

Tagset voor Diachroon corpusmateriaal van het Nederlands (TDN)

Thomas Haga, Katrien Depuydt en Jesse de Does
m.m.v. Roland de Bonth en Dirk Geirnaert

2020

/instituut voor de Nederlandse taal/

Versie 0.9.5

Inhoudsopgave

Woord vooraf	4
1 Inleiding	5
2 Algemene uitgangspunten	6
2.1 Bestaande tagsets	6
2.2 Mogelijkheid tot mapping	7
2.3 Diachrone toepasbaarheid	8
2.4 Wenselijkheid van een kerntagset	8
3 Taggingprincipes	8
3.1 Lexicaal versus functioneel	8
3.2 Token versus woord en overeenkomstige verrijking	9
3.3 Tokenannotaties versus woordannotaties	10
3.4 Vreemde woorden in Nederlandse context	10
3.5 Onderspecificatie	11
4 Beschrijving van de tagset	12
4.1 Hoofdwoordsoort, kenmerken en waardes	12
4.2 Lemmakenmerk versus woordvormkenmerk	13
4.3 Transcategorisaties	13
5 Kenmerken van de woordsoorten	13
5.1 Alle woordsoorten	13
5.2 Naamwoorden	14
5.2.1 Zelfstandig naamwoord	14
5.2.2 Eigennaam	16
5.3 Bijvoeglijk naamwoord-bijwoord	20
5.4 Bijwoord	24
5.5 Werkwoord	27

5.6 Telwoord	31
5.7 Voornaamwoord	35
5.8 Bijzetsel	42
5.9 Voegwoord	44
5.10 Tussenwerpsel	46
5.11 Overblijfsel	47
6 Kerntagset	49
6.1 Alle woordsoorten	49
6.2 Naamwoorden	49
6.2.1 Zelfstandig naamwoord	49
6.2.2 Eigennaam	50
6.3 Bijvoeglijk naamwoord-bijwoord	50
6.4 Bijwoord	50
6.5 Werkwoord	51
6.6 Telwoord	51
6.7 Voornaamwoord	52
6.8 Bijzetsel	53
6.9 Voegwoord	53
6.10 Tussenwerpsel	53
6.11 Overblijfsel	54
6.12 Uitgangen	54
7 Referenties	55
Bijlage 1: overzicht verschillen in lemmatisering	56
Bijlage 2: totaaloverzicht voornaamwoorden	57

Woord vooraf

Dit document is tot stand gekomen op het INT, en is geïnitieerd in de context van het NWO-project CLARIAH+.

Versie 0.9 is besproken door de historisch taalkundigen die partner zijn in het CLARIAH+-project: prof. dr. L.C.J. Barbiers, prof. dr. K.H. van Dalen-Oskam, prof. dr. J.M. van Koppen, dr. M. Rem en prof. dr. N. van der Sijs.

Dit document is versie 0.9.5. Het is momenteel ter review aangeboden aan de volgende personen: dr. G. Bouma, prof. dr. A. Breitbarth, dr. E. Coussé, prof. dr. F. Van Eynde, prof. dr. A.M.S. van Kemenade, dr. M. Kestemont, drs. M. van der Meulen, dr. G.J. Postma, prof. dr. P.Th. van Reenen, dr. G.J. Rutten, T. Struik MA, dr. F. Van de Velde, M. de Vos MPhil, prof. dr. R. Vosters en dr. C. De Wulf.

Leiden, oktober 2020

1 Inleiding

Teksten worden door taalkundige verrijking met woordsoort en lemma beter doorzoekbaar en geschikter voor (taalkundig) onderzoek. Er is een verscheidenheid aan tagsets voor corpora met historisch en hedendaags Nederlands die taalkundig zijn verrijkt, bijvoorbeeld de tagset van het Corpus Gysseling en het Corpus Van Reenen-Mulder en de tagset van het Corpus Gesproken Nederlands en het D-Coi-corpus. Deze tagsets verschillen zowel in mate van detail (hoeveel kenmerken zijn benoemd) als in benoeming (wat in de ene tagset A heet, wordt in een andere B genoemd). Dat maakt de verschillende tagsets moeilijk vergelijkbaar.

In het Nederlab-project¹ is uitgebreid ervaring opgedaan met het verrijken van een groot diachroon corpus. De tagset waarmee gewerkt is, is die van het D-Coi-corpus. Dit heeft onder andere te maken met het feit dat de taalkundige verrijking werd uitgevoerd door *Frog*.² Toen er besloten werd om *Frog* te trainen, is een reductie van de D-Coi-tagset gedefinieerd, om bestaand trainingsmateriaal inzetbaar te kunnen maken.³ De taalkundige verrijking van de gouden standaarden die als trainingsmateriaal gebruikt konden worden, werd naar deze gereduceerde tagset geconverteerd en nieuw evaluatiemateriaal werd met de gereduceerde tagset verrijkt. Uitdagingen hierbij waren het verschil in granulariteit tussen de tagsets bij de verrijking van het beschikbare trainingsmateriaal en het feit dat het beschikbare trainingsmateriaal functioneel getagd is, terwijl de tagset en taggingprincipes van D-Coi uitgaan van een lexicale tagging. Daarnaast werd het duidelijk dat een synchrone tagset voor hedendaags Nederlands niet altijd toepasbaar is op historisch taalmateriaal.

Diachroon onderzoek op basis van taalkundig verrijkte teksten zou beter gefaciliteerd worden als we de beschikking zouden hebben over een tagset die toepasbaar is op alle taalfasen van historisch Nederlands. Hieronder volgt een voorstel voor de Tagset voor Diachroon corpusmateriaal van het Nederlands (TDN). Voor het vaststellen van deze tagset hebben we ons gebaseerd op de tagset van GiGaNT (Groot Geïntegreerd lexicon van de Nederlandse Taal), de corpustagsets CGN/D-Coi, de tagset Corpus Gysseling, de tagset Corpus Van Reenen-Mulder, de tagset bij het Corpus Oudnederlands, het Nederlab-project, het project voor negatie bij de brieven voor P.C. Hooft en de historische woordenboeken: het *Oudnederlands Woordenboek* (ONW) 500-1200, het *Vroegmiddelnederlands Woordenboek* (VMNW) 1200-1300, het *Middelnederlandsch Woordenboek* (MNW) 1250-1550 en het *Woordenboek der Nederlandsche Taal* (WNT) 1500-1976.⁴ We hebben de voorstellen getoetst door een volledige mapping uit te voeren van de tags van CGN/D-Coi, de tags in het Corpus Gysseling en de tags in het Corpus Oudnederlands naar TDN.

Lemmatiseren komt, op enkele opmerkingen na, niet ter sprake. Voor de lemmatisering volgen we de principes zoals geformuleerd en toegepast voor de historische woordenboeken

¹ Zie <https://www.nederlab.nl/onderzoeksportaal>.

² *Frog* is een verzameling van geheugengebaseerde modules van *natural language processing* (NLP) voor het Nederlands. Zie <https://languagemachines.github.io/frog/>.

³ Bij Nederlab is rekening gehouden met de volgende tagsets: de tagset zoals gebruikt in het Corpus Gysseling en het Corpus Van Reenen Mulder (Adelheid), de CGN/D-Coi-tagset, de GiGaNT-tagset, zoals gebruikt in CHN, Brieven als Buit en de Gentse Spelen, en Universal dependencies (UD).

⁴ Een langetermijndoel is ook om de nieuwe tagset beschikbaar te stellen voor de systematische infrastructuur voor historisch Nederlands (zie hiervoor ook Van der Sijs 2020).

van het INT en de computationele lexica van het INT (GiGaNT-Hilex en GiGaNT-Molex).⁵ In Bijlage 1 gaan we kort in op de voornaamste verschillen in lemmatisering met de andere tagsets.

2 Algemene uitgangspunten

Bij het definiëren van de tagset is rekening gehouden met een aantal zaken:

- Bestaande tagsets
- Mogelijkheid tot mapping
- Diachrone toepasbaarheid
- Wenselijkheid van een kerntagset

2.1 Bestaande tagsets

Bij het definiëren van de TDN-tagset is zorgvuldig gekeken naar bestaande tagsets voor historisch en hedendaags Nederlands. Deze worden hieronder kort beschreven. In de beschrijving van de TDN-tagset zelf wordt aangegeven waar deze tagset afwijkt van de bestaande tagsets.

- CGN (2004)⁶ is ontwikkeld voor het Corpus Gesproken Nederlands, opgenomen, getranscribeerd en taalkundig verrijkt tussen 1998 en 2004. Kenmerkend voor de tagset is de mate van detail waardoor de verrijkte woordvorm in principe reconstrueerbaar is met behulp van de combinatie tag en lemma. Voor het referentiecorpus D-Coi (*Dutch Language Corpus Initiative*) (2005),⁷ dat het Nederlands beschrijft van ca. 1990 - ca. 2011, is de CGN-tagset uitgebreid met twee kenmerken voor speciale tokens, namelijk SPEC(afkorting) en SPEC(symbol). De tagsets hebben voor tagging respectievelijk 320 en 322 mogelijkheden. CGN en D-Coi zijn lexicaal getagd. In de rest van het stuk wordt naar deze corpora tegelijk verwezen als CGN/D-Coi, omdat de gebruikte tagset vrijwel identiek is;
- Het Corpus Gysseling is de verzameling van alle dertiende-eeuwse teksten die als bronmateriaal hebben gediend voor het *Vroegmiddelnederlands Woordenboek* (VMNW). De tagset hierbij, tagset Corpus Gysseling, is dezelfde als de tagset voor het Corpus Van Reenen-Mulder (CRM), een verzameling van ongeveer 3000 ambtelijke documenten uit de 14^e eeuw.⁸ Op detailniveau zijn er afwijkingen te vinden in de manier waarop individuele annotatoren de tagset hebben toegepast. Het Corpus Gysseling en het Corpus Van Reenen-Mulder zijn functioneel getagd;⁹

⁵ Zie https://ivdnt.org/images/stories/onderzoek_en_onderwijs/publicaties/TaalbankWorkingpaper4.pdf.

⁶ Zie http://lands.let.ru.nl/cgn/doc_Dutch/topics/version_1.0/annot/pos_tagging/tg_prot.pdf.

⁷ Zie http://lands.let.ru.nl/projects/d-coi/pubs/POS_manual.pdf.

⁸ De tagset is te vinden op: http://adelheid.ruhosting.nl/download/Adelheid1_TagsetManual.pdf.

⁹ De tagset van het Corpus Gysseling heeft in de nieuwe release van de online corpusapplicatie van het Corpus Gysseling een mapping van de oorspronkelijke tags naar de tagnamen van het TDN-voorstel.

- Bij het pilotproject voor negatie bij de brieven van P.C. Hooft uit de periode 1600-1638¹⁰ zijn de teksten eerst automatisch getagd met Adelheid.¹¹ Daarna werd de taalkundige verrijking met woordsoort en lemma gecorrigeerd met behulp van de annotatietool GUSTAVE (Geautomatiseerde Sociolinguïstische en Taalkundige Annotatie van Vroegmoderne Edities).¹² Hierbij is de tagset van CRM gemapt naar CGN/D-Coi. Er zijn ook kenmerken aan toegevoegd, zoals een codering *negcl* voor negatiepartikelen *en* en *ne*, en *nonnom* voor niet-nominatieven. Er is uitgegaan van een lexicale tagging. Deze tagset noemen we in het vervolg van dit stuk GUSTAVE;¹³
- Het Corpus Oudnederlands is de verzameling van het overgeleverde Nederlandse woordmateriaal uit de periode ca. 500 - ca. 1200 dat ten grondslag ligt aan het *Oudnederlands Woordenboek* (ONW). Het corpus is taalkundig verrijkt met woordsoort en lemma. Omdat het corpus relatief klein is, kon de tagset zeer gedetailleerd zijn, zodat bijvoorbeeld bij de werkwoorden naast de meer algemene kenmerken als wijs, tijd, type en getal ook valentie, conjugatietype en werkwoordsklasse zijn aangegeven. Het Corpus Oudnederlands is functioneel getagd. Deze tagset hebben we TaCO genoemd: Tagset Corpus Oudnederlands;¹⁴
- GiGaNT (Groot Geïntegreerd lexicon van de Nederlandse Taal) is een computationeel lexicon van het Nederlands van de zesde eeuw tot nu dat ontwikkeld wordt op het INT.¹⁵ Het bestaat uit een verzameling woorden, woordgroepen en woorddelen, in alle mogelijke spelling- en vormvarianten, telkens gekoppeld aan het bijbehorende moderne Nederlandse lemma. Het lexicon heeft twee hoofdmodules, GiGaNT-Hilex, waarvoor het kernmateriaal uit de historische woordenboeken komt, en GiGaNT-Molex, dat gebaseerd is op modern corpusmateriaal en lexicale data van het INT. De tagset van GiGaNT definieert woordsoorten, kenmerken en mogelijke transcategorisaties per woordsoort en houdt rekening met de diachronie. Hij is gebruikt bij de tagging van het Corpus Hedendaags Nederlands (CHN),¹⁶ het Brieven als Buit-corpus¹⁷ en de Gentse Spelen.¹⁸

2.2 Mogelijkheid tot mapping

Voor het vaststellen van de tagset is er waar mogelijk rekening gehouden met de informatie die volgens de bovengenoemde tagsets wordt geannoteerd. Dat wil zeggen dat deze informatie

¹⁰ Zie voor het project <https://languagedynamics.wp.hum.uu.nl/projects/>.

¹¹ Adelheid is een tagger en lemmatiseerder, door Hans van Halteren ontwikkeld op basis van het CRM. Zie http://adelheid.ruhosting.nl/download/Adelheid1_TagsetManual.pdf.

¹² Zie <https://languagedynamics.wp.hum.uu.nl/wp-content/uploads/sites/251/2015/11/Handleiding-Gustave-juli-2017.pdf>.

¹³ Er is verder geen officiële naam bekend voor deze tagset.

¹⁴ De eerste release van het Corpus Oudnederlands heeft de taalkundige verrijking in TDN-formaat.

¹⁵ Zie https://ivdnt.org/images/stories/onderzoek_en_onderwijs/publicaties/TaalbankWorkingpaper3.pdf en zie voor de totstandkoming hiervan ook Ruitenbergh, De Does en Depuydt (2010).

¹⁶ Het publieke deel van het corpus is voor bij de CLARIN-infrastructuur aangesloten instituten toegankelijk via <http://chn.inl.nl/>.

¹⁷ Dit corpus van 17e-eeuwse en 18e-eeuwse brieven van zeelieden en hun familie is toegankelijk via <http://brievenalsbuit.inl.nl/>.

¹⁸ De Gentse Spelen, een corpus met de teksten van de negentien rederijkerstoneelstukken die bij het Gentse Landjuweel van 1539 opgevoerd werden, zijn op het INT taalkundig verrijkt in de context van een stageproject; dit corpus is nog niet publiekelijk beschikbaar.

vertaalbaar is naar tagging volgens TDN. Daar waar tagsets verschillen in benadering (lexicaal versus functioneel), zou het in principe mogelijk moeten zijn om op basis van de aanwezige informatie vanuit TDN al dan niet met behulp van lexiconinformatie de vertaalslag naar de eigen tagset te maken.¹⁹

2.3 Diachrone toepasbaarheid

De tagset moet toepasbaar zijn op al het Nederlandse tekstmateriaal vanaf het Oudnederlands tot en met het moderne Nederlands, zodat diachroon onderzoek mogelijk wordt. Daarbij moet ook rekening worden gehouden met de ontwikkeling van de Nederlandse taal. Daar waar bijvoorbeeld in de oudste taalfasen een nog volledig naamvalsysteem in voege is, kunnen we voor hedendaags Nederlands alleen restanten daarvan zien, bijvoorbeeld in de vorm van een genitief-s of een versteende datief. De tagset moet dus mogelijkheden bieden om met dit soort verschillen om te gaan, bijvoorbeeld door onderspecificatie.²⁰

2.4 Wenselijkheid van een kerntagset

Tagging van grote hoeveelheden taalmateriaal dat als gouden standaard moet dienen, is niet haalbaar als de tagset heel rijk is. Bovendien zien we bij de bestaande getagde corpora en gouden standaarden een verschil in granulariteit in de taalkundige verrijking. Om het realiseren van een grote hoeveelheid kwalitatief goed getagd materiaal mogelijk te maken, hebben we daarom - na de beschrijving van een complete tagset - ook een voorstel voor een kerntagset waarin de granulariteit van tagging wordt gedefinieerd die we graag in al het diachroon taalkundig verrijkt corpusmateriaal willen hebben. Indien gewenst, kan daar altijd nog meer detail aan worden toegevoegd. De uitgebreide tagset voorziet daarin. Zie verder sectie 6.

3 Taggingprincipes

Voor we overgaan tot de beschrijving van de TDN-tagset willen we ingaan op een aantal principes die we bij het taggen en lemmatiseren willen hanteren.

3.1 Lexicaal versus Functioneel

Een belangrijke keuze die gemaakt moet worden, is of er lexicaal of functioneel getagd wordt. Een woordsoort kan namelijk worden toegekend op grond van twee factoren: het gebruikte woord en de functie in de zin. Bij lexicaal taggen wordt er zoveel mogelijk naar gestreefd wat opgevat wordt als hetzelfde woord in verschillende contexten dezelfde part of speech toe te kennen. Hierbij wordt vaak een woordenboek als referentie gebruikt. Bij functioneel taggen wordt ernaar gestreefd de woordsoort zoveel mogelijk op grond van de functie in de zin toe te

¹⁹ Zie verder sectie 3.1.

²⁰ Zie verder sectie 3.5.

kennen. Een voorbeeld: In de *opgegeten koek* is *opgegeten* ‘lexicaal’ een werkwoord en ‘functioneel’ een bijvoeglijk naamwoord.²¹ Voor TDN kiezen we voor functioneel taggen. De belangrijkste reden hiervoor is dat het vaststellen van de lexicale woordsoort voor een diachroon corpus een moeilijk uitvoerbare taak is. Woordenboeken als referentie gebruiken hiervoor is niet altijd betrouwbaar en voor de oudste taalfasen niet mogelijk, niet alleen vanwege gebrek aan voldoende materiaal, maar ook omdat de oudste historische woordenboeken (ONW en VMNW) geen uitspraak over lexicalisatie doen. Met de keuze voor functioneel taggen sluiten we ook aan bij bestaande gouden standaarden voor historisch Nederlands als bijvoorbeeld het Corpus Oudnederlands, het Corpus Gysseling, CRM en het corpus Brieven als Buit.

