom voorgekoe- deerd. Dat gaat heel snel, we zetten er een 3 (= *eigennaam* in onze invoercode) achter, zie *aemstelre+damme3* in figuur 3 dat daarmee één eigennaam wordt, niet twee. Oorkonden bevatten meestal heel wat eigennamen.

Een speciaal geval vormen de fictieve woorden. Woorden in middel- eeuwse tekst kunnen volledig geassimileerd zijn aan een buurwoord. Dan zijn ze dus compleet verdwenen. Wij voeren ze weer in als fictieve woorden. Dat is nodig voor de syntactische analyse en kan ook bij onderzoek naar assimilatie van pas komen. De fictieve woorden worden gevolgd door \_f (van *fictief*). In *nemet t\_f* (= "neemt het") bijvoorbeeld is de tweede *t* fictief. Het vormt een een- heid met het voorafgaande woord.

### 2.1.8 Discontinue morfemen

Voornaamwoordelijke bijwoorden en werkwoorden komen nogal eens gesplitst in twee, soms zelfs meer, elementen voor (zie figuur 2: *daer ... van* en *droech ... op*). Het inputformaat van deze woorden is vastgelegd in figuur 4.

Het zal duidelijk zijn dat >morfeem moet worden verbonden met <morfeem en morfeem> met <morfeem. Aangezien de posities van ">" en "<" het punt van koppeling aangeven, zal ook duidelijk zijn dat morfeem> en <morfeem al in de volgorde staan waarin ze aan elkaar gekoppeld moeten worden, terwijl dat niet zo is bij >morfeem en morfeem<. Dus zowel begin als eind van een morfeem kunnen als koppelpunt dienen, zodat in het zeer zeldzame geval dat een derde morfeem voorkomt na de twee anderen, het derde morfeem kan worden geno- teerd als <morfeem en het tweede morfeem als <morfeem>.

In de oorkonde van figuur 4 is er nog een tweede complicatie, die welis- waar zelden optreedt, maar waarin toch moet worden voorzien. 1>1droech ... op<1 en daer>2 ... <2van behoren tot twee paren discontinue morfemen waar-

van de domeinen verstrengeld zijn. In dat geval kan ">1" en "<1" worden gebruikt voor het eerste paar morfemen, ">2" en "<2" voor het tweede paar. De twee voorbeelden illustreren tevens beide volgordes. Het voornaamwoordelijk bijwoord *daer*> ... <*van* heeft zijn normale volgorde. Het werkwoord *opdragen* heeft dat niet. Werkwoorden kunnen in twee volgordes voorkomen en een keuze moet worden gemaakt wat als de basisvolgorde wordt beschouwd. De derde persoon enkelvoud <*droech ... op*> in de oorkonde van figuur 3 wijkt dus af van de basisvolgorde.

Notatie	Betekenis
morfeem>	Een morfeem volgt in de tekst dat een geheel vormt met dit morfeem. Dit volgende morfeem is het tweede element in het woord.
<morfeem	Een morfeem gaat vooraf in de tekst dat een geheel vormt met dit morfeem. Dit voorafgaande morfeem is het eerste element in het woord.
>morfeem	Een morfeem volgt in de tekst dat een geheel vormt met dit morfeem. Dit volgende morfeem is het eerste element in het woord.
morfeem<	Een morfeem gaat vooraf in de tekst dat een geheel vormt met dit morfeem. Dit voorafgaande morfeem is het tweede element in het woord.
<morfeem>	Een morfeem gaat vooraf en een morfeem volgt in de tekst die een geheel vormen met dit morfeem. Dit voorafgaande morfeem is het eerste, het volgende morfeem is tweede element in het woord.

Figuur 4: Het invoerformaat van woorden bestaande uit discontinue ele- menten.

### 2.1.9 Aanvullingen op de transcripties: noten

De moderne uitgever heeft soms behoefte tekst en noten toe te voegen aan het middeleeuwse origineel. Noten kunnen bijvoorbeeld betrekking hebben op onzekerheid bij de transcriptie. Noten in de tekst worden voorafgegaan door \_n. Daarachter staat ofwel korte aanvullende informatie ofwel een verwijzing naar elders. Een voorbeeld van korte aanvullende informatie is \_n,u=n? achter *Ricout*. Heel handig is gebruik te maken van \_n,sic voor afwijkende vormen. Het voorkomt telkens opnieuw verifiëren of de transcriptie van het origineel



wel correct was. Informatie over de oorkonden en hun herkomst uit de archieven of over scribenten kan meteen achter het oorkondennummer *\_o* als noot *\_n* worden vermeld, zolang hier geen speciale categorie voor is voorzien.

