

Corpus van veertiende-eeuwse niet-literaire Nederlandse teksten (C14NL)

Technische documentatie van het corpus C14NL

Ron Van den Branden



Centrum voor Teksteditie en Bronnenstudie van de Koninklijke Academie voor
Nederlandse Taal- en Letterkunde, 2009, Gent

© 2009, CTB, Creative Commons Attribution Share Alike



INHOUDSOPGAVE

1. Internationale standaarden voor tekstrepresentatie en -opmaak.....	1
2. Algemene tekststructuur.....	4
3. Het header gedeelte.....	6
3.1. Bibliografische beschrijving: <fileDesc>	6
3.2. Beschrijving van de codering: <encodingDesc>	9
3.3. Beschrijving van contextuele aspecten voor de tekst: <profileDesc>	13
3.4. Beschrijving van wijzigingen aan de elektronische tekst: <revisionDesc>	13
4. Het tekstgedeelte.....	14
4.1. Tekststructuur.....	15
4.2. Transcriptie van tekstfenomenen.....	16
4.3. Editeursingrepen.....	18
4.4. Linguïstische analyse.....	20
4.4.1. Algemeen.....	20
4.4.2. Samengevoegde vormen.....	25
4.4.3. Clitische vormen.....	25
4.4.4. Gescheiden vormen.....	26
Appendix A. Overzicht van attributen en elementen.....	28
Appendix A.1. Overzicht van globale attributen.....	28
Appendix A.2. Overzicht van de elementen die in de teksten kunnen voorkomen.....	28
Appendix B. Overzicht van de morfologische tagset.....	39

Technische documentatie

Dit document beschrijft de technische conventies die werden gevolgd voor de codering van het corpus dat werd ontwikkeld in het project *Databank van veertiende-eeuwse niet-literaire Nederlandse teksten. Opbouw en linguïstisch onderzoek* (C14NL). In dit document wordt geen aandacht besteed aan de opbouw en samenstelling van het corpus; wel aan de structuur van de afzonderlijke elektronische teksten en de coderingen die daarvoor gebruikt zijn.

1. Internationale standaarden voor tekstrepresentatie en -opmaak

Voor de opmaak van het corpus C14NL werden internationale standaarden voor tekstrepresentatie en -opmaak gevolgd. Wat betreft tekstrepresentatie werd gekozen voor het eXtensible Markup Language (XML)¹ formaat. Dit is een internationale standaard die ontwikkeld werd door het World Wide Web Consortium (W3C)², een internationaal consortium (met ongeveer 400 ledenorganisaties waaronder grote industriële spelers als IBM en Microsoft) dat zich richt op de ontwikkeling van standaard webtechnologieën. XML documenten bestaan uit platte tekst waarin allerlei meta-informatie kan worden opgenomen. Die meta-informatie bestaat hoofdzakelijk uit *elementen* en *attributen*, zoals bijvoorbeeld:

```
<?xml version="1.0"?>
<tekst>
  <paragraaf nummer="1">Paragraaf één van
    <titel>De trein der traagheid</titel>.
  </paragraaf>
  <paragraaf nummer="2">Paragraaf twee van
    <titel>De trein der traagheid</titel>.
  </paragraaf>
</tekst>
```

3

Dit voorbeeld bevat 5 elementen, waarvan het bereik wordt aangegeven door tags. Tags worden gemarkeerd door punthaken (< >) en duiden ofwel het begin (starttag) van een element aan, ofwel het einde (eindtag). In starttags wordt binnen de gepunte haakjes de naam van het element opgenomen. In eindtags wordt de naam van het element binnen de gepunte haakjes voorafgegaan door een slash karakter (/). Er kan verdere informatie worden gegeven voor elementen, door gebruik van attributen in hun starttag. Attributen bestaan uit een attribuutnaam, gevolgd door een gelijkheidsteken (=) en attribuutwaarde tussen enkele (') of dubbele (") aanhalingstekens. In dit voorbeeld wordt de eigenlijke tekst omsloten door een <tekst> element. Daarbinnen worden twee <paragraaf> elementen onderscheiden. Die worden verder getypeerd door verschillende waarden voor het **nummer** attribuut. Verder

.....

1 eXtensible Markup Language (XML): <http://www.w3.org/TR/REC-xml/>

2 World Wide Web Consortium (W3C): <http://www.w3c.org/>

3 Hierbij moet worden opgemerkt dat bij de voorbeelden in dit document extra witruimte is toegevoegd om de leesbaarheid te bevorderen. De hoeveelheid witruimte heeft over het algemeen geen betekenis in XML-documenten; alle tekst had net zo goed op één regel kunnen staan.

komen er in de tekst (en dus ook binnen de **<paragraaf>** elementen) nog 2 **<titel>** elementen voor. Om aan te geven dat het om een XML document gaat, moet elk document beginnen met een XML declaratie (in het voorbeeld: `<?xml version="1.0"?>`), met een aanduiding van de versie van de XML standaard waaraan het document beantwoordt. Merk op dat in dit voorbeeld gebruik wordt gemaakt van inzichtelijke elementnamen, maar dat evengoed alle elementen de naam "x" hadden kunnen hebben. De XML standaard bepaalt enkel de syntactische regels waaraan de opmaak voor documenten moet voldoen. Het biedt de volgende voordelen:

- technologische neutraliteit: omdat XML documenten in wezen platte tekst zijn, is de toegankelijkheid niet afhankelijk van bepaalde software, waardoor duurzaamheid wordt gegarandeerd.
- scheiding van structurele *representatie* en visuele *presentatie*: omdat XML documenten niet meer zijn dan platte tekst met structurele meta-gegevens, wordt het mogelijk om de tekst inzichtelijk te beschrijven en pas in latere instantie te bepalen hoe die beschrijving door de computer moet worden geïnterpreteerd.
- semantische uitbreidbaarheid: omdat XML syntactische regels voorschrijft voor de opmaak van teksten, kunnen daarmee verschillende inzichtelijke codeerschema's worden ontwikkeld.

XML voorziet mogelijkheden om de logische structuur van documenten vast te leggen in een Document Type Definition (DTD). Die bevat een structuurgrammatica (definities van de namen, inhoud en distributie van elementen en attributen) waaraan alle documenten van dat type moeten voldoen. Hiermee kunnen dus definities voor de opmaak van verschillende teksttypes formeel worden vastgelegd. Het vorige voorbeeld zou als volgt kunnen worden gestructureerd:

```
<?xml version="1.0"?>
<!DOCTYPE tekst [
  <!ELEMENT tekst (paragraaf+) >
  <!ELEMENT paragraaf (#PCDATA | titel)*>
  <!ATTLIST paragraaf nummer CDATA #IMPLIED>
  <!ELEMENT titel (#PCDATA)>
  <!ENTITY eacute "&#x00E9;" >
]>
<tekst>
  <paragraaf nummer="1">Paragraaf &eacute;&eacute;n van
    <titel>De trein der traagheid</titel>.
  </paragraaf>
  <paragraaf nummer="2">Paragraaf twee van
    <titel>De trein der traagheid</titel>.
  </paragraaf>
</tekst>
```

Zonder in detail te treden ⁴, biedt dit voorbeeld een illustratie van de structuur die wordt voorgeschreven door een Document Type Definition (DTD): een document moet bestaan uit een **<tekst>** element dat een **nummer** attribuut kan hebben en één of meer **<paragraaf>** elementen moet bevatten. Die kunnen op hun beurt platte tekst en **<titel>** elementen bevatten,

.....

⁵ *A gentle introduction to XML*: <http://www.tei-c.org/P4X/SG.html>

⁴ Voor een uitstekende inleiding tot XML en DTD's wordt verwezen naar *A gentle introduction to XML*⁵.

en die laatste bestaan uit platte tekst. XML voorziet in mechanismen om niet-gestandaardiseerde karakters op een neutrale manier te representeren: *character entities*. Omdat de representatie van het karakter *é* niet gestandaardiseerd is, wordt daarvoor een beroep gedaan op een ondubbelzinnige representatie die in de Unicode standaard⁶ is vastgelegd (in dit geval `#x00E9`). Voor die abstracte code kan dan in de DTD een meer inzichtelijke naam worden gedefinieerd (*eacute*), waarmee in het document naar die definitie wordt verwezen met zogenaamde *entity references*. Die beginnen met een ampersand (&) karakter en eindigen met een puntkomma (;), waartussen de naam van de entiteit wordt gegeven. In dit voorbeeld wordt dus vermeden dat de niet-gestandaardiseerde representatie *é* voor het karakter *kleine e met stijgend accent* problemen kan opleveren in computersystemen die een ander systeem gebruiken om dat karakter te representeren. Daarom wordt het gerepresenteerd met de eenduidige verwijzing naar de Unicode definitie van dat karakter (`#x00E9`), met behulp van een inzichtelijker naam voor die representatie (*eacute*). Programma's die het document verwerken vervangen dan het woord `één` door `één` en kunnen zo het karakter *e met stijgend accent* juist interpreteren.

Ook op het vlak van documentdefinities zijn internationale standaarden ontstaan. De belangrijkste inspanning op het gebied van tekstopmaak voor humane wetenschappen werd geleverd door het Text Encoding Initiative (TEI)⁷, een onderzoeksproject dat werd opgericht in 1987 en sinds 2000 is ondergebracht in een consortium. Het heeft als opzet een interdisciplinaire en internationale standaard te ontwikkelen voor de representatie van allerlei types van teksten voor online onderzoek en onderwijs. Sinds 1990 zijn verschillende generaties van TEI codeerschema's en bijhorende richtlijnen gepubliceerd. De laatste editie dateert van 2002, en is gedocumenteerd in *Guidelines for Electronic Text Encoding and Interchange (TEI P4)*⁸. Daarin worden enkele honderden structuurelementen gedefinieerd waarmee opmaakmodellen kunnen worden samengesteld voor zeer uiteenlopende tekstgenres zoals proza, poëzie, drama,... TEI ontwikkelde zelf een voorbeeld-opmaakmodel, namelijk *TEI Lite: An Introduction to Text Encoding for Interchange*⁹. Hoewel dat slechts 144 elementen bevat, bleek dat voor zeer veel projecten al voldoende, zodat het uitgroeide tot een van de meest gebruikte modellen voor de opmaak van teksten. Verder werd op Europees niveau door de Expert Advisory Group on Language Engineering Standards (EAGLES)¹⁰ een standaard ontwikkeld voor de opmaak van tekstcorpora. Deze Corpus Encoding Standard¹¹ is gebaseerd op het TEI model en specifiek bedoeld voor grootschalige tekstcorpora met vooral taaltechnologische toepassingen. Binnen deze toepassings sfeer werden ook inspanningen geleverd tot inhoudelijke standaardisering van linguïstische codes: de belangrijkste zijn de *EAGLES Recommendations for the Morphosyntactic Annotation of Corpora*¹². Die bevatten voorstellen voor een aantal onderscheidingscategorieën en mogelijke waarden die relevant zijn voor de morfosyntactische analyse van Europese talen.

.....

6 Unicode standaard: <http://www.unicode.org/standard/standard.html>

7 Text Encoding Initiative (TEI): <http://www.tei-c.org/>

8 *Guidelines for Electronic Text Encoding and Interchange (TEI P4)*: <http://www.tei-c.org/P4X/>

9 *TEI Lite: An Introduction to Text Encoding for Interchange*: <http://www.tei-c.org/Lite/>

10 Expert Advisory Group on Language Engineering Standards (EAGLES):
<http://www.ilc.cnr.it/EAGLES96/home.html>

11 Corpus Encoding Standard: <http://www.cs.vassar.edu/CES/>

12 *EAGLES Recommendations for the Morphosyntactic Annotation of Corpora*:
<http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>

Voor de transcriptie en linguïstische codering van de veertiende-eeuwse oorkonden in het corpus C14NL werd aansluiting bij deze standaarden onderzocht. Hoewel de CES-aanbevelingen interessant leken omwille van hun specifieke oriëntatie op linguïstische codering van corpora, bleek de centraliteit van zogenaamde *standoff markup* in deze standaard een hinderpaal. In dat model wordt een onveranderlijke referentietekst voorgesteld, waaraan door middel van verwijzingen in externe bestanden de verschillende annotatielagen virtueel worden toegevoegd. Dit zorgt echter voor een veel complexere verwerking en onderhoud van corpusbestanden, waarvoor overigens nog steeds geen algemeen verspreide software bestaat. Voor de ontwikkeling van het corpus C14NL zijn geen middelen voorzien voor interne software-ontwikkeling. Omdat bovendien de annotatielagen beperkt zijn tot lemma- en woordsoortcodes, werd uiteindelijk geopteerd voor aansluiting bij de algemenere TEI-richtlijnen, die vooral gericht zijn op representatie van documenten met ingebede codering. Ook die voorzien middelen voor codering van linguïstische informatie in de teksten zelf, waarvan het gebruik in het British National Corpus (BNC)¹³ een goed voorbeeld kan bieden. Voor het corpus C14NL werd –op basis van de TEI Lite DTD– een projectspecifieke DTD afgeleid met behulp van de interactieve TEI Pizza Chef¹⁴ webpagina op de TEI website. De CAT DTD¹⁵ bevat naast de elementen van TEI Lite ook TEI-conforme voorzieningen voor

- manuscriptbeschrijving: gebaseerd op de voorgestelde *module for manuscript description* voor de nieuwe release van de TEI Guidelines (P5, versie 0.1.10)
- transcriptie van de primaire bron: gebaseerd op de TEI *additional tagset for transcription of primary sources*
- linguïstische codering: gebaseerd op de TEI *additional tagset for simple analysis*

Dit levert een DTD op met 231 elementen, waarvan er in de huidige collectie teksten 70 voorkomen. In de volgende delen van dit document wordt toegelicht hoe deze elementen gebruikt zijn in de opmaak van het corpus C14NL. Daarbij moet worden opgemerkt dat het expliciet niet de bedoeling is om een exhaustieve referentiedocumentatie van de volledige DTD te bieden. Daarvoor wordt verwezen naar de TEI documentatie zelf. Wel wordt de nadruk gelegd op het gebruik van de meest voorkomende elementen in het corpus.

