

# Van het Corpus-Gysseling-INL (CG-INL) naar CGA en CGR13

Pieter van Reenen

May 24, 2013

## 1 Inleiding

De doelstelling van het hier beschreven project is tweeledig:

- Het CG-INL-ambtelijk verbeteren en afcoderen: het CGA.
- Uit het CGA afleiden een tweelingcorpus van CRM14 in Adelheidformaat: het CGR13.

Het CG inclusief een aantal aanvullingen is het vrijwel volledige corpus 13de-eeuws middelnederlands. Het is de afgelopen jaren op het INL gedigitaliseerd tot het CG-INL. Het CG-INL bestaat uit twee delen: ambtelijk taalgebruik en literair taalgebruik. Hieronder gaat het in het vervolg alleen over het CG-INL-ambtelijk, inclusief een aantal aanvullingen. In de toekomst kunnen nog verdere uitbreidingen verwacht worden. Aan het CG-INL zijn twee uiterst belangrijke elementen toegevoegd: een morfologische codering van de middelnederlandse vormen en hun lemmatisering. Op dit punt heeft het INL een formidabele prestatie geleverd, zeker gezien de soms extreme weerbarstigheid en ondoorzichtigheid van het oude taalmateriaal. Blijkens de codering en lemmatisering is de tekst van het CG-INL met grote deskundigheid geanalyseerd. De lemmatisering en codering zelf is echter onvolledig en vaak inconsequent, in enkele gevallen zinloos.

Het verband tussen CG-INL en de gedrukte versie (dan wel de pdf-versie) is het paginanummer, bijvoorbeeld in CG-INL correspondeert regel <L 13:9 met pagina 13 regel 9 in het CG-pdf en CG-boek.

Al eerder was door en onder leiding van Pieter van Reenen en Maaike Mulder het CRM14 aangelegd, een 14de-eeuws corpus van gelokaliseerde en gedateerde originele oorkonden op computer, dat evenals het CG-INL in de loop der jaren is uitgebreid en in de toekomst nog verder zal worden uitgebreid. Het CRM14 wordt sinds enige jaren aangeboden voor taalkundig onderzoek in het zogenaamde Adelheidformaat. Het is de bedoeling dat ook de relevante teksten uit het CG-INL in dat formaat gegoten worden. Daarna vormt het samen met het CRM14 een zeer omvangrijk corpus ouder Nederlands, dat geschikt is voor taalkundig onderzoek, waarbij rekening kan worden gehouden met de plaats van herkomst van het taalmateriaal en de datering ervan. Samen bestrijken de corpora dan de periode van iets voor 1250 tot 1400. De taalkundige beschikt daarmee voor deze periode over een verregaand geordende hoeveelheid taalgegevens voor onderzoek. Gehoopt wordt dat het in de toekomst wordt aangevuld met vergelijkbare gegevens uit later en zo mogelijk eerdere eeuwen.

Het eindproduct, een serie corpora die verregaand op dezelfde leest geschoeid zijn, blijft echter per definitie een halffabrikaat, een hulpmiddel.

Het CG-INL is gecodeerd en gelemmatiseerd volgens de richtlijnen ontwikkeld en gebruikt bij de aanleg van het CRM14. Het CRM14 bevat uitsluitend originele gedateerde en gelokaliseerde oorkonden, waar het CG-INL ook copieën, rekeningen en grafschriften bevat, benevens veel vaak waardevolle informatie in de vorm van dorso-aantekeningen en noten. Deze extra informatie is echter in het corpus niet altijd voldoende duidelijk van de middeleeuwse tekst gescheiden en de codering en lemmatisering zijn niet helemaal af. De hier beschreven LINUX-programma's leiden uit het CG-INL het CGA (A voor ambtelijk) en het CGR13 af. Daarmee bestaat het CG-ambtelijk uiteindelijk in vier varianten:

- Het CG-boek en het CG.pdf die gelijk aan elkaar zijn.
- Het CG-INL inclusief aanvullingen zoals dat op het INL tot stand is gekomen.
- Het CGA.txt. Het CGA-formaat is vooral bedoeld voor historici en letterkundigen. Het directe verband met CG-boek (en CG.pdf) blijft hier gehandhaafd en is zelfs iets vergroot ten opzichte van het CG-INL door invoeging van de informatie 'origineel of niet' en 'handen' uit het CG-boek.
- Het CGR13: een deelverzameling in researchformaat van het CGA, speciaal voor taalkundigen dat wordt aangeboden in Adelheid-formaat. Het bestaat uitsluitend uit originele, gedateerde en gelokaliseerde teksten die plaatselijk taalgebruik representeren.

Parallel aan het CGA.txt zou ook, voorzover nodig, een literair CGL ontwikkeld kunnen worden, mogelijk deels met dezelfde programmatuur.

De indeling van deze overzicht is verder als volgt: In hoofdstuk 2 wordt beschreven hoe met behulp van een serie programma's van het CG-INL het CGA wordt afgeleid. In hoofdstuk 3 hoe van het CGA het CGR13 wordt afgeleid. De namen van de programma's zijn zo gekozen dat ze zo goed mogelijk hun inhoud dekken. Dat maakt het mogelijk bij herzieningen, aanvullingen enz.

- de relevante programma's relatief makkelijk terug te vinden,
- daarin verbeteringen aan te brengen,
- programma-onderdelen toe te voegen of te wijzigen,
- programma-onderdelen te verplaatsen.

Ook de directory DOCUMENTATIE bevat soms nuttige aanvullende, maar niet gebruikte informatie, bijvoorbeeld over niet te automatiseren verbeteringen en HET-en-DE-gebruik. Hoofdstuk 4 beschrijft in grote lijnen de directory en filestructuur. Hoofdstuk 5 geeft een toelichting op het CG-INL en het CGA. Hoofdstuk 6 de codering.

## 2 Van CG-INL naar CGA

Vooraf. Het CG-INL is in een aantal opzichten nog een invoerformaat in plaats van een computerformaat, zie voor deze begrippen Van Reenen e.a. 1985. Een aantal programma's hieronder is erop gericht dit computerformaat tot stand te brengen. Verder is er een grote verscheidenheid aan fouten en problemen die moesten worden opgelost om van het CG-INL naar het CGA te komen. De belangrijkste tekortkomingen zijn:

- Het corpus is niet afgecodeerd.
- "De" en "het" zijn vaak niet uitgesplitst.
- Middelnederlandse woorden kunnen soms niet goed gescheiden worden. Zij worden dan gecodeerd en gelemmatiseerd met een "+" ertussen. Zo is een sequentie als "sercoensone" in principe opgesplitst in vier woorden: 472+000+022+001.DE+HEER+KOEN+ZOON>. In principe, want bij de codering en lemmatisering is de volgorde van de woorden willekeurig, b.v. 472+001+000+022.DE+ZOON+HEER+KOEN>. Herstel van dit probleem was zeer tijdrovend.
- Metawoorden zijn vaak niet onderscheiden van middelnederlandse tekst.
- Dorsotekst is vaak onvolledig opgenomen, en daardoor onbruikbaar.
- De uitsplitsing in de tijd leidt soms tot zinlose corpusrepresentaties.

Hieronder volgt een bespreking van de ontwikkelde LINUX-programma's. Bij de meeste programmaonderdelen is een telprogramma in begrepen. De resultaten van deze tellingen staan in de directory werk/Aantltellr. In de programma's zelf staat vaak ook nog vooral praktische informatie, in de sed-scriptjes, en de awk-programma's in de directory werk/SED waarnaar de programma's verwijzen ook wel eens.

Het is niet gegarandeerd dat de programma's en bijbehorende scripts per definitie geheel foutloos zijn. Bovendien zijn ze niet noodzakelijk volledig. Voor een aantal nog te verbeteren fouten, zie DOCUMENTATIE. Verbeteringen en aanvullingen kunnen in de programma's dan wel in de daar genoemde sed-scripts en awk's worden aangebracht.

### 2.1 Omissies, dubbelzinnigheden en ad hoc verbeteringen

Helemaal vooraf zijn in het CGL-INL vijf problemen opgelost, zie DOCUMENTATIE/PREALABELEVERBETERINGENINCGINL voor details. Het resultaat is corp13new.txt. De problemen betreffen:

- 12 lemma's (types) waarin een trema voorkomt
- 1 lemma met een accent grave (APPEL)
- 5 lokaliseringfouten
- 1 verkeerde ASCII-volgorde
- 5 onjuiste oorkondebestandsnamen

### 2.1.1 0a\_lijnummeringondubbelzinnig

In het CG-INL zijn met twee uitzonderingen tekstregels ondubbelzinnig gedefinieerd door hun regelnummer <L. De eerste uitzondering betreft oa? dorsoteksten zonder <L. Regels zonder <L worden van <L 0:0 voorzien. De tweede uitzondering betreft meer dan eenmaal voorkomende nummers. Zij zijn het gevolg van de aan het CG-boek toegevoegde documenten van elders, met als gevolg dat bijvoorbeeld <L 29:5 3x voorkomt. Een ondubbelzinnige nummering wordt aangebracht, in verband met deze extra toegevoegde teksten. In het Landrecht van Grimbergen wordt <L 229:5 tot <LG 229:5.

Het programma:

1. Voegt regelnummers <L 0:0 toe waar die ontbreken, zodat alle vormen <q> en <C worden voorafgegaan door <L. Daardoor werkt 0j\_dorsomarkeren beter
2. Maakt regelnummers ondubbelzinnig : <L 1:1 in Oudenbiezen > <LO 1:1. <L 1:5 slaat daarmee alleen nog op Gysseling.

- bronAntwObit.awk
- bronLandrechtGrimbergen.awk
- bronOudenbiezen.awk
- bronSchepenbrAnt.awk
- bronTenTongval.awk
- bronWilhem.awk

NB De overige in TenT gepubliceerde oorkonden zouden bij toevoeging naar de auteurs LB Burgers, LC Croenen, LL Leroy, LG Geyrmaerts en Mooyaart, LD Dijkhof ea genoemd kunnen worden.

### 2.1.2 0b\_3cijfercodes

Dit programma vult codes aan tot 3 cijfers door er 0 of 00 voor te zetten. Het maakt van bijvoorbeeld <C 0\_NAAM> <C 000\_NAAM>.

- 3cijfer.sed

### 2.1.3 0c\_voorcorrecties

De input voor het derde programma 0c\_voorcorrecties. De metagegevens van de niet aanwezige "Renterol Deinze" worden hier verwijderd: "<docId>0032a". Eveneens een testbestand. De -X-notatie wordt hier verwijderd. Tenslotte wordt de volgorde van een aantal regels in het CG aangepast, zodat ze niet langer door elkaar staan.