## 2.2 Van invoerformaat naar computerformaat

Uit het invoerformaat wordt volautomatisch het computerformaat afgeleid. Bij het tot stand brengen van het invoerformaat moest alle informatie met zo min mogelijk handwerk voor de taalkundige of leksikograaf worden aangebracht. Voor het tot stand brengen van het computerformaat geldt een andere doelstelling: de verwerking — bevraging en verder verrijking van het materiaal, een interactief proces — zo eenvoudig mogelijk te laten geschieden. Dat kan door het formaat, zodra het op computer is, af te stemmen op de systematiek van de programmatuur. Dat betekent dat we rekening houden met noties als sorteren: een preciese vorm van alfabetiseren die letters, cijfers en leestekens betreft. Aangezien we bij het sorteren geen rekening willen houden met haakjes, hoofdletters, voorcoderingen e.d., voeren we twee grote operaties uit. We zorgen dat (a) de middeleeuwse tekst alleen met kleine letters wordt geschreven; (b) de middeleeuwse woorden intern geen toegevoegde informatie bevatten. Omdat nog steeds de eis geldt dat geen informatie mag verloren gaan, wordt de toegevoegde informatie vanaf dit stadium niet meer in, maar achter het woord opgeslagen. Middeleeuwse tekst is te onderscheiden van toegevoegde tekst door middel van het "\_", zie figuur 3 en vooral figuur 5. Het streepje wordt gevolgd door een aanduiding die het type code aangeeft, bijvoorbeeld *\_g=getal*, *\_l=leesteken*, *\_c=capitaal (hoofdletter)*, *\_o=oorkonde*. De voorcodering 3 wordt hier vervangen door *\_e* (= eigennaam): *Jan3* → *Jan\_e*. Een woord dat de onderzoeker zelf (gedeeltelijk) heeft aangevuld behoort nu in zijn geheel tot de middeleeuwse tekst. Via de code achter het woord kan worden achterhaald dat dit niet werkelijk zo is.

De metatekst, of het commentaar, achter de diverse "\_"s is vaak karaktergeoriënteerd. Dat houdt in dat het betrekking heeft op de positie van de letters in de middeleeuwse tekststring. De positie van die letters wordt aangegeven door middel van twee soorten tellers.

Het eerste type teller geeft binnen het woord de positie aan van de letter waarop het commentaar betrekking heeft. Dit is bij haakjes het geval en bij hoofdletters.

```
poe(n)ni(n)c  -> penninc_p3,6 (p= parentese, haakje)
de[se]        -> dese_v3-4 (v= vierkant haakje)
WJ            -> wj_c1-2 (c= capital)
[J]an         -> jan_c1_v1
```

In *pe(n)ni(n)c* → *penninc\_p3,6* staan haakjes rondom letter 3 en rondom letter 6 van het woord, in *de[se]* staan vierkante haken om letter 3 en 4. In *WJ* → *wj\_c1-2* zijn de letters 1 en 2 hoofdletters.

Het tweede type teller geeft binnen het woord de positie van de letter aan die aan het te signaleren verschijnsel voorafgaat. Dit is bij discontinue elementen het geval, evenals bij koppeling en scheiding van woorden, en sommige speciale tekens.

```
doer$gesteken -> doer_a4 gesteken_a0 (a=punt van aanhechting)
aemstelre+damme -> aemstelredamme_k9 (k=koppelteken)
bru"ns_u3       -> bruns_u3 (u=umlaut)
daer>           -> daer_ba4 (b=begin)
<doer           -> doer_ea0 (e=eind)
>1droech        -> droech_bb0
op<1            -> op_eb2
daer>2          -> daer_bc4
<2van           -> van_ec0
```

In *doer\$gesteken* → *doer\_a4 gesteken\_a0* geeft "a" gevolgd door het cijfer het punt van koppeling aan. Bij *droech op* en *daer van* is er een probleem, omdat *daer* in de tekst tussen *droech* en *op* in staat. Om de juiste woorddelen bij elkaar te krijgen wordt *droech* en *op* daarom gevolgd door *b0* en *b2*, *daer* en *van* door *c4* en *c0*. In *aemstelre+damme* wordt "+" vervangen door "\_k": *aemstelredamme\_k9*. Bij ">" en "<" worden *\_b* (=begin) of *\_e* (=eind) ingevoerd. Bij leestekens en arabische getallen doen we of er geen middeleeuwse tekst aanwezig is: ze komen als geheel achter het "\_" te staan. Bij omzettingen komen ze er als volgt uit te zien:

```
/      -> _ldk (= leesteken duitse komma)
.      -> _lp (= leesteken punt)
8      -> _g8 (= getal 8)
3/8    -> _g3/8
```

## 2.3 Van computerformaat naar onderzoeksformaat: homografen

De tekst in computerformaat wordt op zijn beurt — semi-automatisch, dat wil zeggen interactief — van morfologische codering voorzien. Het resultaat is het onderzoeksformaat: een verzameling teksten, niet alleen gekodeerd in termen van plaats en jaar, maar bovendien voorzien van een taalkundige code per woordvorm.