3.2 Token versus woord en overeenkomstige verrijking

Onder een token verstaan wij een string karakters die gescheiden is van andere tokens door een spatie (en eventueel een leesteken). Een woord daarentegen is voor ons een taalkundige eenheid. Daarbij beschouwen we bijvoorbeeld *te rugh* als een woord. Een woord kan dus bestaand uit een reeks tokens. De reeks kan discontinu zijn: ook de delen van een scheidbaar werkwoord en een voornaamwoordelijk bijwoord vatten we op als een woord. De discontinue tokens krijgen tezamen één lemma. In corpora als CGN en D-Coi wordt het principe toegepast dat één token één woordsoort en één lemma krijgt. Voor historisch taalmateriaal is dit standpunt niet vol te houden. Sommige woorden worden soms geschreven met een spatie erin, bijvoorbeeld *ghe broken* in plaats van *ghebroken*. Sommige tokens bestaan uit meerdere woorden, bijvoorbeeld bij het clitische token *tboec* (= *het + boek*).

Voor het verrijken van diachroon corpusmateriaal laten we het principe één token één woordsoort/lemma in een aantal gevallen los, zoals bij clitische vormen en bij lemmata die uit één woord bestaan maar waarvan de woordvorm (door een spatie) is opgesplitst. Daarentegen worden gescheiden voorkomende vormen van scheidbare werkwoorden en voornaamwoordelijke bijwoorden gekoppeld en van één woordsoort en lemma voorzien, en krijgen ook multi-word units (mwu) één woordsoort en lemma.

Hieronder volgen een paar voorbeelden van de genoemde gevallen:

- Clitische vormen: *tboec* (*het + boek*); *hadstuse* (*had(de)s + tu/du + se*)
- Los geschreven meerledige lemmata: *op dat*; *eighen schappen*; *het welc*; *leer gierighe*
- Scheidbaar voorkomende woorden: *Verbiedende allen anderen tselue niet naer te doene; onder iets door; ossen vleysch ezels of beren vleysche* (= ‘ossenvlees, ezelsvlees of berenvlees’)

²¹ Een derde optie is een dubbele tag wanneer een lexicale benadering afwijkt van een functionele. In enkele gevallen, zoals in *het gebrokene*, is er zelfs een driedubbele tag mogelijk (werkwoord, bijvoeglijk naamwoord en zelfstandig naamwoord). Hoewel alles zo verantwoord wordt, leidt dit er echter al snel toe dat het systeem complex wordt en dat er nog meer tijd nodig is voor handmatige correctie. Bovendien is dergelijke informatie te vinden in bestaande computationele lexica en woordenboeken. In *de gebroken vaas* kan de woordvorm *gebroken* in het lexicon bijvoorbeeld gevonden worden bij het werkwoord *breken*. Derhalve wordt voor de TDN niet geopteerd voor een (drie)dubbele tagging, maar voor slechts één tag.

- Multi-word units: *à volonté; en profil; ave maria; bed and breakfast; weder die dat; New York; Van de Berg*

3.3 Tokenannotaties versus woordannotaties

Zoals in de vorige sectie is besproken, taggen we niet token voor token maar woorden. We taggen dus alleen taalkundige eenheden, die uit één of meerdere tokens bestaan, die ook discontinu kunnen voorkomen. Niettemin is annotatie op tokenniveau voorzien in de volgende gevallen:

- Interpunctie:

Om in een sequentie van tokens aan te geven dat een token geen woord is, maar een interpunctieteken, gebruiken we de tag “PC”.²²

- Scheidbaarheid:

Wanneer een token deel uitmaakt van een scheidbaar werkwoord of voornaamwoordelijk bijwoord, dan geven we dat bij elk gescheiden voorkomend deel aan met “separated=yes”.

Interpunctie en scheidbaarheid geven niet aan welke tokens tot één woord behoren. Daarvoor voorzien we een verwijzingsmechanisme binnen een XML-formaat of tekstformaat. Dat vormt geen onderdeel van de tagset.

3.4 Vreemde woorden in Nederlandse context

Woorden kunnen worden ontleend aan een andere taal, waarbij er drie verschillende vormen van ontlening te onderscheiden zijn: betekenisontlening, vertalende ontlening en leenwoord. Vooral bij de laatste categorie is het lastig te bepalen of een woord (al) Nederlands is, omdat het vaak een vreemd element, zoals afwijkende spelling, met zich meebrengt. Er zijn hierbij verschillende stadia van integratie van een leenwoord in de taal te onderscheiden: (1) het woord wordt gebruikt door een bepaalde groep mensen in een bepaalde context met een specifieke betekenis, (2) het woord wordt door meer groepen gebruikt, in meerdere contexten en kan door adjectieven gespecificeerd worden - het woord kan een leenwoord genoemd worden, maar is nog steeds ‘vreemd’ - (3) het woord raakt ingeburgerd en wordt aangepast aan het Nederlands in klank, spelling en/of woordvorming en (4) het woord is niet meer te herkennen als leenwoord.²³

Het is lastig om zonder uitgebreid onderzoek vast te stellen in welk stadium leenwoorden zijn als ze voorkomen in een volledig Nederlandse context. Is het woord *september* in de zin *ghegheuen te Bruussele den vierden Septembris Anno M.D.L.V.* bijvoorbeeld te beschouwen als ‘vreemd’ vanwege de Latijnse uitgang net als woorden in de zin [...] *een traen die wt enen boem ghelopen is als vten balsem boem van sericus van lupinus ende van lentescus*? Soms raakt een woord met een Latijnse uitgang namelijk ook ingeburgerd,

²² In bijvoorbeeld XML-bestanden in het TEI-formaat (www.tei-c.org) wordt interpunctie al gemarkeerd met het element <pc>. In token-per-line plain text of in bijvoorbeeld XML-bestanden in het FoLiA-formaat (<https://proycon.github.io/folia/index.nl.html>) is deze tag wel noodzakelijk.

²³ Zie Van der Sijs 1996: 10-12.

hoewel het nog steeds een ‘vreemde’ uitgang heeft, zoals *basis*, *matrix*, *musicus*, *museum*, *nomen*, *alumna* en *plurale* (en bovendien kunnen deze woorden ook ‘vreemde’ meervoudsuitgangen hebben, zoals *bases*, *matrices*, *musici*, *musea*, *nomina*, *alumnae* en *pluralia*, hoewel in veel gevallen ook Nederlandse uitgangen voorkomen, zoals *basissen*, *matrixen*, *musicussen*, *museums*).²⁴

Voor TDN worden dergelijke vreemde woorden en multi-word units, zoals *praesens historicum* in dit werkwoord is een *praesens historicum*, behandeld als potentieel Nederlands woord en worden ze dienovereenkomstig gelemmatiseerd en van een woordsoort voorzien.²⁵

Wanneer woorden niet in Nederlandse context staan en ook geen Nederlandse flexie hebben of wanneer een ‘vreemd’ woord in zelfnoemfunctie staat, zoals een woord of cognaat uit een andere taal in de etymologische beschrijving van een woord of in een zin als *pipe is het Franse woord voor pijp*, dan blijven deze woorden vreemd en worden ze als zodanig getagd.²⁶

3.5 Onderspecificatie

Bij het taggen komt het voor dat sommige informatie die met de tagset geannoteerd zou kunnen worden, niet gegeven wordt (*onderspecificatie*). Een reden kan zijn dat er onvoldoende informatie is om tot een precieze en correcte toekenning te komen (*ambigüiteit*). Het kan ook een bewuste keuze zijn om tot onderspecificatie over te gaan, bijvoorbeeld omdat een bepaalde mate van detail niet relevant is voor een bepaalde taalfase of omdat het voor het project waarbinnen de verrijking wordt uitgevoerd niet nodig wordt geacht.²⁷ Typische voorbeelden van onderspecificatie zijn:

- Wanneer het onderscheid tussen meerdere mogelijkheden taalkundig niet relevant is;
- Wanneer de context geen eenduidige keuze toelaat, of is onduidelijk is (*ambigüiteit*);
- Wanneer voor een project een bepaalde mate van detail niet zinvol geacht wordt.²⁸

Om bij een kenmerk aan te geven dat er meerdere kenmerkwaardes mogelijk zijn, gebruikt TDN één of meerdere sluistekens (bijvoorbeeld NOU-C(case=dat|acc) wanneer de annotator geen keuze maakt tussen een accusatief en een datief of NOU-C(case=nom|dat|acc|voc) bij een zelfstandig naamwoord zonder naamvalsuitgangen in het modern Nederlands). Wanneer er bij een kenmerk geen enkele kenmerkwaarde kan worden uitgesloten, wordt de waarde *unclear* gebruikt (bijvoorbeeld NOU-C(gender=uncl) als niet duidelijk is wat het geslacht is van een zelfstandig naamwoord).

Bij de toekenning van de TDN-tags streven we ernaar een kenmerk in een project (afhankelijk van hoeveel detail men kenmerken in een specifiek project wil geven) zo gespecificeerd mogelijk van waardes te voorzien. Dat wil zeggen dat we in principe geen toekenningen krijgen als VRB(person=1|2) waar de werkwoordsvorm an sich ambigu is, maar waarbij de context een eenduidige keuze biedt.

²⁴ Zie [Leenwoorden](#) in de ANS.

²⁵ Let wel, in een zin als *blijven zeuren is my middle name* krijgen de Engelse woorden wel de tag RES(type=for).

²⁶ Ze krijgen de tag RES(type=for). Zie hiervoor ook sectie 5.11.

²⁷ Vergelijk Leech & Wilson 1996: 16-18.

²⁸ De kerntagset die in dit stuk beschreven wordt, is als subset een voorbeeld van ver doorgevoerde onderspecificatie.

4 Beschrijving van de tagset

Om TDN vorm te geven zijn we uitgegaan van de tagset die gebruikt is voor het GiGaNT-lexicon. Voor dit lexicon is er op basis van uitvoerige analyses een tagset samengesteld, waarbij een overzicht is gegeven van de hoofdwoordsoorten, de mogelijke kenmerken en transcategorisaties per woordsoort. De tagset voor dit lexicon is bedoeld voor diachroon taalmateriaal vanaf de zesde eeuw tot nu. De tagset van GiGaNT is er gekomen na uitgebreid onderzoek over de taalkundige verrijking en ook die is uitvoerig besproken en bediscussieerd door de toenmalige taalbankcommissieleden.²⁹

Hier volgt een overzicht van de hoofdwoordsoorten van TDN. In navolging van GiGaNT worden de namen van de tags in het Engels gegeven. De hoofdcategorieën zijn: Zelfstandig naamwoord: soortnaam (NOU-C), Eigennaam (NOU-P), Bijvoeglijk naamwoord/Bijwoord (AA), Bijwoord (ADV), Werkwoord (VRB), Telwoord (NUM), Voornaamwoord (PD), Bijzetsel (ADP), Voegwoord (CONJ), Tussenwerpsel (INT) en Overblijfsel (RES). Bij elke hoofdwoordsoort wordt een overzicht gegeven van de kenmerken die erbij horen.

De voornaamste verschillen tussen TDN en de hierboven genoemde tagsets (zie sectie 2.1) op het niveau van de hoofdwoordsoort zijn het niet apart opnemen van lidwoorden en het samennemen van woorden die zowel bijvoeglijk naamwoord als bijwoord kunnen zijn (zoals dat ook gedaan is in GiGaNT). Daarnaast worden de dialectwoorden van CGN/D-Coi ook niet als aparte categorie behandeld; de woorden van oudere taalfasen hebben sowieso al veel dialectale variatie, omdat er toen nog geen standaardtaal was. Leestekens worden ook niet aangegeven in de Part-of-Speechtagging (zie sectie 3.3).

4.1 Hoofdwoordsoort, kenmerken en waardes

In TDN bestaat een tag uit de vermelding van de hoofdwoordsoort met de daarbij behorende kenmerken en waardes. De kenmerknamen worden ten behoeve van de helderheid uitgeschreven; een voorbeeld: NOU-C(number=sg,case=nom,gender=m). De CGN-/D-Coi-tagging maakt gebruik van kenmerkwaarden die voor twee of meer mogelijke basiswaarden staan, zoals de naamvalswaarde *standaard* voor nominatief of oblique. Wanneer geen enkele waarde kan worden uitgesloten, gebruikt men de kenmerknaam, zoals de waarde *getal* voor een ondergespecificeerde getalsaanduiding. In het eerste geval worden in TDN de mogelijke waardes gegeven en in het tweede geval wordt de waarde *unclear* aan het kenmerk toegekend, omdat alle waardes mogelijk zijn. Zie verder sectie 3.5. We verschillen ook van CGN/D-Coi waar kenmerken die synchroon als één dimensie zouden kunnen worden opgevat, diachroon niet als zodanig kunnen worden behandeld. TDN heeft bijvoorbeeld geen kenmerk PVTIJD dat zowel waardes voor tijd als modus kan bevatten.

²⁹ De working paper van GiGaNT is indertijd uitvoerig becommentarieerd en goedgekeurd door een commissie van deskundigen die bij het INL aan de toenmalige afdeling Taalbank waren verbonden, bestaande uit: prof. dr. K.H. van Dalen-Oskam, prof. dr. F.G.M. de Jong, prof. dr. N. van der Sijs, prof. dr. P.T.J.M. Vossen, prof. dr. L.C.J. Barbiers, prof. dr. A. Berteloot en prof. dr. Frans Daems.

4.2 Lemmakenmerk versus woordvormkenmerk

Bij de taalkundige verrijking hebben kenmerken ofwel betrekking op het lemma ofwel op een voorkomen van het lemma in context (een token of woordvorm). Type is een voorbeeld van een lemmakenmerk, terwijl naamval een woordvormkenmerk is. Bij de beschrijving van TDN wordt dit onderscheid niet geëxpliciteerd, behalve bij de kenmerken getal en geslacht bij voornaamwoorden, omdat deze in een aantal gevallen zowel voor het lemma als de woordvormen een rol spelen. Zie verder sectie 5.7.

4.3 Transcategorisaties

In sectie 5 worden per kenmerk de mogelijke transcategorisaties behandeld. Hierbij wordt aangegeven hoe een woord van woordsoort kan veranderen. Een voorbeeld hiervan is het bijvoeglijk naamwoord *groen* bij de transcategorisaties van bijvoeglijke naamwoorden dat in een constructie als *het groen* als zelfstandig naamwoord wordt opgevat. Dit betekent in de praktijk dat het getranscategoriseerde woord alleen de in de context van toepassing zijnde tag krijgt. In een zin als *het worstelen ging niet zo goed* krijgt *worstelen* dus alleen de tag NOU-C en niet de tag VRB (noch een dubbele tag).

5 Kenmerken van de woordsoorten

De opzet bij elke woordsoort is als volgt. Eerst is er waar nodig een korte inleiding. Daarna worden de kenmerken van de woordsoorten besproken. Ten slotte worden de gevallen van transcategorisatie behandeld. Er wordt verwezen naar andere tagsets als illustratie voor het gebruik van bepaalde kenmerken. We geven ook aan wanneer er significante afwijkingen zijn tussen wat wij voorstellen en wat er in andere tagsets gebeurt.

De formele beschrijving van de tagset waarmee consistentie van de annotatie afgedwongen kan worden, is te vinden in een afzonderlijke bijlage.

5.1 Alle woordsoorten

Geschreven vorm (tags: *abbr*, *mis*, *trunc*)

Er is één kenmerk dat relevant is voor alle woordsoorten. Dit is de geschreven vorm.³⁰ Er wordt onderscheid gemaakt tussen afgekort (*abbreviated*), afgebroken (*truncated*) en drukfout of digitaliseringsfout (*misspelled*). Het bijbehorende moderne lemma wordt altijd in de uitgeschreven en correct gespelde vorm gegeven. Dit kenmerk wordt alleen toegekend als de geschreven vorm op één van de genoemde manieren afwijkt van de reguliere vorm.

Enkele voorbeelden volgen hieronder.

- Afgekort: *a.u.b.*
- Afgebroken: *gebr.*
- Drukfout of digitaliseringsfout: *professpr/profeffor* (in plaats van *professor*)

³⁰ Een nadere specificatie bij de geschreven vorm is alleen relevant bij de telwoorden, zie sectie 5.6

Written Form (WF)		
abbreviated	misspelled	truncated
abbr	mis	trunc

5.2 Naamwoorden (Nouns - tags: NOU-C, NOU-P)

Voor de naamwoorden maken we een onderscheid tussen een gewoon zelfstandig naamwoord (NOU-C(ommon), soortnaam of appellatief) en een eigennaam (NOU-P(roper)).

5.2.1 Zelfstandig naamwoord (Noun - tag: NOU-C)

Zelfstandige naamwoorden die een soortnaam (appellatief) zijn, krijgen als tag NOU-C.

Diminutieven krijgen een eigen lemma. Ze worden opgevat als afleidingen; aangezien er historisch meerdere diminutiefsuffixen (bijvoorbeeld *-lijn* in *cnapelijn* ‘knapelij’ naast *-kijn* in *cnapekijn* ‘knaapje’) zijn, is het niet logisch de diminutief als onderdeel van het paradigma te beschouwen. De consequentie is dus dat het lemma bij *hondje* niet *hond* is, maar *hondje*, zoals ook het lemma bij *vergadering* niet *vergaderen* is. Hierin wijken we af van CGN/D-Coi en GUSTAVE, die als lemma het woord geven waarvan het diminutief is afgeleid. We wijken ook van CGN/D-Coi af door geen kenmerk toe te kennen aan diminutieven (CGN/D-Coi geeft een waarde *diminutief* bij het kenmerk GRAAD).

Naast diminutieven krijgen ook verkortingen zoals *airco*, *kilo* en Middelnederlands *bot* ‘kleine munt’, respectievelijk afgeleid van *airconditioning*, *kilogram* en *botdrager*, een eigen lemma met als tag NOU-C. Ook eigennamen die als soortnaam gebruikt worden, worden als NOU-C beschouwd, zoals *colbert* en *pampers*.

Voor de zelfstandige naamwoorden onderscheiden we de volgende kenmerken:

Getal (tags: sg, pl)

Er wordt onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*).

Number	
singular	plural
sg	pl

Naamval (tags: nom, gen, dat, acc, voc)

Er wordt onderscheid gemaakt tussen vijf naamvallen: de nominatief (*nominative*), de genitief (*genitive*), de datief (*dative*), de accusatief (*accusative*) en de vocatief (*vocative*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Case				
nominative	genitive	dative	accusative	vocative
nom	gen	dat	acc	voc

In hedendaags Nederlands zijn naamvals vormen goeddeels verdwenen. Alleen de genitief en de datief komen sporadisch voor. Zo is de genitief bijvoorbeeld nog zichtbaar bij tijdsaanduidingen als 's *avonds* en versteende constructies als *de heer des huizes is des duivels*. Daarnaast komen we datiefformen tegen in versteende uitdrukkingen als *ten tonele, ter elfder ure* en *ten voeten uit*. CGN/D-Coi tagt bij het enkelvoud van naamwoorden alleen de genitief en datief apart als de naamvalsuitgangen zichtbaar zijn, en alle andere gevallen noemen ze *standaard*.

