

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

/instituut voor de Nederlandse taal/

INT Working Papers 1

Versie 2.0

November 2024

Woord vooraf	5
1 Inleiding	6
2 Algemene uitgangspunten	7
2.1 Bestaande tagsets	7
2.2 Mogelijkheid tot mapping en flexibele toepassing	9
2.3 Diachrone toepasbaarheid	9
2.4 Wenselijkheid van een kerntagset	9
3 Taggingprincipes	9
3.1 Lexicaal versus functioneel	10
3.1.1 Transcategorisatie	11
3.2 Token versus woord en overeenkomstige verrijking	12
3.3 Tokenannotaties versus woordannotaties	13
3.4 Vreemde woorden in Nederlandse context	13
3.5 Onderspecificatie	14
4 Beschrijving van de tagset	15
4.1 Hoofdwoordsoort, kenmerken en waardes	15
4.2 Lemmakenmerk versus woordvormkenmerk	16
4.3 Transcategorisaties	16
5 Kenmerken van de woordsoorten	16
5.1 Alle woordsoorten	16
5.2 Naamwoorden (Nouns - tags: NOU-C, NOU-P)	17
5.2.1 Zelfstandig naamwoord (Noun - tag: NOU-C)	17
5.2.2 Eigennaam (Proper Noun - tag: NOU-P)	20
5.3 Bijvoeglijk naamwoord-Bijwoord (Adjective-Adverb - tag: AA)	23
5.4 Bijwoord (Adverb - tag: ADV)	27
5.5 Werkwoord (Verb - tag: VRB)	30
5.6 Telwoord (Numeral - tag: NUM)	34
5.7 Voornaamwoord (Pronoun-Determiner - tag: PD)	38
5.8 Bijzetsel (Adposition - tag: ADP)	46
5.9 Voegwoord (Conjunction - tag: CONJ)	48
5.10 Tussenwerpsel (Interjection - tag: INT)	50
5.11 Overblijfsel (Residual - tag: RES)	52
6 Kerntagset	54
6.1 Alle woordsoorten	54
6.2 Naamwoorden (Nouns - tags: NOU-C, NOU-P)	54
6.2.1 Zelfstandig naamwoord (Noun - tag: NOU-C)	54
6.2.2 Eigennaam (Proper Noun - tag: NOU-P)	55
6.3 Bijvoeglijk naamwoord-Bijwoord (Adjective-Adverb - tag: AA)	55
6.4 Bijwoord (Adverb - tag: ADV)	55
6.5 Werkwoord (Verb - tag: VRB)	56
6.6 Telwoord (Numeral - tag: NUM)	56

6.7 Voornaamwoord (Pronoun-Determiner - tag: PD)	57
6.8 Bijzetsel (Adposition - tag: ADP)	58
6.9 Voegwoord (Conjunction - tag: CONJ)	58
6.10 Tussenwerpsel (Interjection - tag: INT)	58
6.11 Overblijfsel (Residual - tag: RES)	59
6.12 Uitgangen	59
7 Referenties	60
Bijlage 1: overzicht verschillen in lemmatisering	62
Bijlage 2: totaaloverzicht voornaamwoorden	63
Persoonlijke voornaamwoorden	63
Reflexieve voornaamwoorden	66
Reciproque voornaamwoorden	67
Bezittelijke voornaamwoorden	67
Vragend-betrekkelijke voornaamwoorden	69
Aanwijzende voornaamwoorden	71
Onbepaalde voornaamwoorden	75
Uitroepende voornaamwoorden (exclamatieven)	85
Bijlage 3: formele beschrijving tagset	86

Woord vooraf

Dit document is het resultaat van het werk van het team Historisch Nederlands, bestaande uit taalkundigen Roland de Bonth, Katrien Depuydt, Dirk Geirnaert en Thomas Haga, en computerlinguïst Jesse de Does van het Instituut voor de Nederlandse Taal. Het is in de context van het NWO-project CLARIAH+ uitgevoerd en ook mede gefinancierd door de Nederlandse Taalunie.

Een eerste versie van dit document (versie 0.9, juli 2020) 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.

De versie die daaruit is voortgevloeid (versie 0.9.5, oktober 2020), is voor commentaar voorgelegd 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.

We hebben alle commentaren met zorg bekeken en waar nodig de tekst aangepast en verhelderd. We danken alle eminente deskundigen voor hun feedback en hun steun voor dit initiatief. We hopen hiermee een substantiële stap gezet te hebben naar goed verrijkt diachroon Nederlands corpusmateriaal. Versie 1.0 is in mei 2021 verschenen.

De tagset - meer specifiek de kerntagset - is gebruikt in de training- en evaluatiesets¹ die geproduceerd zijn waarmee de taggers en lemmatiseerders in het GaLaHaD-platform getraind en geëvalueerd zijn.² Naar aanleiding daarvan is de beschrijving op een aantal plaatsen aangescherpt en zijn meer voorbeelden toegevoegd. Daarnaast is het voornaamwoordenoverzicht bijgewerkt.

Katrien Depuydt
Leiden, november 2024

¹ Zie <https://github.com/INL/galahad-corpus-data>.

² Zie <https://portal.clarin.ivdnt.org/galahad/>.

1 Inleiding

Teksten worden door taalkundige verrijking met woordsoort en lemma beter doorzoekbaar en geschikt 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-project³ 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*.⁴ In het kader van een evaluatie van de taalkundige verrijking is een reductie van de D-Coi-tagset gedefinieerd.⁵ 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 van de hoofdwoordsoort. 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) 1500-1976.⁶ 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

³ Zie <https://www.nederlab.nl/onderzoeksportaal>.

⁴ *Frog* is een verzameling van geheugengebaseerde modules van *natural language processing* (NLP) voor het Nederlands. Zie <https://languagemachines.github.io/frog/>.

⁵ 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).

⁶ Een langetermijndoel is ook om de nieuwe tagset beschikbaar te stellen voor de systematische infrastructuur voor historisch Nederlands (zie hiervoor ook Van der Sijs 2019).

van het INT en de computationele lexica van het INT (GiGaNT-Hilex en GiGaNT-Molex).⁷ 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 en flexibele toepassing
- 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)⁸ 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),⁹ 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(symbool). De tagsets hebben voor tagging respectievelijk 320 en 322 mogelijkheden. CGN en D-Coi zijn lexicaal getagd. Dit wil zeggen dat voor de hoofdwoordsoort een lexicale tagging is gekozen. Functionele informatie is opgenomen in woordsoortkenmerken.¹⁰ 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.¹¹ 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;¹²

⁷ Zie Depuydt, Haga & Mooijaart 2024.

⁸ Zie Van den Eynde 2004.

⁹ Zie Van den Eynde 2005.

¹⁰ Bijvoorbeeld het kenmerk POSITIE met de waardes *prenominaal*, *postnominaal*, *nominaal* en *vrij*.

¹¹ De tagset is te vinden op: <https://portal.clarin.nl/node/14316>.

¹² In de nieuwe release van het Corpus Gysseling in 2021 zijn de oorspronkelijke tags van het corpus gemapt naar de TDN (<https://taalmaterialen.ivdnt.org/download/corpus-gysseling-webapp/>).

- Bij het pilotproject voor negatie bij de brieven van P.C. Hooft uit de periode 1600-1638¹³ zijn de teksten eerst automatisch getagd met Adelheid.¹⁴ Daarna werd de taalkundige verrijking met woordsoort en lemma gecorrigeerd met behulp van de annotatietool GUSTAVE (**G**ea**u**tomatiseerde **S**ociolinguïstische en **T**aalkundige **A**nnotatie van **V**roegmoderne **E**ditie).¹⁵ 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;¹⁶
- 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;¹⁷
- 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.¹⁸ 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),¹⁹ het Brieven als Buit-corpus²⁰ en de Gentse Spelen.²¹

¹³ Zie voor het project <https://languagedynamics.wp.hum.uu.nl/projects/>.

¹⁴ Adelheid is een tagger en lemmatiseerder, door Hans van Halteren ontwikkeld op basis van het CRM. Zie <https://portal.clarin.nl/node/14316>.

¹⁵ Zie <https://languagedynamics.wp.hum.uu.nl/wp-content/uploads/sites/251/2015/11/Handleiding-Gustave-juli-2017.pdf>.

¹⁶ Er is verder geen officiële naam bekend voor deze tagset.

¹⁷ In de release van het Corpus Oudnederlands in 2022 zijn de oorspronkelijke tags van het corpus gemapt naar de TDN (<https://taalmaterialen.ivdnt.org/download/corpus-oudnederlands-online/>).

¹⁸ Zie Ruitenberg & Van pellicom 2012 en zie voor de totstandkoming hiervan ook Ruitenberg, De Does en Depuydt 2010.

¹⁹ Het publieke deel van het corpus is voor bij de CLARIN-infrastructuur aangesloten instituten toegankelijk via <https://chn.ivdnt.org/>.

²⁰ Dit corpus van 17e-eeuwse en 18e-eeuwse brieven van zeelieden en hun familie is online toegankelijk (<https://taalmaterialen.ivdnt.org/download/brieven-als-buit-2021/>). Let wel, het gaat hier om de tagging van het originele corpus Brieven als Buit (BAB1); de aanvulling Brieven als Buit-2 (BAB2) is (nog) niet taalkundig verrijkt (<https://taalmaterialen.ivdnt.org/download/brieven-als-buit2/>).

²¹ 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.

2.2 Mogelijkheid tot mapping en flexibele toepassing

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 vertaalbaar is naar tagging volgens TDN. Daar waar tagsets verschillen in benadering (bijvoorbeeld 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.²² De tagset geeft een vocabulaire om taalkundige verschijnselen te annoteren. De fijnmazigheid van annoteren wordt in principe per project bepaald. Door hetzelfde vocabulaire en dezelfde toekenningssystemen te gebruiken beogen we een grotere uniformiteit binnen het taalkundig verrijkt diachroon materiaal.

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.²³

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.

²² Zie verder sectie 3.1.

²³ Zie verder sectie 3.5.

3.1 Lexicaal versus functioneel

Een woordsoort kan worden toegekend op basis van het gebruikte woord of de functie in de zin. Dit kan het beste duidelijk gemaakt worden aan de hand van een voorbeeld: in *de opgegeten koek* kun je zeggen dat *opgegeten* de woordsoort werkwoord heeft, want het is het voltooid deelwoord van het werkwoord *opeten*, zoals ook te vinden is in woordenboeken. Je kunt ook zeggen dat *opgegeten* een bijvoeglijk naamwoord is, omdat *opgegeten* in prenominale positie is aangetroffen.