De codes in het computerformaat beginnen allemaal met een "\_" gevolgd door een specifieke letter. In het onderzoeksformaat wordt alleen nog de morfologische code toegevoegd. Die hoeft dan niet te worden voorafgegaan door een letter om van alle andere coderingen onderscheiden te zijn.

Bij het semi-automatisch koderen van de tekst worden de te koderen vormen telkens vergeleken met een lijst van items die al eerder gelemmatiseerd zijn: werkwoorden, zelfstandige en bijvoeglijke naamwoorden, voornaamwoorden, voegwoorden, telwoorden enz. Elk lemma bestaat uit een lijst van alle verschillende morfologische vormen die er in geschreven vorm zijn, voorzover ze al in het corpus zijn voorgekomen. Naast gekodeerde tekst wordt dus een tweede output gekreëerd: een lijst van gelemmatiseerde items. Bij elkaar vormen deze twee resultaten een zeer krachtig hulpmiddel bij het systematisch onderzoek van Middelnederlands. Ongeveer de helft van ons corpus is inmiddels van een morfologische kode voorzien. Gedurende deze fase wordt de lijst van gelemmatiseerde items telkens aangevuld. Uiteindelijk ziet onze oorkonde eruit als in figuur 5, waarin zogewenst de aard van de verschillende coderingen nader kan worden gespecificeerd.

```
_o:E109`39509
wj_404_c1-2 ricout_020_n:u-ofn?_c1 ians_022_a4 soen_000_a0
ende_800_p3-4 ian_020 bruns_020_u3
zoen_000_ldk0 scepene_011 in_700 aemstelredamme_021_k9_lp0
orconden_204 dat_810 voir_700 ons_404 quam_220 clais_020
godevairds_022_a10 sone_001_a0 ende_800_p3-4 geliede_221 dat_810
hi_403 vircoft_274 heuet_213 pieter_020 minnen_020 enen_484
scepenbrief_000 die_421 sijns_452 vader_000 plach_230 te_700
wesen_250 daer_620_ba4 dese_419_v-3-4 scepenbrief_000_k7
doer_620_ea0 gesteken_274_a0 is_210_ldk0 ende_800_p3-4
droech_220_bb0 hem_423 daer_610_bc4 alle_321 rechten_214
van_700_ec0 op_220_eb2_lp0 ende_800_c1_p3-4 clays_020
godevairds_022_a10_v1-10 sone_001_a0_v1-2 geliede_221_a7
r_600_a0 of_600 betaelt_274 te_700 wesen_250 den_474 lesten_104
penninc_000_p3,6 metten_700_474 eersten_314_ldk0
jn_700_c1 oirkonde_001 desen_414 brieue_001 beseghelt_274
mit_700 onsen_454 segheuen_014 gegheuen_274_c1_p4-8 in_700_a2
t_470_a0 jaer_000_g1000_lp0 driehondert_300_lp0 viue_300
ende_800_a4_p3-4 tneghentich_300_a0
op_700 alre_325_a4 gods_022_a0,4 heiligen_014 dach_000_ls0
```

**Figuur 5: De oorkonde uit figuur 3 omgezet naar onderzoeksformaat.**

Een van de lastigste problemen bij het koderen is het disambigueren van homografen. Het is meestal handwerk. Soms is het handig de code al aan te brengen in het invoerformaat, zoals bij eigennamen: *bakker3* = *bakker\_e* versus *bakker* als beroep. De voorkodering moet zo kort mogelijk zijn, ter voorkoming van extra werk. Disambiguering van homografen kan ook tot stand worden gebracht tijdens het koderen. In de toekomst moet een deel van de homo-

grafen door kontekstgevoelige regels automatisch kunnen worden onderscheiden.