- sed '75225,75232d' # foute <docID>0644TEST weggegooid
- sed '27726,27733d' # Datumcodeweg: text begrijpelijk
- sed 's/999\_ZZZ> ho/250\_HOREN> ho[re]n <VN type=1>/'
- sed '27717,27723d' # Datumcodeweg: text begrijpelijk
- sed '27712,27714d' # Datumcodeweg: text begrijpelijk

- sed '27707,27710d' # Datumcodeweg: text begrijpelijk
- sed 's/<q> [ine] <L 428:14> <C 21\_HENDRIK> He[ine] <VN type=1>@  
/'
- sed '5719,5726d' # lege Renterol Deinze weggegooid
- sed 's/-X//g' # ASART-X wordt ASART ed.
- volgorderegels.sed

#### 2.1.4 Od.tikfouten

In tikfoutjes.sed worden o.a. adhoc enkele fouten van overbodige ”,” in het CG-INL verbeterd. (Dat gebeurt ook wel eens elders ad hoc in de diverse programma's.)

- tikfoutenenz.sed

#### 2.1.5 Oe\_lemmawijzigingen

Dit programma verbetert allerhande min of meer foute lemma's, soms met hun code.

- TEenELKaanv.sed
- DEaanvulling.sed
- HetAanvulling.sed
- RestAanvulling.sed
- WEERWEDER.sed
- DUBBELwENKEL.sed
- EERTANG.sed
- TEINDE.sed
- vraagtekenweg.sed
- lemmafouten.sed

NB Hier eventueel een scriptje toevoegen dat de fouten verbetert die zijn gesignaleerd door Burgers 1993, p 185-186 (TT): vier correcties op dateringen van Gysseling.

#### 2.1.6 Of\_codewijzigingen

- codewijzigingen.sed

### 2.2 Completering codering

Het gaat met name om de vormen voorafgegaan door <q>.

### 2.2.1 1a.mnlPREcontext

### 2.2.2 1b.haakmetawoorden

De ( ) om moderne metatekst zijn verdwenen in de versie CG-INL. Deze zijn in de metatekstprogramma's weer zo goed mogelijk aangebracht. Daarbij zijn in de metatekst geciteerde woorden soms wel soms niet bij de metatekst getrokken. (Maar niet altijd als ze al gecodeerd waren.) Eventueel ontbrekende tekst wordt echter niet aangevuld, al is door het ontbreken de tekst soms onzinnig want de datum of het voorafgaande is weggehaald!! Bijvoorbeeld:

```
<Ldorso 2866:11> <M toegevoegd> (slaat op datum die is weggehaald!!)
<L 274:32> <C 700.IN> in <q> hanekinewerue <q> iii <C 019.LIBRA> lib
( <M toegevoegd> ) Komt ± 16x voor.
```

```
<L 428:14> <C 999_ZZZ> He
<end-datering jaar_tot='1279' jaar_van='1279' tijdsoort='1' tijdstatus='1'>
<datering jaar_tot='1412' jaar_van='1412' tijdsoort='1' tijdstatus='1'>
<q> [ine] <C 1_KALLE> [calle]
<end-datering jaar_tot='1412' jaar_van='1412' tijdsoort='1' tijdstatus='1'>
<datering jaar_tot='1279' jaar_van='1279' tijdsoort='1' tijdstatus='1'>
Zo wordt He[ine] [calle] weergegeven. Maar daar heeft niemand wat aan. Ik heb
daarom datering en end-datering weggehaald, <VN TYPE=1> toegevoegd
Komt nog 2x vlak erna voor: 428:15, 428:16.
```

Alle woorden van het type [hand], [diverse], voorzover van Gysseling worden  
<M (M voor meta).

Voor resten middelnederlands woorden zoals [neghen], zie later.

```
NB <LT 12:10> <q> Ettelgh<A >em</A> <M [...]> <M de> <q> [c]apel<A
>anie</A>
```

Is hier het <M gebruik wel opportuun? metataal van lang geleden!!

NB Getallen als [.x.] zijn als meta gecodeerd, al is dat soms hersteld middelnederlands. Moet nog veranderd!!

NB <M .. beter als leestekens? Ja!!

Overige metagegevens komen aan de beurt in 2.4.

- handhaken.sed
- haakgetal.sed
- haakmetawoorden.sed
- haakromeinsgetal.sed
- LAhandhaak.sed
- haakcontext.sed

### 2.2.3 1c.metatekstscheider

Bevat tekst van Gysseling die niet tussen [ ] staat, soms inclusief geciteerd middelnederlands. Dankzij metaposttext.sed worden ook lidwoorden als "een" meegenomen. Adhoc-gevallen staan in metapretekst.sed.

#### 2.2.4 1d\_latijnfrans

Niet gecodeerde items <q> in het latijn en frans worden als

<C 999\_LATIJN> of <C 999\_FRANS>

geklasseerd. Het gaat om hoge aantallen en soms lange passages latijn, maar altijd bescheiden hoeveelheden frans. Een enkele keer is het niet duidelijk om welke taal het gaat. Voorafgaan kontekstregels die homografen classificeren met behulp van woordseries. In bijvoorbeeld: "<q> pro <q> die" kan "die" geen middelnederlands zijn. Achteraf opnieuw kontekstregels, met als uitgangspunt: als het buurwoord latijn is, dan zal het nog ongecodeerde woord ernaast dat ook wel zijn.

KL en kl 'hier volgt 'klassiek', i.e. latijn' komt alleen in <LA voor:

sed 's/<q> KL <q> /<C 999\_LATIJN> /g'

sed 's/<q> kl <q> /<C 999\_LATIJN> /g'

- CONTEXTFRLAT.sed
- lijstLATIJN.sed
- lijstFRANS.sed
- CONTEXTpost.sed

#### 2.2.5 1e\_getallenRomEnVoluit

Hier worden eerst niet gecodeerde romeinse cijfers als 999\_GETALR> gecodeerd, een tussenstap, voor het kontekstprogramma gladstrijkFLG.sed, dat op basis van 999\_GETALR> FRANS of LATIJN aanmaakt, afhankelijk ervan wat het telwoord ernaast is: FRANS respectievelijk LATIJN. Door dit script te herhalen worden weer nieuwe vormen gevangen. Vervolgens maakt GETALRwordtCIJFER.sed van wat overblijft van GETALR> aan Romeinse cijfers in het lemma Arabische cijfers.

In TALROMenLET.sed worden <q>-getallen die in het Middelnederlands deels romeins deels in letters zijn gespeld overeenkomstig gecodeerd. Maakt ook 't! 0,5 aan. Dat gebeurt hier, anders geeft mlnnr1.sed problemen. Hier ook deels arabische getallen [102] [1298] ed. Maakt van getallen in letters codes in letters.

VoluitGetallenCoderen.sed. Dit script codeert voluitgeschreven getallen.

ArabGetallenCoderen.sed. Dit script codeert arabische getallen.

#### 2.2.6 1f\_namencoderen

Met EigenNAAM.sed worden de eigennamen gecodeerd obv <q>.

Plaatsnamen zijn een woord: SINT-CATHARINA-WEST, NIEUW-VRANKENDIJK

Straatnamen ook: SINT-CATHARINASTRAAT

In andere gevallen (vaak) niet:

<C SINT <C CATHARINA <C ALTAAR

<C 029\_WOUTER> walterus <q> dictus <C 101\_ZOET> soete

029\_WOUTER> walteri <C 102\_SOET> soets <q> dicti

Daarom als er twee vormen zijn opvatten als twee losse namen:

WOUTER> wout<A >er<A> <C 101\_ZOET> soete

WOUTER> wout<A >er<A> <C 104\_ZOET> soete<A >n<A>

WOUTER> wout<A >ra<A> <C 101.ZOET> sote  
 029\_WOUTER> valteru<A >m</A> <C 101.ZOET> sote  
 WOUTER> walteri <C 102.SOET> soets  
 NB1. Correctie van de eigennaamkeuze zou achteraf door een deskundige van het INL doorgelopen kunnen worden.  
 NB2. In CG-INL zijn veel potentiële eigennamen als 00[0-4] e.d. gecodeerd. Dat moet 02[0-4] worden. Dit is inmiddels deels gebeurd. Er blijven echter heel wat lastige en twijfelgevallen over.  
 NB3. WOUTER in CGA.txt is WALTER in CRM14.txt

### 2.2.7 1g\_mnlcoderen

mnlhr1.sed codeert veel afkortingen en rare gevallen als 009 of 019  
 't%, f, s, 't3s, lb, so't3l, 't3d , .t!  
 mnlZZZ.sed codeert als onbegrijpelijk.

- mnlhr1.sed
- mnlBVNMWcode.sed
- mnlBWcode.sed
- mnlMENGcode.sed
- mnlRomRomcode.sed
- mnlTWcode.sed
- mnlVNWcode.sed
- mnlVOEGWcode.sed
- mnlVZcode.sed
- mnlWWcode.sed
- mnlZNWcode.sed
- mnlZZZ.sed

### 2.2.8 1h\_HEBBEN

De niet gecodeerde Hebben-vormen worden gecodeerd. Eerst met context, dan zonder.

### 2.2.9 1i\_vlpleestekens

De leestekens voorafgegaan door <q> worden <l>

## 2.3 Herlemmatisering van getallen

In de volgende programma's gaat het niet om <q>-codes, maar om al aangebrachte codes die verbeterd worden.



### **2.3.1 2a\_Telwoorden**

Splitst duizend en honderd af.

### **2.3.2 2b\_correctevolgordestermat**

Herstelt fouten in codering vwb volgorde van gescheiden woorddelen.

### **2.3.3 2c\_STERRENmatjes**

Overbodige sterren en matjes weg.

### **2.3.4 2d\_restherlemmatisering**

De namen van de scripts spreken voor zich:

- NOGwordtNOCH.sed,
- TweeWordtEen.sed,
- EenWordtTwee.sed,
- KoppelWWlos.sed,
- ZOplits.sed,
- ONTHIER.sed

### **2.3.5 2c\_STERRENmatjes**

### **2.3.6 2d\_restherlemmatisering**

## **2.4 Meta-informatie uitsplitsen en toevoegen**

Niet uit CG-boek overgenomen meta-informatie wordt hier alsnog deels aangebracht. Tezijntijd zou ook de informatie over welke handen welke oorkonden geschreven hebben moeten worden toegevoegd.

### **2.4.1 3a\_origineelofniet**

Voegt origineel, afschrift ed. in obv origstatus.sed, het centrale script. Zie ook 5a\_oorknummer.

### **2.4.2 3b\_dorsomdmarkeren**

Voegt obv statushandkode='md' achter <L dorso in.

### **2.4.3 3c\_nootmarkeren**

Regelnummers van boven de 51 tot 80 betreffen altijd noten. Bijvoorbeeld: <L 3005:63 zou, als het bestond, een noot betreffen.