Bij GUSTAVE is er een onderscheid gemaakt tussen de nominatief en de niet-nominatiefformen.

Case Reduced				
CGN/D-Coi	nominative dative accusative vocative		genitive	dative
	nom dat acc voc		gen	dat
GUSTAVE	nominative	genitive dative accusative vocative		
	nom	gen dat acc voc		

Geslacht (tags: m, f, n)

Er wordt onderscheid gemaakt tussen drie geslachten: mannelijk (*masculine*), vrouwelijk (*feminine*) en onzijdig (*neuter*), en combinaties daarvan. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands. In CGN/D-Coi hebben enkelvoudige zelfstandige naamwoorden een tag zijdig of onzijdig gekregen als ze niet in de genitief of datief staan. Zijdigheid kan worden vertaald naar m/f en onzijdigheid naar n.

Gender		
masculine	feminine	neuter
m	f	n

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

De flexietypes 0 (*geen flexie*), met *-e*, met *-n*, met *-s/-th* en met *-nt* zijn bij CGN/D-Coi toe te passen op nomina en woorden die hier tot de getranscategoriseerde nomina worden gerekend.

De flexietypes met *-e*, met *-n* en met *-r* worden in GUSTAVE gebruikt voor naamwoorden die niet in de nominatief staan. Daarnaast heeft GUSTAVE een aparte flexie-uitgang voor Latijnse vormen die in TDN onder *other* valt. Ook heeft GUSTAVE een flexie-uitgang voor verkleinwoorden, die in TDN een eigen lemma krijgen.

Transcategorisaties

- NOU-C > AA [ADJ]: *een formica tafel; een hardboard kast; een tricot trui; dat scalke wijf* ('de slechte vrouw', waarbij *scalke* eigenlijk 'dienaar' betekende, maar hier de betekenis van de aard van de dienaar overblijft). Net als de attributieve participia functioneren deze woorden als adjectieven, hoewel de lijst beperkter is. Veelal betreft het stofnamen.
- NOU-C > ADP: *richting de markt.*
- NOU-C > ADV: *plankgas reed ik erheen; ik nam de tram retour; maandag(avond) kijk ik naar de tv; ik zie hem begin van de maand; eind van de ochtend bel ik haar.*
- NOU-C > INT: *hemel!; In tijts genouch seyde zij ketijf het en is mensch inde werelt.* Transcategorisaties naar interjecties zijn lastig; zie voor redenen om deze woorden als tussenwerpsels te beschouwen sectie 5.10.

Een verandering van NOU-C naar NOU-P behandelen we niet als transcategorisatie, omdat het in beide gevallen naamwoorden blijven.

5.2.2 Eigennaam (Proper Noun - tag: NOU-P)

Voor eigennamen worden persoonsnamen, plaatsnamen en organisaties onderscheiden.

Eigennamen vormen een bijzondere categorie binnen de naamwoorden. Deze krijgen een eigen tag, NOU-P(roper). In *Jan Klaassen* zijn beide woorden eigennamen. Vanwege de diachronie krijgt de voornaam één lemma met als tag NOU-P en de achternaam ook, zelfs als deze los geschreven staat. In *Ian van den Beerghe* krijgt *Ian* dus NOU-P als tag en *van den Beerghe* eveneens. Dit geldt ook in het geval van toenames, zoals bij *Jan Janszoon van den Berg* waar er een lemma *Jan* en een lemma *Janszoon van den Berg* is, beide met NOU-P als tag.

Eigennamen die als soortnaam gebruikt worden, worden als NOU-C beschouwd, zoals *colbert* en *pampers*. Soms valt niet uit te maken of een toevoeging bij een voornaam beschouwd moet worden als een eigennaam of als een soortnaam. In een zin als *wi doen cont allen den ghenen die nu siin ende wesen sullen dat diederic de appel+manghelare quam vor ons* is niet duidelijk of het hier gaat om een man die voluit *Diederick de Appelmangelaar* heet of om een man die appelmangelaar van beroep is. In dergelijke gevallen beschouwen we het laatste deel van die persoonsaanduiding als soortnaam (NOU-C).

Eigennamen vormen een aanzet tot Named Entity (NE) tagging. Dit wordt in dit voorstel niet apart behandeld, omdat in dit voorstel woordsoortbenoeming centraal staat.³¹

Voor de eigennamen onderscheiden we dezelfde kenmerken als voor de zelfstandige naamwoorden:

Getal (tags: sg, pl)

Er wordt onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*).

Number	
singular	plural
sg	pl

Naamval (tags: nom, gen, dat, acc, voc)

Er wordt onderscheid gemaakt tussen vijf naamvallen: de nominatief (*nominative*), de genitief (*genitive*), de datief (*dative*), de accusatief (*accusative*) en de vocatief (*vocative*).

Case				
nominative	genitive	dative	accusative	vocative
nom	gen	dat	acc	voc

In hedendaags Nederlands zijn naamvalsvormen goeddeels verdwenen. Alleen de genitief en de datief komen sporadisch voor. Zo is de genitief bijvoorbeeld nog zichtbaar bij de bezits-*s* (*Katriens fiets*), de genitief-apostrof (*Thomas' fiets*) en een combinatie van beide (*Frieda's Mercedes*). Daarnaast komen we datiefvormen tegen in versteende (religieuze) uitdrukkingen als *wat den Here toekomt*. CGN/D-Coi tagt bij het enkelvoud van naamwoorden alleen de genitief en datief apart als de naamvalsuitgangen zichtbaar zijn; alle andere gevallen worden *standaard* genoemd.

Bij GUSTAVE is er een onderscheid gemaakt tussen de nominatief en de niet-nominatiefvormen.

³¹ Met *Named Entity tagging* (NE) kan een annotatielaag bovenop de taalkundige verrijking worden gelegd, waarin de namen geïdentificeerd worden en eventueel van een label voorzien. In een dergelijke laag zou een naam *Jan de Bakker* als geheel getagd worden en bijvoorbeeld voorzien van het label persoonsnaam. Namen in allegorische teksten zoals *dit horenkijn heet Beroemen* en *zijn blaasbalg heet ijdele glorie* zouden in de NE-annotatielaag ook als naam gecodeerd worden.

Case Reduced				
CGN/D-Coi	nominative dative accusative vocative		genitive	dative
	nom dat acc voc		gen	dat
GUSTAVE	nominative	genitive dative accusative vocative		
	nom	gen dat acc voc		

Geslacht (tags: m, f, n)

Er wordt onderscheid gemaakt tussen drie geslachten: mannelijk (*masculine*), vrouwelijk (*feminine*) en onzijdig (*neuter*), en combinaties daarvan.³² In CGN/D-Coi hebben enkelvoudige zelfstandige naamwoorden een tag zijdig of onzijdig gekregen als ze niet in de genitief of datief staan. Zijdigheid kan worden vertaald naar m/f en onzijdigheid naar n.

Gender		
masculine	feminine	neuter
m	f	n

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

De flexietyperen 0 (*geen flexie*), met -e, met -n en met -s/-th zijn bij CGN/D-Coi toe te passen op eigennamen.

De flexietyperen met -e, met -n en met -r worden in GUSTAVE gebruikt voor naamwoorden die niet in de nominatief staan. Daarnaast heeft GUSTAVE een aparte flexie-uitgang voor Latijnse vormen die in TDN onder *other* valt. Ook heeft GUSTAVE een flexie-uitgang voor verkleinwoorden, die in dit voorstel een eigen lemma krijgen.

³² Bij de tagging van persoonsnamen wordt uitgegaan van het taalkundig geslacht, niet van het biologisch geslacht. *Jantje* krijgt dus onzijdig (*neuter*) als tag voor geslacht.

Transcategorisaties

- NOU-P > INT: Namen als *Jezus* of *God* die in verwensingen worden gebruikt: *Ooohhhh!*
Jezus, wat is dat? Oh, God! Godverdomme, hè! Maar enfin, wie doet nu zoiets?

Transcategorisaties naar interjecties zijn lastig: zie voor redenen om deze woorden als tussenwerpsels te beschouwen sectie 5.10.

Een verandering van NOU-P naar NOU-C behandelen we niet als transcategorisatie, omdat het in beide gevallen naamwoorden blijven.

5.3 Bijvoeglijk naamwoord-Bijwoord (Adjective-Adverb - tag: AA)

Vrijwel alle adjectieven worden ook als bijwoord gebruikt (zoals *hard* in *de harde hand*; *ik sla hem hard*). Daarom is de koepelcategorie bijvoeglijk naamwoord-bijwoord in het leven geroepen. Ook wordt de dubbelverbonden bepaling (bepaling van gesteldheid) hiermee ondervangen (*ik loop triest naar huis*). De echte bijwoorden worden behandeld in de volgende sectie.³³

Bijwoorden gevormd door de verkleinvorm van bijvoeglijke naamwoorden-bijwoorden met een adverbiale -s worden hier ook gezien als bijvoeglijk naamwoord-bijwoord (bijvoorbeeld *kalmpjes*, *frisjes*, *liefjes* en *netjes*), aangezien deze woorden predicatief kunnen voorkomen (*zij is netjes*; *het is frisjes*) en soms ook attributief (*een netjes en schoon huis*).

Voor de bijvoeglijke naamwoorden-bijwoorden onderscheiden we de volgende kenmerken:

Graad (tags: pos, comp, sup)

Er wordt onderscheid gemaakt tussen de positief (*positive*), de comparatief (*comparative*) en de superlatief (*superlative*), waarbij het lemma de positief als waarde heeft.³⁴

Degree		
positive	comparative	superlative
pos	comp	sup

Positie (tags: prenom, postnom, free (pred, adv))

Er wordt onderscheid gemaakt tussen prenominaal (*prenominal*), postnominaal (*postnominal*) en vrij (*free*). CGN/D-Coi maakt dit onderscheid ook.

Position		
prenominal	postnominal	free
prenom	postnom	free

De positie vrij kan dan nog verder gespecificeerd worden in predicatief (*predicative*) en adverbiaal (*adverbial*) als men een systematisch onderscheid wil maken tussen adverbia en adjectieven. In het Corpus Gysseling en CRM zijn adjectieven bijvoorbeeld systematisch onderscheiden en wordt positie niet verder gespecificeerd.

³³ Zie voor de echte bijwoorden ook de ANS ([Bijwoorden en adjectieven](#)).

³⁴ Wel moet er voor sommige bijzondere comparatief- en superlatiefvormen (bijvoorbeeld *beter*, *best*, *meer*, *meest*, *minder*, *minst*, *nader*, *naast*, *laatst*, *verder*, *verst*) en onechte, dus niet van adjectieven gevormde woorden, (bijvoorbeeld *achterst*, *voorst*, *benedenst*, *bovenst*, *binnenst*, *buitenst*, *onderst*) een apart lemma komen met als waarde comparatief of superlatief.

Position Extended			
prenominal	postnominal	free	
prenominal	postnominal	predicative	adverbial
prenom	postnom	pred	adv

Getal (tags: sg, pl)

Er wordt onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*) ten gevolge van congruentie met het zelfstandig naamwoord. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Number	
singular	plural
sg	pl

Geslacht (tags: m, f, n)

Er wordt onderscheid gemaakt tussen mannelijk (*masculine*), vrouwelijk (*feminine*) en onzijdig (*neuter*) en combinaties daarvan ten gevolge van congruentie met het zelfstandig naamwoord. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Gender		
masculine	feminine	neuter
m	f	n

Naamval (tags: nom, gen, dat, acc, voc)

Er wordt onderscheid gemaakt tussen vijf naamvallen: de nominatief (*nominative*), de genitief (*genitive*), de datief (*dative*), de accusatief (*accusative*) en de vocatief (*vocative*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Case				
nominative	genitive	dative	accusative	vocative
nom	gen	dat	acc	voc

In CGN/D-Coi is het mogelijk om onderspecificatie toe te passen bij naamvallen. De koepelterm *standaard*, van toepassing op vormen zonder naamvalssuffix, duidt een nominatief of een oblique aan. De koepelterm *bijzonder*, van toepassing op vormen met naamvalssuffix, duidt een genitief of datief aan. Per woordsoort wordt bepaald welk niveau van onderscheid wordt aangebracht. Voor de adjectieven met een buigings *-e* wordt in CGN/D-Coi een onderscheid gemaakt tussen *standaard* (nominatief en oblique), wanneer er geen

naamvalsuitgang is, en *bijzonder* (genitief en datief), wanneer er wel een naamvalsuitgang is, zoals bij *zaliger nagedachtenis*, *te goeder trouw* en *in koelen bloede*.

Bij GUSTAVE is er een onderscheid gemaakt tussen de nominatief en de niet-nominatiefvormen.

Case Reduced			
CGN/D-Coi	nominative dative accusative vocative		genitive dative
	nom dat acc voc		genitive dative
GUSTAVE	nominative	genitive dative accusative vocative	
	nom	gen dat acc voc	

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

De flexietypes 0 (*geen flexie*), *met -e*, *met -n* en *met -r/-re* zijn bij CGN/D-Coi toe te passen op bijvoeglijke naamwoorden-bijwoorden en werkwoorden die in TDN tot de bijvoeglijke naamwoorden-bijwoorden worden gerekend.

De flexietypes *met -e*, *met -n*, *met -r* en *met -s* worden in GUSTAVE gebruikt voor adjectieven als die van toepassing zijn.

Transcategorisaties

- AA > ADP: *overeenkomstig mijn mening*.
- AA > INT: *goddeloos, wat een leven!; goed!* Transcategorisaties naar interjuncties zijn lastig: zie voor redenen om deze woorden als tussenwerpsels te beschouwen sectie 5.10.

- AA > NOU-C: *mijn lief; het rood; het Gronings; de beste; ik zie iets moois en veel gek*, *terwijl hij weinig gekkers ziet*,³⁵ *daer een coninc aen ons quam, een fel*.³⁶ In principe functioneren deze woorden als zelfstandige naamwoorden, dus worden ze als NOU-C getagd. Wanneer het uit de onmiddellijke context duidelijk is wat het weggelaten element is, zoals in *de rode ballen en de groene*, dan transcategoriseert het adjectief niet tot zelfstandig naamwoord: *groene* blijft hier dus AA.

³⁵ Adjectieven na onbepaalde pronomina, zoals in *iets groots*, worden gesubstantiveerd (Van de Velde 2009: 106). Het lemma wordt in deze gevallen respectievelijk *mooie*, *gekke* en *gekkere*.

³⁶ In het Middelnederlands kon zonder lidwoord en na *een* en *geen* (later ook na *die/dat*) een gesubstantiveerde adjectief ook zonder -e voorkomen, zoals *een creupel*, *een gierich* en *een sot*. Enkele van die vormen zijn ook zo als volwaardig substantief in het moderne Nederlands terechtgekomen (bijvoorbeeld *het geheim*, *het jong* en *de zot*). Om oude constructies als *een fel* te begrijpen moeten we weten dat *een* oorspronkelijk een zelfstandig onbepaald voornaamwoord was (betekenis: 'iemand') dat een substantief of adjectief als bepaling kon krijgen: *een coninc* ('iemand die koning is'), *een fel* ('iemand die boosaardig is'). Dergelijke constructies werden door de taalgebruiker geherinterpreteerd, zodat substantieven als *zot* en *jong* ontstonden (Duinhoven 1988: 135-137; als geparafraseerd door Decorte 2003).

5.4 Bijwoord (Adverb - tag: ADV)

Hieronder komen woorden aan de orde die enkel als bijwoord voorkomen. Deze bijwoorden zijn in principe onverbuigbaar, hoewel soms toch flexie optreedt, zoals bij *heel* in *de hele mooie film* naast *de heel mooie film*. In deze gevallen wordt *hele* behandeld als variant van *heel*. Sommige woorden kunnen behalve als bijwoord ook als adjectief voorkomen; deze krijgen de tag AA (zie de vorige sectie).

CGN/D-Coi onderscheidt voor bijwoorden geen kenmerken. Het Corpus Oudnederlands, GUSTAVE, het Corpus Gysseling en CRM onderscheiden die wel, hoewel hun classificatie onderling verschilt.³⁷ CRM onderscheidt bijwoorden en voornaamwoordelijke woorden op hoofdwoordsoortniveau, maar in TDN worden voornaamwoordelijke bijwoorden als een type van bijwoorden gezien.

Voor de bijwoorden onderscheiden we de volgende kenmerken:

Type (tags: reg, pron)

Er wordt onderscheid gemaakt tussen een gewoon bijwoord (*regular adverb*), zoals *echter*, en een pronominaal of voornaamwoordelijk bijwoord (*pronominal adverb*), zoals *daarover*. Wanneer de samenstellende delen van een voornaamwoordelijk bijwoord los geschreven staan (*daar sprak ik hem over*), worden ze als een geheel getagd; zie hiervoor sectie 3.2. Gescheiden bijzetselcombinaties (*hij ging onder de brug door*), die soms tot de bijwoorden worden gerekend, worden hier gezien als omzetsels (zie sectie 5.8).

Type	
regular adverb	pronominal adverb
reg	pron

Subtype (tags: dem, gen, indef, inter, neg, pers, prodrop, rel, resump)

Er wordt onderscheid gemaakt tussen algemene/noemende (*general*), aanwijzende (*demonstrative*), betrekkelijke (*relative*), hervattende (*resumptive*), onbepaalde (*indefinite*), ontkennende (*negative*), persoonlijke (*personal*), vragende (*interrogative*) bijwoorden en bijzetselresten (*pronominal drop*) naargelang het type bijwoord. De categorie *other* uit Gysseling is hier niet genoemd, omdat het hier gaat om onduidelijke gevallen en die worden hier anders ondervangen. De speciale negatiepartikelen uit GUSTAVE vallen hier onder de ontkennende bijwoorden of onder de voegwoorden (zie sectie 5.9) en de bijwoordelijke voorzetsels vallen onder de algemene.