Zolang ze tot verschillende woordsoorten behoren, kunnen homografen worden onderscheiden door middel van de taalkundige kode. Zijn het echter allebei zelfstandige naamwoorden, of werkwoorden, dan kan dat niet. Bijvoorbeeld, er zijn twee woorden *grave* nl. "comes" en "gracht". Beide worden als *grave\_000* gecodeerd. Maar ze behoren tot verschillende lemma's. Wij lossen dit probleem als volgt op: *grave\_000\_hl,comes* en *grave\_000\_hm,gracht*, waar de "h" = *homograaf*, en de "l" voor *Latijn* als disambiguerende taal staat, de "n" voor *Nederlands*. De disambiguering moet zo kort mogelijk zijn, en niet in termen van etymologie.

Vergelijken we nu de figuren 1, 2, 3 en 5, dan zien we hoe de metamorfose van archiefstuk tot dokument in onderzoeksformaat in diverse stappen tot stand gekomen is. Verdere verrijkingen van het onderzoeksformaat blijven in de toekomst mogelijk.

### 3. De klinkers van het werkwoord ZULLEN: geografische contrasten

Als het corpus oorkonden eenmaal in onderzoeksformaat (tot op zekere hoogte ook al in computerformaat) beschikbaar is, kan het leksikografisch of taalkundig onderzoek pas goed een aanvang nemen. Veel van dat onderzoek is woordvorm- of spellingsonderzoek. Soms kan dat direct worden vertaald in klankverschillen, zonder dat er veel interpretatie aan te pas komt. Dat zullen wij in deze sectie laten zien aan de hand van de klinkers van het werkwoord *zullen* (dat wij hier altijd met een *z* zullen schrijven: in Van Reenen en Wattel (1992) hebben wij de spelling met *s* of *z* bij o.a. dit werkwoord nader onderzocht). Wij beperken ons tot de infinitief en tot de eerste en derde persoon enkelvoud en meervoud van de tegenwoordige tijd. We vinden hier in het meervoud korte of lange klinker, *o*, *u*, *e*, en in het enkelvoud *zal* of *zel*.

Op basis van een serie kleine programma's hebben wij automatisch tabellen tot stand gebracht waarin het voorkomen van *zal* en *zel*, van *zullen*, *zellen*, *zelen*, *zolen*, *zoelen* is vastgelegd in termen van plaats en tijd, in totaal meer dan 2.000 vormen.<sup>4</sup>

De resultaten kunnen op kaarten worden weergegeven dankzij programmatuur ontwikkeld door E. Wattel, zie Wattel (1990). Voor de weergave op de kaarten heeft elke oorkonde geteld als een getuige. Een oorkonde met meer

<sup>4</sup> Een kleine toelichting op de programma's voor de kenners: Het gaat om korte programmaatjes geschreven onder UNIX in de shell: *egreps* en AWK-programmaatjes selecteren relevante vormen en printen ze op de juiste plaats. Verder wordt gebruik gemaakt van commando's als *tr* en *sort* gecombineerd met *uniq -c* en *join*, en substitutie-commando's in de *ex*. Al deze features zijn standaard onder UNIX beschikbaar. Ze zijn ook op PC beschikbaar: o.a. onder MINIX of onder MS-DOS via de MKS-toolkit.

dan een relevante vorm telt echter zwaarder dan een oorkonde met slecht één relevante vorm: 1 vorm telt voor 1, 4 vormen voor 2, en 9 vormen voor 3 (de vierkantswortel uit het aantal voorkomens per oorkonde). In de tabel in de Appendix telt een oorkonde als een getuige, ongeacht het aantal relevante vormen.

### 3.1 Zel en zal

Een eerste resultaat betreft *zel* en *zal*. Het is te zien in tabel 1. Voor de corresponderende dialectkaart, zie Van Reenen (te verschijnen). *Zel* — ongetwijfeld met een [ɛ]-achtige klank uitgesproken — is westelijk, maar komt ook sporadisch in Groningen voor. Van Loey (1966: § 68) komt aardig in de richting als hij *zel* in Holland plaatst. Toch kent Holland ook *zal*, Dordrecht zelfs overwegend. En verder doet ook Utrecht flink met *zel* mee. In de loop van de eeuw spreidt het gebruik van *zel* licht naar het oosten: Amersfoort, en komt de vorm in het noorden in Groningen op. Volgens de gegevens van De Vriendt-De Man (1958) komt *zel* in de 16de eeuw nog voor in Haarlem, maar vooral in Leiden.<sup>5</sup>

Ook in het Afrikaans is de vorm wel eens gesignaleerd, zie Ponelis (1990: § 4.2). In het Kaapse Afrikaans is [səl] normaal. Tenslotte treffen we ook in de moderne Nederlandse dialecten nog resten van het middeleeuwse kaartbeeld aan, zie kaart 1. Weijnen signaleert de vormen ook in Gouda en Katwijk aan Zee (1958: 373, 374 en 370).