### **2.4.4 3d\_dorsoadmarkeren**

statuskodead.awk markeert passages type 'ad' als dorsoad.

## 2.5 Bestaande codes en lemma's verbeteren

### 2.5.1 4b.anderecode

Dit programma verandert de code van `_AL>` en `_ELK>`: niet `44[0-9]` maar `32[0-9]`.

### 2.5.2 4c.lemmaENcodeomkeren

Woordgroepen zijn in het CG-INL vaak in verkeerde volgorde gekodeerd: Twee eenvoudige voorbeelden:

`_DE+TE>` in plaats van `_TE+DE>` plus bijbehorende codes  
`_HEER+DE>` in plaats van `_DE+HEER>`.

En een ingewikkeld voorbeeld met vele volgordes:

```
472+0+22_DE+HEER+WOUTER>@s<A@>er</A>wout<A@>er</A>mans
472+22+0_DE+WOUTER+HEER> sh<A >er</A>wout<A >er</A>mans
472+22+0_DE+WOUTER+HEER> sherwout<A >er</A>mans
0+22+472_HEER+WOUTER+DE> serwoutermans
1+22+472_HEER+WOUTER+DE> s<A >er</A>ewout<A >er</A>mans
1+472+22_HEER+DE+WOUTER>@tserewout<A@>er</A>mans
22+0+472_WOUTER+HEER+DE> s<A >er</A>woutermans
22+1+472_WOUTER+HEER+DE> s<A >er</A>ewout<A >er</A>mans
22+472+0_WOUTER+DE+HEER> s<A >er</A>woutermans.
22+472+0_WOUTER+DE+HEER> serwout<A >er</A>mans
22+472+0_WOUTER+DE+HEER> tserweitins
```

De vormen worden automatisch teruggezet dankzij een serie van 12 sed-scripts:<sup>1</sup>

- drieplusprobleem.sed  
drie of vier lemma's in diverse verkeerde willekeurige volgordes.
- tweeproblemen.sed  
Tamelijk willekeurige foute volgordes worden hier gecorrigeerd.
- tweepus.sed  
Tamelijk willekeurige foute volgordes worden hier gecorrigeerd.
- groteomkeer.sed  
Specifieke groepen foute volgordes worden gecorrigeerd. Keert bijvoorbeeld `Eigennaam+TE`, `HOEVE+TE` en `Infinities+TE` om.
- substsubstomkeer.sed  
Voorbeeld: `DAG+PETRUS > PETRUS+DAG`. PM kloppen alle `AVE(SOETES)` en `(AVE?)`?
- substadjomkeer.sed  
Voorbeeld: `DIJK+OUD > OUD+DIJK`.
- NE.sed  
Een voorbeeld is `ZULLEN+NE > NE+ZULLEN`.
- Divers1.sed, Divers2.sed, Divers3.sed, Divers4.sed  
meer dan 860 expressies in de `Divers.scripts`, op volgorde code.

---

<sup>1</sup>In dit geval zijn ook de lemma's niet juist. Die worden elders verbeterd.

- probleem.sed

Tamelijk willekeurige foute volgordes worden hier gecorrigeerd, bijvoorbeeld WEDERKOMEN> weder WEDERKOMEN> komen.

NB Heel speciaal is hier het eenmalige " <C 403+406\_ZIJ+ZIJ> ziese " dat zo moet blijven tegenover 8x waar de volgorde moet worden omgekeerd (en waar anders wordt gespeld dan "ziese").

NB TAAIeproblemenvoorlater en foutetc moeten nog aangepakt.

### 2.5.3 4d\_HETbinnenregel

Dit programma reduceert DE/HET tot HET binnen de computerregel voorzover het zelfstandig naamwoord op dezelfde regel staat. Dit is nodig, want het oorspronkelijk corpus is in dit opzicht niet af: DE/HET -> DE ABT en DE(!) KLOOSTER (ook HET KLOOSTER?), DE KERK, maar HET CONVENT etc.

De beste strategie lijkt DE/HET gericht tot HET te reduceren (HET komt aanzienlijk minder voor dan DE) en wat overblijft tot DE.

Voorbeeld:

<C 474\_DE/HET> den <C 104\_VOORZEID> uorseiden  
<VN type=1> <C 004\_VEERTIENNACHT> uirtienachten.  
moet worden

<C 474\_DE> den <C 104\_VOORZEID> uorseiden  
<VN type=1> <C 004\_VEERTIENNACHT> uirtienachten.

NB Woordgeslacht: Manlijk, vrouwelijk en onzijdig is niet altijd uit te maken: Erf lijkt zowel "het"- als "de"-vorm te kunnen zijn. (IN)CARNATIOEN heeft HET en DE.

zie: <L 1843:31> <C 700\_OP> op <C 471\_DE> de <C 001\_ERF> erue <C 421+600\_DIE+ERTOEF> diere+toe <C 203\_BEHOREN> behort <C 800\_EN> en<A >de</A> <C 301\_EEN> ene <C 309\_TIEN> .x. <C 019\_SOLIDUS> 't3s <C 411\_DIE> die <C 210+440\_ZULLEN+MEN> salmen <C 250\_NEMEN> nemen <C 700\_OP> op <C 410\_DAT> dat <C 001\_ERF> erue

Zowel ten als ter WATERLEDE

In zulke gevallen wordt het Modern Nederlands gevolgd of DEadhocnaHET3 als DE-woord.

NB In DOCUMENTATIE staan (ook) nog de uiteindelijk niet-gebruikte DE.sed-lijsten. Ze nemen praktisch zoveel tijd dat ze niet goed bruikbaar zijn. Ze zijn waarschijnlijk ook niet foutloos.

### 2.5.4 4e\_HET3overregel

Dit programma reduceert DE/HET tot HET, als het zelfstandig naamwoord net op de nieuwe computerregel staat. Dit en het volgende programma lossen de HET-gevallen over de regel heen op.

### 2.5.5 4f\_HET6overregel

Idem nog drie plaatsen verder.

### 2.5.6 4g\_DEReductie

Dit programma maakt van alle overige DE/HET-gevallen automatisch DE.

### 2.5.7 4h\_posthoc

Posthoccorrecties: verbetert o.a ZOEN en ZOON.

Output is naast result\_4eposthoc.txt het identieke CGA.txt

Adhoc: maakt LEMMA van <q> dat deels latijn is deels middelnederlands.

Nog nader uit te werken.

## 3 5a\_adhocleestekens

Hier worden leestekens behandeld: scheiden van woorden en in Adelheidformaat omzetten. Nog zeer onvolledig.

## 4 Van CGA naar CGR13: Adelheidformaat

De programma's maken van CGA.txt, dat overeenkomt met het CG-INL-formaat, Adelheidformaat.

### 4.1 6aGR\_oorknr

Staat niet automatisch aan. Zift zo gewenst het corpus Graven uit.

### 4.2 6a\_oorknummer

Origineel-afschrift brengt in CGR13 het verband tussen <docId>, voorzover origineel en lokaliseerbaar, en het Kloekennummer plus de datum en vervangt <docId> door het Kloekennummer plus jaartal. Cruciaal is docidnaaroorknr.sed: semi-automatisch tot stand gebracht. Het probleem dat origineel versus afschrift in CG-INL niet uit CG-boek is overgenomen is daarmee voor een deel opgelost. Maar docidnaaroorknr.sed zou moeten worden gecorrigeerd en aangevuld, met name op regio. Dat zou kunnen op basis van het proefschrift van Marijke Mooijaart. Mogelijk staan er ook nog verkeerde jaartallen afkomstig uit de dorsoteksten. Zie verder DOCUMENTATIE, de files:

- origineelafschrift
- Kloekernrs: plaatsnaam en plaatsnr in CG-INL
- Kloekenrs.sed: koppelt plaatsnr in CG-INL aan Kloekennummer

### 4.3 6b\_ziftoorkuitcorpus

Dit programma gooit allereerst de twee type dorsoteksten weg: het is geen originele tekst. Ook de noten worden hier verwijderd. Ze staan dus niet in CGR.txt. Voor hun inhoud moet men terug naar CGA.txt.<sup>2</sup>

---

<sup>2</sup>Ik heb ze verwijderd omdat ze midden in de tekst staan: VN type='1' (komt op 7194 regels voor) in CGA.txt betekent: zie noot. Daar staat dan bv woord boven regel geschreven of iets dergelijks. Niet alle noten hebben echter een trace type=1, zie p. 61:23 (niet), 61:24 (wel). Ik doe geen poging tot herstel. type=2 komt op 11 regels voor, 10 maal in LT, 1 maal in <L 1809:2. De betekenis lijkt hetzelfde als type=1. Zie ook Codering13ptvr.txt. <L 67:51> <C 279.ZETTEN> ghset <L 67:51 geeft aan dat het om een noot onder aan de pagina gaat. Niet duidelijk is op welke voorafgaande regel de noot betrekking heeft. Idem <L

Vervolgens haalt oorknr.awk de originele, gelokaliseerde en gedateerde oorkonden uit CGA.txt, voornamelijk dankzij oorknr.awk. De relevante informatie zit tussen oorkondennummer, zie 6a., en <header> van de volgende oorkonde. Het programma zoekt op oorknr .o:, <L en <header>. Haalt de originele oorkonden met oorkondennummers eruit

## 4.4 6c\_lemmasopeenregel

Dit programma gooit de regelnummers <L <LA etc. weg en zet elke lemma's op een eigen regel door <Q en <C voor "begin nieuwe regel" geschikt te maken.

### 4.4.1 6d\_afkortingen

Vervangt afkortingen van het type x<A >y</A>z door Adelheidformaat.

## 4.5 6d\_afkortingen

## 4.6 6e\_crmformaat

Hier gebeuren vier dingen:

- alles wat geen oorkondennummer of lijnregel is vervalt
- crm.awk maakt ruw Adelheidformaat
- volgorde code-lemma wordt omgekeerd
- opgeloste afkortingen krijgen het Adelheidformaat

## 4.7 6f\_majusminuscuul

Centraal is hier 'tr' dat hoofdletters van de lemmanotaties vervangt door kleine letters. Eromheen staan awk en join om de juiste kolommen te splitsen en weer samen te voegen.