Als je *opgegeten* in het bovengenoemde voorbeeld de hoofdwoordsoort werkwoord geeft, dan tag je *lexicaal*. Bij *lexicaal taggen* is het streven om aan wat opgevat wordt als hetzelfde woord, in verschillende contexten dezelfde woordsoort toe te kennen. Hierbij wordt vaak een woordenboek of lexicon als referentie gebruikt. Als je *opgegeten* daarentegen de hoofdwoordsoort bijvoeglijk naamwoord geeft, dan tag je *functioneel*. Bij *functioneel taggen* wordt de woordsoort zoveel mogelijk op grond van de functie in de zin toegekend.²⁴

Het grote voordeel van functioneel taggen is dat aan een woordvorm een woordsoort op grond van een concrete context kan worden toegekend, zonder dat een uitspraak gedaan wordt over de mate waarin het gebruik van die woordvorm met die woordsoort is “gelexicaliseerd”. Met lexicalisatie van een woord bedoelen we dat het als zelfstandige eenheid opgenomen is in het veronderstelde/zogeheten mentale lexicon van een taal, waarbij de betekenis niet de facto is af te leiden uit de morfologische structuur.²⁵ Dat betekent dat we geen onderscheid hoeven te maken tussen onderstaande twee situaties:

- (a) *In het begin van het jaar is een bomma gestorven na **het eten van** een flippo* (OpenSoNaR, tag: WW(inf,nom), TDN hoofdwoordsoort NOU-C)
- (b) *neven dat kalf door was een deur en dan moest ge met **het eten van** [het voer van] de koeien en af en toe ja deed dat kalf z'n behoefte in uw klompen* (CGN, tag: N(soort,ev,basis,onz,stan), TDN hoofdwoordsoort NOU-C)

In beide gevallen kent men dan als hoofdwoordsoort zelfstandig naamwoord toe.²⁶ We verliezen dan echter wel het gegeven dat de substantiveringen *werken* en *reinens* in

- (c) *Net als mijn Michelin-ster is deze titel opnieuw een beloning voor **het vele harde werken*** (Corpus Hedendaags Nederlands)

²⁴ Een derde optie is een dubbele tag wanneer een lexicale benadering afwijkt van een functionele. In enkele gevallen, zoals in *het gekregene*, 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*. Daarom wordt voor TDN niet geopteerd voor een (drie)dubbele tagging, maar voor slechts één tag.

²⁵ Vergelijk Brinton & Traugott 2005: 144: “We define lexicalization as: the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the constituents of the construction or the word formation pattern. Over time there may be further loss of internal constituency and the item may become more lexical.”

²⁶ Zie de [Nominalisaties behorend tot type 2](#) in de ANS en vergelijk ook [Nominalisaties behorend tot type 3](#) zonder determiner.

- (d) *plinius seit men siese niet
sonder alst seere reinens pliet*
(Corpus Gysseling, Der Naturen Bloeme, Salamander)

als ‘transcategorisaties’ (zie sectie 3.1.1. hieronder) van het werkwoord gezien kunnen worden, waardoor het behouden van verbale complementen en bepalingen (in dit geval de bijwoordelijke bepalingen *veel* bij *werken* en *seere* bij *reinen*) lastig te duiden is.²⁷ Het taggen als nomen verklaart dan wel weer het voorkomen van het lidwoord *het* en de genitief in *reinens*.

Het is ook mogelijk beide benaderingen te combineren, zoals in de CGN-tagset gebeurt. Daar worden gevallen als (a), (c) en (d) als WW(wvorm=inf,positie=nom) getagd, waarmee zowel aan de werkwoordelijke als aan de nominale eigenschappen recht wordt gedaan. Er moet dan wel een onderscheid gemaakt worden tussen de gelexicaliseerde instantie (b) en de niet-gelexicaliseerde instanties (a), (c) en (d).

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. Vaststellen of iets gelexicaliseerd is binnen een bepaalde taalfase behoeft een uitgebreide analyse, waarbij zowel betekenis als frequentie een rol spelen. Bovendien vindt lexicalisatie niet meteen plaats, maar is het het resultaat van een proces. 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, omdat er onvoldoende taalmateriaal overgeleverd is om dit te kunnen 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 Brieven als Buit-corpus. Het is destijds ook toegepast bij het PAROLE-corpus.²⁸

3.1.1 Transcategorisatie

Bij functioneel taggen zijn er meer woorden die in het corpus in verschillende contexten een verschillende woordsoort kunnen krijgen dan bij lexicaal taggen. We spreken in zulke gevallen van *transcategorisatie*. De term transcategorisatie (synoniemen: syntactische transpositie, conversie) is ontleend aan de morfologie.²⁹ Er is sprake van transcategorisatie als er bij derivatie een verandering van woordklasse optreedt. Nominalisering (a), adjectivering (b) en verbalisering (c) zijn voorbeelden van transcategorisatie:

- (a) *bakken* → *bakker* werkwoord wordt substantief
- (b) *vader* → *vaderlijk* substantief wordt adjectief
- (c) *vader* → *bevaderen* substantief wordt werkwoord

²⁷ Deze informatie kan in deze benadering worden ondergebracht bij de morfologische analyse van de lemmavorm in het lexicon. In het VMNW zijn bijvoorbeeld de tot adjectief getranscategoriseerde voltooid deelwoorden te vinden door in het veld morfologische analyse te zoeken naar “eig. part.perf”.

²⁸ Vergelijk Dutilh & Kruyt 2002.

²⁹ Zie Smedts & Van Belle 1994.

De drie voorbeelden hierboven zijn voorbeelden van expliciete transcategorisatie: de transcategorisatie gebeurt door middel van één of meer in de woordvorm gerealiseerde gebonden morfemen (respectievelijk *-er*, *-lijk* en *be-* & *-en*). Impliciete transcategorisatie gebeurt zonder in de woordvorm gerealiseerde gebonden morfemen (de zogenoemde nulconversie³⁰); bijvoorbeeld:

(d) *zuur (snoepgoed)* → (*het*) *zuur* adjectief wordt substantief

Bij de beschrijving van de TDN-tagset wordt per woordsoort aangegeven wat de mogelijke impliciete transcategorisaties zijn waarmee rekening gehouden moet worden bij het taggen.

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 één woord. Een woord kan dus bestaan 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* (= *dat* + *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, voornaamwoordelijke bijwoorden en omzetsels gekoppeld en van één woordsoort en lemma voorzien. Meerwoordsexpressies bestaande uit minstens één vreemd element dat niet los in het Nederlands voorkomt, krijgen ook één woordsoort en lemma. Hetzelfde geldt voor meerwoordige eigennamen (zie verder sectie 5.2.2).³¹

Hieronder volgen een paar voorbeelden van de genoemde gevallen:

- Clitische vormen: *tboec* (*dat* + *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*; *daar weet ik wel iets voor*; *van jongs af*; *ossen vleysch ezels of beren vleysche* (= ‘ossenvlees, ezelsvlees of berenvlees’)

³⁰ Ook wel nulafleiding of zeroderivatie genoemd.

³¹ Hiermee beogen we de doorzoekbaarheid in corpora te verhogen, bijvoorbeeld voor lexicaal onderzoek. Zie ook Hüning & Schlüker 2015.

- Meerwoordsexpressies: *à volonté; en profil; ave maria; bed and breakfast; New York; Van de Berg*

3.3 Tokenannotaties versus woordannotaties

Zoals in de vorige sectie is besproken, taggen we niet elk token afzonderlijk 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” (PunCtuation).³²
- 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 ontlending te onderscheiden zijn: betekenisontlending, vertalende ontlending 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 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.³³

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, hoewel het nog steeds een ‘vreemde’ uitgang heeft, zoals *basis, matrix, musicus, museum, nomen, alumna* en *plurale* (en bovendien kunnen deze woorden ook ‘vreemde’

³² 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.

³³ Zie Van der Sijs 1996: 10-12.

meervoudsuitgangen hebben, zoals *bases, matrices, musici, musea, nomina, alumnae en pluralia*, hoewel in veel gevallen ook Nederlandse uitgangen voorkomen, zoals *basissen, matrixen, musicussen, museums*).³⁴

Voor TDN worden dergelijke vreemde woorden en meerwoordsexpressies, zoals *praesens historicum* in dit werkwoord is een *praesens historicum*, behandeld als potentieel Nederlands en worden ze dienovereenkomstig gelemmatiseerd en van een woordsoort voorzien.³⁵

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.³⁶

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.³⁷ Typische voorbeelden van onderspecificatie zijn:

- Wanneer het onderscheid tussen meerdere mogelijkheden taalkundig niet relevant is;
- Wanneer de context geen eenduidige keuze toelaat, of onduidelijk is (*ambigüiteit*);
- Wanneer voor een project een bepaalde mate van detail niet zinvol geacht wordt.³⁸

Om bij een kenmerk aan te geven dat er meerdere kenmerkwaardes mogelijk zijn, gebruikt TDN één of meerdere sluittekens (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=2|3) waar de werkwoordsvorm an sich ambigü is, maar waarbij de context een eenduidige keuze biedt.

³⁴ Zie de [Leenwoorden](#) in de ANS.

³⁵ Let wel, in een zin als *blijven zeuren is my middle name* krijgen de Engelse woorden wel de tag RES(type=for).

³⁶ Als lemma wordt de aangetroffen woordvorm herhaald (waarbij de regels voor hoofdlettergebruik in acht worden genomen). Zie verder sectie 5.11.

³⁷ Vergelijk Leech & Wilson 1996: 16-18.

³⁸ 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 taalkundige verrijking en ook die is uitvoerig besproken en bediscussieerd door de toenmalige taalbankcommissieleden.³⁹

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 kenmerkwaardes die voor twee of meer mogelijke basiswaardes 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.

³⁹ 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 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 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 Bijlage 3.

5.1 Alle woordsoorten

Geschreven vorm (tags: abbr, trunc, mis)

Er is één kenmerk dat relevant is voor alle woordsoorten. Dit is de geschreven vorm.⁴⁰ Er wordt onderscheid gemaakt tussen afgekort (*abbreviated*), afgebroken (*truncated*) en drukfout of digitaliseringsfout (*misspelled*). Het onderscheid tussen afgekort en afgebroken is niet altijd helder. Bij TDN kiezen we ervoor om alleen deels weggelaten vormen, die in het hedendaags Nederlands met een afbreekstreepje worden geschreven, te zien als afgebroken en alle andere gevallen van weglatingen als afgekort (zie de voorbeelden hieronder). Gereduceerde vormen van het voornaamwoord (zoals *m'n* voor *mijn*, *'t* voor *het* en *'s* voor *des* en ook in clitische vormen als *gelijkt* = *gelijk* + *het*), de adpositie (zoals *t'* voor *te* in constructies als *t'eeniger tyt*) en telwoorden als *XVIe* worden niet gezien als afgekorte of afgebroken vormen. Het

⁴⁰ Een nadere specificatie bij de geschreven vorm is alleen relevant bij de telwoorden, zie sectie 5.6.

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.*; *J. (de Winter)*; *Ch. (de Vries)*; *enz.*; *binnenl.*; *Febr.*; *voortgetrockē* (nasaalstreep voor een niet geschreven <n>)
- Afgebroken: *appel-* en *perenstroop*, *in-* en *intriest*; *balsport* en *-spel*
- Drukfout of digitaliseringsfout: *professpr/profeffor* (in plaats van *professor*)

Written Form (WF)		
abbreviated	truncated	misspelled
abbr	trunc	mis

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 zijn (bijvoorbeeld *-lijn* in *cnapelijn* ‘knapelijn’ naast *-kijn* in *cnapekijn* ‘knaapje’), is het niet logisch de diminutief als onderdeel van het paradigma te beschouwen.⁴¹ 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:

⁴¹ Hierbij sluiten we aan bij de ANS, die diminutiva ook als afleidingen beschouwt (Zie [De vorming van verkleinwoorden](#)) en De Haas & Trommelen 1993: 279-284. Alle diminutiva worden daarmee op dezelfde manier behandeld, en niet de ene keer als kenmerkwaarde en de andere keer als apart lemma (bijv. *ommetje* dat afgeleid is van het voorzetsel *om* vs. *tweetjes* dat volgens CGN/D-Coi aangegeven wordt als kenmerkwaarde van het telwoord).