2. Algemene tekststructuur

Elke tekst van het corpus is C14NL getranscribeerd in een afzonderlijk elektronisch document. Dat is opgevat als een zelfbeschrijvende eenheid, met naast de tekst zelf ook uitgebreide voorzieningen voor meta-informatie over het document. Elke tekst wordt bevat in een <TEI.2> element, met daarbinnen minimaal een header gedeelte met meta-informatie in

.....

13 British National Corpus (BNC): <http://www.natcorp.ox.ac.uk/>

14 TEI Pizza Chef: <http://www.tei-c.org/pizza.html>

15 CAT DTD: [CAT.dtd](#)

een **<teiHeader>** element, en een tekstgedeelte in een **<text>** element:

```
<TEI.2 id="[oorkondenummer]">
  <teiHeader>
    <fileDesc>...</fileDesc>
    <encodingDesc>...</encodingDesc>
    <revisionDesc>...</revisionDesc>
  </teiHeader>
  <text>
    <front>...</front>
    <body>...</body>
    <back>...</back>
  </text>
</TEI.2>
```

Zoals voorgaand voorbeeld illustreert, krijgt elke tekst in het corpus een unieke identificatiecode voor het **<TEI.2>** element. Daarvoor wordt het ‘oorkondenummer’ gebruikt als waarde voor het **id** attribuut. Het oorkondenummer bestaat uit:

- een plaatsaanduiding in de vorm van een zogenaamde Kloeke-code¹⁶, gevolgd door
- drie cijfers die het jaartal aanduiden (zonder 1000-tal), en
- twee cijfers die het volgnummer van het betreffende document binnen genoemde plaats en jaar aanduiden.

Een tekst die begint met het element **<TEI.2 id="N072p32901">**, kan zo geïdentificeerd worden als de eerste oorkonde (‘01’) in de verzameling die geschreven is in het jaar 1329 (‘329’) en afkomstig uit Ieper in West-Vlaanderen (‘N072p’).

Het **id** attribuut is slechts één van de globale attributen, die kunnen voorkomen op alle elementen. Volgende globale attributen worden gedefinieerd door de DTD:

- **id**: een unieke identificatiecode voor het element. De waarde moet een code zijn die uniek is in het document, begint met een letter en geen spaties mag bevatten.
- **n**: een naam of nummer voor het element. De waarde hoeft niet uniek te zijn in het document.
- **lang**: een aanduiding van de taal voor het betreffende element. De code moet overeenstemmen met een identificatiecode voor een taal die wordt gegeven in een **<language>** element in de header.
- **rend**: een aanduiding van specifieke aspecten van de fysieke realisatie van de inhoud van het element in de brontekst.
- **ana**: een aanduiding van (een) bepaalde interpretatie(s) waaraan dit element wordt verbonden. De waarde moet overeenstemmen met een spatie-begrensde lijst van identificatiecodes van andere elementen in het document.
- **next**: een verwijzing naar een volgend element dat een eenheid vormt met het betreffende element. De waarde moet overeenstemmen met de identificatiecode van een ander element in het document.

.....
16 Kloeke-code: <http://www.meertens.knaw.nl/projecten/mand/ECARTkartografie.html>

- **prev**: een verwijzing naar een vorig element dat een eenheid vormt met het betreffende element. De waarde moet overeenstemmen met de identificatiecode van een ander element in het document.
- **corresp**: een verwijzing naar (een) element(en) waarmee het betreffende element correspondeert. De waarde moet overeenstemmen met een spatie-begrensde lijst van identificatiecodes van andere elementen in het document.

Deze documentatie beschrijft de onderdelen van de header in [3. Het header gedeelte \(p.6\)](#). De gebruikte tekstelementen komen aan bod in [4. Het tekstgedeelte \(p.14\)](#).

3. Het header gedeelte

Binnen het **<teiHeader>** element wordt voor elke tekst meta-informatie gegeven. Dat gebeurt in de volgende elementen:

- **<fileDesc>**: volledige bibliografische beschrijving van het elektronische document.
- **<encodingDesc>**: beschrijving van de relatie tussen de elektronische tekst en de brontekst waarvan die is afgeleid.
- **<profileDesc>**: beschrijving van niet-bibliografische (contextuele) aspecten van de tekst.
- **<revisionDesc>**: beschrijving van de wijzigingsgeschiedenis van het elektronische document.

3.1. Bibliografische beschrijving: <fileDesc>

De beschrijving van het elektronische document in **<fileDesc>** bestaat uit:

- **<titleStmt>**: groepeert elementen voor de beschrijving van de titel van een werk en verantwoordelijkheid voor de totstandkoming.
- **<publicationStmt>**: groepeert informatie over de publicatie of distributie van de tekst.
- **<sourceDesc>**: bibliografische beschrijving van de brontekst waarvan een elektronische tekst is afgeleid.

Elke tekst binnen de tekstenverzameling bevat een aantal vaste gegevens voor de titelbeschrijving. Als titel voor het document wordt het oorkondennummer herhaald binnen een **<title>** element. Vervolgens wordt ruimte voorzien voor identificatie van de auteur van de oorspronkelijke oorkonde in **<author>**. In de huidige tekstenverzameling ontbreekt die identificatie, zodat geen gegevens kunnen worden ingevuld. Daarna wordt de editeur geïdentificeerd in een **<editor>** element. Daarvoor wordt in een **id** attribuut een identificatiecode bepaald waarnaar later kan worden verwezen om bijvoorbeeld de verantwoordelijkheid voor oplossingen van afkortingen aan te geven. Als regel worden hiervoor de initialen van de editeur gebruikt. Andere verantwoordelijken voor de

totstandkoming van de elektronische tekst worden geïdentificeerd in één of meer **<respStmt>** element(en). Daarbinnen geeft **<name>** aan wie verantwoordelijk is voor een aspect van de totstandkoming dat in **<resp>** wordt omschreven. Ook hier worden de initialen van de naam van de verantwoordelijke gebruikt als identificatiecode voor het **id** attribuut van het **<name>** element. Een volgend element is **<principal>**, waarin binnen een **<name>** element de hoofdonderzoeker wordt vermeld onder wiens verantwoordelijkheid de elektronische tekst tot stand kwam. Ten slotte bevat het **<funder>** element gegevens over de financierende instelling voor de aanmaak van de elektronische tekst, met de naam in een **<name>** element en het adres in een **<address>** element. Volgend voorbeeld illustreert de titelgegevens voor een tekst uit het corpus:

```
<titleStmt>
<title>N072p32901</title>
<author>
<name/>
</author>
<editor>
<name id="CDW">Chris De Wulf</name>
</editor>
<respStmt>
<name id="JD">Janneke Diepeveen</name>
<resp>semi-automatische linguïstische codering</resp>
</respStmt>
<respStmt>
<name id="RvdB">Ron Van den Branden</name>
<resp>conversie naar TEI-conform formaat</resp>
</respStmt>
<principal>
<name>Georges De Schutter</name>
</principal>
<funder>
<name>Koninklijke Academie voor Nederlandse Taal- en Letterkunde
(KANTL)</name>
<address>
<addrLine>Koningstraat 18</addrLine>
<addrLine>b-9000 Gent</addrLine>
<addrLine>België</addrLine>
</address>
</funder>
</titleStmt>
```

Na de titelbeschrijving worden details gegeven over de publicatie van de tekst in **<publicationStmt>**. Die bestaan uit een identificatie van de uitgever in **<publisher>**, de plaats van publicatie in **<pubPlace>**, de datum van publicatie in **<date>** en informatie over de beschikbaarheid in **<availability>**. Dat laatste gebeurt door een copyright-vermelding in een paragraaf (**<p>**) en een van de mogelijke waarden voor het **status** attribuut: *restricted*, *free* of *unknown*¹⁷:

```
<publicationStmt>
<publisher>Koninklijke Academie voor Nederlandse Taal- en Letterkunde
(KANTL)</publisher>
```

.....

17 Hierbij moet worden opgemerkt dat de publicatiegegevens voor de huidige tekstenverzameling eerder tentatief en voorlopig zijn zolang het corpus niet opgeleverd is. Daarom is ook geopteerd voor de waarde *restricted* voor het **status** attribuut van **<availability>**.

```

<pubPlace>Gent</pubPlace>
<date value="2004">2004</date>
<availability status="restricted">
  <p>&copy; Copyright 2004, KANTL</p>
</availability>
</publicationStmnt>

```

De brontekst wordt gedetailleerd beschreven in **<sourceDesc>**. Hier is het originele TEI element aangevuld met het element **<msDescription>**. In de voorstellen voor de nieuwe versie van TEI (P5) wordt dit element specifiek geïntroduceerd voor de beschrijving van manuscripten. Voor een volledige beschrijving van dit element wordt verwezen naar hoofdstuk 13 uit de documentatie van TEI P5, versie 0.1.10 die op de TEI Sourceforge project pagina's¹⁸ te vinden is. Binnen de manuscriptbeschrijving is plaats voor:

- **<msIdentifier>**: beschrijving van de lokalisatie van het manuscript in de bewaarplaats.
- **<head>**: bevat een hoofding die op het manuscript aanwezig is.
- **<msContents>**: beschrijving van de intellectuele inhoud van het manuscript.
- **<physDesc>**: beschrijving van fysieke aspecten van het manuscript.
- **<history>**: beschrijving van historische aspecten van het manuscript.
- **<additional>**: beschrijving van bijkomende informatie in verband met de moderne bibliografie, huidige bewaaromstandigheden of andere representaties van een manuscript.
- **<msPart>**: beschrijving van een deel van een samengesteld manuscript.

Voor de teksten in het corpus is momenteel voornamelijk bewarings- en ontstaansinformatie beschikbaar. Daarom wordt binnen **<msDescription>** volstaan met de beschrijving van de lokalisatie en ontstaansgeschiedenis. In **<msIdentifier>** wordt een hiërarchische locatieladder gegeven, gaande van het land waar het manuscript bewaard wordt in **<country>**, de stad of gemeente in **<settlement>**, tot de archiefinstelling in **<repository>**. Vervolgens wordt voor het manuscript beschreven tot welke collectie het behoort in **<collection>**, en hoe het daarbinnen wordt geïdentificeerd in **<idno>**. De ontstaansgeschiedenis wordt binnen het element **<history>** beschreven in **<origin>**. Deze beschrijving is gebaseerd op het oorkondennummer waarmee het document wordt geïdentificeerd. Het **<origin>** element krijgt een **n** attribuut, met als waarde het volgnummer dat voor de betreffende tekst werd bepaald binnen de specifieke ontstaanscontext. Die ontstaanscontext wordt beschreven als paragraaftekst binnen een **<p>** element. Daarbinnen wordt **<placeName>** gebruikt voor de plaatsbeschrijving, en **<origDate>** voor de ontstaansdatum. De plaatsbeschrijving binnen **<placeName>** bestaat uit een **<origPlace>** element dat de formele Kloeke-code bevat, die verder wordt gesitueerd in een **<settlement>** element waarin de ontstaansplaats wordt genoemd, en een **<region>** element dat de bovenliggende regio noemt. Voor de datering in **<origDate>** wordt het jaartal uit het oorkondennummer vermeerderd met 1000, om het jaartal

.....

18 TEI Sourceforge project pagina's: <http://sourceforge.net/projects/tei/>

in vier cijfers weer te geven. Omdat in de transcripties van de corpusteksten de oorkondenummers kunnen worden gevolgd door extra informatie, wordt die in **<additional>** weergegeven. Daarbinnen wordt het **<adminInfo>** element gebruikt, dat bijkomende informatie kan bevatten over huidige bewaaromstandigheden en beschikbaarheid, en over de manuscriptbeschrijving zelf. Het laat ook aanvullende noten toe in **<note>** elementen. Dit werd de meest geschikte plaats geacht voor de extra informatie bij de corpusteksten. Verder werd een leeg **<surrogates>** element voorzien voor de beschrijving van eventuele digitale of fotografische representaties van het manuscript. Binnen het corpusproject wordt een deel van de manuscripten gefotografeerd. Als die later worden beschreven, kan dit in een **<bibl>** element voor bibliografische beschrijving in een **<p>** element binnen **<surrogates>**. Volgend voorbeeld illustreert de beschrijving van een (fictieve) corpustekst in **<sourceDesc>**:

```
<sourceDesc>
  <msDescription>
    <msIdentifier>
      <country>Belgi&euml;</country>
      <settlement>Brugge</settlement>
      <repository>Rijksarchief Brugge</repository>
      <collection>Oorkonden met Blauwe Nummers</collection>
      <idno>3914r</idno>
    </msIdentifier>
    <history>
      <origin n="01">
        <p>
          <placeName>
            <origPlace>N072p</origPlace>
            <settlement>Ieper</settlement>
            <region>West-Vlaanderen</region>
          </placeName>
          <origDate>1329</origDate>
        </p>
      </origin>
    </history>
    <additional>
      <adminInfo>
        <note>o:N072p32901.Ieper.COO-ASW-002.30januari1329ns</note>
      </adminInfo>
      <surrogates>
        <!--beschrijving van mogelijke andere representaties (vb. foto)-->
        <p>
          <bibl/>
        </p>
      </surrogates>
    </additional>
  </msDescription>
</sourceDesc>
```

3.2. Beschrijving van de codering: **<encodingDesc>**

De methodes en editieprincipes die werden gevolgd voor de transcriptie en codering van de tekst worden beschreven in **<encodingDesc>**. Deze beschrijving bevat de volgende onderdelen:

- **<projectDesc>**: beschrijving van projectcontext waarbinnen de elektronische tekst tot stand kwam.