NB. Sommige oorkonden of rekeningen bestaan voornamelijk uit latijn:

```
docId>0025</docId> = .o:H036p26302
docId>0030</docId> = .o:H036p26402
docId>1230</docId> = .o:H036p29306
docId>1329</docId> = .o:H036p29412
docId>1601</docId> = .o:H036p29718
redelijk wat latijn docId>1669</docId> = .o:H036p29807
docId>0001</docId> = .o:I241p23901
docId>0019a</docId> = .o:K244p27101
redelijk wat latijn docId>1923</docId> = .o:I154p30001
docId>0010</docId> = .o:N072p25301
```

---

369:1> <q> witscerres <VN type=1> <L369:51 <q> witscerres In werkelijkheid <L 369:24. Dit lijkt een structureel probleem. <L 2853:59 laat bij vergelijking met het CG-pdf zien dat het een rommeltje is: bij <L 606:21 carnition is de noot: lees carnation niet opgenomen. Of en hoe dit probleem oplosbaar is is mij verre van duidelijk.

#### 4.7.1 6g\_anderecode

Eventueel nog te ontwikkelen: Codes serie 500 tot 800 vereenvoudigen. Ze kunnen worden verwijderd door de lemma's, meestal als 'latijn' gelemmatiseerd, te verwijderen. Soms zijn de eigennamen wel als middelnederlands gecodeerd. Die blijven dan over.

### 4.8 6h\_ring

&ringrepresentatie moet nog ontwikkeld

#### 4.8.1 Splitsing LEMMA+LEMMA

Eventueel nog te ontwikkelen.

## 5 Directory en filestructuur

Hier volgt een overzicht van de directory en filestructuur, afhankelijk van /Users/PietR/GYSSELING/  
Alleen de relatief onveranderlijke paden zijn weergegeven.

### 5.1 CGBoekPdf

CGBoekPdf bevat:

- het CG.pdf ambtelijk.
- het CG.pdf literair.
- 1988\_Corpus van middelnederlandse teksten t/m jaar 1300.pdf
- 1988\_Krantenart over Maurits Gysseling.pdf

### 5.2 Data-2

Data-2 bevat o.a. de files:

- CG1.rtf
- corpus\_codering.txt: de oorspronkelijke toelichting bij het corpus
- plaats.txt.
- regio.txt.

Daarnaast drie directories:

#### 5.2.1 BRON

Bevat het corpus in .zipvorm.

#### 5.2.2 LITERAIR

Heinric van Veldeke, etc.

### 5.2.3 OORKONDEN

Bevat de losse teksten van het CG ambtelijk plus 6 toegevoegde (series) teksten, te weten:

- 0019a.fromdb bevat het Antwerps Obituarium Antwerps Obituarium LA heeft soms licht andere codes: kl KL soms voor latijns woord [hand] op einde regel, volgende pas [P] oid
- 0087a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 0520a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 0612a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 0635a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 1291a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 1300a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 1409a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 1719a.fromdb bevat een van de 8 oorkonden Taal en Tongval 1988
- 0295.fromdb bevat het Goederenregister van Oudenbiezen PM (datering 1280-1290, door ptvr voorlopig op 1285 gezet)
- 0613.fromdb bevat de Schepenbrief Antwerpen van 1285
- 0178.fromdb bevat het Landrecht van Grimbergen
- 0420.fromdb bevat de Oorkonde Corpus Wilhelm

NB 0032a.fromdb Renterol Deinze is leeg, dus kan weg.

## 5.3 werk

werk bevat o.a.:

- alle deelprogramma's met sh aan te roepen. 0a\_ t/m 5h\_
- alle tussenresultaten, bijv. result\_3c.txt
- corp13new.txt
- CGA.txt
- CGR13.txt
- corp14.txt

Daarnaast de directories:

### 5.3.1 Aantltellr

Hier staan de tellingen per programma van de veranderingen die door dat programma tot stand zijn gebracht. De 5[a-h]\_programma's die de omzetting van CGA.txt naar CGR13.txt betreffen tellen niet mee.

### 5.3.2 DOCUMENTATIE

- Kloekenrs
- PREALABELEVERBETERINGENINCGINL bevat:
  - lokaliseringsfouten
  - fatalewoorden: lemma's met te hoge ASCII-waarden, bijvoorbeeld SUFFICIËNT
  - onjuiste filenamen van oorkonden
  - een ASCII-volgordefout
- lokaliseringsgegevens.txt, bevat ook gegevens "origineel", "afschrift", "netschrift" e.d.
- backup13new.txt
- voorbereidingen op sed-scripts
- fouten en problemen
- JOINTTEST
- streep: problemen met en # en \*
- etc

### 5.3.3 PROBEERSELS

Allerhande soorten problemen en fouten waaraan gewerkt wordt of nog gewerkt moet worden.

### 5.3.4 SED

Hierin zitten alle sed-scripts en awk-programmaatjes die worden aangeroepen door de programma's 0a\_ t/m 5h\_.

### 5.3.5 TELLERS

Hier kunnen snel programma's voor tellingen van specifieke vormen worden ontwikkeld, als ze er al niet zijn.

### 5.3.6 VANCGNAARCG13

Hier wordt deze tekst aangemaakt met behulp van LATEX.

## 6 Toelichting bij de bestanden

Deze sectie is gebaseerd op de oorspronkelijke CG-toelichting van het INL: corpus.codering.txt, zie Data-2. Deze is echter onvolledig en deels onjuist. Hier wordt deze tekst aangevuld en verbeterd, per passage. De coderingen zijn afgesplitst en staan in de volgende sectie.



## 6.1 Voorbeeld van een stukje gecodeerde tekst:

<L 44:1> <C 810\_DAT> dat <C 450\_ZIJN> zijn <C 0\_ADEM> adem <C 203\_GAAN> gaet <C 501\_ALDOOR> al+dure  
<L 44:2> <C 500\_OOK> oec <C 440+203\_MEN+VINDEN> vintmen <C 441\_SOMMIG> somege <C 1\_SCHRIFTUUR> scripture  
<L 44:3> <C 421\_DIE> die <C 700\_VAN> van <C 415\_DEZE> des<A >er</A> <C 4\_BEEST> beesten <C 203\_WETEN> weet  
<L 44:4> <C 406+800+421\_ZIJ+EN+DIE> entiese <C 19\_BISON> bisontes <C 213\_HETEN> heet  
<L 44:5> <C 9\_BONASUS> Bonacus <C 210\_ZIJN> es <C 489\_EEN> .i. <C 0\_DIER> dier  
<L 44:6> <C 420\_DAT> dat <C 203\_HEBBEN> heuet  
<C 0+470\_HOOFD+HET> thouet <C 700+474\_NA+DE> naden  
<C 0\_STIER> stier  
<L 44:7> <C 470\_HET> dat <C 0\_LIJF> lijf <C 471+800\_DE+EN> entie  
<C 14\_MAAN> mane<A >n</A> <C 501\_MEDE> mede  
<L 44:8> <C 500\_RECHT> recht <C 475+700\_DE+NA> nader  
<C 11\_PAARD> parde <C 1\_ZEDE> zede

NB In 44:4 moet

406+800+421\_ZIJ+EN+DIE> entiese

800+421+406\_EN+DIE+ZIJ

worden. Deze fout komt zeer veel voor, zie ook HOOFD+HET, DE+EN, DE+NA.

DE+NA -> NA+DE -> NA HET!!!

## 6.2 Metatekst

Metatekst in de vorm van dorsotekst die later is aangebracht en noten en commentaar van Gysseling dient te worden gescheiden van de oorspronkelijke middeleeuwse gedateerde hoofdttekst. In CG-boek is het commentaar van Gysseling in het algemeen duidelijk het geval middels ronde haakjes. Deze zijn in CG-INL echter weggevalen.

NB Alleen <L 1026:25 bevat (>: "<L 1026:25> <q> (?) <q> ipsius" etc. Dat is in tikfoutjes.sed veranderd in [?].

Een extra probleem is dat in het CG-INL deze metatekst soms selectief uit CG-boek is overgenomen. Dit laatste kan leiden tot onbegrijpelijke informatie. Een voorbeeld is in oorkonde 533 de regels 775:11 en 775:12, waar het commentaar van Gysseling (begin 14e e.) niet is opgenomen, maar wel het tweemaal volgende (idem:), maar dan zonder ( ), dat daardoor nu in de lucht hangt.

Metatekst staat vaak ook in de tekst zelf. Bij Gysseling is een en ander duidelijk, maar in het CG-INL is dat niet altijd langer zo. Zo zijn de haakjes bij toevoegingen weggelaten, bijvoorbeeld "(toegevoegd idem:)" wordt "toegevoegd idem:". Daardoor is soms niet duidelijk waar het commentaar van Gysseling begint en ophoudt. Dit leidt tot onnodige ruis in de tekst.

Ook [ ] zijn niet consequent gebruikt en toegevoegd. Cursief [ ] bij Gysseling lijkt modern. Daardoor lopen in het CG-INL gereconstrueerde middelnederlandse woorden en metatekst door elkaar. Metatekst komt in het CG-INL alleen expliciet voor tussen [ ], bijvoorbeeld [hand] en varianten die als <q> zijn

gecodeerd (soms ook als middelnederlands?). Ook middelnederlandse woorden die door de moderne uitgever hersteld zijn staan echter tussen [ ], bijvoorbeeld [vercopen]. Dat is soms lastig onderscheiden. Een vorm als [ver]copen is wel makkelijk van de metavormen te scheiden. De metaprogramma's brengen het onderscheid weer aan.

### 6.3 Handproblemen, deels onderdeel van metaprobleem

Gysseling meldt in principe de handovereenkomst met andere teksten boven de tekst. De 'hand' in CG-pdf verwijst dan naar oorkonden van dezelfde hand, maar deze informatie is in het CG-INL niet opgenomen. Het aanbrengen van deze meta-informatie zou uitbesteed kunnen worden aan een vrijwilliger.

Opmerkingen over de handen staan echter ook vaak in middeleeuwse tekst zelf, met name als de hand binnen een tekst wisselt. Voorbeeld: <L 1:38> <q> [hand] <q> [A] <C 029\_REINIER> Reiner<A >us</A> <q> .filius. Deze opmerkingen moeten van het middelnederlands onderscheiden worden. De handwisselingen zijn door Gysseling tussen haakjes en cursief aangegeven. Deze invoegingen van 'hand' als ook wel andere metatekst zijn echter in de CG-INLversie ten onrechte zonder cursief en ( ) haakjes opgenomen, zie ook boven.

- 1663x [hand], meestal gevolgd door [A], [A?] of [A:], [AA:] [AB:], [Aa], [Ab], [Ai], [B], [E], etc maar niet regel <L 2643:20 waar 1'he (= 1e) volgt en verdere metatekst.

- 13x [hand,] altijd? gevolgd door andere metatekst? Ja, maar altijd fout.

<LA 40:14 <L 161:13 237:8 237:16 237:37 238:1 238:34 240:1 240:40 241:16 241:16 241:29 241:39 is de , Soms, teveel soms omkeren met hand.