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

Het vaststellen van de naamval gebeurt op basis van de functie van het woord in de zin. Met name waar het zelfstandig naamwoord wordt voorafgegaan door een voorzetsel is dit inderdaad geen sinecure. De vorm kan hierbij in sommige gevallen helpen. Als dat niet zo is, moet er voor onderspecificatie gekozen worden.

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* en *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
	nom dat acc voc		gen
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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- 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 *scalk* 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 de woorden in beide gevallen zelfstandige naamwoorden blijven.

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

Eigennamen worden op hoofdwoordsoortniveau onderscheiden en krijgen een eigen tag, NOU-P(roper).

Omdat we (net als bij de appellatieven) in de lemmatisering geen onderscheid willen maken tussen aaneengeschreven en los geschreven vormen van een naam (*Vandenberghes*, *Van den Berghe*) krijgen zowel voornamen als achternamen één lemma met als tag NOU-P, zelfs als deze in meerdere losse delen geschreven staan. In *Ian van den Beerghe* krijgt *Ian* dus NOU-P als tag en *van den Beerghe* eveneens. Datzelfde geldt voor patroniemen. In een voorbeeld als *Jan Janszoon van den Berg* worden *Jan*, *Janszoon* en *van den Berg* als drie afzonderlijke eigennamen getagd en gelemmatiseerd. Ook dubbele voornamen⁴² en dubbele achternamen krijgen één lemma met als tag NOU-P. In een voorbeeld als *Jan Peter Bakker Korff* is *Jan Peter* een NOU-P en *Bakker Korff* een NOU-P.

Gepersonificeerde voorstellingen van begrippen, voorwerpen en dergelijke die bijvoorbeeld als personage in blijspelen voorkomen, beschouwen we ook als eigennamen.⁴³ Plaatsnamen die uit meerdere woorden bestaan, zoals *New York*, krijgen als geheel één lemma (*New York*) en de woordsoort NOU-P. Namen van organisaties en dergelijke bestaande uit meerdere woorden krijgen niet als geheel een tag als NOU-P.⁴⁴ De delen krijgen hun eigen woordsoort. *Kamer van Amsterdam* wordt gelemmatiseerd als *kamer*, NOU-C, *van*, ADP en *Amsterdam*, NOU-P. Indien het om vreemde elementen gaat krijgen de delen RES(type=for).

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, zoals *appel manghelare* in het voorbeeld *wi doen cont allen den ghenen die nu siin ende wesen sullen dat diederic de appel+manghelare quam vor ons*. Omdat het niet duidelijk of we hier al dan niet met een beroepsnaam *appelmangelaar* te maken hebben, worden dergelijke ambigue gevallen getagd als soortnaam (NOU-C).⁴⁵ Bij toenames van adellijke personen, zoals *Karel de Kale* en *Maria van Bourgondië*, lemmatiseren we *de* en *van* los van *kaal* en *Bourgondië*.

⁴² Dit zijn voornamen die opgebouwd zijn uit twee voornamen, zoals *Anna Sophie*, *Jan Peter* en *Gijs Jan*.

⁴³ Zo komen in *Het boeck vanden pelgherym Beroemen*, *Gratie*, *Obstinate* en *Reden* voor als eigennamen.

⁴⁴ Bij de Named Entity Recognition zou *Kamer van Amsterdam* wel als geheel herkend worden als organisatie.

⁴⁵ Het systeem van vaste familienamen is geleidelijk aan tot stand gekomen en is in de oudere taalfasen van bijvoorbeeld CG en CRM niet van toepassing. Er werd gebruikgemaakt van toenames om een specifiek persoon aan te duiden. Dat kon een patroniem zijn, maar ook een plaatsaanduiding, een beroepsnaam of een bijnaam. Zowel CG als CRM heeft gekozen voor het coderen van de losse delen, en omdat er strikt genomen nog geen sprake is van familienamen, is dit ook volgens TDN een correcte tagging. Een Named Entity-laag kan bovenop de TDN-tagginglaag aangebracht worden om bijvoorbeeld voornaam en familienaam aan elkaar te koppelen (*Jan Jansen*, NE person) of om bijvoorbeeld datums te noteren (1 juli, NE other). In het Brieven als Buit-corpus is een classificatie toegekend aan de eigennamen, maar zijn persoonsnamen niet als geheel van een NE-classificatie voorzien.

Voor eigennamen 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*).

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 de bezits-*s* (*Katriens fiets*), de genitief-apostrof (*Thomas' fiets*) en een combinatie van beide (*Frieda's Mercedes*). Daarnaast komen we datievormen 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.

Case Reduced			
CGN/D-Coi	nominative dative accusative vocative		genitive
	nom dat acc voc		gen
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.⁴⁶ 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.

⁴⁶ 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.

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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- 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.

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 de woorden in beide gevallen zelfstandige naamwoorden blijven.

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

Vrijwel alle adjectieven kunnen ook als bijwoord gebruikt worden (zoals *hard* in *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.⁴⁷ Woorden die van oorsprong alleen adjectieven zijn (die in de historische woordenboeken ook vaak (alleen) bijwoorden worden genoemd), worden in TDN altijd beschouwd als AA's. Indien woorden historisch gezien zowel bijwoorden als AA's zijn, herkenbaar aan een suffix (zoals Middelnederlands *-like* naast *-lijk*), dan worden ze onderscheiden zolang de suffixen nog herkenbaar zijn. Wanneer deze suffixen later samenvallen, worden ze enkel nog AA genoemd.

Woorden die gevormd zijn door de verkleinvorm van bijvoeglijke naamwoorden-bijwoorden met een adverbiale -s worden hier gezien als AA, omdat ze zowel predicatief als attributief kunnen voorkomen en als bepaling van gesteldheid, zoals in *zij is netjes; het is frisjes; een netjes en schoon huis; ik vind het maar povertjes*.⁴⁸

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.⁴⁹

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 bijwoorden en adjectieven. In het Corpus Gysseling en CRM worden adjectieven en bijwoorden op systematische wijze onderscheiden en wordt positie bij adjectieven niet verder gespecificeerd.

⁴⁷ Zie voor de echte bijwoorden de [Bijwoorden en adjectieven](#) in de ANS.

⁴⁸ Zie [Verkleinwoorden als bijwoorden](#) in de ANS.

⁴⁹ In het geval van onregelmatige comparatieven en superlatieven, die niet met regels voorspelbaar zijn uit de positief, wordt er een apart lemma toegekend, 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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- 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!; verduiveld!* Transcategorisaties naar interjecties zijn lastig: zie voor redenen om deze woorden als tussenwerpsels te beschouwen sectie 5.10.

- AA > NOU-C: *mijn lief; het rood; het kwaad; het kwade; het Gronings; de beste; ik zie iets moois en veel gek, terwijl hij weinig gekkers ziet;*⁵⁰ *daer een coninc aen ons quam, een fel.*⁵¹ 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. Dit geldt ook voor foutieve samentrekkingen, waar het weggelaten element (in tegenstelling tot het eerder genoemde element) niet overeenkomt in getal, zoals in *jonger dan eenige andere dezer talen, was de Nederduitse*.

⁵⁰ Adjectieven na onbepaalde pronomina, zoals in *iets groots*, worden gesubstantiveerd (Van de Velde 2009: 106). Het lemma wordt in deze voorbeelden respectievelijk *mooie*, *gekke* en *gekkere*. In enkele gevallen worden de vormen op -s later ook in andere constructies gebruikt. In dat geval krijgen ze een ander lemma. In *wat nieuws* krijgt *nieuws* het lemma *nieuwe*, maar in *ik heb het nieuws gezien* krijgt *nieuws* het lemma *nieuws*.

⁵¹ In het Middelnederlands kon zonder lidwoord en na *een* en *geen* (later ook na *die/dat*) een gesubstantiveerd 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)

Dit zijn de zogenaamde ‘echte bijwoorden’ waarvan het bijwoordelijk gebruik domineert en helder af te bakenen is, en waarbij slechts incidenteel een transcategorisatie naar een andere woordsoort plaatsvindt.⁵² Bijwoordelijk gebruikte adjectieven krijgen de tag AA met de kenmerkwaarde *vrij* dan wel de gedetailleerdere kenmerkwaarde *adv* voor het kenmerk positie (zie de vorige sectie). We geven geen kenmerk graad voor die enkele gevallen waarin sprake is van een comparatief of superlatief van echte bijwoorden, zoals *gaerner* en *garenst* bij *gaarne*.

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.⁵³ CRM onderscheidt bijwoorden en voornaamwoordelijke bijwoorden 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 *aldus*, *echter* en *toch*, en een pronominaal of voornaamwoordelijk bijwoord (*pronominal adverb*), zoals *erover*, *hieraan*, *vandaar* en *waarheen*.⁵⁴ 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 van de trap af*), 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

⁵² Onder echte bijwoorden vallen ook de bijwoorden op *-e*, *-jes*, *-lijk*, *-(e)ling(s)*, *-erwijs*, *-gewijs*, *-halve*, *-iter*, *-s*, *-waart(s)* en *-weg*.

⁵³ Het Corpus Oudnederlands onderscheidt net als de ANS ([Indeling 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*.

⁵⁴ In tegenstelling tot de ANS (Zie de [Voornaamwoordelijke bijwoorden](#)) definiëren we een voornaamwoordelijk bijwoord in de traditionele zin als een combinatie van voornaamwoord en adpositie.

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:

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.*)
- 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	
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:

- 0, 'No inflection': *No inflection*
- a, '-a': *Inflection -a*
- e, '-e': *Inflection -e*
- n, '-n': *Inflection -n*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- 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 en waarom; het hier en nu.*

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

5.5 Werkwoord (Verb - tag: VRB)

Voor de beschrijving van de werkwoorden volgen hier de mogelijke kenmerken.⁵⁵ Hierbij moet worden opgemerkt dat *valentie*, *conjugatie* en *werkwoordsklasse* kenmerken zijn die in het Corpus Oudnederlands geannoteerd zijn en opgenomen, zodat bij conversie van dat Corpus Oudnederlands naar TDN geen informatie verloren raakt.⁵⁶ We voorzien niet dat deze kenmerken ook in andere corpora zullen worden geannoteerd.

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

Er wordt onderscheid gemaakt tussen de finiete vorm (*finite*), de infinitief (*infinitive*),⁵⁷ het tegenwoordig/onvoltooid deelwoord (*present participle*) en het voltooid deelwoord (*past participle*).⁵⁸

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 (*imperative*). 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.

⁵⁵ We hebben, in tegenstelling tot tagsets als CGN/D-Coi, geen kenmerk ‘positie’ bij (niet-finiëte) werkwoorden.