Hieronder volgen enkele voorbeelden bij de subtypes van de bijwoorden:

³⁷ Het Corpus Oudnederlands onderscheidt net als de ANS ([Inleiding naar de functie](#) van bijwoorden) vijf functionele categorieën bijwoorden: *noemend*, *aanwijzend*, *onbepaald*, *vragend* en *betrekkelijk* (het persoonlijk voornaamwoordelijk bijwoord wordt bij het Corpus Oudnederlands niet genoemd); Gysseling en CRM onderscheiden respectievelijk acht en zeven functionele categorieën bijwoorden: *demonstrative*, *general*, *indefinite* (P/D), *interrogative*, *negative*, *relative* (P/D), en *resumptive/resuming* en *other* (alleen bij Gysseling); GUSTAVE onderscheidt negen functionele categorieën bijwoorden: *aanwijzend*, *'general'*, *onbepaald*, *vragend*, *persoonlijk voornaamwoordelijk*, *ontkennend* en *betrekkelijk bijwoord*, *bijwoordelijke voorzetsels* en *speciale negatiepartikelen*.

Gewone bijwoorden:

- Algemeen/Noemend: *gewoon, genoeg, altijd*
- Aanwijzend: *daer; aldaer; alzo; dus*
- Betrekkelijk: *daar; danen*
- Hervattend: *dan; so (in hier bi so sal man geven)*
- Onbepaald: *niewer; so waar; ergens*
- Ontkennend: *en/ne; niet; niene*
- Vragend: *waar; hoe*

Voornaamwoordelijke bijwoorden:

- Aanwijzend: *hier/daer + bijzetsel*
- Betrekkelijk: *waar/daer + bijzetsel*
- Bijzetselrest: *vp in plaats van der vp (so wie die porters say iof lakene te wets leghet. jof peneghe vp leend. jof vercochte.)*
- Onbepaald: *(n)ergens/overal + bijzetsel*³⁸
- Persoonlijk: *er + bijzetsel*
- Vragend: *waar + bijzetsel*

Subtype		
<u>subtypes</u>	<u>regular adverb</u>	<u>pronominal adverb</u>
demonstrative	dem	dem
general	gen	
indefinite	indef	indef
interrogative	inter	inter
negative	neg	
personal		pers
pronominal drop		prodrop
relative	rel	rel
resumptive	resump	

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

³⁸ Bij deze categorie worden de twee woorddelen niet aan elkaar geschreven. We beschouwen het als multi-word unit.

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

Transcategorisaties

- ADV > AA [ADJ]: *het vake gebruik; een affe zin; de na beken* ‘de dichtbij zijnde beken’.
- ADV > CONJ: *telkens we elkaar tegenkomen; toen de vader dit tegenwerkt ..., breekt de hertog, in woede ontstoken, de gewichtige staatsbesognes ... plotseling af.*
- ADV > INT: *toe nou!* Transcategorisaties naar interjecties zijn lastig: zie voor redenen om deze woorden als tussenwerpsels te beschouwen sectie 5.10.
- ADV > NOU-C: *het buitenspel werd afgekeurd; in het volgende allegro; het hoe; het hier en nu.*

Zie ook de transcategorisatie van voorzetsel naar bijwoord in sectie 5.8.

5.5 Werkwoord (Verb - tag: VRB)

Voor de werkwoorden onderscheiden we de volgende kenmerken:³⁹

Werkwoordsvorm (tags: fin, inf, prespart, pastpart)

Er wordt onderscheid gemaakt tussen de finiete vorm (*finite*), de infinitief (*infinitive*)⁴⁰, het tegenwoordig/onvoltooid deelwoord (*present participle*) en het voltooid deelwoord (*past participle*).⁴¹

Finiteness			
finite	infinitive	present participle	past participle
fin	inf	prespart	pastpart

Tempus/tijd (tags: pres, past)

Er wordt bij finiete vormen van het werkwoord onderscheid gemaakt tussen de tegenwoordige tijd (*present tense*) en verleden tijd (*past tense*).

Tense	
present	past
pres	past

Modus/wijs (tags: ind, conj, imp)

Er wordt bij finiete vormen van het werkwoord onderscheid gemaakt tussen de indicatief (*indicative*), conjunctief (*conjunctive*) en imperatief (*imperatief*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands en GUSTAVE.

Mood		
indicative	conjunctive	imperative
ind	conj	imp

Bij het Corpus Gysseling en CRM is ervoor gekozen om de imperatief apart te annoteren en de indicatief en de conjunctief tezamen te nemen.

Mood Reduced	
indicative conjunctive	imperative
ind conj	imp

³⁹ We hebben, in tegenstelling tot tagsets als CGN/D-Coi, geen kenmerk positie bij (niet-finiete) werkwoorden, omdat we functioneel en niet lexicaal taggen.

⁴⁰ Een infinitief van TDN kan ook een oud gerundium zijn, namelijk na *te*. Zie verder de transcategorisaties.

⁴¹ In de oorspronkelijke codering van Gysseling en CRM komt het onderscheid tussen participia niet voor; dit wordt met een script in de geconverteerde versie aangebracht.

Getal (tags: sg, pl)

Er wordt bij finiete vormen van het werkwoord onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*) ten gevolge van congruentie met het onderwerp. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands. Voor oudere taalfasen is dit wel lastiger, omdat bijvoorbeeld de tweede persoon meervoud dezelfde vorm heeft als de tweede/derde persoon enkelvoud.

Number	
singular	plural
sg	pl

Persoon (tags: 1, 2, 3)

Er wordt bij finiete vormen van het werkwoord, met uitzondering van de imperatief, onderscheid gemaakt tussen de eerste, de tweede en de derde persoon ten gevolge van congruentie met het onderwerp. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands. Voor oudere taalfasen is dit wel lastiger, omdat bijvoorbeeld de tweede persoon meervoud dezelfde vorm heeft als de tweede/derde persoon enkelvoud.

Person		
1	2	3

Valentie (tags: trans, intr, refl, impers)

Er wordt onderscheid gemaakt tussen het transitieve werkwoord (*transitive verb*),⁴² intransitieve werkwoord (*intransitive verb*), reflexieve werkwoord (*reflexive verb*) en onpersoonlijke werkwoord (*impersonal verb*). Dit onderscheid wordt in context gemaakt en staat dus niet per werkwoord vast. In de zin *ik eet een appel* is *eten* transitief, terwijl het in de zin *ik eet* intransitief is. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Valency			
transitive	intransitive	reflexive	impersonal
trans	intr	refl	impers

Type (tags: mai, cop, aux)

Er wordt onderscheid gemaakt tussen het zelfstandig naamwoord (*main verb*), koppelwerkwoord (*copular verb*) en hulpwerkwoord (*auxiliary verb*).

Type		
main	copular	auxiliary
mai	cop	aux

⁴² Hiertoe behoren ook de ditransitieve werkwoorden, indien deze categorie apart wordt onderscheiden.

Bij het Corpus Gysseling, CRM en GUSTAVE is ervoor gekozen de zelfstandige werkwoorden apart te annoteren en koppelwerkwoorden en hulpwerkwoorden tezamen te nemen.

Type Reduced	
main	copular auxiliary
mai	cop aux

Conjugatie (tags: strong, weak, irreg, pretpres)

Er wordt onderscheid gemaakt tussen het sterke werkwoord (*strong verb*), zwakke werkwoord (*weak verb*), onregelmatige werkwoord (*irregular verb*) en preterito-praesens (*preterite-present*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Conjugation			
strong	weak	irregular	preterite-present
strong	weak	irreg	pretpres

Werkwoordsklasse (tags: 1, 2, 3a, 3b, 4, 5, 6, 7)

Er wordt in het Oudnederlands onderscheid gemaakt tussen de zeven werkwoordsklassen van sterke werkwoorden, waarbij de 3^e klasse wordt onderverdeeld in twee subtypes.⁴³ Dit is niet een typisch kenmerk voor corpustagging, maar het is wel gebruikt voor het Corpus Oudnederlands en daarom wordt het hier in TDN meegenomen.

Verb Class							
1	2	3a	3b	4	5	6	7

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

⁴³ Zie Quak & Van der Horst 2002: 46-47.

De flexietyperen 0 (geen flexie), met -e, met -n en met -t zijn toe te passen bij CGN/D-Coi. De flexietyperen met -e, met -n en met -t worden in GUSTAVE gebruikt bij de finiete vormen van het werkwoord.

Transcategorisaties

- VRB(part) > AA: in de volgende drie gevallen vindt er transcategorisatie plaats:
 1. Bij attributief gebruik: *gedeeld* genoeg; de *gekaapte* buit; een *klagend* kind; *slapende* kinderen; ende *Ostagers ghelevert* tot Burick, tot Borghe, dat *Grave Maurits ontrent Wesel ende de Lippe gheen aenslaghen soude doen; zoott blijft per Acte, ghegheuen te Bruussele den derden Septemb.* In een zin als de boeken *lezende* man wordt *lezende* gezien als een prenominaal adjectief, ook al heeft het een object bij zich.
 2. Als deelwoordsvormen verbogen zijn: *de vrouw is gebogener* dan haar man; *nóg juichender* lijkt...; *vndersuohton unrecht te fuoron suokinda irsuokenussi* ‘ze onderzochten het onrecht, ze gingen zoekende ten onder aan het onderzoek’.
 3. Bij predicatief gebruik van voltooide deelwoorden waarbij de link met het werkwoord niet altijd (meer) duidelijk is. De aanwezigheid van een bepaling bij het voltooid deelwoord is een indicatie: *dat gedrag is niet gepast; de buis is heel erg gebogen.*

Deze AA's kunnen verder transcategoriseren naar nomina: *iets uitdagends; het gebrokene.*

- VRB(part) > ADP: *gezien* zijn gezondheid; *hangende* het onderzoek; *midts gaders van der selfder Ghesanten oordeel, belanghende de wrevelmoedicheyt ende wreetheyt.*
- VRB(part, past) > CONJ (sub): *aangezien* hij had betaald.
- VRB(imperative) > INT: *help!*; *sino, sconu bistu, friundina min - sino, sconu bistu* ‘zie, mooi ben je, mijn vriendin - zie, mooi ben je’. Transcategorisaties naar interjuncties zijn lastig: zie voor redenen om deze woorden als tussenwerpsels te beschouwen sectie 5.10.
- VRB > NOU-C (n, sg): *het leven; het huilen stond hem nader dan het lachen; in ons schrijven delen wij u mee; verzenmaken is een ambacht, zeker minder moeielyk dan ivoordraaien.* In het Middelnederlands komt dit al veelvuldig voor, zoals in: *want du dan leues in enen sterven; ic geue hem al sijn begeren.* Elk werkwoord kan dus als zelfstandig naamwoord voorkomen. Ook in andere functies wordt de infinitief een naamwoord, omdat aan dergelijke constructies naamvalsuitgangen werden toegevoegd. Dit kwam voor in het Middelnederlands (*dese en was nie geuens moede; om te verstane elcs redene moede*), maar ook in moderne uitdrukkingen is dit nog zichtbaar (*een uur gaans; willens en wetens; het is menens; tot ziens*).⁴⁴ Gerundia worden getagd als zelfstandig naamwoord.⁴⁵ Hiermee wijken we af van CGN/D-Coi, omdat de slot-s daar gezien wordt als derivationeel suffix waarmee bijwoorden van infinitieven (en naamwoorden) worden afgeleid. We maken overigens een uitzondering voor het werkwoord na het voorzetsel *te* en na het voorzetsel *om* als *te* verzwegen is.⁴⁶ In Middelnederlands *te geuene* en modern Nederlands *te geven* blijft *geuen* een werkwoord.

⁴⁴ Zie Duinhoven 1997: 192-193.

⁴⁵ Dit onderscheid is overigens alleen systematisch getagd in het Corpus Oudnederlands. In het Corpus Gysseling en CRM zijn infinitieven niet in alle gevallen onderscheiden van gerundia. In het *Oudnederlands Woordenboek* en het *Vroegmiddelnederlands Woordenboek* is gerundium een werkwoordskenmerk dat toegepast wordt op die gevallen die geen voor- of nabepaling kunnen krijgen

⁴⁶ Zie Duinhoven 1997: 195-199 en referenties aldaar.

5.6 Telwoord (Numeral - tag: NUM)

De telwoorden worden onderscheiden in twee types: hoofdtelwoorden en rangtelwoorden. Hiertoe rekenen we de getallen en daarvan afgeleide vormen (*drie mensen; de vierde man*) en andere woorden die een telbare eenheid aangeven, zoals (*alle*)*beide* (namelijk ‘twee’) en *ander* (in de betekenis ‘tweede’). Ook afleidingen hiervan rekenen we tot de telwoorden (*drieërhande; tweeërlei; beiderhande; beiderlei*). Breuken als *vijf achtste* worden net als bij CGN/D-Coi behandeld als een combinatie van een hoofdtelwoord en een rangtelwoord.

Wij rekenen alleen bepaalde telwoorden tot de telwoorden. Onbepaalde telwoorden beschouwen we als onbepaalde voornaamwoorden, omdat deze woorden geen telbare eenheid aangeven en omdat het onderscheid tussen onbepaalde telwoorden en onbepaalde voornaamwoorden niet eenduidig is.⁴⁷ Hier wijken we dus af van het Corpus Gysseling, CRM en het Corpus Oudnederlands, die wel het type onbepaald telwoord hebben.

Hoewel CGN/D-Coi ook geen categorie heeft voor onbepaalde telwoorden, wijken we hier toch van af. CGN/D-Coi rekent namelijk alle woorden met aparte vormen voor hoofd- en rangtelwoorden tot de telwoorden, zoals *twee(de)* en *dertig(ste)*, maar ook woorden als *hoeveel(ste)* en *zoveel(ste)*,⁴⁸ die wij tot de onbepaalde voornaamwoorden rekenen, omdat er geen telbare eenheid is aangegeven.

Telwoorden kunnen ook verkleinvormen hebben, zoals *wij drietjes gaan erheen; ik zie er eentje*. Deze diminutieven krijgen net als de verkleinvormen van zelfstandige naamwoorden een eigen lemma.

Voor de telwoorden onderscheiden we de volgende kenmerken:

Type (tags: card, ord)

Er wordt onderscheid gemaakt tussen hoofdtelwoord (*cardinal*) en rangtelwoord (*ordinal*). Rangtelwoorden krijgen hierbij een eigen lemma (dus *tweede* krijgt als lemma *tweede* en niet *twee*).

Type	
cardinal	ordinal
card	ord

Positie (tags: prenom, postnom, free)

Er wordt onderscheid gemaakt tussen prenominale (*prenominal*), postnominale (*postnominal*) en vrije (*free*) positie. Hieronder volgen enkele voorbeelden:

⁴⁷ Zo rekent de ANS bijvoorbeeld alleen *veel* (ook *zoveel* en *hoeveel*), *weinig* en informeel *tig* tot de onbepaalde telwoorden ([Onbepaalde hoofdtelwoorden](#)), maar behandelen ze andere woorden die in andere grammatica's tot de onbepaalde telwoorden worden gerekend, zoals *enige*, *sommige*, *voldoende* en *zat*, als onbepaalde voornaamwoorden. Van de Velde (2009: 222) noemt het onderscheid tussen de twee categorieën flinterdun of zinloos.

⁴⁸ CGN/D-Coi rekent daarnaast *veel* en *weinig* tot de onbepaalde voornaamwoorden (in tegenstelling tot de ANS), omdat daar geen corresponderend rangtelwoord bij is. Om diezelfde reden hoort ook *beide* tot de onbepaalde voornaamwoorden. Wij rekenen *beide* tot de telwoorden, omdat het een telbare eenheid is.

- Prenominaal: *twee kazen; deze vier broden; de derde man*. Ook gevallen als *twee kilo suiker, vijf pond vlees* en *vijf juli* rekenen wij, in tegenstelling tot CGN/D-Coi, tot de prenominaal telwoorden.
- Postnominaal: *de sterren vijff die eeuwelick gaen dwalen*. Ook gevallen als *de jaren twintig* en *pagina vijftien* rekenen wij, in tegenstelling tot CGN/D-Coi, tot de postnominale telwoorden.
- Vrij: *is het er eentje of zijn het er twee?*; *wij drieën gaan erheen*; (regionaal) *wij gedrieën gaan erheen*; *we voltooien de opdracht met z'n vieren*; *hij is eerste geworden*; *de man overleed op z'n tachtigste*; *zij zaten gedrieën op de bank*; *ik wil er een stuk of drie*; *één, twee, drie!*; *is het tegen zessen of al na zessen?* Hiermee wijken we af van CGN/D-Coi, die gevallen als *met z'n vieren* en *ik zie er eentje* tot de nominale positie rekent en die telwoorden in zinnen als *tegen/na zessen* tot de bijwoorden rekent.

Position		
prenominal	postnominal	free
prenom	postnom	free

Weergave (tags: dig, let, rom, mix-dig, mix-rom)

Er wordt onderscheid gemaakt tussen telwoorden die worden weergegeven met cijfers (*digits*), letters (*letters*), Romeinse cijfers (*roman*), mengvormen met cijfers en letters (*mixed-digits*) en mengvormen met Romeinse cijfers en letters (*mixed-romans*).

Representation				
digits	letters	roman	mixed-digits	mixed-romans
dig	let	rom	mix-dig	mix-rom

Getal (tags: sg, pl)

Er wordt onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*) ten gevolge van congruentie met zelfstandige naamwoorden en voornaamwoorden. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Number	
singular	plural
sg	pl

Geslacht (tags: m, f, n)

Er wordt onderscheid gemaakt tussen mannelijk (*masculine*), vrouwelijk (*feminine*) en onzijdig (*neuter*) en combinaties daarvan ten gevolge van congruentie met zelfstandige naamwoorden en voornaamwoorden. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Gender		
masculine	feminine	neuter
m	f	n

Naamval (tags: **nom, gen, dat, acc, voc**)

Er wordt onderscheid gemaakt tussen vijf naamvallen: de nominatief (*nominative*), de genitief (*genitive*), de datief (*dative*), de accusatief (*accusative*) en de vocatief (*vocative*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Case				
nominative	genitive	dative	accusative	vocative
nom	gen	dat	acc	voc

Voor de preminale telwoorden wordt in CGN/D-Coi een onderscheid gemaakt tussen *standaard* (nominatief en oblique), wanneer er geen naamvalsuitgang is, en *bijzonder* (genitief en datief), wanneer er wel een naamvalsuitgang is, zoals bij *ter elfder ure* en *ten enen male*.

Case Reduced	
nominative dative accusative vocative	genitive dative
nom dat acc voc	gen dat

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

De flexietyperen 0 (*geen flexie*), met -n, met -r/-re en met -s/-th zijn ook toe te passen bij CGN/D-Coi.