- **<samplingDesc>**: beschrijving van de procedures voor selectie van tekstmateriaal in de elektronische tekst.
- **<editorialDesc>**: beschrijving van de gevolgde editiepraktijk bij de elektronische codering van de tekst.
- **<p>**: paragraaf met de beschrijving van de morfologische tagset.

In **<projectDesc>** worden details gegeven over de projectcontext waarin de tekst tot stand kwam. Dat gebeurt als een losse beschrijving in paragrafen (**<p>** elementen). Volgend voorbeeld illustreert de projectbeschrijving voor de Vlaamse toevoegingen aan het corpus C14NL:

```
<projectDesc>
<p>Deze tekst werd verzameld in het kader van het VNC-project
<title>Project: Databank van veertiende-eeuwse niet-literaire Nederlandse
teksten. Opbouw en linguïstisch onderzoek</title>. Dit project liep van
01/01/2004 tot 31/12/2007 en had als doel om lacunes aan te vullen van een
bestaand corpus van veertiende-eeuwse teksten, dat verzameld werd door Van
Mulder en Van Reenen.</p>
<p>Dit deel van het VNC-project werd gefinancierd via FWO-Vlaanderen, onder
kredietnummer 33G029504.</p>
</projectDesc>
```

De procedures die werden gevolgd voor selectie van tekstmateriaal worden beschreven in **<samplingDecl>**. Dat gebeurt als paragraaftekst in **<p>** elementen. Volgend voorbeeld illustreert dit voor de Vlaamse aanvullingen:

```
<samplingDecl>
<p>Integrale tekst</p>
</samplingDecl>
```

Het element **<editorialDecl>** bevat een gedetailleerde beschrijving van de gevolgde codeerpraktijk. In **<correction>** wordt aangegeven of de editor de tekst heeft geëmendeerd, en in welke vorm dat gebeurde. Principes voor normalisatie van het tekstmateriaal worden beschreven in **<normalization>**. Het element **<quotation>** bevat details over de behandeling van aanhalingen en quotaties in de teksten. In **<hyphenation>** wordt de behandeling van geregeleindes en splitsingen verklaard. Niveaus waarop de tekst werd gesegmenteerd worden aangegeven in **<segmentation>**. Bijkomende analyses en interpretatieve informatie die aan het tekstmateriaal werden toegevoegd, worden gekarakteriseerd in het element **<interpretation>**. Volgend voorbeeld illustreert het **<encodingDesc>** element voor een tekst met morfologische annotatie:

```
<editorialDecl>
<correction>
<p>De tekst is letterlijk getranscribeerd. Waar nodig werden mogelijke
fouten
gesignaleerd met het [sic] element, met een suggestie voor verbetering in
het
"corr" attribuut. Ontbrekende tekst werd aangevuld met [supplied],
onduidelijke tekst signaleerd via [unclear]. Afkortingen zijn
```

```

gesignaleerd
  met het [abbr] element, waar mogelijk wordt een oplossing gesuggereerd in
  het
  "expan" attribuut.</p>
</correction>
<normalization>
  <p>De tekst is niet genormaliseerd en volgt zo nauwkeurig mogelijk de
  oorspronkelijke schrijfwijze van de brontekst. Normalisering wordt
  opgevangen
  in de lemmatisering.</p>
</normalization>
<quotation>
  <p>Quotaties worden niet als dusdanig aangeduid; aanhalingstekens en
  apostrofes worden overgenomen zoals ze voorkomen in de brontekst en
  weergegeven als andere leestekens.</p>
</quotation>
<hyphenation>
  <p>Regeleindes werden overgenomen zoals ze voorkomen in de brontekst en
  weergegeven met het [lb/] element. Waar splitsingstekens voorkomen in de
  brontekst, zijn die overgenomen.</p>
</hyphenation>
<segmentation>
  <p>De tekst is gesegmenteerd tot op woordniveau, met elk woord binnen een
  [w]
  element.</p>
</segmentation>
<interpretation>
  <p>De tekst is verrijkt met linguïstische analyses op woordniveau. Elk woord
  is gecodeerd in een [w] element, met codes voor lemmatisering in een
  "lemma"
  attribuut en woordsoort in een "ana" attribuut.</p>
</interpretation>
</editorialDecl>

```

Een laatste element van de beschrijving van de codering wordt gevormd door een verwijzing naar de morfologische tagset die in de teksten wordt gebruikt voor de linguïstische analyse. Dat gebeurt in een paragraaf (<p>), waarbinnen een element <interpGrp> voorkomt. Daarin wordt met de waarde *morfologische tagset* voor het **type** attribuut aangegeven dat dit element een reeks interpretaties bevat die betrekking hebben op de morfologische analyse. Vervolgens wordt elke morfologische code gedefinieerd in een leeg <interp/> element¹⁹. In het attribuut **id** wordt de formele code gegeven waarnaar wordt verwezen bij de morfologische analyse van de woorden in de tekst (door middel van het **ana** attribuut, zie [4.4. Linguïstische analyse \(p. ?\)](#)). In het attribuut **value** wordt omschreven welk fenomeen die code aanduidt. Om te vermijden dat heel de morfologische tagset manueel moet worden ingevoegd bij elk bestand, is die opgeslagen in een afzonderlijk bestand, *interp.ent*. Dat bestand is gedeclareerd als een entity in de *Document Type Declaration* van elke tekst, zodat een verwijzing via een *entity reference* volstaat voor een logische invoeging van deze definities in het document. Volgend voorbeeld illustreert de fysieke constructie:

```
<?xml version="1.0" encoding="iso-8859-1"?>
```

.....

19 Het <interp/> element is een zogenaamd *leeg element*, dat zelf geen inhoud mag bevatten. De XML syntaxis bepaalt dat lege elementen verkort kunnen worden weergegeven met één enkele tag waarin de naam van het element gevolgd wordt door een slash karakter (/): de korte vorm <interp/> is dus gelijk aan de langere versie met start- en eindtag <interp></interp>.

```

<!DOCTYPE TEI.2 PUBLIC "-//CTB//DTD CAT (based on TEI)//NL" "CAT.dtd" [
<!ENTITY interp SYSTEM "interp.ent">
]>
<TEI.2 id="N072p40001">
  <teiHeader>
    <fileDesc>...</fileDesc>
    <encodingDesc>
      ...
    <p>
      &interp;
    </p>
  </encodingDesc>
  <revisionDesc>...</revisionDesc>
</teiHeader>
<text>
  <body>...</body>
</text>
</TEI.2>

```

Wanneer die entity reference wordt opgelost, wordt de definitie van de morfologische tagset toegevoegd in elk document:

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE TEI.2 PUBLIC "-//CTB//DTD CAT (based on TEI)//NL" "CAT.dtd" [
<!ENTITY interp SYSTEM "interp.ent">
]>
<TEI.2 id="N072p40001">
  <teiHeader>
    <fileDesc>...</fileDesc>
    <encodingDesc>
      ...
    <p>
      <interpGrp type="morfologische tagset">
        <interp id="c000" value="zelfstandig_naamwoord enkelvoud geen"/>
        <interp id="c001" value="zelfstandig_naamwoord enkelvoud e"/>
        <interp id="c002" value="zelfstandig_naamwoord enkelvoud s"/>
        <interp id="c003" value="zelfstandig_naamwoord enkelvoud t"/>
        <interp id="c004" value="zelfstandig_naamwoord enkelvoud n"/>
        <interp id="c005" value="zelfstandig_naamwoord enkelvoud r"/>
        <interp id="c006" value="zelfstandig_naamwoord enkelvoud a"/>
        <interp id="c007" value="zelfstandig_naamwoord enkelvoud nt"/>
        <interp id="c009" value="zelfstandig_naamwoord enkelvoud probleem"/>
        ...
      </interpGrp>
    </p>
  </encodingDesc>
  <revisionDesc>...</revisionDesc>
</teiHeader>
<text>
  <body>...</body>
</text>
</TEI.2>

```

Er werd voor dit mechanisme geopteerd omdat op deze manier enerzijds de morfologische codes in de tekst kunnen worden opgenomen in de validering van de teksten. De waarde voor het **ana** attribuut dat de morfologische analyse bevat, moet immers een identificatiecode zijn die voorkomt in het document. Anderzijds wordt zo de mogelijkheid voorzien voor meervoudige analyses, doordat het **ana** attribuut naar verschillende identificatiecodes kan verwijzen.

3.3. Beschrijving van contextuele aspecten voor de tekst: <profileDesc>

In <profileDesc> kunnen allerlei niet-bibliografische gegevens over de tekst worden beschreven, zoals voorkomende talen, deelnemers, tekstclassificatie, situationele context,... In het corpus wordt dit element gebruikt voor teksten waarin verschillende handschriften kunnen worden geïdentificeerd. Die worden beschreven in het element <handList>. Daarbinnen wordt elke hand die verschilt van die van de auteur gedefinieerd in een leeg <hand/> element. Voor elke hand worden een aantal kenmerken beschreven in volgende attributen:

- **id**: een unieke identificatiecode waarnaar kan worden verwezen in elementen die tekst bevatten die door deze hand geschreven is.
- **scribe**: een beschrijving van de hand. Bij gebrek aan gedetailleerder informatie wordt momenteel volstaan met de waarde *andere hand*.
- **first**: een aanduiding of de hand de eerste hand van het document is (*YES*), of niet (*NO*).
- **resp**: een aanduiding van de verantwoordelijkheid voor identificatie van de hand. Hiervoor wordt de identificatiecode gegeven van de transcribeur, editor of andere verantwoordelijken voor de totstandkoming van de elektronische tekst, die elders in de header werden geïdentificeerd.

Volgend voorbeeld illustreert het gebruik van <profileDesc> voor de definitie van een andere hand in de header:

```
<profileDesc>
  <handList>
    <hand id="hand2" scribe="andere hand" first="NO" resp="CDW"/>
  </handList>
</profileDesc>
```

3.4. Beschrijving van wijzigingen aan de elektronische tekst: <revisionDesc>

Het header element <revisionDesc> bevat een logbestand van alle wijzigingen die aan het elektronische document zijn aangebracht. Elke wijziging wordt beschreven in een <change> element. Daarbinnen wordt de datum beschreven in het <date> element, met een **value** attribuut waarin de datum wordt genoteerd in de vorm *jjjj-mm-dd*. Vervolgens worden in een <respStmt> element de naam (<name>) van de verantwoordelijke voor de wijziging, en diens verantwoordelijkheid (<resp>) beschreven. De wijziging zelf wordt beschreven in een <item> element. Als meer dan één wijziging werd aangebracht, worden die binnen dit <item> element opgesomd als een lijst, met een <list> element waarbinnen verschillende <item> elementen kunnen voorkomen. Volgend voorbeeld illustreert het <revisionDesc> element voor een corpustekst:

```
<revisionDesc>
  <change>
```

```

<date value="2005-07-13">2005-07-13</date>
<respStmt>
  <resp>markup</resp>
  <name>Ron Van den Branden</name>
</respStmt>
<item>
  <list>
    <item>toevoeging van @id aan [TEI.2]</item>
    <item>update van header: <list>
      <item>update van [fileDesc/publicationStmt]</item>
      <item>vervanging van [fileDesc/bibl] door [fileDesc/msDescription]</item>
      <item>toevoeging van [encodingDesc]</item></list>
    </item>
  </list>
</item>
</change>
<change>
<date value="2005-05-10">2005-05-10</date>
<respStmt>
  <resp>markup</resp>
  <name>Ron Van den Branden</name>
</respStmt>
<item>wijziging van interpunctie in [w] naar [c]</item>
</change>
<change>
<date value="2005-3-29">2005-3-29</date>
<respStmt>
  <resp>markup</resp>
  <name>Ron Van den Branden</name>
</respStmt>
<item>
  <list>
    <item>invoeging van definities voor woordsoortcodes</item>
    <item>omzetting van codes in @ana attribuuat naar geldige IDREFs</item>
  </list>
</item>
</change>
<change>
<date value="2004-02-05">2004-02-05</date>
<respStmt>
  <resp>markup</resp>
  <name>Ron Van den Branden</name>
</respStmt>
<item>omzetting van platte tekst naar XML</item>
</change>
<change>
<date value="2003">2003</date>
<respStmt>
  <resp>transcriptie</resp>
  <name>Chris de Wulf</name>
</respStmt>
<item>invoer, transcriptie in platte tekst</item>
</change>
</revisionDesc>

```

4. Het tekstgedeelte

De meta-informatie in de header wordt gevolgd door de transcriptie en linguïstische annotatie van de tekst in het **<text>** element. Daarvoor zijn de elementen van TEI Lite overgenomen, aangevuld met elementen uit enkele *additional TEI tagsets* voor transcriptie van primair

bronnenmateriaal en linguïstische segmentering. Hiermee *kunnen* minimaal alle fenomenen die gecodeerd worden in het deelcorpus *Van Reenen – Mulder* in TEI-conform XML formaat worden gerepresenteerd. Omdat ook het uiteindelijke gebruik van deze opmaakelementen in de XML teksten is gebaseerd op de codeerpraktijk in dat corpusdeel, komen uiteraard niet alle elementen uit de DTD voor in de teksten. In de volgende paragrafen wordt daarom vooral aandacht besteed aan het gebruik van opmaakelementen die in de teksten voorkomen. Voor een volledige lijst van elementen die in de teksten *kunnen* voorkomen wordt verwezen naar [Appendix A.2. Overzicht van de elementen die in de teksten kunnen voorkomen \(p.28\)](#) en de relevante TEI documentatie.

4.1. Tekststructuur

Binnen `<text>` wordt de eigenlijke inhoud weergegeven in een `<body>` element. Dit kan eventueel worden voorafgegaan door een voorgedeelte (titelpagina, afkortingenlijsten, inhoudsopgave,...) in een `<front>` element en gevolgd door afsluitende structuurelementen (appendices, indexen,...) in een `<back>` element. In de teksten van het corpus komen `<front>` en `<back>` echter niet voor. Binnen `<body>` is de structurele markup minimaal gehouden tot een enkele paragraaf in een `<p>` element, die de transcriptie van de hele oorkonde bevat. Deze beperking wordt gemotiveerd door het ontbreken van structurele segmentering in de bestaande elektronische transcripties van het corpus. Het onderscheiden van paragrafen in Middelnederlandse manuscripten is een zeer interpretatieve aangelegenheid, niet in het minst door het niet-gestandaardiseerd gebruik van leestekens. Binnen de paragrafen worden wel regelgrenzen aangeduid met het lege element `<lb/>` (line break) en paginagrenzen met het lege element `<pb/>` (page break). De belangrijkste grote structuurelementen binnen `<text>` worden geïllustreerd in de transcriptie van volgend fragment:

```

...
kenlijc sij allen lieden dat ic broeder michiel calconen prior der predicaren in
ypre hebbe ontfanghen van der stede van ypre viue en(de) twintich ponden
parisisen van
den paymente van medewintere bi causen en(de) reco(m)pensacien van der assise
Jn tekene
der warede so hebbe ic broeder michiel voors dese lettere gheseghelt metten
seghele
van miere officie Ghedaen int Jaer ons heren dusentich drie hondert en(de)
neghen
en(de) neghentich den vichtiensten van lauwe maent
...

```

```

<text>
<body>
<p>
kenlijc sij allen lieden dat ic broeder michiel calconen prior der
predicaren in <lb/>
ypre hebbe ontfanghen van der stede van ypre viue <abbr expan="ende"
resp="CDW">en</abbr> twintich ponden parisisen van <lb/>
den paymente van medewintere bi causen <abbr expan="ende"
resp="CDW">en</abbr>
<abbr expan="recompensacien" resp="CDW">recopenciacien</abbr> van der assise
Jn tekene <lb/>

```

```

der warede so hebbe ic broeder michiel <abbr resp="CDW">voors</abbr> dese
lettere gheseghelt metten seghele <lb/>
van miere officie Ghedaen int Jaer ons heren dusentich drie hondert <abbr
expan="ende" resp="CDW">en</abbr> neghen <lb/>
<abbr expan="ende" resp="CDW">en</abbr> neghentich den vichtiensten van
lauwe
maent <lb/>
</p>
</body>
</text>

```

Bij de transcriptie en linguïstische annotatie wordt de tekst verrijkt met elementen voor representatie van tekstuele fenomenen en linguïstische analyse. Hoewel de reeds gecodeerde teksten in het deelcorpus *Van Reenen – Mulder* grotendeels automatisch kunnen worden geconverteerd naar volledig geannoteerde XML versies, is dit voor nieuwe aanvullingen een manueel proces dat gradueel verloopt. Teksten worden eerst getranscribeerd en later linguïstisch geannoteerd. Daarom worden in de volgende paragrafen eerst de elementen besproken voor de transcriptie van de teksten, en vervolgens die voor linguïstische analyse. De bijhorende voorbeelden (inclusief het voorgaande) zijn om redenen van leesbaarheid en duidelijkheid progressief opgevat. In de toelichting van elementen voor de transcriptie van de teksten wordt abstractie gemaakt van linguïstische elementen die waarschijnlijk in een latere fase worden toegevoegd. Voor meer omvattende voorbeelden wordt verwezen naar [4.4. Linguïstische analyse \(p.20\)](#).

4.2. Transcriptie van tekstfenomenen

De tekst van de manuscripten kan typografische afwijkingen bevatten. Die worden aangeduid in een **<hi>** element, met een **rend** attribuut waarmee de aard van de typografische afwijking wordt beschreven. Volgende waarden komen voor in de huidige verzameling corpusteksten:

- *supralineair*: boven de regel geschreven.

```
<hi rend="supralineair">paris</hi>
```

- *sublineair*: onder de regel geschreven.

```
<hi rend="sublineair">nuwpoerd</hi>
```

- *vlag*: vlagje aan het karakter.

```
int iaer on<hi rend="vlag">s</hi> heren
```

- *onvolledig*: karakter is niet volledig gerealiseerd.

```
dusag<hi rend="onvolledig">h</hi>e
```

- *nasaal*: karakter is gerealiseerd via een nasaalteken.

```
co<hi rend="nasaal">u</hi>stume
```

- *scheiding*: aan elkaar geschreven, gescheiden met rechte streep door zelfde hand.

```
<hi rend="scheiding">querema/mach</hi>
```

Op de manuscripten kunnen toevoegingen en verwijderingen voorkomen die gemaakt zijn door de auteur of een andere scribe. Die worden in de transcripties als dusdanig aangeduid. Toegevoegde tekst wordt opgenomen in een **<add>** element. Met behulp van het **place** attribuut kan worden aangegeven waar de toevoeging werd gemaakt. In de huidige tekstenverzameling worden daarvoor de waarden *supralineair* (boven de regel) en *oorspronkelijk* (op de oorspronkelijke plaats) gebruikt. In een **hand** attribuut kan worden aangegeven door wie de toevoeging werd gemaakt, via een verwijzing naar een identificatiecode voor een documentaire hand in de header (zie [3.3. Beschrijving van contextuele aspecten voor de tekst: <profileDesc> \(p. 13\)](#)). De verantwoordelijkheid voor de identificatie van de toevoeging gebeurt door een verwijzing in het **resp** attribuut. Volgend voorbeeld illustreert de aanduiding van een toevoeging door een andere hand:

```
<p>
...
Ende jhans huus en Erve was<lb/>
van <add hand="hand2" place="oorspronkelijk" resp="CDW">Ogi</add> jn dandre
side .
...
</p>
```

Verwijderingen worden aangeduid met het element ****, met een signalering van de verantwoordelijkheid voor de identificatie in een **resp** attribuut. Met het attribuut **rend** wordt aangegeven hoe die verwijdering werd gerealiseerd. In de huidige verzameling corpusteksten gaat het uitsluitend om doorhalingen, waarvoor de waarde *doorstreept* wordt gebruikt, zoals volgend voorbeeld illustreert:

```
<p>
...
de vors <del resp="CDW" rend="doorstreept">steuen</del>
<hi rend="supralineair">piet</hi> ons heeft wel en
soffisantelic betaelt /
...
</p>
```

Verder kunnen nog een aantal fysieke kenmerken van het manuscript gesignaleerd worden met de volgende elementen:

- **<addSpan/>**: leeg element dat het begin aangeeft van een langere toevoeging die structurele grenzen overschrijdt.
- **<delSpan/>**: leeg element dat het begin aangeeft van een langere verwijdering die structurele grenzen overschrijdt.
- **<handShift/>**: leeg element dat een overgang aangeeft naar een andere documentaire hand, of een andere inktsoort.
- **<damage>**: aanduiding van beschadiging van het manuscript.
- **<space/>**: leeg element dat opvallende open ruimte in het manuscript aanduidt.

In de huidige tekstcollectie komen deze elementen echter niet voor.

4.3. Editeursingrepen

Hoewel het zwaartepunt van de codering ligt op de linguïstische analyse van de teksten, worden in de transcriptiefase al fenomenen aangeduid die de latere analyse kunnen vereenvoudigen. Vaak vergt de transcriptie van veertiende-eeuwse manuscripten al heel wat interpretatie, die best expliciet als zodanig wordt gemarkeerd. De corpusteksten bevatten daarom een aantal TEI elementen voor de aanduiding van editeursingrepen. De verantwoordelijkheid van deze editeursingrepen wordt zo veel mogelijk geëxpliciteerd met behulp van een **resp** attribuut voor dergelijke elementen. Daar wordt dan verwezen naar de identificatiecode voor een editeur of andere verantwoordelijke voor de totstandkoming van de elektronische tekst in de header (zie [3.1. Bibliografische beschrijving: <fileDesc> \(p. 6\)](#)).

Middel nederlandse teksten bevatten heel wat afkortingen. Om latere morfologische analyse te kunnen vergemakkelijken worden ze als zodanig gesignaleerd en opgelost waar mogelijk. Dat gebeurt door ze te omsluiten in het element **<abbr>**. Wanneer de editeur een oplossing voorziet voor de afkorting wordt die gegeven als waarde voor een **expan** attribuut. Volgend voorbeeld illustreert respectievelijk een niet-opgeloste en opgeloste afkorting. Merk op hoe de verantwoordelijkheid voor de codering wordt aangeduid door in het **resp** attribuut te verwijzen naar *CDW*, de identificatiecode voor de editeur die in de header is gedefinieerd:

```
<p>
...
mochter die <abbr resp="CDW">vors</abbr> <abbr resp="CDW"
expan="joncvro">joncv</abbr> anes hant an slaen
...
</p>
```

Manifeste onjuistheden worden gesignaleerd in een **<sic>** element. Zoals bij afkortingen wordt de oorspronkelijke inhoud overgenomen binnen dit element, en de editeursingreep aangegeven in een attribuut. Wanneer een correctie voor de fout wordt aangegeven gebeurt dat in een **corr** attribuut, zoals in volgend voorbeeld:

```
<p>
...
den achtersten <sic corr="dach" resp="CDW">dat</sic> van oegstmaende
...
</p>
```

Bij de transcriptie van veertiende-eeuwse manuscripten kunnen heel wat onzekerheden en onduidelijkheden voorkomen, door bijvoorbeeld moeilijk leesbaar schrift, degradatie van het manuscript en dergelijke. Daarvoor bevat de TEI *additional tagset* voor transcriptie van primair bronnenmateriaal²⁰ een reeks voorzieningen. Wanneer een transcribeur niet zeker is van de transcriptie wordt dit gesignaleerd in een **<unclear>** element. In een **reason** attribuut kan worden omschreven waardoor de onduidelijkheid wordt veroorzaakt. In de huidige tekstverzameling komen de volgende waarden voor: *onduidelijk*, *beschadigd*, *afgeschraapt of beschadigd?*, *ontbreekt*, *afgescheurd stuk*, *waterschade* en *niet leesbaar*. Als de onduidelijke tekst onherkenbaar is wordt een *x* getranscribeerd voor elk onduidelijk karakter. Het volgende voorbeeld toont de codering van onduidelijke passages:

```
<p>
...
dat sake dat jan collekin <unclear reason="beschadigd"
resp="CDW">uo</unclear>ers ofte
...
liedekerke <unclear reason="onduidelijk" resp="CDW">x</unclear>elike
vermergriete pieters wijf
...
</p>
```

Wanneer de editeur voldoende aanwijzingen heeft om onduidelijke tekst te reconstrueren wordt die editeurs-aanvulling gemarkeerd door een **<supplied>** element. In een **reason** attribuut kan de reden voor de aanvulling worden aangegeven. Waarden die in de huidige tekstverzameling gebruikt worden zijn: *onduidelijk*, *beschadigd*, *waterschade*, *ontbreekt*, *plooi*, *onzeker* en *niet zichtbaar*. Volgend voorbeeld illustreert door de editeur aangevulde tekst:

```
<p>
...
scepenen zel<supplied reason="beschadigd" resp="CDW">en</supplied>
<supplied reason="beschadigd" resp="CDW">sii</supplied>n te gheroustberghe
...
</p>
```

Bijkomende verklaringen door de editeur kunnen worden gemaakt in het element **<note>**, net na het woord waarop ze van toepassing zijn. Hierbij moet worden opgemerkt dat dit element evenzeer gebruikt wordt voor de representatie van auteursnoten, maar dat die niet voorkomen in de huidige tekstverzameling. Het belangrijkste attribuut is **resp**, waarmee de

.....
20 TEI *additional tagset* voor transcriptie van primair bronnenmateriaal: <http://www.tei-c.org/P4X/PH.html>

verantwoordelijke voor de noot kan worden gesignaleerd en dus onderscheid kan worden gemaakt tussen auteurs- of editeursnoten. Omdat geen uitgebreide niet-linguïstische annotaties gemaakt worden in de teksten, wordt geen typologie aangebracht in de noten (dat is mogelijk met het **type** attribuut). Ook de plaatsbepaling in het **place** attribuut wordt niet nader gespecificeerd: standaard wordt de waarde *unspecified* aangenomen die wordt bepaald door de DTD. Volgend voorbeeld illustreert een verklarende editeursnoot:

```
<p>
...
<abbr resp="CDW">vors</abbr> heinricke yement onghebruic<note resp="CDW"
place="unspecified">diftong?</note> daede ofte doen wilde
...
</p>
```

4.4. Linguïstische analyse

De codering van de linguïstische analyses vormt het belangrijkste deel van de opmaak van het corpus C14NL. Daarvoor werden elementen uit de TEI *additional tagset* voor eenvoudige analytische mechanismen²¹ geselecteerd. Naast enkele algemene analytische elementen bevat die specifieke elementen voor de identificatie van linguïstische segmenten:

- **<s>**: een grammaticale zin.
- **<cl>**: een grammaticale bijzin.
- **<phr>**: een grammaticaal zinsdeel.
- **<w>**: een grammaticaal (niet noodzakelijk orthografisch) woord.
- **<m>**: een grammaticaal morfeem.
- **<c>**: een enkel karakter.

De linguïstische codering van het corpus C14NL is gebaseerd op die van het deelcorpus *Van Reenen – Mulder*. Dat bevat een segmentering op woordniveau, waarbij elk woord wordt voorzien van lemma- en woordsoortcodering. Voor een uitgebreide documentatie van de linguïstische principes die daarvoor werden gehanteerd, wordt verwezen naar een intern werkdokument. In deze documentatie wordt uiteengezet hoe de linguïstische analyse in de XML teksten wordt gerepresenteerd.

4.4.1. Algemeen

Aangezien uitsluitend op woordniveau is gesegmenteerd, worden voor de linguïstische analyse enkel de elementen **<w>** en **<c>** gebruikt voor respectievelijk grammaticale woorden en leestekens. De linguïstische analyse wordt uitgedrukt door volgende attributen van het **<w>** element:

- **lemma**: bevat het lemma voor het betreffende woord, als een losse string (zonder

.....

²¹ TEI *additional tagset* voor eenvoudige analytische mechanismen: <http://www.tei-c.org/P4X/AI.html>

verdere beperkingen).

- **ana**: bevat de morfologische code voor het betreffende woord, als een verwijzing naar één of meer codes voor de linguïstische categorieën die zijn geïdentificeerd in de header.

De morfologische codes die voorkomen in het **ana** attribuut bestaan uit 4 tekens en hebben de volgende opbouw:

- *c*: deze letter gaat elke code vooraf, om aan de eis te voldoen dat identificatiecodes in XML met een letter moeten beginnen.
- identificatie van de woordsoort, door één cijfer van 0 tot 9:
 - 0: zelfstandig naamwoord
 - 1: bijvoeglijk naamwoord
 - 2: werkwoord
 - 3: telwoord
 - 4: voornaamwoord
 - 5: bijwoord
 - 6: voornaamwoordelijk bijwoord
 - 7: voorzetsel
 - 8: voegwoord
 - 9: tussenwerpsel
- woordsoortspecifieke subclassificatie, door één cijfer van 0 tot 9:
 - Substantief (woordsoortcode *c0*):
 - 0: enkelvoud
 - 1: meervoud
 - 2: eigennaam
 - 9: probleem
 - Adjectief (woordsoortcode *c1*):
 - 0: adjectief
 - Werkwoord (woordsoortcode *c2*):
 - 0: persoonsvorm tegenwoordige tijd zelfstandig werkwoord
 - 1: persoonsvorm tegenwoordige tijd hulpwerkwoord / koppelwerkwoord
 - 2: persoonsvorm verleden tijd zelfstandig werkwoord
 - 3: persoonsvorm verleden tijd hulpwerkwoord / koppelwerkwoord

- 4: imperatief
- 5: infinitief
- 7: deelwoord
- 9: probleem

- Telwoord (woordsoortcode *c3*):
 - 0: hoofdtelwoord
 - 1: rangtelwoord
 - 2: onbepaald telwoord
 - 9: probleem

- Voornaamwoord (woordsoortcode *c4*):
 - 0: persoonlijk voornaamwoord
 - 1: aanwijzend voornaamwoord
 - 2: betrekkelijk voornaamwoord
 - 3: vragend voornaamwoord
 - 4: onbepaald voornaamwoord
 - 5: bezittelijk voornaamwoord
 - 6: reflexief voornaamwoord
 - 7: bepaald lidwoord
 - 8: onbepaald lidwoord
 - 9: probleem

- Bijwoord (woordsoortcode *c5*):
 - 0: gewoon bijwoord
 - 1: aanwijzend bijwoord
 - 2: betrekkelijk bijwoord
 - 3: vragend bijwoord
 - 4: onbepaald bijwoord
 - 5: hervattend bijwoord
 - 6: ontkennend bijwoord
 - 9: probleem

- Voornaamwoordelijk bijwoord (woordsoortcode *c6*):
 - 0: persoonlijk voornaamwoordelijk bijwoord

-
- *1*: aanwijzend voornaamwoordelijk bijwoord
 - *2*: betrekkelijk voornaamwoordelijk bijwoord
 - *3*: vragend voornaamwoordelijk bijwoord
 - *4*: onbepaald voornaamwoordelijk bijwoord
 - *9*: probleem

 - Voorzetsel (woordsoortcode *c7*):
 - *0*: voorzetsel

 - Voegwoord (woordsoortcode *c8*):
 - *0*: nevenschikkend voegwoord
 - *1*: onderschikkend voegwoord
 - *2*: voegwoord van vergelijking
 - *3*: *ne/en* in tweede deel van tweeledige zin (balansschikking)
 - *4*: *als(e)* in betekenis 'namelijk'
 - *9*: probleem

 - Tussenwerpsel (woordsoortcode *c9*):
 - *0*: Latijn / totaal probleem
 - *1*: tussenwerpsel

 - uitgangscodes: een cijfer van 0 tot 7, en 9. Voor persoonlijke (woordsoortcode *c40*) en reflexieve (woordsoortcode *c46*) voornaamwoorden codeert dit derde cijfer het getal:
 - *0*: geen uitgang
 - *1*: uitgang -e / eerste persoon enkelvoud voor persoonlijk en reflexief voornaamwoord
 - *2*: uitgang -s / tweede persoon enkelvoud voor persoonlijk en reflexief voornaamwoord
 - *3*: uitgang -t / derde persoon enkelvoud voor persoonlijk en reflexief voornaamwoord
 - *4*: uitgang -n / eerste persoon meervoud voor persoonlijk en reflexief voornaamwoord
 - *5*: uitgang -r / tweede persoon meervoud voor persoonlijk en reflexief voornaamwoord
 - *6*: uitgang -a / derde persoon meervoud voor persoonlijk en reflexief voornaamwoord

- 7: uitgang -nt
- 9: probleemgeval

In het originiele TEI model kan <w> uitsluitend tekst bevatten. Deze beperking bleek echter te stringent voor woorden die interne opmaak bevatten, zoals typografisch gemarkeerde letters (aangeduid met <hi>), of door de editeur toegevoegde letters (aangeduid met <supplied>), of voor gescheiden samengestelde woorden waarin een deel als afkorting is gemarkeerd (<abbr>). Daarom werd het TEI element <w> geherdefinieerd in de DTD voor het corpus C14NL, zodat het zelf nog andere elementen kan bevatten. Waar leestekens voorkomen in het origineel worden ze overgenomen en gecodeerd in het element <c>. Volgend voorbeeld illustreert hoe volgend fragment wordt gecodeerd:

```
...
e(nde) gherard
den parmentier scepen scepenen sforseits profts / orconden der jhan prochipape
van mo-
[r]enghe(m)
...
```

```
<p>
...
<w lemma="en" ana="c800">
  <abbr resp="CDW" expan="ende">e</abbr>
</w>
<w lemma="gerhard" ana="c020">gherard</w>
<lb/>
<w lemma="de parmentier" ana="c020">den parmentier</w>
<w lemma="scheiden" ana="c010">
  <del resp="CDW" rend="doorstreept">scepen</del>
</w>
<w lemma="scheiden" ana="c014">scepenen</w>
<w lemma="de" ana="c472" rend="space(0)">s</w>
<w lemma="voorgezegd" ana="c102">forseits</w>
<w lemma="proost" ana="c002">profts</w>
<c>/</c>
<w lemma="oorkonden" ana="c204">orconden</w>
<w lemma="de heer" ana="c470 c000">der</w>
<w lemma="johannes" ana="c020">
  <hi rend="supralineair">jhan</hi>
</w>
<w lemma="parochiepaap" ana="c001">prochipape</w>
<w lemma="van" ana="c700">van</w>
<w lemma="moregem" ana="c020">
  <abbr resp="CDW" expan="morengem">mo-  

  <supplied resp="CDW" reason="onduidelijk">r</supplied>enghe</abbr>
</w>
...
</p>
```

Voor inhoudelijke details over de gehanteerde principes bij de linguïstische analyse wordt verwezen naar de interne documentatie. Wel worden hier nog enkele formele bijzonderheden toegelicht voor codering van bijzondere gevallen. Die hebben te maken met segmentering van:

- samengevoegde vormen
- clitische vormen
- gescheiden vormen

4.4.2. Samengevoegde vormen

Middel nederlandse teksten bevatten heel wat woorden die toevallig aan mekaar geschreven zijn. Wanneer dit gebeurt zonder klankverlies, worden de vormen linguïstisch beschouwd als afzonderlijke zelfstandige woorden. Bij de linguïstische segmentering worden ze als dusdanig gecodeerd in afzonderlijke `<w>` elementen, elk met hun eigen analyse in **lemma** en **ana** attributen. Ook in de codering worden die `<w>` elementen niet gescheiden door een spatie. Omdat witruimte in XML teksten kwetsbaar kan zijn, werd besloten om dit aspect van de realisatie van deze woorden te representeren in een attribuut. Aangezien het om presentationele aspecten gaat, is hiervoor het globale **rend** attribuut gebruikt. Voor woorden of leestekens die niet gevolgd worden door een spatie, wordt in het **rend** attribuut de waarde *space(0)* bepaald. Volgend voorbeeld illustreert hoe de toevallige eenheid van twee woorden die in het manuscript voorkomen als *biden* er na linguïstische codering uit ziet:

```
<p>
...
<w lemma="bij" ana="c700" rend="space(0)">bi</w>
<w lemma="de" ana="c474">den</w>
...
</p>
```

4.4.3. Clitische vormen

Wanneer woordvormen samengevoegd voorkomen, waarbij klankverlies optreedt bij minstens één van de samenstellende delen, worden deze woorden beschouwd als *clitische vormen*. Omdat door de klankverandering geen duidelijke splitsingsgrenzen kunnen worden bepaald, worden ze niet verder gesegmenteerd. Ze worden dus gecodeerd als één `<w>` element, maar krijgen wel aparte waarden voor de **lemma** en **ana** attributen, gescheiden door spaties. Volgend voorbeeld bevat een illustratie van de codering van de clitische vorm die in het manuscript voorkomt als *vondemer*. Verder wordt het verschil duidelijk met toevallig aan mekaar geschreven vormen als *dieme*:

```
<p>
...
<w lemma="vinden men er" ana="c220 c440 c500">vondemer</w>
<w lemma="meer" ana="c320">mee</w>
<w lemma="schuld" ana="c001">schulde</w>
<w lemma="die" ana="c421" rend="space(0)">die</w>
<w lemma="men" ana="c440">
  <abbr resp="CDW" expan="men">me</abbr>
</w>
<w lemma="redelijk" ana="c500">redenlike</w>
```

```
<w lemma="togen" ana="c250">toeghen</w>
<w lemma="mogen" ana="c231">mochte</w>
...
</p>
```

4.4.4. Gescheiden vormen

Omgekeerd is het mogelijk dat woorden die afzonderlijk voorkomen in het manuscript als een linguïstische eenheid worden beschouwd, met één lemma en één woordsoortcode. Wanneer deze gescheiden delen van één woord continu voorkomen (enkel gescheiden door spaties), worden ze gecodeerd in één **<w>** element, met één waarde voor de **lemma** en **ana** attributen. Het volgende voorbeeld illustreert hoe de woorden *daer wtgaede* gesegmenteerd worden tot de woordvormen *daer wt* en *gaede*:

```
<p>
...
<w lemma="daaruit" ana="c610" rend="space(0)">daer wt</w>
<w lemma="gaan" ana="c271" rend="space(0)">
  <abbr resp="CDW" expan="gaende">gaede</abbr>
</w>
<c>.</c>
...
</p>
```

Wanneer delen van een woord discontinu voorkomen (gescheiden door andere woorden), is het onmogelijk om ze binnen één **<w>** element te coderen. Daarom wordt voorgesteld om de discontinue delen binnen aparte **<w>** elementen op te nemen, maar slechts het eerste deel een linguïstische codering te geven in **lemma** en **ana** attributen. Verder wordt gesuggereerd om de fragmentaire status van de delen aan te geven via het specifieke **part** attribuut voor linguïstische segmenten. Dat kan door volgende waarden:

- *I* (Initial): eerste fragment van een samenstelling
- *M* (Medial): mediaal fragment van een samenstelling
- *F* (Final): laatste fragment van een samenstelling

Bovendien wordt aan elk fragment een identificatiecode in het globale **id** attribuut gegeven. Om de uniciteit binnen het corpus te garanderen wordt voorgesteld om die te vormen door samenvoeging van het oorkondenummer, een punt en een rangnummer van het aantal discontinue fragmenten binnen die oorkonde. Van die identificatiecodes wordt dan gebruik gemaakt voor verwijzingen binnen die fragmenten, met behulp van de globale attributen voor het verbinden van elementen. Met het attribuut **next** wordt in niet-finale fragmenten verwezen naar de identificatiecode van het onmiddellijk volgende fragment. Met het attribuut **prev** wordt in niet-initiële fragmenten verwezen naar het onmiddellijk voorgaande fragment. Volgend hypothetisch voorbeeld toont hoe de vorm *wasser ... toe* zou kunnen worden gecodeerd²²:

.....