### 3.2 Zellen, Zelen, Zo(e)len, lange klinker en korte klinkers

De frekwenties van de klinkers van *zullen* (infinitief of meervoud) staan in tabel 2 per (deel van een) provincie geordend. *Zellen* hoort bij *zel* te horen, maar vergelijking van tabel 1 en 2 laat zien dat dat geografisch niet helemaal klopt. De vorm is namelijk vooral in Utrecht te vinden, veel minder in Holland, en een

<sup>5</sup> De Vriendt-De Man (1958) heeft 540 vormen verzameld waarvan 18 *zel*. Het is opmerkelijk hoe sceptisch ze staat tegenover de betrouwbaarheid van haar gegevens: "De taal der ambtelijke documenten ... wemelt van traditionele formules en archaïstische zinswendingen die dikwijls klakkeloos werden overgeschreven uit een ouder, gelijkaardig stuk". Zou zij gelijk hebben, dan heeft haar onderzoek geen zin gehad. Men vraagt zich af waarom zij het heeft uitgevoerd. Nu is vaststellen of een tekst van een 16de-eeuwse oorkonde al dan niet klakkeloos is overgeschreven geen eenvoudige opgave. Zij heeft voor haar scepticisme dan ook geen overtuigend bewijs gevonden. Tot het tegendeel is aangetoond beschouwen wij daarom de informatie die zij aan deze ambtelijke documenten ontleent als betrouwbaar. Wel delen wij haar scepsis ten aanzien van de literaire teksten waar zij zich op baseert, zolang die niet nader in tijd en ruimte zijn geplaatst.

enkele vorm in Noord-Brabant. In Noord-Holland lijkt *zellen* in de loop van de eeuw wat af te nemen, in Zuid-Holland en Utrecht wat toe. De vorm komt op in Groningen. Raadplegen we Van Loey (1966: § 68), dan vinden we vooral in Holland en Utrecht andere accenten. *Zellen* zou Noordhollands zijn (p.88) en de vorm zou ook wel eens in Brabant voorkomen (p.90). Het beeld gesuggereerd in Van den Berg en Vermeulen (1973) — in belangrijke mate gebaseerd op een kleine deelverzameling van de gegevens die wij ook hebben gebruikt — klopt beter. In haar 16de-eeuwse gegevens heeft De Vriendt-De Man (1958) de vormen niet meer aangetroffen.

Naast *zellen* is er nog een andere vorm met een e-achtige klinker. Kaart 2 laat zien dat *zelen* een typisch Belgisch Brabantse vorm is met als centrum Brussel. Dat vinden we ook bij Van Loey en Van den Berg en Vermeulen en Goossens (1980: 62) terug. Blijkens De Vriendt-De Man (1958: 213-232) komen ze ook in de 16de eeuw nog veelvuldig in Brussel en omgeving voor: 50 op de 109 vormen.

Uit tabel 2 blijkt verder dat *zoelen* en *zolen* oostelijke vormen zijn, zie ook kaart 3, waar *zoelen* en *zolen* samen tegenover de andere vormen zijn gekarteerd. Van Loey (1966: 88, 91) beschouwt *solen* als Limburgs en Hollands. In het licht van onze 14de-eeuwse kaart moet in Holland die vorm dan wel de laatste der Mohikanen zijn geweest, wellicht ook de eerste. De informatie van Van Loey is op dit punt zeker niet representatief. Voor 13de-eeuwse gegevens, zie ook Van Loey (1952).

Tenslotte valt uit tabel 2 de tegenstelling tussen *zullen* met een enkele *-l-* en *zullen* met een dubbele *-ll-*, af te lezen, een verschil dat algemeen wordt opgevat als tussen *lange klinker* en *korte klinker*. De vorm met dubbel *-ll-* (korte klinker) is duidelijk aan de winnende hand, vooral in Noord-Brabant en in het Noorden, zo leert een nadere analyse van de gegevens.

### 3.3 Discussie

Uit de kaarten en tabellen die wij hierboven hebben gegeven komt een algemeen verschil naar voren met bestaande kennis. Daar worden vanouds de dialectverhoudingen in algemene termen aangegeven, vaak vaag, een enkele keer ook te precies. Betekent een uitspraak: "*Zellen* komt in Holland voor", dat *zullen* daar niet voorkomt? Of dat *zellen* alleen in een deel van Holland voorkomt? En is dat dan weleens of altijd? Zelfs wanneer Van Loey een vorm met een datering geeft, kunnen we niet weten of die vorm representatief is. Verder worden veel Nederlandse dialectgebieden niet of nauwelijks vermeld: Utrecht, Gelderland en Zeeland schitteren vaak door afwezigheid. De wel vermelde streken zijn soms wel erg groot en niet altijd precies afgebakend: Holland, Brabant, Limburg en Vlaanderen.