Transcategorisaties

- NUM(ord) > ADV: eerst ging hij naar huis.
- NUM(card) > NOU-C: een zes; een zeven; de acht. Aangezien niet-attributieve hoofdtelwoorden naar hun syntactische positie niet goed te onderscheiden zijn van nomina, nemen we alleen transcategorisatie aan voor de niet-tellende betekenis, zoals een dikke een; hij heeft twee zevens in zijn hand.
- NUM(ord) > NOU-C: de eerste die hard wegloopt, is de sjaak; hij is de tweede; dat is een derde van de bevolking. Wanneer het uit de onmiddellijke context duidelijk is wat het weggelaten element is, zoals in *de vijfde man en de zesde*, dan transcategoriseert het rangtelwoord niet tot zelfstandig naamwoord: *zesde* blijft hier dus rangtelwoord.

5.7 Voornaamwoord (Pronoun-Determiner - tag: PD)

Voornaamwoorden en determiners worden samen behandeld omdat ze een groot aantal kenmerken gemeen hebben. Over enkele woorden verschilt men van mening of het tot de categorie PD moet worden gerekend of tot een andere woordsoort. We beginnen daarom met de beschrijving van die specifieke gevallen en hoe TDN deze gevallen behandelt.

Zoals al besproken in sectie 5.6 worden onbepaalde telwoorden hier tot de onbepaalde voornaamwoorden gerekend, omdat ze geen telbare eenheid aangeven. In tegenstelling tot CGN/D-Coi, het Corpus Gysseling, CRM, GUSTAVE en TaCO behandelen we lidwoorden als voornaamwoorden, omdat ze in de vroegste fasen van historisch Nederlands niet puur op vorm te onderscheiden zijn. *Dat* kan bijvoorbeeld zowel een aanwijzend voornaamwoord zijn als een bepaald lidwoord.⁴⁹ CGN/D-Coi heeft een categorie adv-pron. In TDN zijn dat bijwoorden of delen van voornaamwoordelijke bijwoorden. *Er* is volgens TDN een bijwoord, behalve in de kwantitatieve functie (zoals in *Rozen? Ik heb er tien*), dan is het een persoonlijk voornaamwoord.⁵⁰ In een aantal gevallen ten slotte is er geen eenduidigheid over het feit of er gesproken moet worden over een adjectief of een voornaamwoord, zoals *dezelfde*, *dergelijk(e)* en *zodanig(e)*.

Om helder te krijgen wat de mogelijke voornaamwoorden zijn, is een uitgebreide analyse gedaan van de historische woordenboeken van het Nederlands. We hebben de aanwezige voornaamwoorden geïnventariseerd, en daar waar nodig de woordsoortaanduiding gecorrigeerd. Deze uitvoerige lijst is achteraan het voorstel opgenomen in Bijlage 2: totaaloverzicht voornaamwoorden.

Voor de voornaamwoorden onderscheiden we de volgende kenmerken:

Type (tags: d-p (dem, rel,) excl, indef, pers, poss, recip, refl, w-p (inter, rel))

Er wordt onderscheid gemaakt tussen bezittelijke voornaamwoorden (*possessive*), d-pronomina (*d-pronomina*), onbepaalde voornaamwoorden (*indefinite*), persoonlijke voornaamwoorden (*personal*), reciproque voornaamwoorden (*reciprocal*), reflexieve voornaamwoorden (*reflexive*), uitroepende voornaamwoorden (*exclamative*) en w-pronomina (*w-pronomina*). Hieronder volgen enkele voorbeelden van deze acht types:

- Bezittelijk: *zijn vriend*; [*dan*] *sael ic comen vriendeken mijn*; *al ware al de weereit dijn*. In CGN/D-Coi zijn *mijnen/onzen* ook bezittelijk in constructies als *de mijnen* en *ten onzent*, maar dat zijn in TDN naamwoorden (zie de transcategorisaties aan het einde van deze sectie).
- D-pronomina: *de*; *die*; *dat*; *deze*; *dit*; *dusdanige*; *gene*; *zodanige*. Dit zijn (aanwijzende) deiktische en betrekkelijke pronoun/determiners die voor oudere taalfases niet puur op grond van de vorm kunnen worden onderscheiden. Deze categorie is genoemd naar de beginletter van de meest voorkomende woorden in deze categorie. De waarde d-pronomen

⁴⁹ Zie Duinhoven 1988: 223.

⁵⁰ *Er* is eigenlijk een oude, versteende naamval van het persoonlijk voornaamwoord van de derde persoon meervoud in alle geslachten. Omdat deze vorm min of meer tot een zelfstandig woord is geworden, krijgt het toch het lemma *er*.

wordt alleen toegekend als er systematisch geen onderscheid wordt gemaakt tussen aanwijzende en betrekkelijke voornaamwoorden.

- Onbepaald: *veel; weinig; al(le); geen; enige; sommige; een* (lidwoord).
- Persoonlijk: *ik loop daar; Roland ziet haar; jijzelf hebt dat gedaan; daar gaat ie*. In navolging van de ANS worden ook het niet-verwijzende *het* en *men* tot de persoonlijke voornaamwoorden gerekend, zoals in *het regent; het schijnt dat men weet van de overval*.⁵¹
- Reciproque: *Jesse en Katrien zien elkaar*.
- Reflexief: *Dirk wast zich; wij wassen ons(zelf); hij wast hem*.⁵²
- Uitroepend: *welk een dwaasheid*.
- W-pronomen: *wie, welke, hoedanig*. Dit zijn vragende en betrekkelijke pronoun/determiners die voor oudere taalfases niet puur op grond van de vorm kunnen worden onderscheiden. Deze categorie is genoemd naar de beginletter van de meest voorkomende woorden in deze categorie. De waarde w-pronomen wordt alleen toegekend als er systematisch geen onderscheid wordt gemaakt tussen vragende en betrekkelijke voornaamwoorden.

Type							
d-pronoun	exclamative	indefinite	personal	possessive	reciprocal	reflexive	w-pronoun
d-p	excl	indef	pers	poss	recip	refl	w-p

D-pronomen kunnen worden onderscheiden in aanwijzende voornaamwoorden (*demonstrative*) en betrekkelijke voornaamwoorden (*relative*) die beginnen met een d-, zoals *die*, terwijl w-pronomen kunnen worden onderverdeeld in de overige betrekkelijke voornaamwoorden (*relative*) en de vragende voornaamwoorden (*interrogative*). Hierdoor zijn er in totaal negen types voornaamwoorden (zie de onderstaande tabel).

CGN/D-Coi heeft zowel vragend-betrekkelijke (*wie*) als vragende (*watte*) en betrekkelijke (*die*) voornaamwoorden en een aparte categorie voor persoonlijke reflexieven (zoals *je(zelf)*) en de reflexieven *zich* en *zichzelf*. De andere tagsets onderscheiden geen exclamatieven. Reflexief en reciproque zijn tezamen genomen in GUSTAVE, het Corpus Gysseling en CRM.⁵³ Verder vallen de negatieve onbepaalde voornaamwoorden van GUSTAVE hier onder de onbepaalde voornaamwoorden.

Type Extended								
demonstrative	exclamative	indefinite	interrogative	personal	possessive	reciprocal	reflexive	relative
dem	excl	indef	inter	pers	poss	recip	refl	rel

⁵¹ Zie de [Inleiding](#) van het persoonlijk voornaamwoord.

⁵² Dit geldt alleen voor die gevallen waarbij *hem* dezelfde betekenis heeft als *zich*. Dit is het geval in vroegere fasen van het Nederlands voordat *zich* als reflexief werd gebruikt en het komt ook tegenwoordig in dialecten nog voor.

⁵³ Overigens is in de Adelheidmapping van CRM alleen de term reflexief gebruikt.

Subtype lidwoord (tags: art, oth)

Het bepaald lidwoord wordt gezien als een subtype van het d-pronomen of als een subtype van het aanwijzend voornaamwoord als dat apart wordt onderscheiden. Evenzo wordt het onbepaald lidwoord beschouwd als een subtype van het onbepaald voornaamwoord. Bepaaldheid hoeft niet te worden aangegeven, omdat een lidwoord als subtype van D-P/DEM automatisch bepaald is en omdat een lidwoord als subtype van INDEF automatisch onbepaald is. Ook is het vanwege het lemma zelf al redundant. De niet-lidwoorden worden bij deze voornaamwoorden aangegeven met de waarde *other*. De toekenning van dit subtype gebeurt alleen als de betreffende voornaamwoorden in prenominale positie staan.

Subtype			
d-pronoun / demonstrative pronoun		indefinite pronoun	
article	non-article	article	non-article
art	oth	art	oth

Positie (tags: prenom, postnom, free)

Er wordt onderscheid gemaakt tussen prenominaal (*prenominal*), postnominaal (*postnominal*) en vrij (*free*). Sommige voornaamwoorden komen alleen voor in vrije positie (met name persoonlijke, reciproque en reflexieve voornaamwoorden), terwijl andere voornaamwoorden ook in de andere posities kunnen voorkomen. We wijken hierbij af van CGN/D-Coi door de niet-attributieve posities - de nominale positie⁵⁴ en de vrije positie - samen te nemen tot de positie vrij. Ook wijken we af door een positie postnominaal aan te nemen, zowel omdat er enkele hedendaagse voorbeelden van te vinden zijn als omdat het in oudere fases van het Nederlands met regelmaat te vinden is.⁵⁵ Hieronder volgen enkele voorbeelden:

- Prenominaal: *mijn vriend*; *welke man*; *die stoel*; *elk huis*.
- Postnominaal: *het is de oude man zelf*;⁵⁶ *kindeke mijn*; *[dan] sael ic comen vriendeken mijn*; *fiunda sina ertha leccon sulun* ‘zijn vijanden zullen de aarde likken’; *begian thi folc alla* ‘mogen alle volkeren je openlijk belijden’.
- Vrij: *ik loop daar*; *Roland ziet haar*; *Dirk wast zich*; *Jesse en Katrien zien elkaar*; *welk een dwaasheid*; *de man die daar staat?*; *hetgeen je daar ziet*; *waar is Thomas straks?*; *wiens/diens hoed*; *mijns gelijke*; *al die mensen*; *ze kregen er elk twee*; *al ware al de weerelt dijn*.

Position		
prenominal	postnominal	free
prenom	postnom	free

⁵⁴ Niet alle voornaamwoorden die in CGN/D-Coi nominaal zijn, rekenen wij tot de voornaamwoorden. In sommige gevallen transcategoriseert het voornaamwoord in TDN naar een zelfstandig naamwoord, bijvoorbeeld in gevallen als *de mijnen*, *de minsten* en *te mijnent*. Zie ook de transcategorisaties aan het einde van deze sectie.

⁵⁵ CGN/D-Coi noemt het postnominale gebruik wel, maar zegt dat dit zo bijzonder is dat er geen voorzieningen voor worden getroffen.

⁵⁶ *Zelf* wordt hier net als bij Van de Velde (2009: 301) tot de voornaamwoorden gerekend.

Getal en geslacht: onderscheid lemmakenmerk en woordvormkenmerk

Bij de voornaamwoorden is er een onderscheid te maken tussen het getal (*wij* en *hun* zijn meervoud) en het geslacht (*hij* en *hem* zijn mannelijk) van het lemma en het getal en geslacht van de woordvorm in context. Het lemmakenmerk wordt toegekend aan de persoonlijke, bezittelijke, reflexieve en reciproque voornaamwoorden en het contextkenmerk aan de andere voornaamwoorden en de bezittelijke voornaamwoorden; alleen bij de bezittelijke voornaamwoorden kunnen dus beide kenmerken worden toegekend.

Niet alle tagsets maken dit onderscheid. CGN/D-Coi maakt hier een onderscheid op basis van positie. Attributieve voornaamwoorden krijgen agreementkenmerken en niet-attributieve, voor zover toepasbaar, kenmerken van de referent, met uitzondering van de preminale bezittelijke voornaamwoorden die allebei krijgen. Dit standpunt is voor oudere taalfases niet werkbaar omdat de zelfstandig gebruikte voornaamwoorden dezelfde woordvormkenmerken (naamval, getal, geslacht) hebben als de attributief gebruikte. Voorbeelden hiervan zijn: (*De giraf*) *kert em om desen ende om dien om dattet wil sijn besien; uan themo richedume sint unse wort. zo themo ist unse mût gekart.* ‘Over die rijkdom handelen onze woorden, daarop is onze geest gericht.’

De lemmakenmerken en woordvormkenmerken worden hieronder als aparte kenmerken behandeld.

Lemmakenmerk Getal (tags: sg, pl)

Er wordt bij de persoonlijke, bezittelijke, reflexieve en reciproque voornaamwoorden onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*). In het Corpus Oudnederlands, het Corpus Gysseling en CRM is getal alleen een kenmerk van de persoonlijke en de reflexieve/reciproque voornaamwoorden.

Lemma Number (LN)	
singular	plural
sg	pl

Lemmakenmerk Persoon (tags: 1, 2, 3)

Er wordt bij de persoonlijke, bezittelijke, reflexieve en reciproque voornaamwoorden onderscheid gemaakt tussen de eerste, tweede en derde persoon. In het Corpus Oudnederlands, het Corpus Gysseling en CRM en in GUSTAVE is persoon alleen een kenmerk van de persoonlijke en de reflexieve/reciproque voornaamwoorden.⁵⁷

Het verschil dat CGN/D-Coi maakt tussen pronomina die een persoonlijke en een onpersoonlijke referent vereisen (bijvoorbeeld *wie*; *iedereen* naast *wat*; (*n*)*iets*) en tussen vertrouwelijke en beleefde vormen bij de tweede persoon (*je*; *jij*; *jou*; *jullie* naast *u*) wordt in TDN niet gecodeerd.

⁵⁷ GUSTAVE heeft één kenmerk persoon, waarbij het bij de persoonlijke en reflexieve voornaamwoorden gaat om de inherente persoon, maar waarbij het bij de aanwijzende en bezittelijke voornaamwoorden altijd om de derde persoon gaat vanwege congruentie met het zelfstandig naamwoord. In onze interpretatie hebben voornaamwoorden alleen een inherente persoon.

Lemma Person (LP)		
1	2	3

Lemmakenmerk Geslacht (tags: m, f, n)

Er wordt bij de persoonlijke, bezittelijke, reflexieve en reciproque voornaamwoorden voor de derde persoon enkelvoud onderscheid gemaakt tussen mannelijk (*masculine*), vrouwelijk (*feminine*) en onzijdig (*neuter*). In het Corpus Gysseling, CRM en GUSTAVE is dit lemmakenmerk niet toegekend en in het Corpus Oudnederlands niet systematisch.

Lemma Gender (LG)		
masculine	feminine	neuter
m	f	n

Woordvormkenmerk Getal (tags: sg, pl)

Er wordt bij de voornaamwoorden, met uitzondering van de persoonlijke, de reflexieve en reciproque, onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*). In tegenstelling tot het Corpus Oudnederlands is dit woordvormkenmerk in het Corpus Gysseling en CRM niet toegekend. In CGN/D-Coi worden, op basis van overeenkomst, getal en geslacht gecombineerd in één tag, zoals bijvoorbeeld bij het bezittelijk voornaamwoord *ons*, dat in combinatie met het zelfstandig naamwoord *paard* een kenmerk met waarde *evon* (enkelvoud onzijdig) krijgt. Getal is ook toegekend aan de aanwijzende en bezittelijke voornaamwoorden in GUSTAVE.

Number	
singular	plural
sg	pl

Woordvormkenmerk Geslacht (tags: m, f, n)

Er wordt bij de voornaamwoorden, met uitzondering van de persoonlijke, de reflexieve en de reciproque, onderscheid gemaakt tussen mannelijk (*masculine*), vrouwelijk (*feminine*) en onzijdig (*neuter*). In tegenstelling tot het Corpus Oudnederlands is dit woordvormkenmerk in het Corpus Gysseling, CRM en GUSTAVE niet toegekend. In CGN/D-Coi worden, op basis van overeenkomst, getal en geslacht gecombineerd in één tag, zoals bijvoorbeeld bij het bezittelijk voornaamwoord *ons*, dat in combinatie met het zelfstandig naamwoord *paard* een kenmerk met waarde *evon* (enkelvoud onzijdig) krijgt.

Gender		
masculine	feminine	neuter
m	f	n

Naamval (tags: nom, gen, dat, acc, voc)

Er wordt onderscheid gemaakt tussen vijf naamvallen: de nominatief (*nominative*), de genitief (*genitive*), de datief (*dative*), de accusatief (*accusative*) en de vocatief (*vocative*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Case				
nominative	genitive	dative	accusative	vocative
nom	gen	dat	acc	voc

In hedendaags Nederlands zijn naamvals vormen goeddeels verdwenen. Bij de voornaamwoorden kunnen soms de nominatief, de genitief en/of de datief worden onderscheiden. In CGN/D-Coi is het mogelijk om onderspecificatie toe te passen bij naamvallen. De koepelterm *standaard*, van toepassing op vormen zonder naamvalssuffix, duidt een nominatief of een oblique aan. De koepelterm *bijzonder*, van toepassing op vormen met naamvalssuffix, duidt een genitief of datief aan. Per woordsoort wordt bepaald welk niveau van onderscheid wordt aangebracht. De mogelijke onderscheiden bij de voornaamwoorden hebben we in een overzicht gezet, met een vertaling van de waardes naar naamvallen.⁵⁸

Case Reduced					
Waardes van CGN/D-Coi	<i>Standaard</i> (in geen enkele naamval een suffix)			<i>Bijzonder</i> (met suffix)	
	<i>Standaard</i>	Nominatief	Oblique	Genitief	Datief
Article (def.)	nom dat acc voc			gen	dat
Article (indef.)	nom dat acc voc			gen	
Demonstratives	nom dat acc voc			gen	dat
Exclamatives	nom dat acc voc				
Indefinites	nom dat acc voc			gen	dat
Interrogatives	nom dat acc voc				
Personals	nom dat acc voc	nom voc	dat acc	gen	
Possessives	nom dat acc voc			gen	dat
Reflexives			dat acc		
Reciprocals			dat acc	gen	
Relatives	nom dat acc voc			gen	
W-Pronouns	nom dat acc voc		dat acc	gen	

⁵⁸ De adverbiale pronomina, in CGN/D-Coi gemarkeerd met de waarde adv-pron, zijn in dit overzicht niet opgenomen, omdat ze in TDN een andere woordsoort krijgen.

Bij GUSTAVE is er een onderscheid gemaakt tussen de nominatief en de niet-nominatiefvormen.

Case Reduced	
nominative	genitive dative accusative vocative
nom	gen dat acc voc

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

De flexietyes 0 (*geen flexie*), *met -e*, *met -n*, *met -r/-re* en *met -s/-th* zijn toe te passen bij CGN/D-Coi.

De flexietyes *met -e*, *met -n* en *met -r/-re* worden in GUSTAVE gebruikt voor aanwijzende, onbepaalde en bezittelijke voornaamwoorden en het flexietye *met -s* wordt in GUSTAVE gebruikt om bij alle voornaamwoorden de genitief aan te duiden.