Idem regel 261 <L 6:15 <LA 37:15 <L 63:12 163:22 173:8 181:13 496:31 497:10 497:14 497:21 501:42 503:2 510:9 512:8 514:24 514:25 staat er een soort teken tussen [hand] en [B] dat is allemaal fout: op p. 6 G: [hand B] soort -teken - 11x hand met verdere tekst van Gysseling, zie hand.\* lemma

- 5x [handen:], altijd voorafgegaan door [diverse]

- 3x hand: , met enige tekst vooraf van Gysseling

- 1x [toegevoegd, andere hand:] [hand:] <164:33

- 1x [hnd] [P:] drukfout Antw Ob (naast [hand]) Verbeterd in 0a

- 1x [hand.V:] Antw Obituarium <docId>0019a. Verbeterd in 0a

- 1x [handA] Inconsequente notatie: Verbeterd in 0a

### 6.4 Coderingalslatijnoffrans

Bij de codering als latijn of frans zijn grensgevallen: a) romeinse getallen en b) eigennamen. Of romeinse getallen als LATIJN of in arabische cijfers gecodeerd zijn maakt niet veel uit. Hebben eigennamen latijnse uitgangen dan heeft een codering LATIJN de voorkeur. Maar eigennamen zijn al in het CG-INL niet altijd consequent zo opgelost. Overhaupt is de eigenaamtoekenning mogelijk niet altijd optimaal. Een deskundige, Tanneke Schoonheim?, zou daar mogelijk tezijnertijd wat aan kunnen doen. Dit lijkt voorlopig geen ernstig probleem.

## 6.5 Coderingsformaat

De codering van PoS, lemma en token is:

<C (tagnummers)\_(lemmata)> (token).

Een token kan meerdere woordsoort-tags en lemmata hebben:

<C 0+470\_HOOFD+HET> thouet

Hierbij hoeft de volgorde van de lemmadelen (zoals hier HOOFD en HET) niet te kloppen met het token, (dit wordt HET+HOOFD dankzij 4c\_lemmaENcodeomkeren) bij meer dan een +, er zijn er maximaal vier, worden alle volgordes gebruikt.

Houd verder rekening met de <A>-tag die voor uitgeschreven afkortingen gebruikt wordt:

<C 415\_DEZE> des<A >er</A>

staat voor een afkorting uitgeschreven tot "deser"

Scheidbare woorden worden in de tags gelinkt met gebruik van de tekens '#', '\*' en '@': Voorbeeld (koppeling "daer" en "an", beide delen krijgen woordsoort 610=voornaamwoordelijk bijw.):

<L 4:11> <C 800\_WANT> Want <C 610\*\_DAAR-> daer <C 560\_NE> ne  
<C 210\_MOGEN> mach <C 440\_GEEN> ghen <C 0\_VERLIES> verlies < C  
610\_-AAN> an <C 260\_WEZEN> weese<A >n</A>

## 6.6 Opmerkingen bij scheidbare woorden

- @ komt niet voor. Wat wel voorkomt is: %, &, ?
- Kruistekentje komt 3x voor: geeft †
- + betekent 4x ± in jaartal, bv [%+1272:] in <L 238:1
- 't% 19x ¶(soort paragraafteken)
- & 1x in 2728:42 betekent £
- % in terr%a en primerossa% is een tikfout in CG-INL, verbeterd in a0\_voorcorrecties.
- & komt voor als middelnederlands.
- ? komt ook voor in lemma's en geïsoleerd, zie bv Id1486 waar CG (?) heeft.
- Andere coderingen tussen punthaken, zoals: VOORBEELD.

## 6.7

<header>

<docId>7000</docId>

NB Onder deze regel is dankzij 3a\_origineelofniet toegevoegd.

<origstatus>Origineel</origstatus>

toegevoegd

<genre>W</genre>

<bron>

```

<bron_afk>Bern.</bron_afk>
<bron_oms>Glossarium Bernense</bron_oms>
</bron>
</header>

```

## 6.8 Atlaskenmerk

```

<atlas_kenmerk atlas_kenmerkkode='A.S.' Degene die gekodeerd heeft?

```

## 6.9 tijdstatus

```

tijdstatus='1', tijdstatus='2', tijdstatus='5'

```

## 6.10 datering

```

<datering jaar_tot='1240' jaar_van='1240' tijdsoort='1' tijdstatus='1'>
NB tijdsoort='1' klopt vaak en is zinnig
NB tijdsoort='2' (t/m 6? docId 1933) betreft in dorso, cf 9 Gent 1253
NB docId 0021 tijdsoort='1' 1344 moet dat niet tijdsoort='2' zijn?
NB betekenis tijdstatus='2' is onduidelijk
<lokalisering kanselarij='0' plaats_1=" plaats_2=" regio_1='12' regio_2=" regiostatus='1'>
Hoe datering van latere opmerkingen scheiden van hoofdtekst?

```

## 6.11 statushand

```

<statushand statushandkode='mn'>
geven informatie over lokalisering, datering, bron
<atlas_lokalisering> komt van de atlas van Marijke Mooijaart.
Haar lokaliserings zijn in tegenstelling tot die van Gysseling betrouwbaar.
In haar atlas is veel van Brugge niet opgenomen, maar in het CG-INL heeft zij
ze wel gelokaliseerd.
Ze heeft ook afschriften gelokaliseerd cf 0002B, 0011. Of die kloppen?
Ik geef ze geen Kloekenr
De nummering van Mooijaart is licht afwijkend in gevallen als 0009 0009A
0009AA Gent, waar zij 0009 geeft en ik 0009A en 0009AA beide opneem.
Heeft zij alleen de A versie?
Ik deel regionrs uit voorlopig volgens eigen systeem bv: in plaats van Leefdaal
e.o. zoals Mooijaart, geef ik Kloekenr van Kortenbergh P076
NB waar Gysseling zegt ± 1294 is dat geworden 1294.
Vaak opmerking in de relevante lokaliseringsfiles.
Kan ik Mooijaarts OOST-VL NOORD-WEST-systeem gebruiken?
Moeten Sijsele H041 en Brugge H043 (beter?) gescheiden?
Zie hierover Leroy 1993, p.209 (TT 45).
In bijvoorbeeld docId 92, 93, 94 is docId niet herhaald, eenmaal lokalisering
Gent en zeer veel wisselende dateringen. Een project op zich om deze knoop te
ontwarren.

```

## 6.12

dubbeltelling:

184 regio\_1=" te hoog voor lokalisatie; Willem Rooms koning docId0012

1258 regio\_1='1' Hgk Hollandse graf. kanselarij

704 regio\_1='10' Zuid-Brabant

12 regio\_1='11' Limburg

40 regio\_1='12' Belg. Limburg, in Mooijaart onder 11

256 regio\_1='2' West-Holland

98 regio\_1='3' Oost-Holland: bij Mooijaart vaak/soms Utrecht e.o.

8 regio\_1='4' Utrecht

110 regio\_1='5' Zeeland

120 regio\_1='6' Noord-Brabant

24 regio\_1='7' West-Vlaanderen met Gent

4844 regio\_1='8' Oost-Vlaanderen met Brugge

1144 regio\_1='9' West-Vlaanderen met Gent

## 6.13

<statushand> met het attribuut 'statushandcode' geeft informatie over de handwisselingen. Invullingen statushandcode:

- 'ad': eerdere documentregels, niet van de hoofdhand.  
NB Is bedoeld "volgend op 'md'"? Niet altijd in elk geval. Eerder dan Gysseling? Nader de betekenis proberen te bepalen. In docId 0021 regel 25 noot uit later cartularium, zou 'md' moeten zijn? Idem docId 0022 <L 77:13 opm uit 15e eeuw. Idem docId 0024 <L 79:10 opm uit 14e eeuw.
- 'an': eerdere documentregels van de hoofdhand
- 'md': latere documentregels, niet van de hoofdhand NB Dit lijkt om dorsotekst ed te gaan, zie 0d\_dorsomarkeren.  
NB Voor de verhouding 'mn' en 'ad' zie 462:11 en 462:12. Beide dorso, maar 'ad' kan ook midden in tekst staan.  
Voor de verhouding 'md' en 'ad' zie 476:39, waar 'ad' niks doet want geen <L. <C of <q aan begin regel als kapstok
- 'mn': latere documentregels van de hoofdhand.  
NB Afschrift docId 0011 naar Facsimile docId 0010 naar moderne uitgave? Opmerking: het 'eerder' en 'later' is m.b.t. de tijd van schrijven. Voorbeeld: Een oorkonde kon later voorzien worden van aantekeningen boven of onder de al geschreven tekst. Dit kon gedaan worden door degene die de oorkonde had geschreven of door iemand anders.  
Betekenis van <Ldorsoad 1569:23>  
<C 500.ITEM<sub>i</sub> Jtem <C 472+002.DE+ZATERDAG> sat<A >er</A >daghes <C 700.NA> na <C 024.SINKSEN> sinxenen. en volgende: gelijktijdige latere hand, dus niet dorso

Aantal voorkomens:

- 3540 statushandcode='ad'>

- 5304 statushandkode='an'> lijkt te gaan om de gezochte oorkonde: origineel of kopie
- 3118 statushandkode='md'>
- 974 statushandkode='mn'>

## 6.14

De code <VN type=1> komt 7194x voor, betekent dat voor de transcriptie in de tekst gekozen is, maar dat daar wat mee is:

Voorbeelden:

- 410\_DIT> dit <VN type=1> in noot: dit of det andere voorbeelden: lees armer voor amer, o verbeterd in u,
- van de m alleen derde beentje over, d onzeker De voorbeelden komen van CG-boek p. 97 regel 28 tot 4, 98 regel 2
- 1809:2> <C 001.STEIGER> steghere <VN type=2> Dit type komt maar 11x voor
- 10x in LT en 1x in afschrift B van de oorkonde staat steghre
- Ook: type=1 29:24 regel 1318 blijkt op noot te slaan. Idem 1312 dach van het type boven de regel bijgeschreven.

Om van het corpus plain tekst met tag en lemma te maken kan bijvoorbeeld iets als het volgende perl script gebruikt worden:

Nog uitwerken!

## 7 Codering

NB1 De codering is ontleend aan die van CRM14 en daarmee voor het grootste deel identiek.