⁵⁶ Deze kenmerken zijn - waar aanwezig - niet op lemmaniveau toegekend, maar aan woordvormen in context.

⁵⁷ Een infinitief van TDN kan ook een oud gerundium zijn, namelijk na *te*. Zie verder de transcategorisaties.

⁵⁸ In de oorspronkelijke codering van Gysseling en CRM wordt het onderscheid tussen beide participia niet gemaakt; we hebben dit in de online versie van het Corpus Gysseling aangebracht op grond van formele kenmerken.

Mood Reduced	
indicative conjunctive	imperative
ind conj	imp

Getal (tags: sg, pl)

Er wordt bij finiete vormen van het werkwoord onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*) vanwege congruentie met het onderwerp. Dit is bijvoorbeeld gedaan voor het Corpus Oudnederlands. Voor oudere taalfasen is dit lastiger, omdat de tweede persoon meervoud bijvoorbeeld 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*),⁵⁹ 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*).

⁵⁹ Hiertoe behoren ook de ditransitieve werkwoorden, indien deze categorie apart wordt onderscheiden.

Type		
main	copular	auxiliary
mai	cop	aux

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. De waarde *onregelmatig* wordt toegekend aan sterke of zwakke werkwoorden die niet het regelmatige patroon van sterke respectievelijk zwakke werkwoorden vertonen.⁶⁰

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 zeven werkwoordsklassen van sterke werkwoorden, waarbij de 3^e klasse wordt onderverdeeld in twee subtypes.⁶¹ 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:

⁶⁰ Zie ook De Smet & Van de Velde 2019.

⁶¹ Zie Quak & Van der Horst 2002: 46-47.

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

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

Transcategorisaties

- VRB(part) > AA: hier volgen een aantal voorbeelden met de reden voor transcategorisatie erachter: *gedeeld* *genoegen*; de *gekaapte* *buit* (attributief); *haar toon wordt vriendelijker* (verbuiging); *dat gedrag is niet gepast*; de *buis is heel erg gebogen* (bijwoordelijke voorbepaling); *de dief wil ik dood of levend hebben* (nevenschikking bij AA's); *een brandend hete dag* (bijwoordelijk gebruik); *hij bleef als vernietigd staan* (als-bepaling).⁶² Deze AA's kunnen verder transcategoriseren naar nomina: *iets uitdagends*; *het gebrokene*.
- VRB(part) > ADP: *gezien zijn gezondheid*; *hangende het onderzoek*.
- VRB(part, past) > CONJ (sub): *aangezien hij had betaald*.
- VRB(imperative) > INT: *help!*; *Ik zal u, verdoemd! wel leren armen en beenen roeren*", *zwoer hij, en zelfs nog wat grover*, Transcategorisaties naar interjecties 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 het Middelnederlands komt dit al veelvuldig voor, zoals in: *want du dan leves in enen sterven*; 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 (*om te verstanen elcs redene moede*), maar ook in moderne uitdrukkingen is dit nog zichtbaar (*een uur gaans*; *tot ziens*).⁶³ Gerundia worden getagd als zelfstandig naamwoord.⁶⁴ Hiermee wijken we af van CGN/D-Coi, omdat de slot-s daar gezien wordt als derivationeel suffix waarmee bijwoorden van infinitieven worden afgeleid. We maken overigens een uitzondering voor het werkwoord na het voorzetsel *te* en na het voorzetsel *om* als *te* verzwegen is.⁶⁵ In Middelnederlands *te gevene* en modern Nederlands *te geven* blijft *geven* een werkwoord.

⁶² Wanneer het participium niet attributief gebruikt wordt, kan het onderscheid tussen werkwoord en adjectief moeilijk te maken zijn. Bij twijfel worden deze participia als werkwoord getagd.

⁶³ Zie Duinhoven 1997: 192-193.

⁶⁴ 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 werkwoordskensmerk dat toegepast wordt op die gevallen die geen voor- of nabepaling kunnen krijgen.

⁶⁵ Zie Duinhoven 1997: 195-199 en referenties aldaar.

5.6 Telwoord (Numeral - tag: NUM)

De telwoorden worden onderscheiden in twee types: hoofdtelwoorden en rangtelwoorden. Breuken als *vijf achtste* worden net als bij CGN/D-Coi behandeld als een combinatie van een hoofdtelwoord en een rangtelwoord.⁶⁶ Woorden als *miljoen*, *miljard* en *biljoen* zijn voor TDN zelfstandig naamwoorden, hoewel we ons bewust zijn van de overgang die deze woorden aan het maken lijken te zijn naar telwoord.⁶⁷

Wij rekenen alleen de 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.⁶⁸ Hier wijken we dus af van het Corpus Gysseling, CRM en het Corpus Oudnederlands, die wel een categorie 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)*,⁶⁹ die wij tot de onbepaalde voornaamwoorden rekenen, omdat er geen telbare eenheid is aangegeven. In tegenstelling tot CGN/D-Coi rekenen we ook *een* en *ander* tot de telwoorden in constructies als: *het een & (het) ander*, *de een & de ander(e)*, *om de andere dag* en *de eerste*, *de andere*, *de derde*. *Ander* is hierbij altijd een rangtelwoord.

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 (het blijven telwoorden).

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

⁶⁶ In vrije positie transcategoriseert het rangtelwoord naar zelfstandig naamwoord.

⁶⁷ Vergelijk Van der Horst & Van de Velde 2016.

⁶⁸ Zo rekent de ANS bijvoorbeeld *veel* (ook *zoveel(ste)* en *hoeveel(ste)*), *weinig* (ook *minder* en *minste*), *enkele*, *ettelijke*, *meerdere*, *verscheidene*, *verschillende*, *talloze*, *middelste*, *laatste*, *voorlaatste* en informeel *tig(ste)* tot de onbepaalde telwoorden (zie [Het telwoord](#)), 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.

⁶⁹ 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.

Positie (tags: prenom, postnom, free)

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

- Prenominaal: *twee kazen*; *deze vier broden*; *de derde man*; *ik heb toch zeven achtste van de pizza opgegeten*. Ook gevallen als *twee kilo suiker*, *vijf pond vlees* en *vijf juli* rekenen wij, in tegenstelling tot CGN/D-Coi, tot de prenominale telwoorden.
- Postnominaal: *de sterren vijff die eeuwelick gaen dwalen*. Ook gevallen als *de jaren twintig*, *pagina vijftien* en *Willem II* rekenen wij, in tegenstelling tot CGN/D-Coi, tot de postnominale telwoorden.
- Vrij: *vijf dusent volcx haddi*; *is het er eentje of zijn het er twee?*; *wij vijven gaan erheen*; (Belgisch-Nederlands) *wij gedrieën gaan erheen*; *we voltooien de opdracht met z'n vieren*; *hij is eerste geworden*; *zij zaten ge vijven 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-roman*).

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

Getal (tags: sg, pl)

Er wordt onderscheid gemaakt tussen enkelvoud (*singular*) en meervoud (*plural*) vanwege 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 prenominale 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 flexietypes 0 (*geen flexie*), met -n, met -r/-re en met -s/-th zijn ook toe te passen bij CGN/D-Coi.

Transcategorisaties

- 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; ik heb toch zeven achtste van de pizza opgegeten*. 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. Dit geldt ook voor foutieve samentrekkingen, waar het weggelaten element (in tegenstelling tot het eerder genoemde element) niet overeenkomt in getal, zoals in *Er zijn vier honden, waarvan de eerste en de tweede een staart hebben*.

5.7 Voornaamwoord (Pronoun-Determiner - tag: PD)

Voornaamwoorden en determiners worden samen behandeld omdat ze een groot aantal kenmerken gemeen hebben. In tegenstelling tot andere benaderingen (de CGN-tagset, Universal Dependencies⁷⁰) maken we in de tagging geen onderscheid tussen ‘determiners’ en ‘voornaamwoorden/pronomina’. Met de term ‘voornaamwoord’ duiden we beide woordsoorten aan, zoals dat in Nederlandstalige literatuur (zoals de ANS⁷¹) ook gebruikelijk is. De reden om het onderscheid tussen pronomina en determiner niet expliciet te maken is, naast praktische complicaties, de diachrone instabiliteit van dit onderscheid.⁷² In plaats daarvan biedt de tagset wel de mogelijkheid door middel van het kenmerk ‘positie’ aan te geven dat het voornaamwoord als deel van een naamwoordelijke constituent functioneert.

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.⁷³ 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.⁷⁴ 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-pronoun*), onbepaalde voornaamwoorden (*indefinite*), persoonlijke

⁷⁰ Zie <https://universaldependencies.org/nl/index.html>.

⁷¹ Zie het [Voornaamwoord](#) in de ANS.

⁷² Zie Van de Velde 2010. Als criterium voor determinerschap geldt onder andere dat het woord niet predicatief gebruikt kan worden. Dat zou inhouden dat we bij het taggen van een possessief als *mijn* zouden moeten inschatten of het woord in het taaleigen van het te taggen document ook in contexten als *het boek is mijn* mogelijk is.

⁷³ Zie Duinhoven 1988: 223.

⁷⁴ *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*.

voornaamwoorden (*personal*), reciproque voornaamwoorden (*reciprocal*), reflexieve voornaamwoorden (*reflexive*), uitroepende voornaamwoorden (*exclamative*) en w-pronomen (*w-pronoun*). Hieronder volgen enkele voorbeelden van deze acht types:

- Bezittelijk: *zijn vriend*; *[dan] sael ic comen vriendeken mijn*; *al ware al de weerelt 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-pronomen: *datgene*; *de*; *die*; *dat*; *deze*; *dit*; *dusdanige*; *gene*; *hetgeen*; *zodanige*; *het* (*lidwoord*). Dit zijn (aanwijzende) deictische 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* 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*.⁷⁵
- Reciproque: *Jesse en Katrien zien elkaar*.
- Reflexief: *Dirk wast zich*; *wij wassen ons(zelf)*; *hij wast hem*.⁷⁶
- Uitroepend: *welk een dwaasheid*; *zulk een lef heeft die vent, het is ongelooflijk!*; *wat al banaliteit*.
- W-pronomen: *dewelke*; *hetwelk*; *hoedanig*; *wat*; *welk*; *wie*. 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*) als betrekkelijke (*die*) voornaamwoorden en een aparte categorie voor persoonlijke reflexieven

⁷⁵ Zie de [Inleiding](#) van het persoonlijk voornaamwoord in de ANS.

⁷⁶ 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.

(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.⁷⁷ 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

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 preminale 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 preminaal (*preminial*), postminaal (*postminial*) 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 positie⁷⁸ en de vrije positie - samen te nemen tot de positie vrij. Ook wijken we af door een positie postminaal 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.⁷⁹ Hieronder volgen enkele voorbeelden van de drie posities:

⁷⁷ Overigens is in de Adelheidmapping van CRM alleen de term reflexief gebruikt.

⁷⁸ Niet alle voornaamwoorden die in CGN/D-Coi minaal 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*, *het onze* en *te mijnent*. Zie ook de transcategorisaties aan het einde van deze sectie.

⁷⁹ CGN/D-Coi noemt het postminale gebruik wel, maar zegt dat dit zo bijzonder is dat er geen voorzieningen voor worden getroffen.