Transcategorisaties

- PD > AA: *niet*, zoals in: *dade hi (God) sine hoede daer af (van het geschapene) iet, het soude weder al werden niet, want het van niet quam algader; daeromme eist al niet, dattu in einigher manieren twivelen menghes.*
- PD > ADV: *hij is al tachtig jaar; de zon schijnt al; hij werkt veel; Altion (l. Alcion) es i uoghelkin meere dan die mussce siin jn die zee es uele sijn ganc.*
- PD > CONJ: *noweder*, zoals in: *no weder man no wijf; na dat gelof so bleef lutgart Van allen rowe so verclart Dat dinc in ertrike engheene En was no weder groet no cleene Die mochte ontpointen haren moet.*
- PD > NOU-C: *een hij; het pas ontdekte zelf; de grote onbekende men; de zijnen; ten onzent; de meesten.*

5.8 Bijzetsel (Adposition - tag: ADP)

Onder de bijzetsels vallen de voorzetsels, achterzetsels en omzetsels. In navolging van de nieuw te verschijnen e-ANS kiezen we ervoor om, in tegenstelling tot andere tagsets, combinaties als *onder iets door* als omzetsels te beschouwen, omdat zowel *onder* als *door* van belang zijn voor de betekenis van deze bijzetsels. De versmolten voorzetsels *ten* en *ter* van CGN/D-Coi worden hier gezien als een combinatie van een voorzetsel en een verbogen vorm van het lidwoord.

Voor de bijzetsels onderscheiden we de volgende kenmerken:

Type (tags: pre, post, circ)

Er wordt onderscheid gemaakt tussen voorzetsel (*preposition*), achterzetsel (*postposition*) en omzetsel (*circumposition*). Hieronder volgen enkele voorbeelden van deze drie types:

- Voorzetsel: *in de tuin*; *op de fiets*; *achter het huis*. Hiertoe rekenen we, net als CGN/D-Coi, ook *te* en in ouder Nederlands *om* (zonder *te*) die aan infinitieven voorafgaan. Ook voorzetsels voorafgaand aan *te*, zoals *om*, *met* en *door*, worden gezien als voorzetsels.
- Achterzetsel: *oec ne mach hi niet leuen hine si den watre beneuen* ('beneven').
- Omzetsel: *als men hem van jongs afleert*; *hij loopt onder de trap door*.

Type		
preposition	postposition	circumposition
pre	post	circ

'Government' (tags: gen, dat, acc)

Er wordt bij adposities die een bepaalde naamval vereisen onderscheid gemaakt tussen +genitief (+*genitive*), +datief (+*dative*) en +accusatief (+*accusative*). Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands.

Government		
+genitive	+dative	+accusative
gen	dat	acc

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

Transcategorisaties

- ADP > ADV:⁵⁹ *ik zag hem van achter; boven zag ik haar niet; ik ga naar buiten.*
- ADP > CONJ: *je moet nablijven tot je hebt nagedacht!; niettegenstaande het feit dat...; voor hij vertrok, was zij al weg.*
- ADP > NOU-C: *(de) voor(s) en tegen(s); (de) pro('s) en contra('s).*

⁵⁹ Het is soms lastig te zien of de richting van transcategorisatie van bijzetsel naar bijwoord is of andersom.

5.9 Voegwoord (Conjunction - tag: CONJ)

Voegwoorden maken het verband tussen woorden of zinnen duidelijk. Voor de voegwoorden onderscheiden we de volgende kenmerken:

Type (tags: coor, sub, comp, neg, expl, rel, qual)

Er wordt onderscheid gemaakt tussen nevenschikkende voegwoorden (*coordinating conjunctions*) en onderschikkende voegwoorden (*subordinating conjunctions*).

Type	
coordinating	subordinating
coor	sub

Alle tagsets hebben nevenschikkende en onderschikkende voegwoorden. Hiernaast hebben GUSTAVE en het Corpus Oudnederlands het type vergelijkend en hebben het Corpus Gyseling en CRM explicatieve, ontkennende, relatief-adverbiale, vergelijkende voegwoorden, voegwoorden van hoedanigheid en een aparte categorie voor *alse* + te.

In TDN wordt onderscheid gemaakt tussen explicatieve (*explicative*), nevenschikkende (*coordinating*), onderschikkende (*subordinating*), ontkennende (*negative*), relatief-adverbiale (*relative*), vergelijkende (*comparative*) voegwoorden en voegwoorden van hoedanigheid (*quality*). Alle types worden hier overgenomen, behalve het type *alse* + te, omdat *alse* hier gezien wordt als een voorzetsel.

Hieronder volgen enkele voorbeelden uit het Corpus Gyseling:

- (1) Nevenschikkend: *ende; iof; sonder (so wie garen coept met iemene sonder met sinen bordate ... hi sal verburen...); bedi (bedi hets valsc); want; newaere (ne waere dese moreide moeten wesen ghelijst binnen egghen)*
- (2) Onderschikkend: *dat; of; als(e); bedi; hoe (dit sijn poynte ... hoemen halden sal); des, dies (alsoe vele ... dessi die rente moghen wulcoepen (dessi: ‘zodat zij’))*
- (3) Vergelijkend: *dan (niet dan eene maeltijt); ghelike dat; ghelikerwijs dat (gheliker wijs dat hier voren gescreven es); dat/iof (laken van alsulker varwe dat hie wille)*
- (4) Ontkennend: *ne/en (niemene ne moet erue copen in M. hine hore ten schependome; dī den anderen mint hīne sal vme nit vntsīn. hīne sūl vme sechen dat vme erlik is.); alse ... niet*
- (5) Explicatief: *alse (camen voer ons alse voer scepnen)*
- (6) Relatief-adverbiaal: *dat/als (int jaer ons heren dat/alse men scriuet)*
- (7) van Hoedanigheid: *als (Pietre vten Houe alse prouiseur ende voersire van den huse van sente Obbrecht)*

Type Extended						
coordinating	subordinating	comparative	negative	explicative	relative	of quality
coor	sub	comp	neg	expl	rel	qual

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

Transcategorisaties

- CONJ > ADP: *mits hardwerkend kan hij voor het examen slagen; naargelang de wens van de regering blijft men binnen.*
- CONJ > ADV: *dat was de bedoeling althans.*
- CONJ > NOU-C: *de althansen van de dagvaarding; een maar weten te vinden.*

5.10 Tussenwerpsel (Interjection - tag: INT)

Tot de tussenwerpsels of interjecties rekenen we de woorden die syntactisch gezien op zichzelf staan en die geen relaties hebben met andere woorden. Ook woorden die tot andere woordsoorten behoren, zoals *hemel* (NOU-C), *Jezus* (NOU-P), *goed* (AA), *toe* (ADV) en *help* (VRB), kunnen soms als zelfstandige taaluitingen optreden en zijn dus functioneel gezien interjecties. Hierin volgen we de ANS.⁶⁰ We wijken hierin af van CGN/D-Coi, die geen transcategorisatie van andere woordsoorten naar tussenwerpsels toelaat, omdat daarin lexicaal getagd wordt.

Er zijn geen verdere kenmerken bij tussenwerpsels, aangezien ze altijd op zichzelf staan voorkomen en er geen syntactische relaties zijn met andere woorden.

Voor de tussenwerpsels onderscheiden we alleen uitgangen:

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

Transcategorisaties

- INT > NOU-C: *op de onthulling van de winnaar volgde een aha; het adieu eindigt een gesprek, terwijl een hallo dat begint; het vaarwel valt hem zwaar; toen ging ik naar den diender toe, en die stopte me in de kiekeboe.*
- INT > AA: *ik vind het voorstel oké; dat is best wel een oké idee van je; hopelijk houdt de hond zich koest.*
- INT > ADV: *ze wisten wel: op zoo een eersten schouwdag gaat alles nog koest.*

⁶⁰ Zie de [Algemene inleiding](#) van tussenwerpsels.

5.11 Overblijfsel (Residual - tag: RES)

Tot de overblijfsels rekenen we de woordvormen die niet tot één van de hierboven genoemde categorieën behoren. Hierin wijken we af van CGN/D-Coi, CRM en GUSTAVE.

CGN, D-Coi, tagset Corpus Gysseling, CRM, GUSTAVE en TaCO divergeren hier dusdanig in de (hoeveelheid) tags dat de verschillen tussen de tagsets ter verduidelijking zijn weergegeven in een tabel.

Overblijfsels uit de verschillende tagsets						
Type	CGN	D-Coi	Gysseling	CRM	GUSTAVE	ONL
achtergrondgeluid	(achter)	(achter)				
afbreking	(afgebr)	(afgebr)				
afkorting		(afk)				
commentaar	(comment)	(comment)				
deeleigen	(deeleigen)	(deeleigen)				
meta	(meta)	(meta)				
onverstaanbaar	(onverst)	(onverst)				
onduidelijk			onduidelijk	(unclear)	onduidelijk	
symbool		(symb)				
vreemd	(for)	(for)	niet getagd	(Latin)	vreemd	Gr. (Grieks) Lat. (Latijn)

Voor de overblijfsels onderscheiden we de volgende kenmerken:

Type (tags: aff, for, form, meta, symb, uncl, oth)

De volgende types (grotendeels uit het bovenstaande) achten wij relevant voor de overblijfsels.⁶¹ Er wordt onderscheid gemaakt tussen affixen (*affix*), formules (*formula*), onduidelijke woorden (*unclear*), symbolen (*symb*), vreemde woorden (*foreign*), woorden in zelfnoemfunctie (*meta*) en overig (*other*). Hieronder zijn enkele voorbeelden gegeven van de types die we onderscheiden, en waar nodig een uitleg.

- Affix: *isme; ultra; super; de onscheidbaere voorzetsels be, ge, er, her, ver, ont, zyn ten allen tyde kort, en valt daerom de klank op de volgende lettergreep; ferro-, in de chemie de oude maar nog veel gebruikte term, ter aanduiding van 2-waardig ijzer.*
- Formule: $E=mc^2$
- Onduidelijk: dit zijn zowel gevallen waarbij het niet uit te maken is wat de woordsoort is, of omdat de context onduidelijk is of omdat het woord/de woordvorm zelf onduidelijk is:

⁶¹ Achtergrondgeluid is niet relevant voor corpustagging, afkortingen en afbrekingen zijn relevant voor alle woordsoorten, commentaar hoort niet in de tagging en deeleigen is hier een deel van een NOU-P.

Item van, drien vaten wijns ieghens [...]en .lx. marc; sijn si rentemeestere. baliuwe. scoutaten. G[...]el[...]ren. wie[datse] s[ijn]; Abraham woende dar amre. In dat dal dat hiet nambre⁶²

- Symbool: \$; %; @
- Vreemd: dit zijn de woorden die niet in Nederlands syntactisch verband staan: *Distichon heroicum; Sapphicum carmen; de goudsbloem (Calendula officinalis) is een plant die behoort tot de composietenfamilie (Compositae of Asteraceae)*⁶³
- Zelfnoemfunctie: *lopen is een werkwoord; help is een veelgebruikte uitroep in penibele situaties; het teken ±*
- Overig: dit zijn alle gevallen die niet aan één van de genoemde typeringingen voldoen.

Type						
affix	foreign	formula	meta	symbol	unclear	other
aff	for	form	meta	symb	uncl	oth

Uitgang

De uitgang kan benoemd worden, los van verdere specificatie. Bij sommige woordsoorten duidt dit op verbuiging of vervoeging. Deze methode is gebruikt bij het taalkundig verrijken van het Corpus Gysseling en CRM.

Hieronder volgt een lijst met mogelijke uitgangen:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- nt, '-nt': *Inflection -nt*
- other, 'Other': *Other values*

Transcategorisaties

- RES(Aff) > NOU-C: *hij heeft weer een nieuw isme bedacht.*
- RES(Aff) > ADV: *hij zit weer super te bepalen; ik vind haar super.*

⁶² Dit zijn de woorden in het Corpus Gysseling en CRM die beginnen met een 9 met uitzondering van de woorden die beginnen met 91, want dit zijn de tussenwerpsels.

⁶³ In de zin *de calendula officinalis behoort tot de asteraceae of compositae* zijn deze woorden niet als vreemd getagd, maar als naamwoorden, omdat ze in Nederlands zinsverband staan.

6 Kerntagset

Hierboven is het uitgebreide vocabulaire besproken van TDN. In de kerntagset (zie ook sectie 2.4) wordt gedefinieerd wat onzes inziens al het taalkundig verrijkt historisch materiaal minimaal aan informatie zou moeten bevatten. Het trainings- en evaluatiemateriaal dat in de context van CLARIAH+ wordt geproduceerd, zal verrijkt worden met de kerntagset.

Bij het vaststellen van de kerntagset hebben we met de volgende zaken rekening gehouden:

- Het moet haalbaar zijn om met de kerntagset binnen een acceptabel tijdsbestek grote hoeveelheden handmatig nagekeken trainings- en evaluatiemateriaal te maken;
- De kerntagset moet die elementen bevatten die bruikbaar zijn voor allerlei onderzoek met behulp van diachrone corpora;
- De kerntagset moet aanknopingspunten bieden voor mapping naar andere gebruikte tagsets.

6.1 Alle woordsoorten

Geschreven vorm (tags: **abbr**, **mis**, **trunc**)

Er wordt onderscheid gemaakt tussen afgekort (*abbreviated*), afgebroken (*truncated*) en drukfout of digitaliseringsfout (*misspelled*). Dit kenmerk wordt alleen toegekend als de geschreven vorm op één van de genoemde manieren afwijkt van de reguliere vorm.

Written Form (WF)		
abbreviated	misspelled	truncated
abbr	mis	trunc

6.2 Naamwoorden (Nouns - tags: **NOU-C**, **NOU-P**)

6.2.1 Zelfstandig naamwoord (Noun - tag: **NOU-C**)

Voor de naamwoorden onderscheiden we voor de kerntagset het volgende kenmerk:

Getal (tags: **sg**, **pl**)

Er wordt onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*).

Number	
singular	plural
sg	pl

Getal is doorgaans te herkennen aan de uitgang *-en/-s* in het meervoud en het ontbreken van deze uitgang in het enkelvoud; deze uitgang mag niet verward worden met vergelijkbare naamvalsuitgangen.

6.2.2 Eigennaam (Proper Noun - tag: NOU-P)

Voor de eigennamen onderscheiden we voor de kerntagset geen kenmerken.

6.3 Bijvoeglijk naamwoord-Bijwoord (Adjective-Adverb - tag: AA)

Voor de bijvoeglijke naamwoorden-bijwoorden onderscheiden we voor de kerntagset de volgende kenmerken:

Graad (tags: pos, comp, sup)

Er wordt onderscheid gemaakt tussen de positief (*positive*), de comparatief (*comparative*) en de superlatief (*superlative*), waarbij het lemma de positief als waarde heeft.

Degree		
positive	comparative	superlative
pos	comp	sup

Graad is doorgaans te herkennen aan de uitgang *-er/-re* in de comparatief en aan de uitgang *-st* in de superlatief.

Positie (tags: prenom, postnom, free)

Er wordt onderscheid gemaakt tussen prenominaal (*prenominal*), postnominaal (*postnominal*) en vrij (*free*).

Position		
prenominal	postnominal	free
prenom	postnom	free

Hiermee kan het attributieve gebruik (prenominaal & postnominaal) gescheiden worden van het niet-attributieve gebruik (vrij: predicatief & adverbiaal).

6.4 Bijwoord (Adverb - tag: ADV)

Voor de bijwoorden onderscheiden we voor de kerntagset het volgende kenmerk:

Type (tags: reg, pron)

Er wordt onderscheid gemaakt tussen een gewoon bijwoord (*regular adverb*), zoals *echter*, en een pronominaal of voornaamwoordelijk bijwoord (*pronominal adverb*), zoals *daarover*.

Type	
regular adverb	pronominal adverb
reg	pron

6.5 Werkwoord (Verb - tag: VRB)

Voor de werkwoorden onderscheiden we voor de kerntagset de volgende kenmerken:

Werkwoordsvorm (tags: fin, inf, prespart, pastpart)

Er wordt onderscheid gemaakt tussen de finiete vorm (*finite*), de infinitief (*infinitive*), het tegenwoordig/onvoltooid deelwoord (*present participle*) en het verleden/voltooid deelwoord (*past participle*).

Finiteness			
finite	infinitive	present participle	past participle
fin	inf	prespart	pastpart

Tempus/tijd (tags: pres, past)

Er wordt bij finiete vormen van het werkwoord onderscheid gemaakt tussen de tegenwoordige tijd (*present tense*) en verleden tijd (*past tense*).

Tense	
present	past
pres	past

6.6 Telwoord (Numeral - tag: NUM)

Voor de telwoorden onderscheiden we voor de kerntagset de volgende kenmerken:

Type (tags: card, ord)

Er wordt onderscheid gemaakt tussen hoofdtelwoord (*cardinal*) en rangtelwoord (*ordinal*).

Type	
cardinal	ordinal
card	ord

Rangtelwoorden zijn doorgaans te herkennen aan de uitgang *-de/-ste*.

Positie (tags: prenom, postnom, free)

Er wordt onderscheid gemaakt tussen prenominale (*prenominal*), postnominale (*postnominal*) en vrije (*free*) positie.

Position		
prenominal	postnominal	free
prenom	postnom	free

Hiermee kan het attributieve gebruik (prenominaal & postnominaal) van het niet-attributieve gebruik (vrij) worden onderscheiden.

Weergave (tags: **dig, let, rom, mix-dig, mix-rom**)

Er wordt onderscheid gemaakt tussen telwoorden die worden weergegeven met cijfers (*digits*), letters (*letters*), Romeinse cijfers (*roman*), mengvormen met cijfers en letters (*mixed-digits*) en mengvormen met Romeinse cijfers en letters (*mixed-romans*).

Representation				
digits	letters	roman	mixed-digits	mixed-romans
dig	let	rom	mix-dig	mix-rom

6.7 Voornaamwoord (Pronoun-Determiner - tag: PD)

Voor de voornaamwoorden onderscheiden we voor de kerntagset de volgende kenmerken:

Type (tags: **d-p, excl, indef, pers, poss, recip, refl, w-p**)

Er wordt onderscheid gemaakt tussen bezittelijke voornaamwoorden (*possessive*), d-pronomina (*d-pronomina*), onbepaalde voornaamwoorden (*indefinite*), persoonlijke voornaamwoorden (*personal*), reciproque voornaamwoorden (*reciprocal*), reflexieve voornaamwoorden (*reflexive*), uitroepende voornaamwoorden (*exclamative*) en w-pronomina (*w-pronomina*).