```
<p>
...
<w lemma="zijn ertoe" ana="c230 c600" id="0598r39501.1" next="0598r39501.2"
part="I">wasser</w>
...
<w id="0598r39501.2" prev="0598r39501.1" part="F">toe</w>
...
</p>
```

-
- 22 Bij de linguïstische codering van 157 corpusteksten aan de KANTL was dit aspect nog problematisch. Daarom zijn discontinue woorddelen vooralsnog niet gecodeerd; ze hebben geen **lemma** en geen **ana** attribuut. Het gaat om 399 gevallen of maximaal 200 linguïstische woorden, waarvan volgens het huidige voorstel dus een 100-tal nog een volledige linguïstische codering moeten krijgen.

Appendix A. Overzicht van attributen en elementen

Appendix A.1. Overzicht van globale attributen

ana	een aanduiding van (een) bepaalde interpretatie(s) waaraan dit element wordt verbonden. De waarde moet overeenstemmen met een spatie-begrensde lijst van identificatiecodes van andere elementen in het document.
corresp	een verwijzing naar (een) element(en) waarmee het betreffende element correspondeert. De waarde moet overeenstemmen met een spatie-begrensde lijst van identificatiecodes van andere elementen in het document.
id	een unieke identificatiecode voor het element. De waarde moet een code zijn die uniek is in het document, begint met een letter en geen spaties mag bevatten.
lang	een aanduiding van de taal voor het betreffende element. De code moet overeenstemmen met een identificatiecode voor een taal die wordt gegeven in een <language> element in de header.
n	een naam of nummer voor het element. De waarde hoeft niet uniek te zijn in het document.
next	een verwijzing naar een volgend element dat een eenheid vormt met het betreffende element. De waarde moet overeenstemmen met de identificatiecode van een ander element in het document.
prev	een verwijzing naar een vorig element dat een eenheid vormt met het betreffende element. De waarde moet overeenstemmen met de identificatiecode van een ander element in het document.
rend	een aanduiding van specifieke aspecten van de fysieke realisatie van de inhoud van het element in de brontekst.

Appendix A.2. Overzicht van de elementen die in de teksten kunnen voorkomen

<ab>	‘semantisch neutraal’ structurelement op het zelfde niveau als een paragraaf.
-------------------	---

<abbr>	afkorting waarvan de oplossing kan worden gegeven in het expan attribuut.
<accMat>	beschrijving van begeleitend of nauw gerelateerd materiaal bij manuscripten.
<acquisition>	beschrijving van aspecten van het verwervingsproces van een manuscript door de bewarende instelling.
<add>	toevoeging.
<additional>	beschrijving van bijkomende informatie in verband met de moderne bibliografie, huidige bewaaramstandigheden of andere representaties van een manuscript.
<additions>	beschrijving van significante toevoegingen in een manuscript, zoals marginalia of andere annotaties.
<address>	adresgegevens.
<addrLine>	een enkele lijn in een adres.
<addSpan/>	leeg element dat het begin aangeeft van een langere toevoeging die structurele grenzen overschrijdt.
<adminInfo>	beschrijving van gegevens in verband met de huidige bewaring en beschikbaarheid van een manuscript, en met de manuscriptbeschrijving zelf.
<altIdentifier>	beschrijving van een alternatieve of vroegere structurele code waarmee een manuscript werd of wordt geïdentificeerd.
<altName>	beschrijving van een alternatieve of vroegere naam waarmee een manuscript werd of wordt geïdentificeerd.
<anchor/>	leeg element waarmee een punt in het document kan worden geïdentificeerd voor verdere verwijzing.
<argument>	beschrijving (formeel of proza) van de onderwerpen die aan bod komen in een onderdeel van een tekst.
<author>	naam van een auteur van een werk.
<authority>	naam van een persoon of instelling verantwoordelijk voor de terbeschikkingstelling van het elektronische document, behalve een uitgever of distributeur.
<availability>	beschrijving van gegevens in verband met beschikbaarheid van een document.
<back>	afsluitende structurelementen (appendices, indexen,...) die volgen op en niet behoren tot de eigenlijke inhoud van een tekst.

<bibl>	bibliografische beschrijving, al dan niet gestructureerd.
<biblFull>	volledig gestructureerde bibliografische beschrijving.
<biblScope>	aanduiding van het bereik van een bibliografische beschrijving (pagina's, tekstdivisies,...).
<binding>	beschrijving van de band van een manuscript.
<bindingDesc>	beschrijving van aspecten van de band van een manuscript, als een reeks paragrafen of een reeks <binding> elementen.
<body>	eigenlijke inhoud van een tekst.
<byline>	aanduiding van de verantwoordelijkheid voor een werk, zoals die gegeven wordt op een titelpagina, voor- of achteraan een werk.
<c>	karakter.
<catDesc>	beschrijving van een categorie in een taxonomie of teksttypologie, in de vorm van een korte prozabeschrijving.
<category>	een individuele beschrijvende categorie in een gebruikersgedefinieerde taxonomie.
<catRef/>	leeg element waarmee verwezen wordt naar een of meer definities van categorieën in een taxonomie of teksttypologie.
<cb/>	leeg element voor de aanduiding van een kolomgrens in een tekst.
<cell>	cel in een tabel.
<change>	beschrijving van een wijziging die werd aangebracht aan een elektronisch document.
<cit>	citaat uit een externe bron, met een bibliografische verwijzing naar de bron.
<cl>	grammaticale bijzin.
<classCode>	bevat een classificatiecode voor de tekst in een standaard referentiesysteem.
<classDecl>	bevat een of meer taxonomieën van classificatiecodes die elders in de tekst worden gebruikt.
<closer>	groepeert afsluitende structuurelementen (slotformule, datum, ...) van tekstdivisies.
<collation>	beschrijving van de collatie van een manuscript.
<collection>	aanduiding van de collectie waartoe een manuscript behoort.

<colophon>	bevat de tekst van een colofon in een manuscript, met gegevens over datum, plaats, publicatie,...
<condition>	beschrijving van de fysieke conditie van een manuscript.
<corr>	editeurscorrectie van een manifeste fout, die kan worden weergegeven in het sic attribuut.
<correction>	beschrijving van de gevolgde praktijk in verband met correcties bij de transcriptie van een tekst.
<country>	de naam van een land.
<creation>	beschrijving van gegevens over het ontstaan van een tekst.
<custEvent>	beschrijving van één gebeurtenis in de bewaringsgeschiedenis van het manuscript.
<custodialHist>	beschrijving van aspecten van de bewaringsgeschiedenis van een manuscript.
<damage>	aanduiding van beschadiging van een manuscript.
<date>	een datum.
<dateline>	beschrijving van plaats, datum, tijdstip van ontstaan een het document, zoals die voorkomt in de opening of de afsluiting.
<decoDesc>	beschrijving van decoratieve elementen in een manuscript.
<decoNote>	beschrijving van een decoratief element of een homogene groep decoratieve elementen in een manuscript.
	een verwijdering.
<delSpan/>	leeg element dat het begin aangeeft van een langere verwijdering die structurele grenzen overschrijdt.
<depth>	aanduiding van de diepte van een manuscript.
<dimensions>	aanduiding van de afmetingen van een manuscript.
<distributor>	aanduiding van de verdeler van een tekst.
<div>	algemene tekstdivisie binnen het <front> , <body> of <back> gedeelte van een tekst, of een andere tekstdivisie.
<div0>	hiërarchisch hoogste tekstdivisie binnen een genummerde aanduiding van tekstdivisies in het <body> gedeelte van een tekst.
<div1>	op één na hoogste hiërarchische tekstdivisie binnen een genummerde aanduiding van tekstdivisies in <front> ,

	< body > of < back >, of een hogere < div0 > tekstdivisie.
< div2 >	hiërarchisch lagere tekstdivisie binnen < div1 >.
< div3 >	hiërarchisch lagere tekstdivisie binnen < div2 >.
< div4 >	hiërarchisch lagere tekstdivisie binnen < div3 >.
< div5 >	hiërarchisch lagere tekstdivisie binnen < div4 >.
< div6 >	hiërarchisch lagere tekstdivisie binnen < div5 >.
< div7 >	hiërarchisch lagere tekstdivisie binnen < div6 >.
< divGen >	aanduiding van de plaats waar een automatisch gegenereerde tekstdivisie in een document moet worden geplaatst.
< docAuthor >	de auteur van een tekst, zoals die vaak voorkomt op een titelpagina.
< docDate >	datering van een tekst, zoals die vaak voorkomt op een titelpagina.
< docEdition >	editiebeschrijving van een tekst, zoals die vaak voorkomt op een titelpagina.
< docImprint >	publicatiegegevens zoals die vaak voorkomen op een titelpagina.
< docTitle >	titel van een tekst zoals die vaak voorkomt op een titelpagina.
< edition >	beschrijving een editie van een tekst.
< editionStmnt >	beschrijving van gegevens in verband met de editie van een tekst.
< editor >	aanduiding van de verantwoordelijke (persoon of instelling) voor samenstelling, editie vertaling,... van een tekst.
< editorialDecl >	beschrijving van de gevolgde editiepraktijk bij de elektronische codering van een tekst.
< emph >	tekst die typografisch gemarkeerd is voor linguïstisch of retorisch effect.
< encodingDesc >	beschrijving van de relatie tussen een elektronische tekst en de brontekst waarvan die is afgeleid.
< epigraph >	citaat aan het begin van een tekstdivisie.
< explicit >	bevat de tekst van het 'explicit', de slotwoorden van een tekst(structuur).
< extent >	beschrijving van de omvang van een elektronische tekst op een drager.
< figDesc >	beschrijving van een figuur in een tekst.

<figure>	aanduiding van een plaats waar een figuur moet worden ingevoegd in de elektronische tekst.
<fileDesc>	volledige bibliografische beschrijving van het elektronische document.
<filiation>	beschrijving van informatie over de relatie van een manuscript met andere manuscripten, foto's en apografen.
<foliation>	beschrijving van het nummeringsschema voor de bladzijden in een bundel.
<foreign>	aanduiding van een woord in een andere taal dan de omringende tekst.
<formula>	bevat een wiskundige formule in een eigen notatie.
<front>	inleidende structurelementen (titelpagina, opdracht, voorwoord,...) die voorafgaan aan en niet behoren tot de eigenlijke inhoud van een tekst.
<funder>	aanduiding van de verantwoordelijke (persoon of instelling) voor de financiering van een project of tekst.
<fw>	lopende hoofding (koptekst, voettekst), motto, of soortgelijke tekst op de huidige pagina.
<gap/>	leeg element voor de aanduiding van weggelaten tekst bij de codering.
<gloss>	verklaring van een term.
<group>	groept een verzameling van enkelvoudige teksten (<text>).
<hand/>	leeg element voor de definitie van een documentaire hand.
<handDesc>	beschrijving van de verschillende schrijfstijlen in een manuscript.
<handList>	bevat een lijst van definities voor de verschillende documentaire handen in een document.
<handNote>	beschrijving van een specifieke schrijfstijl of handschrift in een manuscript.
<handShift/>	leeg element dat een overgang aangeeft naar een andere documentaire hand, of een andere inktsoort.
<head>	hoofding of titel van een tekststructuur.
<height>	aanduiding van de hoogte van een manuscript.
<heraldry>	wapenspreuk of formule als deel van een wapenschild op een manuscript.

<hi>	typografisch gemarkeerde tekst.
<history>	beschrijving van historische aspecten van een manuscript.
<hyphenation>	beschrijving van de gevolgde praktijk in verband met gereleides bij de transcriptie van een tekst.
<idno>	identificatiecode voor een bibliografisch object.
<imprint>	bevat gegevens over de publicatie of distributie van een bibliografisch object.
<incipit>	bevat de tekst van de aanhef, aan de hand waarvan de tekst kan worden geïdentificeerd.
<index/>	leeg element voor de lokalisatie van een tekstelement dat moet worden opgenomen in een index.
<institution>	de naam van een organisatie.
<interp/>	leeg element voor de definitie van een interpretatieve annotatie.
<interpGrp>	bevat een groep van definities voor interpretatieve annotaties.
<interpretation>	beschrijving van analytische of interpretatieve informatie die werd toegevoegd aan de transcriptie van een tekst.
<item>	een element in een lijst.
<join/>	leeg element dat fragmenten van tekstonderdelen reconstrueert tot een virtuele eenheid.
<joinGrp>	bevat een reeks <join/> elementen.
<keywords>	een lijst van termen of korte beschrijvingen van het onderwerp van een tekst.
<l>	een regel in een gedicht.
<label>	een hoofding voor een element in een lijst, of een term die in een verklarende woordenlijst wordt verklaard.