- Prenominaal: *mijn vriend; welke man; die stoel; elk huis.*
- Postnominaal: *het is de oude man zelf;*⁸⁰ *kindeke mijn; [dan] sael ic comen vriendeken mijn; fiunda sina ertha leccon sulun* ‘zijn vijanden zullen de aarde likken’.
- Vrij: *ik loop daar; Roland ziet haar; Dirk wast zich; Jesse en Katrien zien elkaar; de man die daar staat; hetgeen je ziet; diens hoed heeft vier deuken; mijns gelijke; al die mensen; ze kregen er elk twee; het was zo’n vis; al ware al de weerelt dijn; gheloeft den ghenen niet* (twee keer vrij); *inden eywn der welker saet* (twee keer vrij).

Position		
prenominal	postnominal	free
prenom	postnom	free

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 en reflexieve 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 prenominal 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 zijn besien; uan themo richedume sint unse wort. zo themo ist unse mūt 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 het lemmakenmerk getal alleen een kenmerk van de persoonlijke en de reflexieve/reciproque voornaamwoorden.

Lemma Number (LN)	
singular	plural
sg	pl

⁸⁰ *Zelf* wordt hier net als bij Van de Velde (2009: 301) tot de voornaamwoorden gerekend.

Lemmakenmerk Persoon (tags: 1, 2, 3)

Er wordt bij de persoonlijke, bezittelijke en reflexieve 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.⁸¹

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.

Lemma Person (LP)		
1	2	3

Lemmakenmerk Geslacht (tags: m, f, n)

Er wordt bij de persoonlijke, bezittelijke en reflexieve 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, behalve bij 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 bij het bezittelijk voornaamwoord *ons*, dat in combinatie met het zelfstandig naamwoord *paard* een kenmerk met waarde *evon* (enkelvoud onzijdig) krijgt. Het woordvormkenmerk getal is ook toegekend aan de aanwijzende en bezittelijke voornaamwoorden in GUSTAVE.

Number	
singular	plural
sg	pl

⁸¹ 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.

Woordvormkenmerk Geslacht (tags: m, f, n)

Er wordt bij de voornaamwoorden, behalve bij 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 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 het hedendaags Nederlands zijn naamvalsvormen 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.⁸²

⁸² 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.

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	

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*
- nt, '-nt': *Inflection -nt*

- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- other, 'Other': *Other values*

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

De flexietypes met -e, met -n en met -r/-re worden in GUSTAVE gebruikt voor aanwijzende, onbepaalde en bezittelijke voornaamwoorden en het flexietype 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 erterike 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. De afbakening van bijzetsels is niet altijd evident. *Richting* in het voorbeeld *richting centrum* is voor het hedendaags Nederlands geaccepteerd als voorzetsel, terwijl dat (nog) niet het geval is voor *hartje* in het voorbeeld *hartje centrum*.⁸³ Tot de voorzetsels 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 *door*, *met*, *na* en *om* worden gezien als voorzetsels, zelfs als er woorden of zinsdelen tussen staan (*het is te koud om in de winter maar uren en uren te wandelen*). Ten slotte worden ook woorden die voorafgaan aan *dat* (vaak gevolgd door een bijzin), zoals in *tegen dat hij komt* als voorzetsel beschouwd.

De ANS behandelt twee delen die om hun complement heen staan, zoals *van iemand af* en *om de vrouw heen* als omzetsels.⁸⁴ Dit zijn in de meeste gevallen meerwoordsexpressies. Wij beperken ons tot alleen die gevallen die ook als voorzetsel voorkomen. 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*; *vanaf Leiden tot Leuven*; (Belgisch-Nederlands) *doorheen de tijd*.
- Achterzetsel: *de trap af*; *dag in, dag uit*; *oec ne mach hi niet leuen hine si den watre beneuen* ('beneven').
- Omzetsel: *als men hem van jongs af leert*; *door de tijd heen*.

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

⁸³ Zie [Moeilijkheden bij het afbakenen van de categorie voorzetsels](#) in de ANS.

⁸⁴ Zie de [Omzetsels \(circumposities\)](#) in de ANS.

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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- other, 'Other': *Other values*

Transcategorisaties

- ADP > ADV:⁸⁵ *ik zag hem van achter; boven zag ik haar niet; ik ga naar buiten.*
- ADP > CON: *je moet nablijven tot je hebt nagedacht!; voor hij vertrok, was zij al weg.*
- ADP > NOU-C: *(de) voor(s) en tegen(s); (de) pro('s) en contra('s).*

⁸⁵ 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 Gysseling 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 onderschikkend voegwoord.

Hieronder volgen enkele voorbeelden uit het Corpus Gysseling:

- (1) Nevenschikkend: *ende; iof; sonder (so wie garen coept met iemene sonder met sinen brodhate ... 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: *no (‘noch’) (Hine heeft ooc niemene so lief, no den coninc minen here, hine wilde [...]); ne/en: (a) Sine mochten in. Sine wilden har arnasch of doen. (‘Ze mochten niet naar binnen als ze hun harnas niet af wilden doen’); (b) Bachten hem ganc in allen tiden, Hine hietie gaen neuen siere siden, (‘Loop altijd achter hem, tenzij hij je uitnodigt naast hem te lopen’); (c) Mar ioas dat vule quaet. Ne dochte om die eere niet. Die hem die vader beriet. Hine steende zacharias ghinder. Ende met hem al sine kinder. (‘Maar Ioas, die vuile schurk, dacht niet aan de eer die de vader [d.w.z. Jojada, de vader van Zacharias] hem bewezen had, maar stenigde Zacharias, en al zijn kinderen met hem.’).⁸⁶*

⁸⁶ We taggen, in overeenstemming met het VMNW en het MNW, *ne* als ontkennend voegwoord in gevallen waar de betekenis van de context niet te verklaren is door *ne* simpelweg als een ontkennend bijwoord (en dus logisch gezien als een eenplaatsige operatie) te duiden. Dit is om het onderscheid te kunnen maken met het gebruik van *ne* als (deel van een) ontkennend bijwoord. Het gaat hier steeds om een combinatie van twee zinsdelen, waarbij het tweede deel formeel gekenmerkt is door een *ne* zonder *niet*, met het werkwoord op de derde plaats in de zin. Meestal is de context van de vorm $\neg A \neg B$, met als basisbetekenis de tweeplaatsige logische operatie $\neg B \Rightarrow \neg A$

- (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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- 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.*

(a), iets zeldzamer van de vorm $A \neg B$ met als basisbetekenis $\neg B \Rightarrow A$ (b) en soms ook (betekenis 3 in het woordenboekartikel) van de vorm $A \neg B$ met een contrastieve betekenis A maar B (c). Van der Horst (2008: 517) vergelijkt deze constructie met de latere balansschikking. Zie voor deze constructies ook verder Postma 2002, Breitbarth 2014 en Witzenhausen 2019.

5.10 Tussenwerpsel (*Interjection* - tag: INT)

Tot de tussenwerpsels of interjecties rekenen we de woorden die syntactisch gezien op zichzelf staan en die (vrijwel) geen relatie hebben met andere woorden. De ANS onderscheidt drie types tussenwerpsels naar hun functie: klanknabootsende tussenwerpsels (*haha, miauw, bam*), expressieve tussenwerpsels (*au, ach, tsjonge*) en communicatieve tussenwerpsels (*foei, tja, sorry*).⁸⁷ Wij volgen de ANS, met een paar uitzonderingen, waarin wij de historische woordenboeken volgen.⁸⁸

- (1) Woorden die als een elliptisch antwoord kunnen worden opgevat (zoals *zeker* en *natuurlijk*), blijven bijwoord of adjectief.
- (2) *ja* en *nee(n)* worden alleen als tussenwerpsel opgevat wanneer ze een expressieve lading hebben.

Ook woorden die tot andere woordsoorten behoren, zoals *hemel* (NOU-C), *Jezus* (NOU-P), *gaaf* (AA), *toe* (ADV) en *help* (VRB), kunnen soms als zelfstandige taaluitingen optreden en zijn dus functioneel gezien interjecties. We wijken hierin af van CGN/D-Coi, die geen transcategorisatie van andere woordsoorten naar tussenwerpsels toelaat, en van de ANS, die alleen transcategorisatie toelaat als het meestal of altijd op die manier gebruikt wordt.

Er zijn geen verdere kenmerken bij tussenwerpsels. 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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- 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.*

⁸⁷ Zie de [Het tussenwerpsel \(de interjectie\)](#) in de ANS.

⁸⁸ Zowel de historische woordenboeken van het Nederlands als bijvoorbeeld de Oxford English Dictionary.

- INT > AA: *ik vind het voorstel oké; dat is best wel een oké idee van je; hopelijk houdt de hond zich koest.*
- INT > ADP: *De grondt mijns heyls, het stutsel van mijn gang Is Godt de Heer. puff alle noodt en quaden ('ondanks').*
- INT > ADV: *hij sprong pardoes het water in.*

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 (CG), CRM, GUSTAVE en tagset Corpus Oudnederlands (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	CG	CRM	GUSTAVE	TaCO
achtergrond-geluid	(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, bre, form, meta, symb, uncl, oth)

De volgende types (grotendeels uit het bovenstaande) achten wij relevant voor de overblijfsels.⁸⁹ Er wordt onderscheid gemaakt tussen affixen (*affix*), *bounded root element* (*bre*), 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*
- Bounded root element: *in godsnaam; in voege* (uit Frans en vogue)

⁸⁹ De beschrijving van achtergrondgeluid en commentaar onderscheiden we niet met behulp van woordsoorttags. Afkortingen en afbrekingen zijn relevant voor alle woordsoorten en deeleigen is een deel van een NOU-P (zie hiervoor sectie 5.2.2)

- Formule: $E=mc^2$
- Onduidelijk: dit zijn zowel gevallen waarbij het niet uit te maken is wat de woordsoort is, ofwel omdat de context onduidelijk is ofwel omdat het woord/de woordvorm zelf onduidelijk is: *sijn si rentemeestere. baliuwe. scoutaten. G[...]/el[...]/ren. wie[datse] s[sijn]; Abraham woende dar amre. In dat dal dat hiet nambre*⁹⁰
- 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)*⁹¹), dan wel niet-Nederlandse frases binnen Nederlands zinsverband: *ik heb De ratione studii van Erasmus gelezen, nadat ik Das Leben der Anderen heb gezien*
- Zelfnoemfunctie: *lopen is een werkwoord; help is een veelgebruikte uitroep in penibele situaties; het teken ±; Aspiratie oft H en es gheen letter; Ooc tusschen, ENDE, Coniunctie ende, Hende Nomen, en maeckt hy ooc gheen diuer sitheid; sermoen is het gewone woord voor eene preek; dat het woord apoplexie eigenlijk geen pathologisch-anatomisch begrip uitdrukt*
- Overig: dit zijn alle gevallen die niet aan één van de bovengenoemde typeringenvoldoen.