Overigens worden Gelderland, Overijssel, Drente en Groningen niet altijd tot het Middelnederlandse taalgebied gerekend, al hebben wij ze wel onder-

zoekt. Ook Friesland valt erbuiten. Onze kaarten suggereren dat in Friesland *zel-* en *zellen-*achtige vormen moeten hebben voorgekomen. Raadpleging van Vries (1977) leert dat dit klopt. Het middeleeuwse Fries kent *skel* en *sel*, *skellen* en *sellen*.

#### 4. Conclusie

Deze studie valt uiteen in twee delen. Het eerst deel had tot doel te laten zien hoe grote bestanden middeleeuwse teksten op computer kunnen worden opgeslagen en zinvol worden ingericht. De inrichting van de bestanden maakt duidelijk hoe de taalgegevens betrekkelijk eenvoudig kunnen worden gebruikt met het oog op o.a. lexicografisch onderzoek. In het tweede deel werd daarom kort een ander aspect van het taalmateriaal belicht dat onvermijdelijk, bij gebrek aan een standaardtaal, door alle taal- en lexicografisch onderzoek uit die periode heen speelt: de spreiding der gegevens in de ruimte (en de tijd). Getoond werd — aan de hand van de klinkers van het werkwoord *zullen* — wat dit type onderzoek aan resultaten kan opleveren. Door grote aantallen gegevens te destilleren uit gedateerde en plaatsgebonden taalgetuigen en die te systematiseren op dialectkaarten en in tabelvorm, kan onze kennis van het Middelnederlands niet alleen taalkundig maar ook lexicografisch met sprongen toenemen.

#### Verwijzingen

- De Vriendt-De Man, M.J. 1958. *Bijdrage tot de kennis van het gebruik en de flexie van het werkwoord ZULLEN in de 16de eeuw*. Koninklijke Vlaamse Academie voor Taal- en Letterkunde, Reeks VI, Nr. 81. Gent.
- Erens, M.A. 1950, 1952. *De oorkonden der Abdij van Tongerlo* II, III. Tongerlo: St. Norbertsdrukkerij.
- Goossens, J. 1980. *Middelnederlandse vocaalsystemen*. VMA, afl. 2.
- Gysseling, M. 1977. *Corpus van Middelnederlandse Teksten (tot en met het jaar 1300)*. Reeks I: *Ambtelijke Bescheiden*. 's-Gravenhage: Martinus Nijhoff.
- Huber, O.C. 1989. *The Construction of Lexically Analyzed Text Corpora on the Computer*. Corpus Based Word Analysis, Report #1, Vakgroep Taalkunde, Vrije Universiteit, Amsterdam.
- Moors, Joseph. 1952. *Bouwstoffen tot de historische taalgeografie van het Nederlands, II, De oorkonden-taal in Belgisch-Limburg*. Belgisch Interuniversitair centrum voor Neerlandistiek.
- Ponelis, F.A. 1990. *Historiese klankleer van Afrikaans*. Universiteit van Stellenbosch, Annale 1.
- Van den Berg, B. en A.G.M. Vermeulen. 1973. 'Zullen' als proefgeval. *De Nieuwe Taalgids* 66: 445-452.
- Vangassen, H. 1964. *Bouwstoffen tot de historische taalgeografie van het Nederlands, VIII, Noordhol-landse charters*. Belgisch Interuniversitair centrum voor Neerlandistiek.
- Van Loey, A. 1952. Een en ander over Mnl. *Sullen*. VMA: 919-922.

- Van Loey, A. 1966. *Middelnederlandse spraakkunst I*. Vormleer, 5de druk. Groningen: Wolters.
- Van Reenen, P. Te verschijnen. *The Study of Medieval Language in the Low Countries: the Good, the Bad and the Future*. *Medieval Dialectology, Workshop Taxonomy and Typology in Medieval Dialect Studies*. Edinburgh, April 1992.
- Van Reenen, P. en E. Wattel. 1992. De uitspraak van /s/ en /z/ voor klinker in het Nederlands: zes eeuwen variatie. Bennis, H.J. en J. de Vries. (Red.). 1992. *De binnenbouw van het Nederlands*: 291-309. Dordrecht: ICG Publications.
- Vries, O. 1977. *Oudfriesse oorkonden*. Vierde deel. (Oudfriesse taal- en rechtsbronnen uitgeven door dr. P. Sipma. Veertiende deel.) 's-Gravenhage: Martinus Nijhoff.
- Wattel, E. 1990. *A Mathematical Model to Describe the Boundary of a Vowel Shift in the Past on the Basis of Data from Contemporary Dutch Dialects*. Rapportnr. WS -359, Mei 1990, Vrije Universiteit Amsterdam.
- Weijnen, A. 1958. *Nederlandse dialectkunde*. Assen: Van Gorcum.