<language>	beschrijving van een taal die in een tekst wordt gebruikt.
<langUsage>	groepeert de beschrijving van verschillende talen die in een tekst worden gebruikt.
<layout>	beschrijving van aspecten van de lay-out van een manuscript.
<layoutDesc>	groepeert een reeks lay-out beschrijvingen voor een manuscript.
<lb/>	leeg element dat een gereleide aanduidt.

<lg>	een lijngroep in een gedicht.
<link/>	leeg element voor de definitie van een verband tussen tekstelementen.
<linkGrp>	groepeert een reeks definities van verbanden tussen tekstelementen.
<list>	een lijst.
<listBibl>	een lijst van bibliografische beschrijvingen.
<locus>	identificeert een locatie in een manuscript, meestal als een reeks verwijzingen naar folio's.
<m>	een grammaticaal morfeem.
<material>	beschrijving van het materiaal waarop een manuscript is geschreven.
<mentioned>	aanduiding van tekst die illustratief vermeld wordt.
<milestone/>	leeg element voor de aanduiding van een begrenzing tussen verschillende secties van een document die worden onderscheiden volgens een standaard referentiesysteem.
<msContents>	beschrijving van de intellectuele inhoud van een manuscript.
<msDescription>	beschrijving van een manuscript.
<msIdentifier>	beschrijving van de lokalisatie van een manuscript in de bewaarplaats.
<msItem>	beschrijving van een individueel aspect van de intellectuele inhoud van een manuscript.
<msPart>	beschrijving van een deel van een samengesteld manuscript.
<musicNotation>	beschrijving van een muzieknoot in een manuscript.
<name>	een eigennaam.
<normalization>	beschrijving van de gevolgde praktijk in verband met normalisatie bij de transcriptie van een tekst.
<note>	een noot of annotatie.
<notesStmt>	groepeert noten met bijkomende informatie over een tekst, die niet bevat is in andere delen van de bibliografische beschrijving.
<num>	een getal.
<objectDesc>	beschrijving van de fysieke samenstellende delen van het object dat wordt beschreven.
<opener>	groepeert inleidende structurelementen

	(openingsformule, motto, datum, ...) van tekstdivisies.
<orig>	aanduiding van de originele vorm van een woord, waarvoor een normalisatie kan worden bepaald als waarde voor het reg attribuut.
<origDate>	de datering van een manuscript.
<origin>	beschrijving van aspecten in verband met het ontstaan van een manuscript.
<origPlace>	de plaatsbepaling van een manuscript.
<p>	een paragraaf.
<pb/>	leeg element voor de aanduiding van een pagina-einde.
<phr>	een grammaticaal zinsdeel.
<physDesc>	beschrijving van fysieke aspecten van een manuscript.
<placeName>	een plaatsnaam.
<principal>	identificatie van de hoofdonderzoeker die verantwoordelijk is voor totstandkoming van een elektronische tekst.
<profileDesc>	beschrijving van niet-bibliografische (contextuele) aspecten van een tekst.
<projectDesc>	beschrijving van de projectcontext waarbinnen een elektronische tekst tot stand kwam.
<provenance>	beschrijving van een afzonderlijke gebeurtenis in de geschiedenis van een manuscript na de creatie, maar voor de verwerving door de bewarende instelling.
<ptr/>	leeg element waarmee een verwijzing wordt gemaakt naar een ander geïdentificeerd punt in hetzelfde document.
<publicationStmnt>	groepeert informatie over de publicatie of distributie van een tekst.
<publisher>	aanduiding van de verantwoordelijke (persoon of instelling) voor publicatie of distributie van een bibliografisch element.
<pubPlace>	de plaats van publicatie.
<q>	een aanhaling.
<quotation>	beschrijving van de gevolgde praktijk in verband met aanhalingstekens bij de transcriptie van een tekst.
<recordHist>	beschrijving van aspecten in verband met de bron en de wijzigingsgeschiedenis van de manuscriptbeschrijving zelf.

<ref>	een verwijzing naar een een ander geïdentificeerd punt in hetzelfde document.
<refsDecl>	beschrijving van de opbouw van referentiesystemen die gebruikt worden in de tekst.
<reg>	bevat een genormaliseerde vorm van een woord, waarvoor het origineel kan worden weergegeven in een orig attribuut.
<region>	de naam van een regio.
<rendition>	bevat informatie over de standaardweergave van een of meer elementen in een tekst.
<repository>	de naam van een bewaarplaats van een manuscript.
<resp>	aanduiding van de verantwoordelijkheid van een persoon.
<respStmt>	beschrijving van een aspect van de verantwoordelijkheid voor de totstandkoming van een tekst.
<restore>	aanduiding van herstelling van een tekstfragment naar een vorige staat door opheffing van een eerdere auteurs- of editeurswijziging.
<revisionDesc>	beschrijving van de wijzigingsgeschiedenis van een elektronische tekst.
<row>	een rij in een tabel.
<rs>	een niet nader bepaalde verwijzing naar een benoemde entiteit.
<rubric>	de tekst van een ‘rubriek’ of hoofding in een manuscript.
<s>	een grammaticale zin.
<salute>	een openingsbegroeting aan het begin of einde van een tekst(divisie).
<samplingDecl>	beschrijving van de procedures voor selectie van tekstmateriaal in een elektronische tekst.
<seal>	beschrijving van een afzonderlijk zegel op een manuscript.
<sealDesc>	groepeert een reeks beschrijvingen van zegels op het document in de header.
<secFol>	openingswoord(en) op het tweede blad van een manuscript (gebruikt in Middeleeuwse catalogusbeschrijvingen om individuele kopieën te identificeren).

<code><seg></code>	‘semantisch neutraal’ structurelement dat binnen paragrafen kan voorkomen.
<code><segmentation></code>	beschrijving van de gevolgde praktijk in verband met segmentering bij de linguïstische codering van een tekst.
<code><seriesStmt></code>	groepeert informatie over de serie waarvan een publicatie deel uitmaakt.
<code><settlement></code>	de naam van een gemeente of stad.
<code><sic></code>	aanduiding van distantie van de editeur ten opzichte van een manifeste fout in de bron, waarvoor een correctie kan worden gegeven in het corr attribuut.
<code><signatures></code>	beschrijving van de gebruikte pagina-markeringen op de bladzijden van een manuscript.
<code><signed></code>	een slotbegroeting aan het begin of einde van een tekst(divisie).
<code><soCalled></code>	tekst waarvan de editeur zich enigszins distantieert.
<code><source></code>	aanduiding van de originele bron voor de informatie in een manuscriptbeschrijving.
<code><sourceDesc></code>	bibliografische beschrijving van de brontekst waarvan een elektronische tekst is afgeleid.
<code><sp></code>	een individuele tekstlijn in een dramatekst.
<code><space/></code>	leeg element dat opvallende witruimte in de brontekst signaleert.
<code></code>	identificatie van een tekstdeel met een bepaalde interpretatie.
<code><spanGrp></code>	groepeert een reeks tekstdelen met een bepaalde interpretatie.
<code><speaker></code>	identificatie van een bepaalde spreker in een tekstlijn van een dramatekst.
<code><sponsor></code>	aanduiding van een sponsor-instelling of organisatie voor de totstandkoming van een tekst.
<code><stage></code>	een regie-aanduiding in een dramatekst.
<code><summary></code>	een samenvatting van de intellectuele inhoud van een manuscript.
<code><supplied></code>	editeurstoevoeging.
<code><support></code>	beschrijving van het materiaal waarop een manuscript is geschreven.
<code><supportDesc></code>	groepeert elementen voor de fysieke beschrijving van materiaal waarop een manuscript is geschreven.

<surrogates>	beschrijving van informatie over digitale of fotografische representaties van een manuscript.
<table>	een tabel.
<tagsDecl>	beschrijving van de elementen die voorkomen in een SGML of XML tekst.
<tagUsage>	beschrijving van het voorkomen en gebruik van een bepaald element in een SGML of XML tekst.
<taxonomy>	definitie van een typologie voor de classificatie van een tekst.
<TEI.2>	bevat een TEI-conform elektronisch document.
<teiHeader>	de gedetailleerde meta-beschrijving van een TEI-conform elektronisch document.
<term>	een technische term.
<text>	de tekst van een TEI-conform elektronisch document.
<textClass>	groepeert elementen voor de beschrijving van het onderwerp van een tekst.
<textLang>	beschrijving van het gebruik van talen en schriftsystemen in een manuscript.
<time>	een tijdsaanduiding.
<title>	een titel van een werk.
<titlePage>	een titelpagina.
<titlePart>	een subtitel van een werk zoals die voorkomt op een titelpagina.
<titleStmt>	groepeert elementen voor de beschrijving van de titel van een werk en verantwoordelijkheid voor de totstandkoming.
<trailer>	een afsluitende titel of voettekst aan het einde van een tekstdivisie.
<unclear>	aanduiding van onduidelijke tekst in het origineel.
<w>	een grammaticaal woord.
<watermark>	beschrijving van een watermerk in een manuscript.
<width>	aanduiding van de breedte van een manuscript.
<xptr/>	leeg element waarmee een verwijzing wordt gemaakt naar een externe bron.
<xref>	een verwijzing naar een externe bron.

Appendix B. Overzicht van de morfologische tagset

c000	zelfstandig_naamwoord enkelvoud geen
c001	zelfstandig_naamwoord enkelvoud e
c002	zelfstandig_naamwoord enkelvoud s
c003	zelfstandig_naamwoord enkelvoud t
c004	zelfstandig_naamwoord enkelvoud n
c005	zelfstandig_naamwoord enkelvoud r
c006	zelfstandig_naamwoord enkelvoud a
c007	zelfstandig_naamwoord enkelvoud nt
c009	zelfstandig_naamwoord enkelvoud probleem
c010	zelfstandig_naamwoord meervoud geen
c011	zelfstandig_naamwoord meervoud e
c012	zelfstandig_naamwoord meervoud s
c013	zelfstandig_naamwoord meervoud t
c014	zelfstandig_naamwoord meervoud n
c015	zelfstandig_naamwoord meervoud r
c016	zelfstandig_naamwoord meervoud a
c017	zelfstandig_naamwoord meervoud nt
c019	zelfstandig_naamwoord meervoud probleem
c020	zelfstandig_naamwoord eigennaam geen
c021	zelfstandig_naamwoord eigennaam e
c022	zelfstandig_naamwoord eigennaam s
c023	zelfstandig_naamwoord eigennaam t
c024	zelfstandig_naamwoord eigennaam n
c025	zelfstandig_naamwoord eigennaam r
c026	zelfstandig_naamwoord eigennaam a
c027	zelfstandig_naamwoord eigennaam nt
c029	zelfstandig_naamwoord eigennaam probleem
c090	zelfstandig_naamwoord probleem geen
c091	zelfstandig_naamwoord probleem e
c092	zelfstandig_naamwoord probleem s
c093	zelfstandig_naamwoord probleem t
c094	zelfstandig_naamwoord probleem n
c095	zelfstandig_naamwoord probleem r

c096	zelfstandig_naamwoord probleem a
c097	zelfstandig_naamwoord probleem nt
c099	zelfstandig_naamwoord probleem probleem
c100	bijvoeglijk_naamwoord enkelvoud geen
c101	bijvoeglijk_naamwoord enkelvoud e
c102	bijvoeglijk_naamwoord enkelvoud s
c103	bijvoeglijk_naamwoord enkelvoud t
c104	bijvoeglijk_naamwoord enkelvoud n
c105	bijvoeglijk_naamwoord enkelvoud r
c106	bijvoeglijk_naamwoord enkelvoud a
c107	bijvoeglijk_naamwoord enkelvoud nt
c109	bijvoeglijk_naamwoord enkelvoud probleem
c110	bijvoeglijk_naamwoord meervoud geen
c111	bijvoeglijk_naamwoord meervoud e
c112	bijvoeglijk_naamwoord meervoud s
c113	bijvoeglijk_naamwoord meervoud t
c114	bijvoeglijk_naamwoord meervoud n
c115	bijvoeglijk_naamwoord meervoud r
c116	bijvoeglijk_naamwoord meervoud a
c117	bijvoeglijk_naamwoord meervoud nt
c119	bijvoeglijk_naamwoord meervoud probleem
c190	bijvoeglijk_naamwoord probleem geen
c191	bijvoeglijk_naamwoord probleem e
c192	bijvoeglijk_naamwoord probleem s
c193	bijvoeglijk_naamwoord probleem t
c194	bijvoeglijk_naamwoord probleem n
c195	bijvoeglijk_naamwoord probleem r
c196	bijvoeglijk_naamwoord probleem a
c197	bijvoeglijk_naamwoord probleem nt
c199	bijvoeglijk_naamwoord probleem probleem
c200	werkwoord pv_tt_zelfst geen
c201	werkwoord pv_tt_zelfst e
c202	werkwoord pv_tt_zelfst s
c203	werkwoord pv_tt_zelfst t

c204	werkwoord pv_tt_zelfst n
c205	werkwoord pv_tt_zelfst r
c206	werkwoord pv_tt_zelfst a
c207	werkwoord pv_tt_zelfst nt
c209	werkwoord pv_tt_zelfst probleem
c210	werkwoord pv_tt_hulp_koppel geen
c211	werkwoord pv_tt_hulp_koppel e
c212	werkwoord pv_tt_hulp_koppel s
c213	werkwoord pv_tt_hulp_koppel t
c214	werkwoord pv_tt_hulp_koppel n
c215	werkwoord pv_tt_hulp_koppel r
c216	werkwoord pv_tt_hulp_koppel a
c217	werkwoord pv_tt_hulp_koppel nt
c219	werkwoord pv_tt_hulp_koppel probleem
c220	werkwoord pv_vt_zelfst geen
c221	werkwoord pv_vt_zelfst e
c222	werkwoord pv_vt_zelfst s
c223	werkwoord pv_vt_zelfst t
c224	werkwoord pv_vt_zelfst n
c225	werkwoord pv_vt_zelfst r
c226	werkwoord pv_vt_zelfst a
c227	werkwoord pv_vt_zelfst nt