Type							
affix	bre	foreign	formula	meta	symbol	unclear	other
aff	bre	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*
- nt, '-nt': *Inflection -nt*
- r/re, '-r/-re': *Inflection -r or -re*
- s/th, '-s/-th': *Inflection -s or -th*
- t, '-t': *Inflection -t*
- other, 'Other': *Other values*

Transcategorisaties

- RES(Aff) > NOU-C: *hij heeft weer een nieuw isme bedacht.*

⁹⁰ 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.

⁹¹ 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. Ook voor de kerntagset geldt dat er gebruik gemaakt kan worden van de in 3.5 beschreven mechanismen voor onderspecificatie.

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**, **trunc**, **mis**)

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	truncated	misspelled
abbr	trunc	mis

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

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-roman*).

Representation				
digits	letters	roman	mixed-digits	mixed-roman
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-pronomen (*d-pronoun*), onbepaalde voornaamwoorden (*indefinite*), persoonlijke voornaamwoorden (*personal*), reciproque voornaamwoorden (*reciprocal*), reflexieve voornaamwoorden (*reflexive*), uitroepende voornaamwoorden (*exclamative*) en w-pronomen (*w-pronoun*).

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 een 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 prenominal 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, bre, form, meta, symb, uncl, oth)

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

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

Door de verschillende types te onderscheiden kunnen tokens van allerlei 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 die niet bij de andere types kunnen worden ondergebracht.

6.12 Uitgangen

De tagset waarmee het Corpus Gysseling en CRM werd 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 getagd, 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 is het mogelijk onderzoek te doen naar bepaalde vormen door gerichte zoekopdrachten te geven.⁹²

⁹² In het online [Corpus Gysseling](#): [\[word=".*s\\$" & lemma != ".*s" & pos="NOU-C"\]](#) 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

- Algemene Spraakkunst Nederlands (e-ANS) Version 3.1 [Online Service]. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-t5>.
- Breitbarth, A. (2014). *The History of Low German Negation*. Oxford: Oxford University Press.
- Brinton, L., & Traugott, E. (2005). *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Decorte, S., m.m.v. Kruyt, T., Ruitenberg, T. en Depuydt, K. (2003). *Aanzet tot een tagset en tagtoekenningsmethode voor de Geïntegreerde taalbank*. (intern discussiestuk)
- Depuydt, K., Haga, T. & Mooijaart, M. m.m.v. De Bonth, R., Depoorter, G., Van pellicom, K. & Verheij, B. (2024). *Lemmatiseerprincipes voor GiGaNT, het centrale lexicon van het INT* (Uitgebreide en herziene versie van Mooijaart M. (2002). Het lemma in het GiGaNT-lexicon. INL Working Papers - Taalbank Nederlands 4). INT Working Papers 2 (Instituut voor de Nederlandse taal).
- 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: Synchroon en diachroon. Deel 1: De naamwoordgroep*. Leiden: Nijhoff.
- Duinhoven, A.M. (1997). *Middel nederlandse syntaxis: Synchroon en diachroon. Deel 2: De werkwoordgroep*. Groningen: Nijhoff.
- Dutilh, T. and Kruyt, T. (2002). Implementation and Evaluation of PAROLE PoS in a National Context. In: *Proceedings of the third international conference on language resources and evaluation (LREC-2002)*. Las Palmas, 1615–1621.
- 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, project internal version.
- Haas, W de & Trommelen, M. (1993). *Morfologisch Handboek van het Nederlands*. 's-Gravenhage: SDU Uitgeverij.
- Horst, J.M. van der & Van de Velde, F. (2016). Miljoen. In: *Leuvense Bijdragen: Tijdschrift voor Germaanse Filologie* 100, 410-424.
- Hüning, M., & Schlücker, B. (2015). Multi-word expressions. In Muller, P.O., Ohnheiser, I., Olsen, S., & F. Rainer (Eds.), *Word-formation: An international handbook of the languages of Europe*, 450–467. De Gruyter Mouton.

- Leech, G., & Wilson, A. (1996). Recommendations for the morphosyntactic annotation of corpora. *Relatório técnico, Expert Advisory Group on Language Engineering Standard (EAGLES)*.
- Postma, G.J. (2002). De enkelvoudige clitische negatie in het Middelnederlands en Jespersen-cyclus. In: *Nederlandse Taalkunde* 7 (1), 44-82.
- 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, & Depuydt, K. (2010). Developing GiGaNT, a lexical infrastructure covering 16 centuries. In: *Proceedings of the XIV Euralex International Congress*, Ljouwert: Afûk, 468-476.
- 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 (2019). Historische taalkunde en Digital Humanities: samen naar een mooie toekomst. *Tijdschrift voor Nederlandse Taal-en Letterkunde* 135 (4), 384-405.
- Smedts, W. & Van Belle, W. (1994), *Taalboek Nederlands*. Kapellen: Uitgeverij Pelckmans.
- De Smet, I. & Van de Velde, F. (2019). Reassessing the evolution of West Germanic preterite inflection. In: *Diachronica* 36 (2), 139-180.
- Van de Velde, F. (2009). *De nominale constituent: structuur en geschiedenis*. Leuven University Press.
- Van de Velde, F. (2010). The emergence of the determiner in the Dutch NP*. In: *Linguistics* 48 (2), 263-299.
- Witzenhausen, E. (2019). *Negation – Exception – Contrast. The post-cyclical development of ne/en in Middle High German, Middle Low German and Middle Dutch* (Doctoral dissertation, Ghent University).

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).⁹³ 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, in tegenstelling tot GUSTAVE, de positief, hoewel er uitzonderingen zijn voor bijzondere comparatief- en superlatiefvormen als *beter*, *minst* en *benedenst* (zie verder sectie 5.3);
- 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;
- *1 t/m 100*, honderdtallen (zoals *200*, *1300*) en duizendtallen (zoals *3000*, *15000*) krijgen een uitgeschreven lemma. Een uitzondering hierop vormen jaartallen, die altijd het getal als lemma krijgen. Een dergelijke systematische aanpak is bijvoorbeeld niet terug te vinden in het Corpus Gysseling;
- 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* en *te+dat*, *ter* als *te+de*;⁹⁴
- Woorden waarvan de woordsoort onduidelijk is en die de tag RES(type=uncl) krijgen, hebben in tegenstelling tot bijvoorbeeld GUSTAVE, een liggend streepje (-) als lemma;
- Woorden waarvan het lemma onduidelijk is, maar de woordsoort niet, krijgen, in tegenstelling tot bijvoorbeeld in het corpus Gysseling, UNKNOWN (met hoofdletters) als lemma.

⁹³ Zie Depuydt, Haga & Mooijaart 2024.

⁹⁴ In enkele gevallen zoals *ten achteren* en *ten onder* is *ten* gewoon *te* (de <n> is alleen gebruikt als hiaatdelging).

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 kleine kapitalen (bijv. IKZELF bij persoonlijke voornaamwoorden) staan de voornaamwoorden die in de historische woordenboeken niet van een woordsoort zijn voorzien, met daarachter het lemma waar ze onder opgenomen zijn. Daarnaast zijn voor de volledigheid ook de verwijsslemmata (afgekort VWL) opgenomen, omdat ze vaak verwante spellingen weergeven.⁹⁵ Omdat deze verwijsslemmata naar meerdere lemmata kunnen verwijzen, hebben ze soms meerdere moderne lemmata.

Persoonlijke voornaamwoorden

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

Lemma	ONW	VMNW	MNW	WNT
du	thū	du	du dich	du(I) dij(III) (VWL)
er	-	-	-	er(I) der(I) (VWL)
gij	gi	ghi	gi ju jou ure(III) us(II) ou(IV) uwer, uwes jo(II) u(III) uus(III)	gij u(II) uw(II) (VWL) uwe(II) (VWL) uwes(I) (VWL) je(II) (VWL)
gij / u / uw	-	-		ou(III) (VWL)
gij / uw	-	-	hu(II)	au(III), auw (VWL)
gijlieden	-	ghiliede	ulieden uwerlieden	gijlieden

⁹⁵ Voor zover ze gevonden zijn. Er kan hier (nog) niet gemakkelijk op worden gezocht.

het	hit	het	het, bet. I & II et	het(I) 't(I) (VWL)
hij	hi	hi	hi imme(III) om(II) ome(II) omme(I) un ye(II) -ene -ne hon(I) (VWL)	hij hum(II) zijn(IV) apart van de andere <i>hij</i> , want onzekere etymologie -en(IV) (VWL) -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)	ik me(I) (VWL) mijn(IV) (VWL)
ikzelf	-	-	-	IKZELF (ik)
ir	-	ir	-	-
jij	-	-	-	jij ju je(I) (VWL) jou(I) (VWL)
jouwerliefde	-	-	-	JOUWERLIEFDE (jouw(II))
jullie	-	-	-	jullie (VWL) jelui (VWL)
men	man(II)	men	men me(III) mi(IV)	men me(III) (VWL)
U Ed.	-	-	-	UE
ulieden	-	-	-	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) us(I) use wer(IV) uus(I) wir wijr (VWL) wie(V) (VWL) wij (VWL)	wij(I) ons(II) (VWL) me(II) (VWL) wi(II) (VWL) we (VWL) wie(III) (VWL)
wijmekaar	-	-	-	WIJMEKAAR (wij(I))
wijzelf	-	-	-	WIJZELF (wij(I))
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)
zezij	-	-	-	Z'HUN (zij(I))
zich	-	-	-	zich(I), bet. 10
zij	siu sia(II)	si(II) si(III)	si(I) se soe hender(II) hun(II) ore(III) sou see(II) hare(VI) (zie ook <i>haar</i>) oor(I) (zie ook	zij haar(VII) (VWL) '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)

			<i>haar</i> hore(V) (zie ook <i>haar</i>) -er(II) -ge haer(VIII) (VWL)	
zij / haar	-	-	hoir (VWL)	heur (VWL)
zijlieden	-	haerliede siliede	haerlieden hemlieden siliede (VWL)	zijlieden haarlieden, haarlui (VWL) hunlieden, hunlui (VWL) zijlder, zijle(VWL) zijlude(n), zijlui(den)(VWL) zijnder(I) (VWL) zilder (VWL) zulder (VWL) zunder(II) (VWL) zullie (VWL)
zijmekaar	-	-	-	ZIJMEKAAR (zij(I))
zijzelf	-	-	-	ZIJZELF (zij(I))

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) (VWL) 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) au(III), auw (VWL)
uzelf	-	-	-	UZELF (u(II))
zich	sik(I)	sich(I)	sich(II) si(III)	zich(I)
zichzelf	-	-	-	zichzelf
zijzelf	-	-	-	zijzelf

Reciproque voornaamwoorden

Lemma	ONW	VMNW	MNW	WNT
elkander	-	elcanderen	elcander (VWL)	elkander
haerlijkander	-	haerlijcander	-	-
hen	hin	hem(IV)	-	-
iegelijkander	-	ieghelijcander	-	-
malkander	-	manlijcanderen	malcander, malcanderen mancander	malkaar, malkander(VWL) mekaar (VWL)
u	-	u(III)	-	-
zich	sik(II)	sich(II)	-	zich(I), bet. 11