NB2. Als bij hercodering of verbetering van de codering aarzeling bestond tussen vorm en functie, dan is gekozen voor (code) vorm. Een voorbeeld is

<C 500\_HET-EERST> Terst wordt opgesplitst in <C 470+310\_HET+EERST> Terst

Van bijwoord naar lidwoord plus telwoord. Als beter opgesplitst kon worden in twee woorden, is dat meestal gedaan: HET en EERST. Omdat HET en EERST hier niet te splitsen zijn, is de +-notatie gebruikt. NB3. In het CG-INL verwijzen woorden naar elkaar, waar dat in het CRM14 niet gebeurt. Dit systeem is, vaak na de nodige correctie, gehandhaafd als volgt: Het eerste woord krijgt een # het tweede een #. De twee (soms drie) woorden worden niet voorzien van -, aangezien ze nooit aan elkaar vast voorkomen. Dus \_ZO>, niet \_ZO->. Voorbeelden:

<C 800\_BEIDE> bede .... <C 800#\_EN> ende \_ZO> .... #\_ZO> \_WEDER> .... #\_OF> 700\*\_TE> the .... <C 700#\_WAART>

NB4. Soms komt met twee ## voor.

<C 210+403\_ZIJN+HET> ist <C 001\_ERF> erue <C 210+403#\_ZIJN+HET> ist <C 000\_LEEN> leen <C 210+403#\_ZIJN+HET> ist <C 000\_PACHT>

[pa]cht

Soms \* en \* plus ##.

<C 251.AF-> of <C 800.EN> ende <C 251.OP-> vp <C 700.TE> te <C 251##.-VAREN> vaerne

De betekenis van de tags is:

0 zelfstandig naamwoord enkelvoud, geen uitgang

1 zelfstandig naamwoord enkelvoud, uitgang -e

2 zelfstandig naamwoord enkelvoud, uitgang -s/-th

3 zelfstandig naamwoord enkelvoud, uitgang -t

4 zelfstandig naamwoord enkelvoud, uitgang -n

5 zelfstandig naamwoord enkelvoud, uitgang -r/-re

6 zelfstandig naamwoord enkelvoud, uitgang -a

9 zelfstandig naamwoord enkelvoud, uitgang latijn, onduidelijk of evident fout geschreven

# Deze codes zijn aangevuld met 00 ervoor.

10 zelfstandig naamwoord meervoud, geen uitgang

11 zelfstandig naamwoord meervoud, uitgang -e

12 zelfstandig naamwoord meervoud, uitgang -s/-th

13 zelfstandig naamwoord meervoud, uitgang -t

14 zelfstandig naamwoord meervoud, uitgang -n

15 zelfstandig naamwoord meervoud, uitgang -r/-re

16 zelfstandig naamwoord meervoud, uitgang -a

19 zelfstandig naamwoord meervoud, uitgang latijn, onduidelijk of evident fout geschreven

20 eigennaam, geen uitgang

21 eigennaam, uitgang -e

22 eigennaam, uitgang -s/-th

23 eigennaam, uitgang -t

24 eigennaam, uitgang -n

25 eigennaam, uitgang -r/-re

26 eigennaam, uitgang -a

29 eigennaam, uitgang latijn, onduidelijk of evident fout geschreven

90 zelfstandig naamwoord probleemgeval, geen uitgang

91 zelfstandig naamwoord probleemgeval, uitgang -e

92 zelfstandig naamwoord probleemgeval, uitgang -s/-th

93 zelfstandig naamwoord probleemgeval, uitgang -t

94 zelfstandig naamwoord probleemgeval, uitgang -n

95 zelfstandig naamwoord probleemgeval, uitgang -r/-re

96 zelfstandig naamwoord probleemgeval, uitgang -a

99 zelfstandig naamwoord probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven

# Deze codes zijn aangevuld met 0 ervoor.

100 bijvoeglijk naamwoord enkelvoud, geen uitgang

101 bijvoeglijk naamwoord enkelvoud, uitgang -e

102 bijvoeglijk naamwoord enkelvoud, uitgang -s/-th

103 bijvoeglijk naamwoord enkelvoud, uitgang -t

104 bijvoeglijk naamwoord enkelvoud, uitgang -n

105 bijvoeglijk naamwoord enkelvoud, uitgang -r/-re

106 bijvoeglijk naamwoord enkelvoud, uitgang -a

109 bijvoeglijk naamwoord enkelvoud, uitgang latijn, onduidelijk of evident fout  
 geschreven  
 110 bijvoeglijk naamwoord meervoud, geen uitgang  
 111 bijvoeglijk naamwoord meervoud, uitgang -e  
 112 bijvoeglijk naamwoord meervoud, uitgang -s/-th  
 113 bijvoeglijk naamwoord meervoud, uitgang -t  
 114 bijvoeglijk naamwoord meervoud, uitgang -n  
 115 bijvoeglijk naamwoord meervoud, uitgang -r/-re  
 116 bijvoeglijk naamwoord meervoud, uitgang -a  
 119 bijvoeglijk naamwoord meervoud, uitgang latijn, onduidelijk of evident fout  
 geschreven  
 190 bijvoeglijk naamwoord probleemgeval, geen uitgang  
 191 bijvoeglijk naamwoord probleemgeval, uitgang -e  
 192 bijvoeglijk naamwoord probleemgeval, uitgang -s/-th  
 193 bijvoeglijk naamwoord probleemgeval, uitgang -t  
 194 bijvoeglijk naamwoord probleemgeval, uitgang -n  
 195 bijvoeglijk naamwoord probleemgeval, uitgang -r/-re  
 196 bijvoeglijk naamwoord probleemgeval, uitgang -a  
 199 bijvoeglijk naamwoord probleemgeval, uitgang latijn, onduidelijk of evident  
 fout geschreven  
 200 persoonsvorm o.t.t. (zelfst.ww.), geen uitgang  
 201 persoonsvorm o.t.t. (zelfst.ww.), uitgang -e  
 202 persoonsvorm o.t.t. (zelfst.ww.), uitgang -s/-th  
 203 persoonsvorm o.t.t. (zelfst.ww.), uitgang -t  
 204 persoonsvorm o.t.t. (zelfst.ww.), uitgang -n  
 205 persoonsvorm o.t.t. (zelfst.ww.), uitgang -r/-re  
 206 persoonsvorm o.t.t. (zelfst.ww.), uitgang -a  
 209 persoonsvorm o.t.t. (zelfst.ww.), uitgang latijn, onduidelijk of evident fout  
 geschreven  
 210 persoonsvorm o.t.t. (koppel-/hulpww.), geen uitgang  
 211 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang -e  
 212 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang -s/-th  
 213 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang -t  
 214 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang -n  
 215 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang -r/-re  
 216 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang -a  
 219 persoonsvorm o.t.t. (koppel-/hulpww.), uitgang latijn, onduidelijk of evi-  
 dent fout geschreven  
 220 persoonsvorm o.v.t. (zelfst.ww.), geen uitgang  
 221 persoonsvorm o.v.t. (zelfst.ww.), uitgang -e  
 222 persoonsvorm o.v.t. (zelfst.ww.), uitgang -s/-th  
 223 persoonsvorm o.v.t. (zelfst.ww.), uitgang -t  
 224 persoonsvorm o.v.t. (zelfst.ww.), uitgang -n  
 225 persoonsvorm o.v.t. (zelfst.ww.), uitgang -r/-re  
 226 persoonsvorm o.v.t. (zelfst.ww.), uitgang -a  
 229 persoonsvorm o.v.t. (zelfst.ww.), uitgang latijn, onduidelijk of evident fout  
 geschreven  
 230 persoonsvorm o.v.t. (koppel-/hulpww.), geen uitgang  
 231 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang -e  
 232 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang -s/-th



233 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang -t  
 234 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang -n  
 235 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang -r/-re  
 236 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang -a  
 239 persoonsvorm o.v.t. (koppel-/hulpww.), uitgang latijn, onduidelijk of evident fout geschreven  
 240 imperatief, geen uitgang  
 241 imperatief, uitgang -e  
 242 imperatief, uitgang -s/-th  
 243 imperatief, uitgang -t  
 244 imperatief, uitgang -n  
 245 imperatief, uitgang -r/-re  
 246 imperatief, uitgang -a  
 249 imperatief, uitgang latijn, onduidelijk of evident fout geschreven  
 250 infinitief (zelfst.ww.), geen uitgang  
 251 infinitief (zelfst.ww.), uitgang -e  
 252 infinitief (zelfst.ww.), uitgang -s/-th  
 253 infinitief (zelfst.ww.), uitgang -t  
 254 infinitief (zelfst.ww.), uitgang -n  
 255 infinitief (zelfst.ww.), uitgang -r/-re  
 256 infinitief (zelfst.ww.), uitgang -a  
 259 infinitief (zelfst.ww.), uitgang latijn, onduidelijk of evident fout geschreven  
 260 infinitief (koppel-/hulpww.), geen uitgang  
 PM 260 bestaat niet in CRM14  
 261 infinitief (koppel-/hulpww.), uitgang -e  
 262 infinitief (koppel-/hulpww.), uitgang -s/-th  
 263 infinitief (koppel-/hulpww.), uitgang -t  
 264 infinitief (koppel-/hulpww.), uitgang -n  
 265 infinitief (koppel-/hulpww.), uitgang -r/-re  
 266 infinitief (koppel-/hulpww.), uitgang -a  
 269 infinitief (koppel-/hulpww.), uitgang latijn, onduidelijk of evident fout geschreven  
 270 deelwoord (zelfst.ww.), geen uitgang  
 271 deelwoord (zelfst.ww.), uitgang -e  
 272 deelwoord (zelfst.ww.), uitgang -s/-th  
 273 deelwoord (zelfst.ww.), uitgang -t  
 274 deelwoord (zelfst.ww.), uitgang -n  
 275 deelwoord (zelfst.ww.), uitgang -r/-re  
 276 deelwoord (zelfst.ww.), uitgang -a  
 279 deelwoord (zelfst.ww.), uitgang latijn, onduidelijk of evident fout geschreven  
 280 deelwoord (koppel-/hulpww.), geen uitgang  
 PM 280 bestaat niet in CRM14  
 281 deelwoord (koppel-/hulpww.), uitgang -e  
 282 deelwoord (koppel-/hulpww.), uitgang -s/-th  
 283 deelwoord (koppel-/hulpww.), uitgang -t  
 284 deelwoord (koppel-/hulpww.), uitgang -n  
 285 deelwoord (koppel-/hulpww.), uitgang -r/-re  
 286 deelwoord (koppel-/hulpww.), uitgang -a  
 289 deelwoord (koppel-/hulpww.), uitgang latijn, onduidelijk of evident fout geschreven  
 290 werkwoord probleemgeval, geen uitgang