Appendix

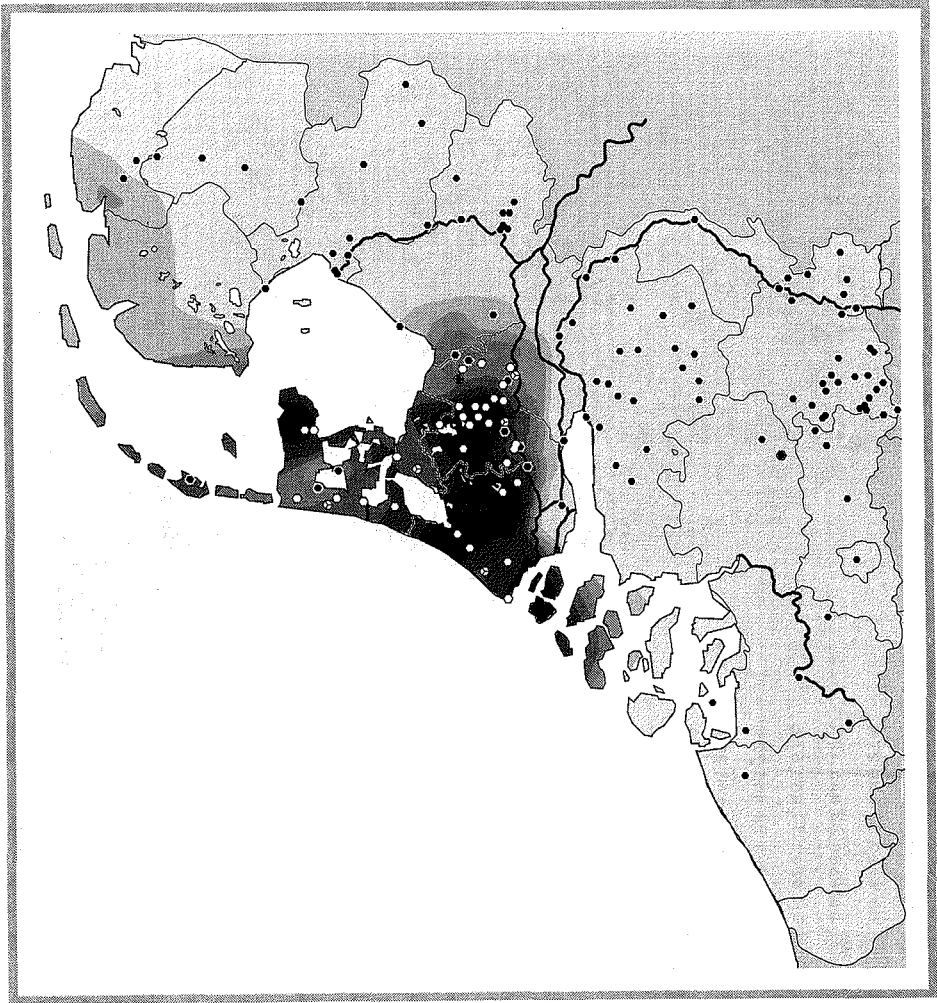
1	2	3	4
	zal	zel	provincie
01	6.5	1.5	01 = Groningen/Drente
02	10	0	02 = Overijssel
03	11	0	03 = Gelderland
04	11	17	04 = Utrecht
05	8.5	11.5	05 = Noord-Holland
06	4	6	06 = Zuid-Holland West
07	3	2	07 = Zuid-Holland Oost
08	1	0	08 = Zeeland
09	6	1	09 = Noord-Brabant West
10	18	0	10 = Noord-Brabant Oost
11	7	0	11 = Limburg
12	1	0	12 = West-Vlaanderen
13	2	0	13 = Oost-Vlaanderen
14	3	0	14 = Antwerpen
15	4	0	15 = Brabant
16	23	0	16 = Belgisch Limburg

Tabel 1: Frekwentie van zal en zel in het 14de-eeuws Middelnederlands.