Type							
d-pronoun	exclamative	indefinite	personal	possessive	reciprocal	reflexive	w-pronoun
d-p	excl	indef	pers	poss	recip	refl	w-p

Subtype lidwoord (tags: **art, oth**)

Wanneer een lidwoord duidelijk kan worden onderscheiden, dan wordt het lidwoord aangegeven als subtype van d-pronomen in het geval van een bepaald lidwoord en als subtype van een onbepaald voornaamwoord in het geval van een onbepaald lidwoord. De niet-lidwoorden worden bij deze voornaamwoorden aangegeven met de waarde *other*. De toekenning van dit subtype gebeurt alleen als de betreffende voornaamwoorden in prenominale positie staan.

Subtype			
d-pronoun / demonstrative pronoun		indefinite pronoun	
article	non-article	article	non-article
art	oth	art	oth

Positie (tags: prenom, postnom, free)

Er wordt onderscheid gemaakt tussen prenominaal (*prenominal*), postnominaal (*postnominal*) en vrij (*free*).

Position		
prenominal	postnominal	free
prenom	postnom	free

Hiermee kan het attributieve gebruik (prenominaal & postnominaal) van het niet-attributieve gebruik (vrij) wordt onderscheiden.

6.8 Bijzetsel (Adposition - tag: ADP)

Voor de bijzetsels onderscheiden we voor de kerntagset het volgende kenmerk:

Type (tags: pre, post, circ)

Er wordt onderscheid gemaakt tussen voorzetsel (*preposition*), achterzetsel (*postposition*) en omzetsel (*circumposition*).

Type		
preposition	postposition	circumposition
pre	post	circ

6.9 Voegwoord (Conjunction - tag: CONJ)

Voor de voegwoorden onderscheiden we voor de kerntagset het volgende kenmerk:

Type (tags: coor, sub)

Er wordt onderscheid gemaakt tussen nevenschikkende voegwoorden (*coordinating conjunctions*) en onderschikkende voegwoorden (*subordinating conjunctions*).

Type	
coordinating	subordinating
coor	sub

6.10 Tussenwerpsel (Interjection - tag: INT)

Voor de tussenwerpsels onderscheiden we voor de kerntagset geen kenmerken.

6.11 Overblijfsel (Residual - tag: RES)

Voor de overblijfsels onderscheiden we voor de kerntagset het volgende kenmerk:

Type (tags: **aff, for, form, meta, symb, uncl, oth**)

De volgende types achten wij van toepassing op de overblijfsels. Er wordt onderscheid gemaakt tussen affixen (*affix*), formules (*formula*), onduidelijke woorden (*unclear*), symbolen (*symb*), vreemde woorden (*foreign*), woorden in zelfnoemfunctie (*meta*) en overig (*other*).

Type						
affix	foreign	formula	meta	symbol	unclear	other
aff	for	form	meta	symb	uncl	oth

Door de verschillende types te onderscheiden kunnen tokens van allerhande aard toch van een tag worden voorzien, wanneer ze niet bij één van de andere woordsoorten onder te brengen zijn. Het type *other* kan hierbij toegekend worden aan tokens waarvoor we geen type hebben bedacht.

6.12 Uitgangen

De tagset waarmee het Corpus Gysseling en CRM werden geannoteerd, geeft de uitgangen als kenmerk aan bij iedere woordsoort. Een uitgebreide analyse van de manier waarop de tagset is toegepast in de verschillende corpora en wat de interne en onderlinge verschillen zijn, heeft geleerd dat het taggen van de uitgang niet altijd consequent is toegepast en dat de regels niet overal even helder zijn geformuleerd. Dezelfde woordvormen worden met of zonder uitgang getagd: Gysseling heeft *naest* zowel met uitgang-0 als met uitgang-t. En *na(e)r* met uitgang-0 en uitgang-r. In CRM is het overgrote deel van de gevallen uitgangsleloos, terwijl daar wel vormen zijn die een uitgang hebben. Het is ook niet altijd evident: *parochie* bijvoorbeeld moet een uitgang-e krijgen omdat er ook vormen aangetroffen worden zonder -e, bijvoorbeeld *prochi*.

Het voorstel is om uitgangen in de kerntagset niet te taggen. De belangrijkste redenen zijn dat de regels hiervoor lastig te definiëren en toe te passen zijn en dat de kerntagset minder betrouwbaar wordt, omdat uitgangen door verschillende annotatoren anders zullen worden gecodeerd. Bovendien kunnen de resultaten vaak worden benaderd door zoekopdrachten als [\[word=".*s\\$" & lemma != ".*s" & pos="NOU-C"\]](#) voor een zelfstandig naamwoord met genitief-s of [\[pos="NOU-C" & pos_number="sg" & word=".*e" & lemma != ".*e"\]](#) voor een zelfstandig naamwoord in het enkelvoud met uitgang-e.

7 Referenties

- Decorte, S., m.m.v. Kruyt, T., Ruitenbergh, T. en Depuydt, K. (2003). Aanzet tot een tagset en tagtoekenningsmethode voor de Geïntegreerde taalbank, intern discussiestuk.
- Dietz, F., Koppen, M. van, & Kramer, I. (2017). Intra-author variation in negation: the letters of PC Hooft.
- Duinhoven, A.M. (1988). *Middel nederlandse syntaxis: Synchron en diachroon. Deel 1: De naamwoordgroep*. Leiden: Nijhoff.
- Duinhoven, A.M. (1997). *Middel nederlandse syntaxis: Synchron en diachroon. Deel 2: De werkwoordgroep*. Groningen: Nijhoff.
- Eynde, F. Van (2004). Part of speech tagging en lemmatisering van het Corpus Gesproken Nederlands. KU Leuven.
- Eynde, F. Van (2005). Part of speech tagging en lemmatisering van het D-COI corpus. Intermediate, projectinternal version.
- Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J., & Van den Toorn, M. (1997). *Algemene Nederlandse Spraakkunst*. Tweede, geheel herziene druk. Groningen: Martinus Nijhoff.
- Leech, G., & Wilson, A. (1996). Recommendations for the morphosyntactic annotation of corpora. *Relatório técnico, Expert Advisory Group on Language Engineering Standard (EAGLES)*.
- Mooijaart, M. (2012). Het lemma in het GiGaNT-lexicon. *INL Working Papers - Taalbank Nederlands 4*. Leiden.
- Quak, A., & Horst, J.M. van der (2002). *Inleiding Oudnederlands*. Leuven University Press.
- Rem, M., & Halteren, H. van (2007). Tagging and Lemmatization Manual for the corpus van Reenen-Mulder and the Adelheid 1.0 Tagger-Lemmatizer. Radboud University Nijmegen.
- Ruitenbergh, T., Does, J. de, en Depuydt, K. (2010). Developing GiGaNT, a lexical infrastructure covering 16 centuries. In *Proceedings of the XIV Euralex International Congress* (pp. 468-76). Ljouwert: Afûk.
- Ruitenbergh, T. & Van pellicom, K. (2012). De morfositactische module van het GiGaNT-lexicon. *INL Working Papers - Taalbank Nederlands 3*. Leiden.
- Sijs, N. van der (1996). *Leenwoordenboek. De invloed van andere talen op het Nederlands*. Den Haag: Sdu.
- Sijs, N. van der (2020). Historische taalkunde en Digital Humanities: samen naar een mooie toekomst. *Tijdschrift voor Nederlandse Taal-en Letterkunde*, 135(4), 384-405.
- Velde, F. Van de (2009). *De nominale constituent: structuur en geschiedenis*. Leuven University Press.

Bijlage 1: overzicht verschillen in lemmatisering

We volgen de lemmatiseerprincipes zoals geformuleerd en toegepast voor de historische woordenboeken van het INT en de computationele lexica van het INT (GiGaNT-Hilex en GiGaNT-Molex).⁶⁴ Dit betekent wel dat er op bepaalde punten structureel wordt afgeweken van wat er voor andere projecten, corpora en tagsets gebeurd is. Hieronder worden de meest significante afwijkingen in lemmatisering kort weergegeven.

- Diminutieven krijgen een eigen lemma, met het verkleinwoord als ingang. Hierin wijken we af van bijvoorbeeld GUSTAVE en CGN/D-Coi, waar alleen uitzonderingen als *meisje* en *toetje* een eigen lemma krijgen;
- Het lemma van adjectieven - in TDN AA's - is in de regel altijd, in tegenstelling tot GUSTAVE, de positief, hoewel er uitzonderingen zijn voor bijzondere comparatief- en superlatiefvormen als *beter*, *minst* en *benedenst*; zie daarvoor ook voetnoot 34;
- Bijvoeglijk gebruikte deelwoorden - in TDN AA's - krijgen, in tegenstelling tot bij bijvoorbeeld GUSTAVE en CGN/D-Coi, de basisvorm van het deelwoord en niet de infinitief als lemma;
- Rangtelwoorden krijgen, in tegenstelling tot bij CGN/D-Coi, hun eigen lemma en niet het lemma van het corresponderende hoofdtelwoord;
- Voornaamwoorden krijgen, in tegenstelling tot bij CGN/D-Coi, bij suppletie (bijvoorbeeld nominatief *ik* versus oblique *me*) en bij alle gevallen van reductie de onverbogen standaardvorm als lemma (CGN/D-Coi doet dit bij gevallen zonder klinker, zoals *z'n* (met lemma *zijn*) en *'t* (met lemma *het*), maar niet bij gevallen met klinker, zoals *me* (met lemma *me*));
- De twee versmolten voorzetsels *ten* en *ter* van CGN/D-Coi worden hier behandeld als combinatie van voorzetsel en (verbogen) lidwoord; *ten* wordt gelemmatiseerd als *te+de*, *te+het*, *te+die* en *te+dat*, *ter* als *te+de*, *te+dat* en *te+die*;
- Vreemde woorden die getagd worden met RES(type=for) en woorden waarvan de woordsoort onduidelijk is en die de tag RES(type=uncl) krijgen, hebben in tegenstelling tot bij bijvoorbeeld GUSTAVE, een liggend streepje (-) als lemma.

⁶⁴ Zie: https://ivdnt.org/images/stories/onderzoek_en_onderwijs/publicaties/TaalbankWorkingpaper4.pdf.

Bijlage 2: totaaloverzicht voornaamwoorden

Op basis van het materiaal in de vier historische woordenboeken van het Nederlands - het *Oudnederlands Woordenboek* (ONW), het *Vroegmiddelnederlands Woordenboek* (VMNW), het *Middelnederlandsch Woordenboek* (MNW) en het *Woordenboek der Nederlandsche Taal* (WNT) - is er een inventarisatie gemaakt van de daarin voorkomende voornaamwoorden. We onderscheiden acht verschillende voornaamwoorden en geven per voornaamwoord het modern Nederlands lemma en per woordenboek het historisch lemma dat met dat voornaamwoord correspondeert. Dat kan betekenen dat een historisch lemma uit een woordenboek op meerdere plekken kan verschijnen. Dit komt met name voor in het MNW en het WNT, omdat in deze woordenboeken lemmata meerdere woordsoorten kunnen hebben.

In het overzicht zijn voor de volledigheid ook de verwijlslemmata (afgekort VWL) opgenomen, omdat ze vaak verwante spellingen weergeven.⁶⁵ Omdat deze verwijlslemmata naar meerdere lemmata kunnen verwijzen, hebben ze soms meerdere moderne lemmata.

Persoonlijke voornaamwoorden

N.B. In navolging van de ANS vallen ook het niet-verwijzende *het* en *men* hieronder.

Lemma	ONW	VMNW	MNW	WNT
du	thū	du	du dich	du dij(III) (VWL)
er	-	-	-	er(I) der (VWL)
gij	gi	ghi	gi ju ure(III) us(II) ou(IV) hu(II) uwer, uwes jo(II) u(III) uus(III) (VWL)?	gij u(II) je(I) (VWL) uw(II) (VWL) uwe(II) (VWL) uwes(I) (VWL) je(II) (VWL) ou(III) (VWL) au(III), auw (VWL)
gijlieden	-	ghiliede	-	gijlieden
het	hit	het	het et	het(I) 't(I)
hij	hi	hi	hi imme(III) om(II) ome(II) omme(I)	hij hum(II) zijn(IV) -en(IV) (VWL) -i (VWL)

⁶⁵ Voor zover ze gevonden zijn. Er kan hier (nog) niet gemakkelijk op worden gezocht.

			un ye(II) sijn(II) -ene -ne hon(I) (VWL)	
hij / het	-	-	sire(III)	zijns(I)
hij / ik / men	-	-	-	'm, m' (VWL)
hij / zij	-	-	on(II) in(II)	hem(II) (VWL)
ik	ik	ic	ic ich mi(I) mie mijns mien(II) (VWL)	ik me(I) (VWL) mijn(IV) (VWL)
ir	-	ir	-	-
jj	-	-	jou	jj jou ju
jullie	-	-	-	jullie jelui
men	man	men	men me(III) mi(IV)	men me(III) (VWL)
U Ed.	-	-	-	UE
ulieden	-	-	ulieden uwerlieden	ulieden ulder(II) (VWL) ulie (VWL) ulieder (VWL) ulle(I) (VWL) ullie(I) (VWL) ului(den) (VWL)
ulieden / uw~liefde	-	-	-	UL (VWL)
u	-	-	-	u(III) uw(III) (VWL) uwe(III) (VWL) uwes(II) (VWL)
wij	wi	wi	wi ons(I)	wij(I) ons(II) (VWL)

			us(I) wer(IV) uus(I) wir wjr (VWL) wie(V) (VWL) wij (VWL)	me(II) (VWL) wi(II) (VWL) we (VWL) wie(III) (VWL)
wijlieden	-	wiliede onslieden	wiliede	wijlieden wielu (VWL) wijder(II) (VWL) wijlder (VWL) wijnder (VWL) wilder (VWL) wuili(e) (VWL) wulder (VWL) wullie (VWL) wunder(I) (VWL)
zij	siu sia	si(II) si(III)	si(I) se soe hender hun(II) ore(III) see(II) hare(VI) oor hore(V) -er(II) -ge sou(VWL)? haer(VIII) (VWL)	zij haar(VII) 'r(III) hen(II) (VWL) hun(II) (VWL) zoe(I, II) (VWL) zo(I, II) (VWL) zoo(III, IV) (VWL) zu (VWL) er(III) (VWL)
zij / haar	-	-	hoir (VWL)	heur (VWL)
zijlieden	-	haerliede siliede	haerlieden hemlieden	zijlieden haarlieden, haarlui (VWL) hunlieden, hunlui (VWL) zijlder, zijle(VWL) zijlude(n), zijlui(den)(VWL) zijnder(I) (VWL) zijlder (VWL) zulder (VWL) zunder(II) (VWL) zullie (VWL)

Reflexieve voornaamwoorden

Lemma	ONW	VMNW	MNW	WNT
dij	thi	di	-	-
elkhaar	-	elkhare	-	-
haar	-	haer(II) haer(III)	-	-
hem	himo(I) himo(II)	hem(I) hem(II) hem(III)	umme(II) hi (onder hij) -ne (onder hij)	-
hemlieden	-	hemlieden	-	-
mij	mi	mi	-	-
ons	uns	ons	-	-
u	iu	u(II)	-	u(II) je(II) (VWL) uw(II) (VWL) uwe(II) (VWL) uwes(I) (VWL) ou(III) (VWL) au(III), auw (VWL)
zich	sik(I)	sich(I)	sich(II) si(III)	zich
zichzelf	-	-	-	zichzelf
zijzelf	-	-	-	zijzelf

Reciproque voornaamwoorden

Lemma	ONW	VMNW	MNW	WNT
elkander/elkaar	-	elcanderen	-	elkander
haarlijkander	-	haerlijcander	-	-
hen	hin	hem	-	-
iegelijkander	-	ieghelijcander	-	-
malkander / mekaar / malkaar	-	manlijcanderen	malcander, malcanderen mancander	malkaar, malkander (VWL)
u	-	u(III)	-	-
zich	sik(II)	sich(II)	-	-

Bezittelijke voornaamwoorden

Lemma	ONW	VMNW	MNW	WNT
dijn	thīn	dijn	dijn	dijn
haar	hiro(I) hiro(II)	-	hore hoere(II) ure(V) ur hare(VI) oor haer(IX) (VWL) heur (VWL)	haar(VIII) 'r(IV) er(IV) (VWL)
hunlieden	-	-	hemlieden	zijlieden haarlieden, haarlui (VWL) hunlieden, hunlui (VWL) zijlder, zijle(VWL) zijlude(n), zijlui(den)(VWL) zijnder(I) (VWL) zijlder (VWL) zulder (VWL) zunder(II) (VWL) zullie (VWL)
hun	-	-	hun(III)	hun(III)

			hen(I) hon(II) (VWL) honne (VWL)	hen(III) (VWL)
jouw	-	-	-	jouw je(III) (VWL)
jouw / uw	-	-	-	juw
mijn	mīn	mijn	mijn mi(II) me(II) mien(III) (VWL) miere (VWL)	mijn(III)
ons	unsa	onse	onse uus(II) use (VWL)? ons(II) (VWL)	ons(III)
U Ed.	-	-	-	UE
uwlieden	-	-	uwerlieden	ulieden ulder(II) (VWL) ulie (VWL) ulieder (VWL) ulle(I) (VWL) ullie(I) (VWL) ului(den) (VWL)
uw	iuwa	uwe	u(II) ure(IV) us(III) hu(II) uw	uw(I) u(IV) (VWL) uw(III) (VWL) uwe(III) (VWL) ou(III) (VWL) au(III), auw(VWL)
zij / haar	-	-	hoir (VWL)	heur (VWL)
zijn	sīn	sijn	sijn(II) si(II) sire(II) (VWL)?	zijn(III) zijn(V) zijns(II) (VWL)

Vragend-betrekkelijke voornaamwoorden

De vragende en betrekkelijke voornaamwoorden zijn tezamen genomen; niet altijd zijn de betreffende lemmata zowel vragend als betrekkelijk.