291 werkwoord probleemgeval, uitgang -e  
 292 werkwoord probleemgeval, uitgang -s/-th  
 293 werkwoord probleemgeval, uitgang -t  
 294 werkwoord probleemgeval, uitgang -n  
 295 werkwoord probleemgeval, uitgang -r/-re  
 296 werkwoord probleemgeval, uitgang -a  
 299 werkwoord probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven  
 300 hoofdtelwoord, geen uitgang  
 301 hoofdtelwoord, uitgang -e  
 302 hoofdtelwoord, uitgang -s/-th  
 303 hoofdtelwoord, uitgang -t  
 304 hoofdtelwoord, uitgang -n  
 305 hoofdtelwoord, uitgang -r/-re  
 306 hoofdtelwoord, uitgang -a  
 309 hoofdtelwoord, romeins getal, uitgang latijn, onduidelijk of evident fout geschreven  
 310 rangtelwoord, geen uitgang  
 311 rangtelwoord, uitgang -e  
 312 rangtelwoord, uitgang -s/-th  
 313 rangtelwoord, uitgang -t  
 314 rangtelwoord, uitgang -n  
 315 rangtelwoord, uitgang -r/-re  
 316 rangtelwoord, uitgang -a  
 319 rangtelwoord, romeins getal, uitgang latijn, onduidelijk of evident fout geschreven  
 320 onbepaald telwoord, geen uitgang  
 321 onbepaald telwoord, uitgang -e  
 322 onbepaald telwoord, uitgang -s/-th  
 323 onbepaald telwoord, uitgang -t  
 324 onbepaald telwoord, uitgang -n  
 325 onbepaald telwoord, uitgang -r/-re  
 326 onbepaald telwoord, uitgang -a  
 329 onbepaald telwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 390 telwoord probleemgeval, geen uitgang  
 391 telwoord probleemgeval, uitgang -e  
 392 telwoord probleemgeval, uitgang -s/-th  
 393 telwoord probleemgeval, uitgang -t  
 394 telwoord probleemgeval, uitgang -n  
 395 telwoord probleemgeval, uitgang -r/-re  
 396 telwoord probleemgeval, uitgang -a  
 399 telwoord probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven  
 401 persoonlijk voornaamwoord, 1e persoon enkelvoud  
 402 persoonlijk voornaamwoord, 2e persoon enkelvoud  
 403 persoonlijk voornaamwoord, 3e persoon enkelvoud  
 404 persoonlijk voornaamwoord, 1e persoon meervoud  
 405 persoonlijk voornaamwoord, 2e persoon meervoud  
 406 persoonlijk voornaamwoord, 3e persoon meervoud  
 409 probleemgeval  
 410 aanwijzend voornaamwoord, geen uitgang  
 411 aanwijzend voornaamwoord, uitgang -e

412 aanwijzend voornaamwoord, uitgang -s/-th  
 413 aanwijzend voornaamwoord, uitgang -t  
 414 aanwijzend voornaamwoord, uitgang -n  
 415 aanwijzend voornaamwoord, uitgang -r/-re  
 416 aanwijzend voornaamwoord, uitgang -a  
 419 aanwijzend voornaamwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 420 betrekkelijk voornaamwoord, geen uitgang  
 421 betrekkelijk voornaamwoord, uitgang -e  
 422 betrekkelijk voornaamwoord, uitgang -s/-th  
 423 betrekkelijk voornaamwoord, uitgang -t  
 424 betrekkelijk voornaamwoord, uitgang -n  
 425 betrekkelijk voornaamwoord, uitgang -r/-re  
 426 betrekkelijk voornaamwoord, uitgang -a  
 429 betrekkelijk voornaamwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 430 vragend voornaamwoord, geen uitgang  
 431 vragend voornaamwoord, uitgang -e  
 432 vragend voornaamwoord, uitgang -s/-th  
 433 vragend voornaamwoord, uitgang -t  
 434 vragend voornaamwoord, uitgang -n  
 435 vragend voornaamwoord, uitgang -r/-re  
 436 vragend voornaamwoord, uitgang -a  
 439 vragend voornaamwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 440 onbepaald voornaamwoord, geen uitgang  
 441 onbepaald voornaamwoord, uitgang -e  
 442 onbepaald voornaamwoord, uitgang -s/-th  
 443 onbepaald voornaamwoord, uitgang -t  
 444 onbepaald voornaamwoord, uitgang -n  
 445 onbepaald voornaamwoord, uitgang -r/-re  
 446 onbepaald voornaamwoord, uitgang -a  
 449 onbepaald voornaamwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 450 bezittelijk voornaamwoord, geen uitgang  
 451 bezittelijk voornaamwoord, uitgang -e  
 452 bezittelijk voornaamwoord, uitgang -s/-th  
 453 bezittelijk voornaamwoord, uitgang -t  
 454 bezittelijk voornaamwoord, uitgang -n  
 455 bezittelijk voornaamwoord, uitgang -r/-re  
 456 bezittelijk voornaamwoord, uitgang -a  
 459 bezittelijk voornaamwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 460 wederkerend/wederkerig voornaamwoord  
 461 wederkerend/wederkerig voornaamwoord, 1e persoon enkelvoud  
 462 wederkerend/wederkerig voornaamwoord, 2e persoon enkelvoud  
 463 wederkerend/wederkerig voornaamwoord, 3e persoon enkelvoud  
 464 wederkerend/wederkerig voornaamwoord, 1e persoon meervoud  
 465 wederkerend/wederkerig voornaamwoord, 2e persoon meervoud  
 466 wederkerend/wederkerig voornaamwoord, 3e persoon meervoud  
 469 probleemgeval  
 470 bepaald lidwoord, geen uitgang  
 471 bepaald lidwoord, uitgang -e  
 472 bepaald lidwoord, uitgang -s/-th

473 bepaald lidwoord, uitgang -t  
 474 bepaald lidwoord, uitgang -n  
 475 bepaald lidwoord, uitgang -r/-re  
 476 bepaald lidwoord, uitgang -a  
 479 bepaald lidwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 480 onbepaald lidwoord, geen uitgang  
 481 onbepaald lidwoord, uitgang -e  
 482 onbepaald lidwoord, uitgang -s/-th  
 483 onbepaald lidwoord, uitgang -t  
 484 onbepaald lidwoord, uitgang -n  
 485 onbepaald lidwoord, uitgang -r/-re  
 486 onbepaald lidwoord, uitgang -a  
 489 onbepaald lidwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 490 voornaamwoord probleemgeval, geen uitgang  
 491 voornaamwoord probleemgeval, uitgang -e  
 492 voornaamwoord probleemgeval, uitgang -s/-th  
 493 voornaamwoord probleemgeval, uitgang -t  
 494 voornaamwoord probleemgeval, uitgang -n  
 495 voornaamwoord probleemgeval, uitgang -r/-re  
 496 voornaamwoord probleemgeval, uitgang -a  
 499 voornaamwoord probleemgeval, uitgang latijn, onduidelijk of evident fout  
 geschreven  
 500 bijwoord, geen uitgang  
 # 501-506 reduceren tot 500, zie onder 700  
 501 bijwoord, uitgang -e  
 502 bijwoord, uitgang -s/-th  
 503 bijwoord, uitgang -t  
 504 bijwoord, uitgang -n  
 505 bijwoord, uitgang -r/-re  
 506 bijwoord, uitgang -a  
 509 bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 510 aanwijzend bijwoord, geen uitgang  
 511 aanwijzend bijwoord, uitgang -e  
 512 aanwijzend bijwoord, uitgang -s/-th  
 513 aanwijzend bijwoord, uitgang -t  
 514 aanwijzend bijwoord, uitgang -n  
 515 aanwijzend bijwoord, uitgang -r/-re  
 516 aanwijzend bijwoord, uitgang -a  
 519 aanwijzend bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 520 betrekkelijk bijwoord, geen uitgang  
 521 betrekkelijk bijwoord, uitgang -e  
 522 betrekkelijk bijwoord, uitgang -s/-th  
 523 betrekkelijk bijwoord, uitgang -t  
 524 betrekkelijk bijwoord, uitgang -n  
 525 betrekkelijk bijwoord, uitgang -r/-re  
 526 betrekkelijk bijwoord, uitgang -a  
 529 betrekkelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 530 vragend bijwoord, geen uitgang  
 531 vragend bijwoord, uitgang -e  
 532 vragend bijwoord, uitgang -s/-th

533 vragend bijwoord, uitgang -t  
 534 vragend bijwoord, uitgang -n  
 535 vragend bijwoord, uitgang -r/-re  
 536 vragend bijwoord, uitgang -a  
 539 vragend bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 540 onbepaald bijwoord, geen uitgang  
 541 onbepaald bijwoord, uitgang -e  
 542 onbepaald bijwoord, uitgang -s/-th  
 543 onbepaald bijwoord, uitgang -t  
 544 onbepaald bijwoord, uitgang -n  
 545 onbepaald bijwoord, uitgang -r/-re  
 546 onbepaald bijwoord, uitgang -a  
 549 onbepaald bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 550 hervattend bijwoord, geen uitgang  
 551 hervattend bijwoord, uitgang -e  
 552 hervattend bijwoord, uitgang -s/-th  
 553 hervattend bijwoord, uitgang -t  
 554 hervattend bijwoord, uitgang -n  
 555 hervattend bijwoord, uitgang -r/-re  
 556 hervattend bijwoord, uitgang -a  
 559 hervattend bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 560 ontkennend bijwoord, geen uitgang  
 561 ontkennend bijwoord, uitgang -e  
 562 ontkennend bijwoord, uitgang -s/-th  
 563 ontkennend bijwoord, uitgang -t  
 564 ontkennend bijwoord, uitgang -n  
 565 ontkennend bijwoord, uitgang -r/-re  
 566 ontkennend bijwoord, uitgang -a  
 569 ontkennend bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 590 bijwoord probleemgeval, geen uitgang  
 591 bijwoord probleemgeval, uitgang -e  
 592 bijwoord probleemgeval, uitgang -s/-th  
 593 bijwoord probleemgeval, uitgang -t  
 594 bijwoord probleemgeval, uitgang -n  
 595 bijwoord probleemgeval, uitgang -r/-re  
 596 bijwoord probleemgeval, uitgang -a  
 599 bijwoord probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven  
 600 voornaamwoordelijk bijwoord, geen uitgang  
 601 voornaamwoordelijk bijwoord, uitgang -e  
 602 voornaamwoordelijk bijwoord, uitgang -s/-th  
 603 voornaamwoordelijk bijwoord, uitgang -t  
 604 voornaamwoordelijk bijwoord, uitgang -n  
 605 voornaamwoordelijk bijwoord, uitgang -r/-re  
 606 voornaamwoordelijk bijwoord, uitgang -a  
 609 voornaamwoordelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 610 aanwijzend voornaamwoordelijk bijwoord, geen uitgang  
 611 aanwijzend voornaamwoordelijk bijwoord, uitgang -e  
 612 aanwijzend voornaamwoordelijk bijwoord, uitgang -s/-th  
 613 aanwijzend voornaamwoordelijk bijwoord, uitgang -t