Bezittelijke voornaamwoorden

Lemma	ONW	VMNW	MNW	WNT
dijn	thīn	dijn	dijn	dijn
gij / u / uw	-	-	hu(II)	ou(III) (VWL)
gij / uw	-	-	-	au(III), auw (VWL)

haar	hiro(I) hiro(II)	-	hore(V) (zie ook <i>zij</i>) hoere(II) ure(V) ur hare(VI) (zie ook <i>zij</i>) oor(I) (zie ook <i>zij</i>) haer(IX) (VWL) heur (VWL)	haar(VIII) 'r(IV) er(IV) (VWL)
hun	-	-	hun(III) hen(I) hon(II) (VWL) honne (VWL)	hun(III) hen(III) (VWL)
jouw	-	-	-	jouw juw je(III) (VWL)
mijn	mīn	mijn	mijn mi(II) me(II) mien(III) (VWL) miere(I) (VWL)	mijn(III)
ons	unsa	onse	onse(I) uus(II) ons(II) (VWL)	ons(III)
U Ed.	-	-	-	UE
uw	iuwa	uwe	u(II) ure(IV) us(III) uw	uw(I) u(IV) (VWL) uw(III) (VWL) uwe(III) (VWL)
zij / haar	-	-	hoir (VWL)	heur (VWL)
zich	-	-	-	zich(I), bet. 12 (en 13?)
zijn	sīn(II)	sijn(II)	sijn(II) si(II) sire(II) (VWL) ser(II)	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)
datwelk	-	-	datwelc	-
dewelke	-	dewelke	diewelke dewelke	dewelke
die	thie(IV)	die(III)	die(II) de(II)	die, bet. II
die / dat	-	-	des(II) (VWL)	-
hetgeen	-	-	-	hetgeen hetgunt (VWL)
hetwelk	-	-	hetwelc	hetwelk
hoedaan	-	-	hoedaen	hoedaan
hoedanig	-	hoedanich	hoedanich	hoedanig
hoegedaan	-	hoeghedaen	hoegedaen	-
hoegedanig	-	hoeghedanich	hoegedanich	-
hoezulk	-	-	hoezulk	-
wat	wat	wat(I) wes(I)	wat(I) wes(I) watte (VWL)?	wat(I) wes(I) wa(III) (VWL) watte(I) (VWL) wis(III) (VWL)
weder	-	-	-	weder(IV)
wederlei	-	-	-	WEDERLEI (weder(IV))
wel	-	-	wel(II) weel(II) wele(I) weele(I) (VWL)?	wel(IV) wiele(II) (VWL)
welk	wilik	welc	welc(I) wilc(I) walc(II) willeg	welk(II) 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	-	= ONBEP. VNW	welkertiere	-
welver	-	-	-	WELVER (wel(IV))
wesgelijk	-	-	-	WESGELIJK (wes(I))
wie	wie(I)	wie(II)	wie(I) wime(I), wieme wer(III) wis(IV) wime(III) wi(IV) (VWL) wies(II) (VWL) wim (VWL) win(IV) (VWL)	wie(I) wij(III,IV)(VWL) weem(II) (VWL) weme(n)(II)(VWL) wemes (VWL) wiemes (VWL) wien(I) (VWL) wiene (VWL)
wiedaan	-	-	wiedaen	WIEDAAN (wie(II))
wiedanig	-	-	wiedanich	WIEDANIG (wie(II))
wiegedaan	-	wieghedaen	wiegedaen	-
wie veel	-	-	-	WIEVEEL (wie(II))
woedaan	-	-	woedaen (VWL)	WOEDAAN (woe)
woedanig	-	-	woedanich	-
woegedaan	-	-	woegedaen	-
woeveel	-	-	woevele	-
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	-	-
alzozulk	-	-	alsosulc	-
alzulk	-	alsulc alsesulc (VWL) alsosulc (VWL)	alselc alsesulc alsolijc	alzulk
alzulkdanig	-	alsulcdanich	alsulcdanich	alzulkdanig
alzulks	-	-	-	alzulks
alzusdaan	-	alsusdaen alsuster (VWL)	alsustaen	-
alzusdanig	-	alsusdanich	alsustanich	-
alzusgedaan	-	-	alsusgedaen	-
dat	that(I) that(II) = LID	dat(I) dat(III) = LID	dat(I) dos(II) s(III)	dat (VWL)
datgene	-	-	tgint tgene, tgent(VWL)	-
de	thie(III) = LID	die(IV) = LID	de(I) = LID & VNW die	de = LID & VNW dat; bet.

				1(VNW) & 3(LID)
degene	-	-	degene diegene	degene diegene
degenige	-	-	diegenige	-
dergelijk	-	diereghelike	derregelike (VWL)	dergelijk diergelijk (VWL)
desgelijk	-	desghelike, diesghelike(I) desghelikes, diesghelikes(I)	desgelijc	desgelijk diesgelijk
deze	these	dese	dese	deze
dezelfde	-	-	-	dezelfde diezelfde
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)	-
dit	thit, thitti	dit	dit	(zie onder deze)
ditzelfde	-	-	-	ditzelfde
dulk	-	-	dulc	-
dus	-	-	-	dus
dusdaan	-	dusdaen	dusdaen	dusdaan
dusdanig	-	dusdanich	dusdanich	dusdanig
dusgedaan	-	dusghedaen	dusgedaen	dusgedaan (VWL)
dusgedanig	-	-	dusgedanich	-
dusk	-	-	dusc	dusch
duskgedaan	-	-	dusschedaen	-

duswaard	-	-	duswert	-
duszulk	-	dussulc	-	-
eadem	-	eadem	-	-
gene	-	ghene(II)	gene gint(I) goont (VWL)	gene gunt (VWL)?
gewelk	= ONBEP. VNW	ghewelc	-	-
het	-	-	het, bet. III = LID the(II) = LID	het(II) = LID 't(II) = LID 't(III) = LID(VWL)
hetgeen	-	-	-	hetgeen hetgunt (VWL)
hetzelfde	-	-	-	hetzelfde
hetzelfste	-	-	-	hetzelfste
hetzelve	-	-	-	hetzelve
hoogstdezelve	-	-	-	hoogstdezelve
idem	-	idem	-	idem (marginaal)
proper	-	= BNW & BW	proper, bet. III	= BNW, BW
wat	-	wat(I) wes(I)	wat(I)	-
zelf	self selvo(I)	selve	selve(I) selves, bet. II	zelf(II) zelft(I,II) (VWL) zelve(r)(I,II,III) (VWL) zelven(s) (VWL) zelvers (VWL)
zelfde	-	-	selfde	zelfde
zelfdig	-	-	-	ZELFDIG (zelfde)
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.
zulk	sulik	sulc	sulc solijc selc (VWL) solc (VWL) suilc (VWL) swelc (VWL)	zulk
zulkdaan	-	-	sulcdaen	-
zulkdanig	-	sulke danighe(VWL)	sulcdanich	zulkdanig
zulkerhande	-	sulkerhande	sulkerhande	-
zulkerlei	-	-	-	ZULKERLEI (zulk)
zulkgedaan	-	-	sulcgedaen	-
zulkgedanig	-	sulcghedanich	sulcgedanich	-
zulks	-	-	sulcs	zulks
zulksdergelijk	-	-	-	ZULKSDEGELIJK (zulks)
zulkswat	-	-	-	ZULKSWAT (zulks)
zus	-	-	-	zus(II)
zusdaan	-	-	susdaen	ZUSDAAN (zus(II))
zusdanig	-	susdanich	susdanich	ZUSDANIG (zus(II))
zusgedaan	-	susghedaen(I)	susgedaen	-
zusgedanig	-	-	susgedanich	-
zusk	-	-	-	zusk
zuslijk	-	-	suslijc	-
zustig	-	-	-	ZUSTIG (zus(II))

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* worden allemaal gezien als onbepaalde voornaamwoorden, ongeacht of het eerste lid een telwoord of zelf een onbepaald voornaamwoord is.

N.B. 4 Ook (*alle*)*beide*, *beidegader* en woorden gevormd met *ge-* en *-n*, zoals *gedrieën*, worden gezien als onbepaald voornaamwoord, omdat net zoals bij *alle* en *sommige* in de betekenis wordt uitgedrukt dat het gaat over de hele verzameling en in tweede instantie pas om uit hoeveel elementen deze verzameling bestaat.

Lemma	ONW	VMNW	MNW	WNT
achterhande	-	-	-	achterhande
achterlei	-	-	-	achterlei
al	al(II)	al(II)	al(I) alles (VWL) ol(I)	al(I) alle (VWL)
algelijk	-	-	algelijc	-
algader	-	algader(II)	algader(II) allegader (VWL)	allegaar algader (VWL) allegader (VWL)
allebeide	-	-	-	allebeide allebei (VWL)
allemaal	-	-	-	allemaal(I) sup.
alleman	-	alleman	alman	alleman(II)
allerhande	-	alrehande(I)	alrehande (VWL) alrande (VWL)	allerhande
allerkunne	-	-	alreconne, alrekunne	-
allerlei	-	-	alreleide (VWL)	allerlei
allermalk	-	-	alremalc, alremallijc	-
allermannelijk	-	alremanlijc	-	-
allerslachte	alleroslahta	alreslachte	alreslachte (VWL)	-
allertieren	-	alretiere(n)	alretiere (VWL)	allertier ALLERTIER (tier(I))
alles	-	-	-	alles(I)
allijkeveel	-	-	allikeveel	allijkeveel
alluttel	-	-	allettel(I)	-

			aluttel, alluttelken, alluttelkijn(VWL)	
altegader	-	altegadere(I)	= BW	= BW
altemaal	-	-	-	altemaal sup.
alzomenig	-	alsomenich	-	-
alzowat	-	alsowat	-	-
alzulkerhande	-	-	alselkerhande	-
ander	andar(II)	ander	ander aer(IV)	ander
anderhande	-	-	anderhande	-
anderlei	-	-	anderleie (VWL)	ANDERLEI (ander)
anderlieden	-	-	-	ANDERLIEDEN (lieden)
anderman	-	anderman	-	anderman
andertieren	-	-	andertiere	-
beide	bētho(I)	beide(II)	beide(II) bee, beede (VWL)	beide
beidegader	-	beidegader	beidegader	beidegaar
beiderhande	-	-	-	beiderhande
beiderlei	-	-	-	beiderlei
drieërhande	-	-	drierhande	drieërhande
drieërlei	-	-	drierleye	drieërlei
duizenderhande	-	-	-	DUIZENDERHAN DE (duizend)
duizenderlei	-	-	-	duizenderlei
duizentigsterhande	-	-	dusentichsterande, dusentichsterhande	-
een	ēn(II) ēn(III) = LID	een(II) een(III) = LID	een(III) een(II) = LID ien (VWL)	een(I), bet. IV & II = LID van onbepaaldheid
een iegelijk	-	eenieghelijc	-	-
eenweder	-	-	eenweder	-
eenwelk	ēnwilik	-	-	-
el	-	el(I)	el(I) els. bet. 1 & 2	el(II)