c229	werkwoord pv_vt_zelfst probleem
c230	werkwoord pv_vt_hulp_koppel geen
c231	werkwoord pv_vt_hulp_koppel e
c232	werkwoord pv_vt_hulp_koppel s
c233	werkwoord pv_vt_hulp_koppel t
c234	werkwoord pv_vt_hulp_koppel n
c235	werkwoord pv_vt_hulp_koppel r
c236	werkwoord pv_vt_hulp_koppel a
c237	werkwoord pv_vt_hulp_koppel nt
c239	werkwoord pv_vt_hulp_koppel probleem
c240	werkwoord imp geen
c241	werkwoord imp e

c242	werkwoord imp s
c243	werkwoord imp t
c244	werkwoord imp n
c245	werkwoord imp r
c246	werkwoord imp a
c247	werkwoord imp nt
c249	werkwoord imp probleem
c250	werkwoord inf geen
c251	werkwoord inf e
c252	werkwoord inf s
c253	werkwoord inf t
c254	werkwoord inf n
c255	werkwoord inf r
c256	werkwoord inf a
c257	werkwoord inf nt
c259	werkwoord inf probleem
c270	werkwoord deelwoord geen
c271	werkwoord deelwoord e
c272	werkwoord deelwoord s
c273	werkwoord deelwoord t
c274	werkwoord deelwoord n
c275	werkwoord deelwoord r
c276	werkwoord deelwoord a
c277	werkwoord deelwoord nt
c279	werkwoord deelwoord probleem
c290	werkwoord probleem geen
c291	werkwoord probleem e
c292	werkwoord probleem s
c293	werkwoord probleem t
c294	werkwoord probleem n
c295	werkwoord probleem r
c296	werkwoord probleem a
c297	werkwoord probleem nt
c299	werkwoord probleem probleem

c300	telwoord hoofd geen
c301	telwoord hoofd e
c302	telwoord hoofd s
c303	telwoord hoofd t
c304	telwoord hoofd n
c305	telwoord hoofd r
c306	telwoord hoofd a
c307	telwoord hoofd nt
c309	telwoord hoofd probleem
c310	telwoord rang geen
c311	telwoord rang e
c312	telwoord rang s
c313	telwoord rang t
c314	telwoord rang n
c315	telwoord rang r
c316	telwoord rang a
c317	telwoord rang nt
c319	telwoord rang probleem
c320	telwoord onbepaald geen
c321	telwoord onbepaald e
c322	telwoord onbepaald s
c323	telwoord onbepaald t
c324	telwoord onbepaald n
c325	telwoord onbepaald r
c326	telwoord onbepaald a
c327	telwoord onbepaald nt
c329	telwoord onbepaald probleem
c390	telwoord probleem geen
c391	telwoord probleem e
c392	telwoord probleem s
c393	telwoord probleem t
c394	telwoord probleem n
c395	telwoord probleem r
c396	telwoord probleem a

c397	telwoord probleem nt
c399	telwoord probleem probleem
c400	voornaamwoord persoonlijk geen
c401	voornaamwoord persoonlijk enkelvoud_1
c402	voornaamwoord persoonlijk enkelvoud_2
c403	voornaamwoord persoonlijk enkelvoud_3
c404	voornaamwoord persoonlijk meervoud_1
c405	voornaamwoord persoonlijk meervoud_2
c406	voornaamwoord persoonlijk meervoud_3
c409	voornaamwoord persoonlijk probleem
c410	voornaamwoord aanwijzend geen
c411	voornaamwoord aanwijzend e
c412	voornaamwoord aanwijzend s
c413	voornaamwoord aanwijzend t
c414	voornaamwoord aanwijzend n
c415	voornaamwoord aanwijzend r
c416	voornaamwoord aanwijzend a
c417	voornaamwoord aanwijzend nt
c419	voornaamwoord aanwijzend probleem
c420	voornaamwoord betrekkelijk geen
c421	voornaamwoord betrekkelijk e
c422	voornaamwoord betrekkelijk s
c423	voornaamwoord betrekkelijk t
c424	voornaamwoord betrekkelijk n
c425	voornaamwoord betrekkelijk r
c426	voornaamwoord betrekkelijk a
c427	voornaamwoord betrekkelijk nt
c429	voornaamwoord betrekkelijk probleem
c430	voornaamwoord vragend geen
c431	voornaamwoord vragend e
c432	voornaamwoord vragend s
c433	voornaamwoord vragend t
c434	voornaamwoord vragend n
c435	voornaamwoord vragend r

c436	voornaamwoord vragend a
c437	voornaamwoord vragend nt
c439	voornaamwoord vragend probleem
c440	voornaamwoord onbepaald geen
c441	voornaamwoord onbepaald e
c442	voornaamwoord onbepaald s
c443	voornaamwoord onbepaald t
c444	voornaamwoord onbepaald n
c445	voornaamwoord onbepaald r
c446	voornaamwoord onbepaald a
c447	voornaamwoord onbepaald nt
c449	voornaamwoord onbepaald probleem
c450	voornaamwoord bezittelijk geen
c451	voornaamwoord bezittelijk e
c452	voornaamwoord bezittelijk s
c453	voornaamwoord bezittelijk t
c454	voornaamwoord bezittelijk n
c455	voornaamwoord bezittelijk r
c456	voornaamwoord bezittelijk a
c457	voornaamwoord bezittelijk nt
c459	voornaamwoord bezittelijk probleem
c460	voornaamwoord reflexief geen
c461	voornaamwoord reflexief enkelvoud_1
c462	voornaamwoord reflexief enkelvoud_2
c463	voornaamwoord reflexief enkelvoud_3
c464	voornaamwoord reflexief meervoud_1
c465	voornaamwoord reflexief meervoud_2
c466	voornaamwoord reflexief meervoud_3
c469	voornaamwoord reflexief probleem
c470	voornaamwoord lidwoord_bepaald geen
c471	voornaamwoord lidwoord_bepaald e
c472	voornaamwoord lidwoord_bepaald s
c473	voornaamwoord lidwoord_bepaald t
c474	voornaamwoord lidwoord_bepaald n

c475	voornaamwoord lidwoord_bepaald r
c476	voornaamwoord lidwoord_bepaald a
c477	voornaamwoord lidwoord_bepaald nt
c479	voornaamwoord lidwoord_bepaald probleem
c480	voornaamwoord lidwoord_onbepaald geen
c481	voornaamwoord lidwoord_onbepaald e
c482	voornaamwoord lidwoord_onbepaald s
c483	voornaamwoord lidwoord_onbepaald t
c484	voornaamwoord lidwoord_onbepaald n
c485	voornaamwoord lidwoord_onbepaald r
c486	voornaamwoord lidwoord_onbepaald a
c487	voornaamwoord lidwoord_onbepaald nt
c489	voornaamwoord lidwoord_onbepaald probleem
c490	voornaamwoord probleem geen
c491	voornaamwoord probleem e
c492	voornaamwoord probleem s
c493	voornaamwoord probleem t
c494	voornaamwoord probleem n
c495	voornaamwoord probleem r
c496	voornaamwoord probleem a
c497	voornaamwoord probleem nt
c499	voornaamwoord probleem probleem
c500	bijwoord gewoon geen
c501	bijwoord gewoon e
c502	bijwoord gewoon s
c503	bijwoord gewoon t
c504	bijwoord gewoon n
c505	bijwoord gewoon r
c506	bijwoord gewoon a
c507	bijwoord gewoon nt
c509	bijwoord gewoon probleem
c510	bijwoord aanwijzend geen
c511	bijwoord aanwijzend e
c512	bijwoord aanwijzend s

c513	bijwoord aanwijzend t
c514	bijwoord aanwijzend n
c515	bijwoord aanwijzend r
c516	bijwoord aanwijzend a
c517	bijwoord aanwijzend nt
c519	bijwoord aanwijzend probleem
c520	bijwoord betrekkelijk geen
c521	bijwoord betrekkelijk e
c522	bijwoord betrekkelijk s
c523	bijwoord betrekkelijk t
c524	bijwoord betrekkelijk n
c525	bijwoord betrekkelijk r
c526	bijwoord betrekkelijk a
c527	bijwoord betrekkelijk nt
c529	bijwoord betrekkelijk probleem
c530	bijwoord vragend geen
c531	bijwoord vragend e
c532	bijwoord vragend s
c533	bijwoord vragend t
c534	bijwoord vragend n
c535	bijwoord vragend r
c536	bijwoord vragend a
c537	bijwoord vragend nt
c539	bijwoord vragend probleem
c540	bijwoord onbepaald geen
c541	bijwoord onbepaald e
c542	bijwoord onbepaald s
c543	bijwoord onbepaald t
c544	bijwoord onbepaald n
c545	bijwoord onbepaald r
c546	bijwoord onbepaald a
c547	bijwoord onbepaald nt
c549	bijwoord onbepaald probleem
c550	bijwoord hervattend geen

c551	bijwoord hervattend e
c552	bijwoord hervattend s
c553	bijwoord hervattend t
c554	bijwoord hervattend n
c555	bijwoord hervattend r
c556	bijwoord hervattend a
c557	bijwoord hervattend nt
c559	bijwoord hervattend probleem
c560	bijwoord ontkennend geen
c561	bijwoord ontkennend e
c562	bijwoord ontkennend s
c563	bijwoord ontkennend t
c564	bijwoord ontkennend n
c565	bijwoord ontkennend r
c566	bijwoord ontkennend a
c567	bijwoord ontkennend nt
c569	bijwoord ontkennend probleem
c590	bijwoord probleem geen
c591	bijwoord probleem e
c592	bijwoord probleem s
c593	bijwoord probleem t
c594	bijwoord probleem n
c595	bijwoord probleem r
c596	bijwoord probleem a
c597	bijwoord probleem nt
c599	bijwoord probleem probleem
c600	voornaamwoordelijk_bijwoord persoonlijk geen
c601	voornaamwoordelijk_bijwoord persoonlijk e
c602	voornaamwoordelijk_bijwoord persoonlijk s
c603	voornaamwoordelijk_bijwoord persoonlijk t
c604	voornaamwoordelijk_bijwoord persoonlijk n
c605	voornaamwoordelijk_bijwoord persoonlijk r
c606	voornaamwoordelijk_bijwoord persoonlijk a
c607	voornaamwoordelijk_bijwoord persoonlijk nt

c609	voornaamwoordelijk_bijwoord persoonlijk probleem
c610	voornaamwoordelijk_bijwoord aanwijzend geen
c611	voornaamwoordelijk_bijwoord aanwijzend e
c612	voornaamwoordelijk_bijwoord aanwijzend s
c613	voornaamwoordelijk_bijwoord aanwijzend t
c614	voornaamwoordelijk_bijwoord aanwijzend n
c615	voornaamwoordelijk_bijwoord aanwijzend r
c616	voornaamwoordelijk_bijwoord aanwijzend a
c617	voornaamwoordelijk_bijwoord aanwijzend nt
c619	voornaamwoordelijk_bijwoord aanwijzend probleem
c620	voornaamwoordelijk_bijwoord betrekkelijk geen
c621	voornaamwoordelijk_bijwoord betrekkelijk e
c622	voornaamwoordelijk_bijwoord betrekkelijk s
c623	voornaamwoordelijk_bijwoord betrekkelijk t
c624	voornaamwoordelijk_bijwoord betrekkelijk n
c625	voornaamwoordelijk_bijwoord betrekkelijk r
c626	voornaamwoordelijk_bijwoord betrekkelijk a
c627	voornaamwoordelijk_bijwoord betrekkelijk nt
c629	voornaamwoordelijk_bijwoord betrekkelijk probleem
c630	voornaamwoordelijk_bijwoord vragend geen
c631	voornaamwoordelijk_bijwoord vragend e
c632	voornaamwoordelijk_bijwoord vragend s
c633	voornaamwoordelijk_bijwoord vragend t
c634	voornaamwoordelijk_bijwoord vragend n
c635	voornaamwoordelijk_bijwoord vragend r
c636	voornaamwoordelijk_bijwoord vragend a
c637	voornaamwoordelijk_bijwoord vragend nt
c639	voornaamwoordelijk_bijwoord vragend probleem
c640	voornaamwoordelijk_bijwoord onbepaald geen
c641	voornaamwoordelijk_bijwoord onbepaald e
c642	voornaamwoordelijk_bijwoord onbepaald s
c643	voornaamwoordelijk_bijwoord onbepaald t
c644	voornaamwoordelijk_bijwoord onbepaald n
c645	voornaamwoordelijk_bijwoord onbepaald r

c646	voornaamwoordelijk_bijwoord onbepaald a
c647	voornaamwoordelijk_bijwoord onbepaald nt
c649	voornaamwoordelijk_bijwoord onbepaald probleem
c690	voornaamwoordelijk_bijwoord probleem geen
c691	voornaamwoordelijk_bijwoord probleem e
c692	voornaamwoordelijk_bijwoord probleem s
c693	voornaamwoordelijk_bijwoord probleem t
c694	voornaamwoordelijk_bijwoord probleem n
c695	voornaamwoordelijk_bijwoord probleem r
c696	voornaamwoordelijk_bijwoord probleem a
c697	voornaamwoordelijk_bijwoord probleem nt
c699	voornaamwoordelijk_bijwoord probleem probleem
c700	voorzetsel geen
c701	voorzetsel e
c702	voorzetsel s
c703	voorzetsel t
c704	voorzetsel n
c705	voorzetsel r
c706	voorzetsel a
c707	voorzetsel nt
c709	voorzetsel probleem
c790	voorzetsel probleem geen
c791	voorzetsel probleem e
c792	voorzetsel probleem s
c793	voorzetsel probleem t
c794	voorzetsel probleem n
c795	voorzetsel probleem r
c796	voorzetsel probleem a
c797	voorzetsel probleem nt
c799	voorzetsel probleem probleem
c800	voegwoord nevenschikkend geen
c801	voegwoord nevenschikkend e
c802	voegwoord nevenschikkend s
c803	voegwoord nevenschikkend t

c804	voegwoord nevenschikkend n
c805	voegwoord nevenschikkend r
c806	voegwoord nevenschikkend a
c807	voegwoord nevenschikkend nt
c809	voegwoord nevenschikkend probleem
c810	voegwoord onderschikkend geen
c811	voegwoord onderschikkend e
c812	voegwoord onderschikkend s
c813	voegwoord onderschikkend t
c814	voegwoord onderschikkend n
c815	voegwoord onderschikkend r
c816	voegwoord onderschikkend a
c817	voegwoord onderschikkend nt
c819	voegwoord onderschikkend probleem
c820	voegwoord vergelijking geen
c821	voegwoord vergelijking e
c822	voegwoord vergelijking s
c823	voegwoord vergelijking t
c824	voegwoord vergelijking n
c825	voegwoord vergelijking r
c826	voegwoord vergelijking a
c827	voegwoord vergelijking nt
c829	voegwoord vergelijking probleem
c830	voegwoord ne en geen
c831	voegwoord ne en e
c832	voegwoord ne en s
c833	voegwoord ne en t
c834	voegwoord ne en n
c835	voegwoord ne en r
c836	voegwoord ne en a
c837	voegwoord ne en nt
c839	voegwoord ne en probleem
c840	voegwoord else geen
c841	voegwoord else e

c842	voegwoord alse s
c843	voegwoord alse t
c844	voegwoord alse n
c845	voegwoord alse r
c846	voegwoord alse a
c847	voegwoord alse nt
c849	voegwoord alse probleem
c890	voegwoord probleem geen
c891	voegwoord probleem e
c892	voegwoord probleem s
c893	voegwoord probleem t
c894	voegwoord probleem n
c895	voegwoord probleem r
c896	voegwoord probleem a
c897	voegwoord probleem nt
c899	voegwoord probleem probleem
c900	rest latijn probleem geen
c901	rest latijn probleem e
c902	rest latijn probleem s
c903	rest latijn probleem t
c904	rest latijn probleem n
c905	rest latijn probleem r
c906	rest latijn probleem a
c907	rest latijn probleem nt
c909	rest latijn probleem probleem
c910	tussenwerpsel geen
c911	tussenwerpsel e
c912	tussenwerpsel s
c913	tussenwerpsel t
c914	tussenwerpsel n
c915	tussenwerpsel r
c916	tussenwerpsel a
c917	tussenwerpsel nt
c919	tussenwerpsel probleem