614 aanwijzend voornaamwoordelijk bijwoord, uitgang -n  
 615 aanwijzend voornaamwoordelijk bijwoord, uitgang -r/-re  
 616 aanwijzend voornaamwoordelijk bijwoord, uitgang -a  
 619 aanwijzend voornaamwoordelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 620 betrekkelijk voornaamwoordelijk bijwoord, geen uitgang  
 621 betrekkelijk voornaamwoordelijk bijwoord, uitgang -e  
 622 betrekkelijk voornaamwoordelijk bijwoord, uitgang -s/-th  
 623 betrekkelijk voornaamwoordelijk bijwoord, uitgang -t  
 624 betrekkelijk voornaamwoordelijk bijwoord, uitgang -n  
 625 betrekkelijk voornaamwoordelijk bijwoord, uitgang -r/-re  
 626 betrekkelijk voornaamwoordelijk bijwoord, uitgang -a  
 629 betrekkelijk voornaamwoordelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 630 vragend voornaamwoordelijk bijwoord, geen uitgang  
 631 vragend voornaamwoordelijk bijwoord, uitgang -e  
 632 vragend voornaamwoordelijk bijwoord, uitgang -s/-th  
 633 vragend voornaamwoordelijk bijwoord, uitgang -t  
 634 vragend voornaamwoordelijk bijwoord, uitgang -n  
 635 vragend voornaamwoordelijk bijwoord, uitgang -r/-re  
 636 vragend voornaamwoordelijk bijwoord, uitgang -a  
 639 vragend voornaamwoordelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 640 onbepaald voornaamwoordelijk bijwoord, geen uitgang  
 641 onbepaald voornaamwoordelijk bijwoord, uitgang -e  
 642 onbepaald voornaamwoordelijk bijwoord, uitgang -s/-th  
 643 onbepaald voornaamwoordelijk bijwoord, uitgang -t  
 644 onbepaald voornaamwoordelijk bijwoord, uitgang -n  
 645 onbepaald voornaamwoordelijk bijwoord, uitgang -r/-re  
 646 onbepaald voornaamwoordelijk bijwoord, uitgang -a  
 649 onbepaald voornaamwoordelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 650 bijwoordelijk deel van voornaamwoordelijk bijwoord, geen uitgang  
 651 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang -e  
 652 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang -s/-th  
 653 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang -t  
 654 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang -n  
 655 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang -r/-re  
 656 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang -a  
 659 bijwoordelijk deel van voornaamwoordelijk bijwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 690 voornaamwoordelijk bijwoord probleemgeval, geen uitgang  
 691 voornaamwoordelijk bijwoord probleemgeval, uitgang -e  
 692 voornaamwoordelijk bijwoord probleemgeval, uitgang -s/-th  
 693 voornaamwoordelijk bijwoord probleemgeval, uitgang -t  
 694 voornaamwoordelijk bijwoord probleemgeval, uitgang -n  
 695 voornaamwoordelijk bijwoord probleemgeval, uitgang -r/-re  
 696 voornaamwoordelijk bijwoord probleemgeval, uitgang -a  
 699 voornaamwoordelijk bijwoord probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven

Opmerking ten aanzien van de codes 701-706. Deze blijven gehandhaafd in het CGA.txt, maar kunnen eventueel worden vervangen door 700 in CGR13 1-6 gaan over de syntactische valentie van naamwoorden, dus hier niet. Bovendien zijn de codes 701-706 niet consequent toegepast, bijvoorbeeld <C 703\_NAAST> naest 30x versus <C 700\_NAAST> naest 56x. 700\_TUSSEN net zo iets: met 700 en 704, NAAR ook met 700 en 705. Het gaat bij 70[1-6] om 6844 veranderingen. Regels met bijvoorbeeld 130 keer 701\_OM ombe, 20 tegenover 62084 keer 700. 709 blijft (vaak latijn of rare vormen). 79[0-9] blijft. "Preter" inconsequent (?) gecodeerd als 709 en 790.

Dit verhaal ook bij 500, 600 en 800, controleren

700 voorzetsel, geen uitgang  
701 voorzetsel, uitgang -e  
702 voorzetsel, uitgang -s/-th  
703 voorzetsel, uitgang -t  
704 voorzetsel, uitgang -n  
705 voorzetsel, uitgang -r/-re  
706 voorzetsel, uitgang -a  
709 voorzetsel, uitgang latijn, onduidelijk of evident fout geschreven  
790 voorzetsel (Latijn)  
791 voorzetsel (Latijn)  
792 voorzetsel (Latijn)  
793 voorzetsel (Latijn)  
794 voorzetsel (Latijn)  
795 voorzetsel (Latijn)  
796 voorzetsel (Latijn)  
797 voorzetsel (Latijn)  
798 voorzetsel (Latijn)  
799 voorzetsel (Latijn)  
800 nevenschikkend voegwoord, geen uitgang  
801 nevenschikkend voegwoord, uitgang -e  
802 nevenschikkend voegwoord, uitgang -s/-th  
803 nevenschikkend voegwoord, uitgang -t  
804 nevenschikkend voegwoord, uitgang -n  
805 nevenschikkend voegwoord, uitgang -r/-re  
806 nevenschikkend voegwoord, uitgang -a  
809 nevenschikkend voegwoord, uitgang latijn, onduidelijk of evident fout geschreven  
810 onderschikkend voegwoord, geen uitgang  
811 onderschikkend voegwoord, uitgang -e  
812 onderschikkend voegwoord, uitgang -s/-th  
813 onderschikkend voegwoord, uitgang -t  
814 onderschikkend voegwoord, uitgang -n  
815 onderschikkend voegwoord, uitgang -r/-re  
816 onderschikkend voegwoord, uitgang -a  
819 onderschikkend voegwoord, uitgang latijn, onduidelijk of evident fout geschreven  
820 vergelijkend voegwoord, geen uitgang  
821 vergelijkend voegwoord, uitgang -e  
822 vergelijkend voegwoord, uitgang -s/-th  
823 vergelijkend voegwoord, uitgang -t  
824 vergelijkend voegwoord, uitgang -n

825 vergelijkend voegwoord, uitgang -r/-re  
 826 vergelijkend voegwoord, uitgang -a  
 829 vergelijkend voegwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 830 ontkennend voegwoord, geen uitgang  
 831 ontkennend voegwoord, uitgang -e  
 832 ontkennend voegwoord, uitgang -s/-th  
 833 ontkennend voegwoord, uitgang -t  
 834 ontkennend voegwoord, uitgang -n  
 835 ontkennend voegwoord, uitgang -r/-re  
 836 ontkennend voegwoord, uitgang -a  
 839 ontkennend voegwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 840 explicatief voegwoord, geen uitgang  
 841 explicatief voegwoord, uitgang -e  
 842 explicatief voegwoord, uitgang -s/-th  
 843 explicatief voegwoord, uitgang -t  
 844 explicatief voegwoord, uitgang -n  
 845 explicatief voegwoord, uitgang -r/-re  
 846 explicatief voegwoord, uitgang -a  
 849 explicatief voegwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 850 relatief voegwoord, geen uitgang  
 851 relatief voegwoord, uitgang -e  
 852 relatief voegwoord, uitgang -s/-th  
 853 relatief voegwoord, uitgang -t  
 854 relatief voegwoord, uitgang -n  
 855 relatief voegwoord, uitgang -r/-re  
 856 relatief voegwoord, uitgang -a  
 859 relatief voegwoord, uitgang latijn, onduidelijk of evident fout geschreven  
 860 voegwoord van hoedanigheid, geen uitgang  
 861 voegwoord van hoedanigheid, uitgang -e  
 862 voegwoord van hoedanigheid, uitgang -s/-th  
 863 voegwoord van hoedanigheid, uitgang -t  
 864 voegwoord van hoedanigheid, uitgang -n  
 865 voegwoord van hoedanigheid, uitgang -r/-re  
 866 voegwoord van hoedanigheid, uitgang -a  
 869 voegwoord van hoedanigheid, uitgang latijn, onduidelijk of evident fout geschreven  
 870 also + te + infinitief, geen uitgang  
 871 also + te + infinitief, uitgang -e  
 872 also + te + infinitief, uitgang -s/-th  
 873 also + te + infinitief, uitgang -t  
 874 also + te + infinitief, uitgang -n  
 875 also + te + infinitief, uitgang -r/-re  
 876 also + te + infinitief, uitgang -a  
 879 also + te + infinitief, uitgang latijn, onduidelijk of evident fout geschreven  
 881 prevoegwoord + voegwoord, uitgang -e  
 882 prevoegwoord + voegwoord, uitgang -s/-th  
 883 prevoegwoord + voegwoord, uitgang -t  
 884 prevoegwoord + voegwoord, uitgang -n  
 885 prevoegwoord + voegwoord, uitgang -r/-re  
 886 prevoegwoord + voegwoord, uitgang -a



889 prevoegwoord + voegwoord, uitgang latijn, onduidelijk of evident fout  
 geschreven  
 890 voegwoord probleemgeval, geen uitgang  
 891 voegwoord probleemgeval, uitgang -e  
 892 voegwoord probleemgeval, uitgang -s/-th  
 893 voegwoord probleemgeval, uitgang -t  
 894 voegwoord probleemgeval, uitgang -n  
 895 voegwoord probleemgeval, uitgang -r/-re  
 896 voegwoord probleemgeval, uitgang -a  
 899 voegwoord probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven  
 900 woordsoort probleemgeval, geen uitgang  
 901 woordsoort probleemgeval, uitgang -e  
 902 woordsoort probleemgeval, uitgang -s/-th  
 903 woordsoort probleemgeval, uitgang -t  
 904 woordsoort probleemgeval, uitgang -n  
 905 woordsoort probleemgeval, uitgang -r/-re  
 906 woordsoort probleemgeval, uitgang -a  
 909 woordsoort probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven  
 910 tussenwerpsel, geen uitgang  
 911 tussenwerpsel, uitgang -e  
 912 tussenwerpsel, uitgang -s/-th  
 913 tussenwerpsel, uitgang -t  
 914 tussenwerpsel, uitgang -n  
 915 tussenwerpsel, uitgang -r/-re  
 916 tussenwerpsel, uitgang -a  
 919 tussenwerpsel, uitgang latijn, onduidelijk of evident fout geschreven  
 990 woordsoort probleemgeval, geen uitgang  
 991 woordsoort probleemgeval, uitgang -e  
 992 woordsoort probleemgeval, uitgang -s/-th  
 993 woordsoort probleemgeval, uitgang -t  
 994 woordsoort probleemgeval, uitgang -n  
 995 woordsoort probleemgeval, uitgang -r/-re  
 996 woordsoort probleemgeval, uitgang -a  
 999 woordsoort probleemgeval, uitgang latijn, onduidelijk of evident fout geschreven