			elx(I) (VWL)	
eliemand	-	eliemene	-	-
elk	-	elc	elc	elk
elkeen	-	-	-	elkeen
elkerhande	-	-	elkerhande	-
elkerlijk	-	elkerlijc	elkerlijc	elkerlijk
elkermalk	-	-	elkermalc	-
elkertieren	-	-	elkertiere	-
elkman	-	-	elcman	-
elks	-	-	elx(II)	-
elnegeen	-	elnegheen	-	-
enerhande	-	eenrehande, bet. 1	eenrehande	eenerhande
enerlei	-	-	-	eenerlei
enig	ēnig(II)	enich(II)	enich(I) eenic	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) inkel(II) (VWL)?	enkel(II) inkel(II) (VWL)?
enthoeveel	-	-	-	enthoeveel
entwat	-	-	-	entwat (VWL)
entwie	-	-	-	entwie
ettelijk	-	-	itlijc	ettelijk
eweinig	-	-	ewenich	-
evenveel	-	-	-	evenveel
gedrieën	-	-	-	gedrieën
getienen	-	-	-	getienen

getwaalven	-	-	-	getwaalven
getweeen	-	-	-	getweeen
getwintigen	-	-	-	getwintigen
gevieren	-	-	-	gevieren
gevijven	-	-	-	gevijven
gezevenen	-	-	-	gezevenen
geen	gēn	gheen	geen(II)	geen
gelijk	-	-	gelijc(V)	-
gelijkveel	-	-	= BNW	gelijkeveel
generhande	-	ghenerehande	geenrehande	geenerhande
generkunne	-	-	geenreconne geinreconne(VWL)	-
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	-	hoemenich	-	hoemenig
hoemenigste	-	-	-	hoemenigste
hoeveel	-	hoevele	hoeveel	hoeveel
hoeveelde	-	-	-	hoeveelde
hoeveelste	-	-	-	hoeveelste
honderderlei	-	-	-	honderderlei
ieder	-	-	ieder	ieder
iedereen	-	-	-	iedereen
iederman	-	-	-	iederman
iegelijk	iegilīk	ieghelijc	iegelijc jegenlijc	iegelijk

iegewelk	-	ieghewelc	iegewelc	-
ielijk	-	-	ielijc illijc	-
iemand	ieman	ieman, iemant iemant (VWL)	ieman	iemand
iet	iewiht	iewet(I) iewent (VWL)	iet(I) iewent	iet (VWL)
iets	-	-	-	iets
ietwat	-	-	-	ietwat
ietwes	-	-	-	ietwes
iewelk	-	-	iewelc	-
kooi	-	-	-	kooi(IV)
loene	-	-	-	loene(I)
lou	-	-	-	louw(III)
lou loene	-	-	-	louw loene
luttel	luttel(I) (<i>her lutcel êren</i>)	luttel(III)	luttel(II)	luttel(II) luttel lutter(II)
luttelgoed	-	luttelgoed(I)	luttelgoed	-
malk	-	-	malc mallic (VWL) mallinc(II) (VWL)	-
malkmalk	-	-	malcmalc (VWL)	-
mannelijk	-	manlijc(II)	manlijc(II) manlike(II) (VWL)	-
meen	-	-	mene(IV)	-
meer	mēro(III)	mee(I) mere(IV)	meer(IX) mere(IX) meere(IV)(VWL) mee(II) (VWL) me(I) (VWL)	meer(VI)
meerder	-	= BNW	= BNW	meerder
meest	= BNW	meest(II)	meest	meest
menig	manig manigiro	menich menighere	menich manich(I) manech (VWL) menichste	menig
menigeen	-	menicheen	menicheen (VWL)	menigeen
menigerhande	-	menigherhande	menigerhande	menigerhande

menigerkunne	-	-	menigerconne	-
menigerlei	-	-	menigerleye menigerleide(VWL)	menigerlei
meniglieden	-	menigheliede	-	-
meniglijk	-	-	minclijc(II)	-
menigerslachte	manigerslahta	menigherslachte(n)	-	-
menigertieren	-	menighertiere(n)(I)	menigertiere, menigertieren	menigertier
min	= BW	min(I)	min(I)	min(IV)
minder	= BW	minre (II)	= BNW	minder
misselijkerhande	-	-	misselikerhande	-
neen	nēn	-	neen(II)	-
negeen	nehēn	negheen	negeen engeen (VWL) egeen (VWL)	negeen egeen (= GEEN LID)
negenderhande	-	-	negenrehande	negenderhande
negenderlei	-	-	negenderleye	negenderlei
negenerhande	-	neghenerhande	negeenrehande	-
negenerkunne	-	-	negeenreconne	-
negenerlei	-	-	negeenreleye	-
negenerslachte	-	negheenreslachte	negeenreslachte, negeenreslachten	-
negenertieren	-	negheenretiere	negeenretiere geretiere (VWL)	-
negenerhande / negenertieren / negenerwijs	-	-	negerehande, negeretiere, negerewijs(VWL)	-
negiemand	-	-	negiemant	-
nenerlei	-	-	nenerleye	-
neweder	newethar(I)	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

niemendal	-	-	-	niemendal
nienegeen	-	nienegheen	nienegeen	-
niet	= BW	niewet(I) niewent(I)	niet(I) niewet niewent	niet(II) nie(II) (VWL)
niets	niewiht(I), niet 1	niets	niets	niets
nietwes	-	-	-	nietwes
nieweder	-	-	nieweder	-
nihil	-	-	nihil nichel(I) (VWL)	
niks	-	-	-	niks
nimmeer	-	nemmee(I)	nemmee, bet. II nemmeer(I) (VWL) nemmer(I) (VWL)	-
noeg	-	-	noech	-
nochweder	-	nochweder	=VOEGWOORD	-
nop	-	-	nop(III) (VWL)	-
noppes	-	-	-	noppes
noweder	-	=VOEGWOORD	noweder & VW	-
nuist	-	-	nuyst	-
onmenig	unmanig	onmenich	onmenich	-
onzerlijk	-	-	onserlijc	-
overgenoeg	-	-	-	overgenoeg
overveel	-	-	overvele(I)	overveel
plenty	-	-	-	plentie
som	-	som	som(I) somme(III)	som(III) zom
somlieden	-	-	somliede	-
sommelijk	sumalīk	somelijc	somelijc	-
sommenk	-	-	-	sommenk
sommig	-	somich	somich	sommig
sommigeen	-	somicheen	= als verbinding onder sommig	-
tel	-	-	tel	- NIET ALS VNW

tienderhande	-	-	tienderhande	tienderhande
tienderlei	-	-	-	tienderlei
twaaifderhande	-	-	-	TWAAIFDERHANDE (twaalf(I))
twaaifderlei	-	-	-	TWAAIFDERLEI (twaalf(I))
tweeheidjes	-	-	-	tweeheidjes
tweeërhande	-	-	tweerhande	tweederhande
tweeërlei	-	-	tweerleye	tweederlei
tweërtieren	-	-	tweertiere	-
twink	-	-	twinc	-
twint	-	-	twint	twint(I)
twintigerlei	-	-	-	TWINTIGERLEI (twintig)
uist	-	-	vist	-
uwerlijk	-	-	uwerlijc	-
veel	filo(II)	vele(I)	vele(I) vele(II) vole(II) (VWL) ? voel(I) (VWL) volle(VII) (VWL)	veel(IV) veul(I) (VWL)
veertienderlei	-	-	-	VEERTIENDERLEI (veertien)
veertigerhande	-	-	-	VEERTIGERHAND E (veertig)
veertigerlei	-	-	-	VEERTIGERLEI (veertig)
velerhande	-	-	velerhande	velerhande veelderhande(VWL)
velerhande / velerlei	-	-	veelrehande, veelreleye (VWL)	
velerlei	-	-	-	velerleie veelderlei (VWL)
velertieren	-	-	velertiere	-
verscheiden		-	-	verscheiden(IIIC)
verscheidenerhande	-	-	-	VERSCHEIDENERHANDE (verscheiden(III))
verscheidenerlei	-	-	-	VERSCHEIDENERLEI (verscheiden(III))

verscheiderlei	-	-	-	VERSCHEIDERLEI (verscheiden(III))
verschillende	-	-	-	verschillend, bet. 4
verschillenderhande	-	-	-	VERSCHILLENDERHAN DE (verschillend)
verschillenderlei	-	-	-	VERSCHILLENDERL EI (verschillend)
vierderhande	-	-	vierderhande	vierderhande
vierderlei	-	-	vierderleye	vierderlei
vijfderhande	-	-	vijfterande, vijfterhande	vijfderhande
vijfderlei	-	-	vijfterleye	vijfderlei
vijftienderhande	-	-	-	VIJFTIENDERHAN DE (vijftien)
vijftienderlei	-	-	-	VIJFTIENDERLEI (vijftien)
vijftigerhande	-	-	-	VIJFTIGERHANDE (vijftig)
vo	*fō	-	-	-
vrijwat	-	-	-	vrijwat
wat	-	wat(I) wes(I)	wat(I) wes(I)	wat(I) wa(III) (VWL) watte(I) (VWL)
watkunne	-	-	watconne	-
watterhande	-	-	watterhande	-
watterlei	-	-	watterleye	-
weder	wethar(II)	weder(III)	weder(IV)	NIET ONBEP.
wee	-	-	wee(III)	-
weinig	= BNW	= BW	weinich	weinig weenig (VWL) wennig (VWL)
weinigerhande	-	-	-	WEINIGERHANDE (weinig)
wel	-	-	wele(I)	wel(IV) wiele(II) (VWL)
welk	-	welc	welc(I) wilc(I) walc(II) willeg	welk(II) 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	-	-
wonderveel	-	-	-	WONDERVEEL (wonder(II))
zat	= BNW	= BNW & BW	= BNW & BW	zat(I) zatter (VNW)
zeker		seker(II), bet. 3	seker(I)	zeker(I)
zesderhande	-	-	sesterhande	zesderhande
zesderlei	-	-	-	zesderlei
zestienderhande	-	-	-	ZESTIENDERHANDE (zestien)
zestienderlei	-	-	-	ZESTIENDERLEI (zestien)
zevenderhande	-	-	sevenderhande	zevenderhande
zevenderlei	-	-	-	zevenderlei
zeventienderhande	-	-	-	ZEVENTIENDERHANDE (zeventien)
zeventienderlei	-	-	-	ZEVENTIENDERLEI (zeventien)
zeventigerhande	-	-	-	ZEVENTIGERHANDE (zeventig)
zeventigerlei	-	-	-	ZEVENTIGERLEI (zeventig)
ziezoveel	-	-	-	ZIE-ZOO-VEEL (zien(I))
zoiet	-	-	-	ZOOGET (zoo(II))
zoiets	-	-	-	ZOOIETS (zoo(II))
zomenig	-	somenich	-	-
zoveel	-	sovele(I)	-	zooveel
zoveelde	-	-	-	ZOVEELDE (zooveel)

zoveelste	-	-	-	zooveelste
zowat	-	so wat	-	zoowat
zulk	-	sulc	sulc selc (VWL) solc (VWL) suilc (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*.⁹⁶ 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*. *Zo 'n* komt niet als lemma voor, alleen in het WNT bij *zoo* in de combinatie *zoo + een*, maar het wordt hier aanwijzend genoemd.

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)
zulk	-	-	-	zulk, bij I, (c.)?

⁹⁶ Zie de [Exclamatieven](#) in de ANS.

Bijlage 3: formele beschrijving tagset