Lemma	ONW	VMNW	MNW	WNT
dat	that(III)	dat(II)	dat(II)	dat (VWL) > die
dewelke	-	dewelke	diewelke	dewelke
die	thie(IV)	die(III)	die(II) de(II)	die, bet. II
die / dat	-	-	des(II) (VWL)	-
hetgeen	-	-	-	hetgeen hetgunt
hetwelk	-	-	hetwelc datwelc	hetwelk
hoedaan	-	-	hoedaen	hoedaan
hoedanig	-	hoedanich	hoedanich	hoedanig
hoegedaan	-	hoeghedaen	hoegedaen	-
hoegedanig	-	hoeghedanich	hoegedanich	-
hoezulk	-	-	hoezulk	-
waarbij	-	waerbi	= BW	= BW
wanenaf	-	= BW	= BW	wanof
wat	wat	wat wes	wat wes	wat(I) wes wa(III) (VWL) watte(I) (VWL)
watkunne	-	-	watconne	-
watterhande	-	-	watterhande	-
watterlei	-	-	watterleye	-
weder	-	-	-	weder(IV)
wel	-	-	wel(II) weel(II) weele(I) (VWL)? wele(I) (VWL)?	wel(IV) wiele(II) (VWL)

welk	wilik	welc	welc wilc walc(II) willeg	welk welkst wilk(I) (VWL) wek(II) (VWL) weuk (VWL) wuk (VWL)
welkdanig	-	-	-	welkdanig
welker	-	welkhare	wilcoor	welker
welkerhande	-	welkerhande	welkerhande	-
welkerkunne	-	-	welcerconne	-
welkerlei	-	-	welkerleie	-
welkerlijk	-	-	welcerlijc	-
welkertieren	-	welcertiere(n)	welkertiere	-
wie	wie	wie	wie wime, wieme wer(III) wis(IV) wime(III)(VWL)? wi(IV) (VWL) wies(II) (VWL) wim (VWL)? win(IV) (VWL)?	wie wij(III,IV)(VWL) weem(II) (VWL) weme(n)(II)(VWL) wemes (VWL) wiemes (VWL) wien(I) (VWL) wiene (VWL)
wiedaan	-	-	wiedaen (VWL)?	-
wiedanig	-	-	wiedanich	-
wiegedaan	-	wieghedaen	wiegedaen(VWL)?	-
woedaan	-	-	woedaen (VWL)	-
woedanig	-	-	woedanich(VWL)?	-
woegedaan	-	-	woegedaen	-
zo	-	-	-	zoo(II)
zulks	-	-	-	zulks

Aanwijzende voornaamwoorden

N.B. Bij woorden die (ook) lidwoorden zijn, staat '= LID'.

Lemma	ONW	VMNW	MNW	WNT
aldanig	-	-	aldanich	-
aldulk	-	-	aldulc	-
aldusdaan	-	aldusdaen(I)	aldusdaen	-
aldusdanig	-	aldusdanich	aldusdanich	aldusdanig
aldusgedaan	-	aldusghedaen	aldusgedaen	-
aldusgedanig	-	-	aldusgedanich	-
alduszulk	-	-	aldussulc	-
alsgedaan	-	-	alsegedaen	-
alzodaan	-	alsodaen(I)	alsodaen	-
alzodanig	-	alsodanich	alsodanich	alzoodanig
alzogedaan	-	alsoghedaen	alsogedaen	-
alzogedanig	-	alsoghedanich	-	-
alzulk	-	alsulc alsesulc (VWL) alsosulc (VWL)	alselc alsesulc alsolijc alsosulc	alzulk
alzulldanig	-	alsulcdanich	alsulcdanich	alzulldanig
alzulks	-	-	-	alzulks
alzusdaan	-	alzusdaen alsuster	alsustaen	-
alzusdanig	-	alzusdanich	alsustanich	-
alzusgedaan	-	-	alzusgedaen	-
dat	that(I) that(II) = LID	dat(I) dat(III) = LID	dat(I)	dat (VWL)
datgene	-	-	tgint	-
de	thie(III) = LID	die(IV) = LID	de(I) = LID & VNW	de = LID & VNW; bet. 1(VNW) & 3(LID) te(IV) = LID

degene	-	-	degene	degene
dergelijk	-	diereghelike	= BW	dergelijk diergelijk
desgelijk	-	desghelike, diesghelike desghelikes, diesghelikes	desgelijc	desgelijk diesgelijk
deze	these	deze	deze	deze
dezelfde	-	-	-	dezelfde
dezelfste	-	-	-	dezelfste
dezelve	-	-	dieselve	dezelve
dezelvige	-	-	-	dezelvige
dezulke	-	-	deselke dieselke	dezulke
die	thie(II)	die(II)	die(I) die(III) -re	die
die / dat	-	-	des(I) (VWL)	-
diegene	-	-	diegene	diegene
diezelfde	-	-	-	diezelfde
dit	thit / thit, thitti	dit	dit	(zie onder deze)
ditzelfde	-	-	-	ditzelfde
dulk	-	-	dulc (VWL)?	-
dusdaan	-	dusdaen	dusdaen	dusdaan
dusdanig	-	dusdanich	dusdanich	dusdanig
dusgedaan	-	dusghedaen	dusgedaen	dusgedaan(VWL)
duskgedaan	-	-	dusschedaen	-
dusgedanig	-	-	dusgedanich	-
dusk	-	-	dusc	dusch
duswaard	-	-	duswert	-
duszulk	-	dussulc	-	-

eadem	-	eadem	-	-
eigen	-	-	-	eigen, bet. IV
eigens	-	-	-	eigens
gene	-	ghene	gene goont	gene gunt
gewelk	= ONBEP. VNW	ghewelc	-	-
het	-	-	het = LID/AANW s(III) the(II) = LID	het(II) = LID 't(II) = LID 't(III) = LID(VWL)
hetgeen	-	-	-	hetgeen hetgunt (VWL)
hetzelfde	-	-	-	hetzelfde
hetzelve	-	-	-	hetzelve
hoogstdezelve	-	-	-	hoogstdezelve
idem	-	idem	-	idem (marginaal)
proper	-	= BNW & BW	proper bet. III	= BNW, BW
wat	-	wat wes	wat	-
zelf	self selvo	selve	selve(I)	zelf
zelfde	-	-	selfde	zelfde
zelfs	-	-	selfde	-
zelfste	-	-	selvest	zelfste zelveste (VWL) zelfst (VWL)
zelvig	-	-	selvich	zelvig
zo	-	-	-	zoo(II), bet. III
zodaan	-	-	sodaen	zoodaan
zodanig	-	sodanich	sodanich	zoodanig
zogedaan	sōgidān	soghedaen	sogedaen	-
zogedanig	-	-	sogedanich	-

zoveel	-	-	-	zooveel, bep.aank.
zozulk	-	so sulc	sosulc	-
zulk	sulik	sulc	sulc solijc selc (VWL) swelc (VWL)	zulk
zulkdaan	-	-	sulcdaen	-
zulkdanig	-	sulke danighe(VWL)	sulcdanich	zulkdanig
zulkgedaan	-	-	sulcgedaen	-
zulkerhande	-	sulkerhande	sulkerhande	-
zulkgedanig	-	sulcghedanich	sulcgedanich	-
zulks	-	-	sulcs	zulks
zus	-	-	-	zus(II)
zusdaan	-	-	susdaen	- (kop. bij zus)
zusdanig	-	susdanich	susdanich	- (kop. bij zus)
zusgedaan	-	susghedaen(I)	susgedaen	-
zusgedanig	-	-	susgedanich	-
zusk	-	-	-	zusk
zuslijk	-	-	suslijc	-

Onbepaalde voornaamwoorden

N.B. 1 Bij woorden die (ook) lidwoorden zijn, staat ‘= LID’.

N.B. 2 Alle onbepaalde telwoorden zijn hier onder onbepaalde voornaamwoorden gerekend.

N.B. 3 Samengestelde woorden op *-hande*, *-lei*, *-kunne*, *-tiere* en *-slachte* krijgen de woordsoort van het eerste deel; die is meestal onbepaald, maar af en toe ook aanwijzend of vragend-betrekkelijk. Zie ook aldaar.

Lemma	ONW	VMNW	MNW	WNT
al	al(II)	al(II)	al(I) alles (VWL)?	al(I)
allegader	-	algader	algader	allegaar
alleman	-	alleman	alman	alleman
allerhande	-	alrehande(I)	alrande (VWL)?	allerhande
allerkunne	-	-	alreconne, alrekunne	-
allerlei	-	-	-	allerlei
allermannelijk	-	alremanlijk	-	-
allerslachte	alleroslahta	alreslachte	alreslachte (VWL)	-
allertieren	-	alretiere(n)	alretiere (VWL)?	-
alles	-	-	-	alles
allijkeveel	-	-	allikeveel	allijkeveel
altegader	-	altegadere	altegader	= BW
alzomenig	-	alsomenig	-	-
alzowat	-	alsowat	-	-
alzulkerhande	-	-	alselkerhande	-
ander	andar(II)	ander	ander	ander
anderhande	-	-	anderhande	-
anderman	-	anderman	-	-
andertieren	-	-	andertiere	-
een	ēn(II) ēn(III) = LID	een(II) een(III) = LID	een(III) een(II) = LID ien (VWL)	een, bet. II = LID van onbepaaldheid
eender	-	-	eender	= BNW, BW
eeniegelijk	-	eeniegelijk	-	-
eenweder	-	-	eenweder	-

eenwelk	ēnwilik	-	-	-
el	-	el	el	el
eliemand	-	eliemene	-	-
elk	-	elc	elc	elk
elkeen	-	-	-	elkeen
elkerhande	-	-	elkerhande	-
elkerlijk	-	elkerlijc	elkerlijc	elkerlijk
elkermalk	-	-	elkermalc	-
elkertieren	-	-	elkertiere	-
elkman	-	-	elcman	-
elks	-	-	elx	-
elnegeen	-	elnegheen	-	-
els	-	-	els	-
enerhande	-	eenrehande, bet. 1	eenrehande	= BNW
enig	ēnig	enich	enich	eenig
enigerhande	-	enigherhande	enigerhande	eenigerhande
enigerkunne	-	-	enigerconne	-
enigerlei	-	-	enigerleie	eenigerlei
enigerslachte	ēnigerslahta	enigherslachte	enigerslacht	-
enigertieren	-	enighertiere	enigertiere	-
enigewelk	-	-	enigewelc	-
enigman	-	-	enichman	-
enkel	-	-	enkel engel(III)	enkel inkel
enthoeveel	-	-	-	enthoeveel
entwat	-	-	-	entwat (VWL)
entwie	-	-	-	entwie
ettelijk	-	-	itlijc	ettelijk
eenweinig	-	-	ewenich	-
geen	gēn	geen	geen	geen
gelijk	-	-	gelijc(V)	-
gelijkveel	-	-	gelikevele (VWL)	gelijkeveel

generhande	-	ghenerehande	geenrehande	geenerhande
generlei	-	-	geenreleie	geenerlei
genertieren	-	gheneretiere	-	-
genig	-	-	genich	geenig
genoeg	ginuog	ghenoech(I)	genoech gnoech (VWL)	genoeg
gewelk	giwilik	= AANW. VNW	-	-
giddeswelk	githeswelik	-	-	-
giemand	-	-	giemant	-
haargelijk	-	haerghelijc	haergelijc	-
haarlijk	-	haerlijc	haerlijc (VWL)	
hoemenig	-	hoemenig	-	hoemenig
hoemenigste	-	-	-	hoemenigste
hoeveel	-	hoevele	hoeveel	hoeveel
hoeveelde	-	-	-	hoeveelde
hoeveelste	-	-	-	hoeveelste
ieder	-	-	ieder	ieder
iedereen	-	-	-	iedereen
iederman	-	-	-	iederman
iegelijk	iegilīk	ieghelijc	iegelijc	iegelijk
iegewelk	-	ieghewelc	iegewelc	-
ielijk	-	-	ielijc illijc	-
iemand	ieman	ieman, iemant	ieman	iemand
iet	iewiht	iewet	iet	iet
iets	-	-	-	iets ietwes
ietwat	-	-	-	ietwat
iewelk	-	-	iewelc	-
letje	-	-	-	letje
loene	-	-	-	loene
lou loene	-	-	-	louw loene
luttel	= BNW & BW	luttel(III)	luttel & als	luttel

			BNW & BW	lattel
lutter	-	-	-	lutter
malk	-	-	malc mallic (VWL) mallinc(II) (VWL)	-
mannelijk	-	manlijc	manlijc manlike(II) (VWL)	-
meen	-	-	mene(IV)	-
meer	mēro(III)	mee(I) mere(IV)	meer(IX) mere(IX) (VWL)? meere(IV)(VWL) mee(II) (VWL) me(I) (VWL)	meer(VI) meer(IX)
meerder	-	= BNW	= BNW	meerder
meest	= BNW	meest	meest	meest
menig	manig manigiro	menich menighere	menich manich manech (VWL)	menig
menigeen	-	menicheen	menicheen (VWL)	menigeen
menigerhande	-	menigherhande	menigerhande	menigerhande
menigerkunne	-	-	menigerconne	-
menigerlei	-	menigheliede	menigerleye	menigerlei
menigerslachte	manigerslahta	menigherslachte(n)	- (1 vb bij slacht)	-
menigertieren	-	menighertiere(n)	menigertiere, menigertieren	menigertier
meniglieden	-	menigheliede	-	-
min	= BW	min(I)	min	min(V)??
minder	= BW	minre, (II) & (I)?	= BNW?	minder
minst	-	minst	minst (VWL) minnest (VWL)	minst
meniglijk	-	-	minclijc	-
misselijkerhande	-	-	misselikerhande	-
neen	nēn	-	neen	-
negeen	nehēn	negheen	negeen engeen (VWL) egeen (VWL)	negeen egeen = LID (bet.B)
negenerhande	-	neghenerhande	negeenrehande	-
negenerkunne	-	-	negeenreconne	-

			geenreconne geinreconne(VWL)	
negenerslachte	-	negheenreslachte	negeenreslachte, negeenreslachten	-
negenertieren	-	negheenretier	negeenretiere geretiere (VWL)	-
negererhande / negenertieren / negererwijs	-	-	negererhande, negeretiere, negererwijs(VWL)	-
negiemand	-	-	negiemant	-
nenerlei	-	-	nenerleye	-
neweder	newethar	neweder	neweder	-
niemand N.B. bij het MNW is er een verwijslemma van zeven woorden dat begint met 'niemand')	nieman	nieman, nietmant niemant (VWL) niemene (VWL) niemens (VWL)	nieman nietmant niemand, niemans, niemant, niemen, niemend, niemende, niemene (VWL)	niemand
niemand~el	-	nieman,niemant el	-	-
niemendal	-	-	-	niemendal
nienegeen	-	nienegeen	nienegeen	-
niet / niets	= BW	niewet(I)	niewet niet(I)	niet
niet / niets	niewiht(I)	niets	niets	niets nietwes
nieweder	-	-	nieweder	-
niewent	-	niewent & BW	niewent	-
niks	-	-	-	niks
noeg	-	-	noech	-
nochweder	-	nochweder	=VOEGWOORD	-
noppes	-	-	-	noppes
noweder	-	=VOEGWOORD	noweder & VW	-
nuist	-	-	nuyst	-
onmenig	unmanig	onmenich	onmenich	-
overgenoeg	-	-	-	overgenoeg

plenty	-	-	-	plentie
som	-	som	som somme	som zom
somlieden	-	-	somliede	-
sommelijk	sumalīk	somelijc	somelijc	-
sommenig	-	-	-	sommenk
sommig	-	somich	somich	sommig
sommigeen	-	somicheen	= als verbinding onder sommig	-
talrijk	-	-	-	talrijk
tel	-	-	tel	- NIET ALS VNW
uist	-	-	vist	-
uwerlijk	-	-	uwerlijc	-
veel	filo(II)	vele	vele(I) voel (VWL) vole(II) (VWL) volle(VII) (VWL)	veel(IV) veul (VWL)
velerhande	-	-	velerhande	velerhande veelderhande
velerhande / velerlei	-	-	veelrehande, veelreleye (VWL)	
velerlei	-	-	-	velerleie veelderleie
velertieren	-	-	velertiere	-
verscheiden		verscheiden(II)	verscheiden(II)	verscheiden(III)
verschillend	-	-	-	verschillend(bet.4)
vo	*fō	-	-	-
vrijwat	-	-	-	vrijwat
wat	-	wat wes	wat wes	wat(I) wa(III) (VWL) watte(I) (VWL)
watterhande	-	-	watterhande	-
watterlei	-	-	watterleye	-
weder	wethar(II)	weder(III)	weder(IV)	NIET ONBEP.
wee	-	-	wee	-

weinig	= BNW	= BW	weinich	weinig
wel	-	-	wele(I)	wel(IV) wiele(II) (VWL)
welk	-	welc	welc wilc walc(II) willeg	welk wilk(I) (VWL) wek(II) (VWL) weuk (VWL) wuk (VWL)
welkdanig	-	-	-	welkdanig
welkerhande	-	= VR. VNW	welkerhande	-
welkertieren	-	welkertiere(n)	= VR. VNW	-
wichtgelijk	wiht-gilik	-	-	-
wie	-	-	wie(I)	wie(I)
wiemenigste	-	wiemenichste	-	-
wie~zo	-	wie so	-	-
zat	= BNW	= BNW & BW	= BNW & BW	zat(I)
zatter	-	-	-	zatter
zeker		seker(II)	seker(I)	zeker(I)
zoveel	-	sovele(I)	-	zooveel
zoveelste	-	-	-	zooveelste
zowat	-	so wat	-	zoowat
zowelk	-	so welc	sowelc	-
zowie	-	so wie	sowie	-
zozulk	-	sosulc	= AANW. VNW	-
zo~hoemenig	-	so hoemenich	-	-
zo~hoeveel	-	so hoevele	-	-
zo welkertieren~dat	-	so welkertiere dat	-	-
zulk	-	sulc	sulc selc (VWL) swelc (VWL)	zulk

Uitroepende voornaamwoorden (exclamatieven)

Een laatste categorie betreft de uitroepende voornaamwoorden. De ANS onderscheidt *wat* en (in mindere mate) *welk*, *zo'n* en *zulk*.⁶⁶ Van alle woordenboeken wordt er alleen in het WNT aan gerefereerd, hoewel zeer sporadisch, bij *wat*(I) en *welk*(II) en in minder duidelijke bewoordingen bij *zulk*. Er moet nog gekeken worden wat we hier uiteindelijk mee gaan doen. *Zo'n* komt niet als lemma voor, alleen in het WNT bij *zoo* in de combinatie *zoo + een*. Het wordt daar aanwijzend genoemd, maar het kan dus ook een uitroepend voornaamwoord zijn.

Lemma	ONW	VMNW	MNW	WNT
wat	-	-	-	wat(I), zie bij B. wa(III) (VWL) watte(I) (VWL)
welk	-	-	-	welk(II), zie II. wilk(I) (VWL) wek(II) (VWL) weuk (VWL) wuk (VWL)
zo	-	-	-	zoo(II), zie III.
zulk	-	-	-	zulk, zie I.

Mogelijk is deze categorie uit te breiden met historische lemmata in deze betekenissen.

⁶⁶ Zie de [Exclamatieven](#).