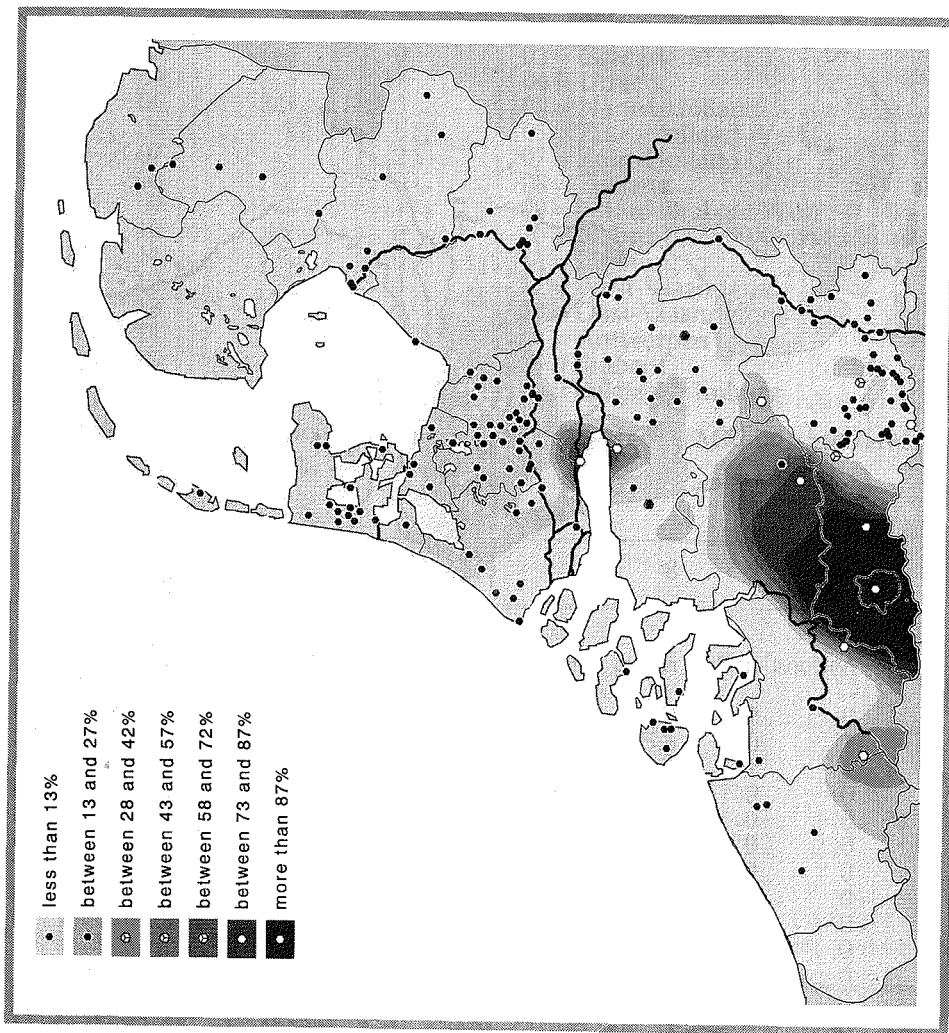
1	2	3	4	5	6	7	8	9	10
prov	zullen	zellen	zelen	zoelen	zolen	zuelen	zulen	zollen	zoellen
01	26.6	6.2	0.5	6	33.5	0	0	1.2	0
02	18.07	0	0	33.87	29.56	0	5.5	0	1
03	2.83	1	0	15.25	38.31	19.48	6.13	1	0
04	11.33	110.17	0	18	1	0.5	0	0	0
05	49.83	15.17	0	0	0	0	0	0	0
06	20.64	8.36	0	0	0	0	0	0	0
07	40.47	2.53	0	0	1	0	0	0	0
08	50	0	0	0	0	0	0	0	0
09	12.89	4.11	7	4	1	0	0	0	0
10	29.39	0	2	24.22	15.39	0	1	1	0
11	10.5	0	0	2.62	18.88	20	2	0	0
12	71	0	0	0	0	0	0	0	0
13	32.25	1	0.75	0	0	0	0	0	0
14	0	0	7.5	2.83	2.67	1	0	0	0
15	1	0	40	0	2	0	0	0	0
16	19	0	3.5	46.32	44.76	9.42	1	0	1

Tabel 2: De klinkers van het werkwoord zullen in 14de-eeuws Middelnederlands

Kaart 1: Verdeling zel (donker gebied) versus zal in 14de-eeuws Middelnederlands.



Kaart 2: Spelling *zelen* (donker) versus andere vormen van *zullen*.



Kaart 3: Spelling *zolen* of *zoelen* (donker) versus andere vormen van *zullen*.

