

Lassy Syntactische Annotatie

Revision : 21780

Gertjan van Noord, Ineke Schuurman, Gosse Bouma

26 april 2021

Inhoudsopgave

1	Introductie	7
1.1	Inleiding	7
1.2	Uitgangspunten van de syntactische annotatie	8
1.3	Dependentiestructuren	9
1.4	XML-formaat	12
1.4.1	DTD	12
1.4.2	Kanonieke dependentiestructuren	14
1.4.3	Representatie van POSTAG en LEMMA	15
1.5	Relatie met CGN syntactische annotatie	16
1.5.1	Illustraties	16
1.5.2	XML formaat	18
1.5.3	Verschillen in de annotatie zelf	18
1.6	XPATH in plaats van TigerSearch	20
1.6.1	Inleiding	20
1.6.2	Meer voorbeelden	23
1.6.3	Valkuilen	24
1.6.4	De voorbeelden van Lai en Bird	25
1.6.5	Voorbeelden met XPATH 2.0	27
2	Lokale dependentiestructuren	29
2.1	Het verbale domein	29
2.1.1	Werkwoordelijk hoofd	29
2.1.2	Categorielabels	30
2.1.2.1	Declaratieve zin	31

2.1.2.2	Werkwoordsinitiële zin	31
2.1.2.3	Overige categorieën	32
2.1.3	Onderwerp, voorlopig onderwerp	33
2.1.3.1	Plaatsonderwerp	34
2.1.4	Direct object, voorlopig direct object	34
2.1.5	Verplicht reflexief object	35
2.1.6	Secundair object: meewerkend, belanghebbend, onder- vindend voorwerp	36
2.1.7	Loos <i>het</i>	38
2.1.8	Predicatief complement	39
2.1.8.1	Clefts	42
2.1.9	Verbale complementen	44
2.1.9.1	Hulpwerkwoorden	44
2.1.9.2	Passief	48
2.1.9.3	AcI	51
2.1.9.4	Beknopte bijzinnen	52
2.1.9.5	Finiete bijzinnen	54
2.1.10	Vorzetselvoorwerp	55
2.1.11	Maatcomplement	56
2.1.12	Locatief/directioneel complement	57
2.1.13	Scheidbaar deel van samengesteld werkwoord	58
2.1.13.1	Partikels	58
2.1.13.2	Idiomatisch deel van werkwoordelijke uitdruk- king	62
2.1.13.3	Postposities	71
2.1.14	Modificerende elementen in het verbale domein	71
2.1.14.1	MOD en PREDM	71
2.1.14.2	Door-bepaling in passieve zinnen	72
2.1.14.3	Oordeelspartikels	72
2.1.14.4	Aanhechting van modificerende elementen in het verbale domein	72
2.1.15	Kiezen tussen modificeerder en complement	73

2.1.15.1	Argumenten en modifiers bij ‘verzwegen’ werkwoord	73
2.1.15.2	De rol van PP bij werkwoorden	74
2.1.15.3	Werkwoorden met <i>mee</i>	75
2.1.15.4	Het werkwoord <i>zijn</i>	75
2.2	Extensies van verbale domeinen: complementeerdere	76
2.2.1	Bijzinnen ingeleid met onderschikkend voegwoord	76
2.2.1.1	Bijzinnen met <i>dat, of, om</i>	76
2.2.1.2	Bijzinnen met <i>van</i>	78
2.2.1.3	Bijzinnen als modificeerder	79
2.2.1.4	Modificatie van complementeerdere zoals <i>voordat, nadat</i>	81
2.2.1.5	Niet-finiete bijzinnen als complement	82
2.2.2	Complexe voegwoorden	86
2.2.3	Of-zinnen	88
2.2.4	Als het ware	91
2.2.5	Dat is goed te doen	92
2.2.6	Laat staan	93
2.2.7	Onderschikkende voegwoorden zonder verbaal vervolg	94
2.3	Het adjectiefdomein	96
2.3.1	PC, VC, OBJ1	97
2.3.2	Reflexief complement	99
2.3.3	Comparatieven	99
2.3.4	Superlatieven	101
2.3.5	Maatcomplement	102
2.3.6	Andere complementen	103
2.3.7	Modificatie	103
2.3.8	Het is ADJ om/dat	104
2.3.9	Leuk om aan terug te denken	107
2.3.10	’s Morgens vroeg	109
2.4	Het prepositionele domein	109
2.4.1	Prepositionele hoofden	109
2.4.2	Reflexief object van een prepositie	110

2.4.3	(Voorlopig) object van een prepositie	110
2.4.4	Absolute <i>met</i> -constructies	112
2.4.5	Modificerende elementen	113
2.4.6	Voorzetseluitdrukkingen	115
2.4.7	Als, zoals	115
2.5	Het nominale domein	116
2.5.1	De nominale kern	116
2.5.2	De determinator	117
2.5.2.1	Complexe determinatoren	118
2.5.2.2	Wat dan ook	125
2.5.2.3	De partitieve <i>van</i> -constructie	127
2.5.3	Complementen binnen de nominale constituent	127
2.5.4	Modificerende elementen	128
2.5.5	Bijstellingen: APP	132
2.5.6	Deverbale nominalisaties	133
2.5.7	VP modificatie met missend argument	135
2.5.8	Vocatieven	136
2.5.9	Jan zijn fiets	136
2.5.10	Voor NP aan NP	137
2.5.11	Kwantitatief <i>er</i>	138
2.5.12	Problemen hebben/krijgen met	138
2.5.13	Projectie van voornaamwoorden	139
2.5.14	Iets van, rond de, meer dan + telwoord	139
2.6	Het domein van de bijwoorden	140
2.6.1	Complementen en bepalingen bij bijwoorden	140
2.6.2	Terug, geleden	141
2.6.3	OBCOMP bij bijwoorden	142
2.7	Named Entities	150
2.7.1	Namen, tijdsuitdrukkingen, data, adressen	150
2.7.2	Gemodificeerde multi-word-units	151
2.8	Geschreven taal fenomenen	152
2.8.1	Afkortingen en dergelijke	152
2.8.2	Niet-talig en niet-Nederlands materiaal	154

2.8.3	huis-, tuin- en keukenproblemen	156
2.8.4	Materiaal tussen haakjes	156
3	Niet-lokale dependenties	159
3.1	Introductie	159
3.2	Relatieve zinnen	161
3.2.1	Relatieve zinnen als modificeerders bij zelfstandig naam- woorden	161
3.2.2	Relatieve zin met ingesloten antecedent	164
3.2.3	Relatieve zin als modificeerder van werkwoorden	165
3.2.4	Verkorte relatieve zinnen	167
3.3	Constituentvraag: hoofd- en bijzin	168
3.4	Er zijn er/Je hebt er	169
3.5	Verkorte vraagzinnen	170
3.6	Vraagzinnen zonder persoonsvorm	172
3.7	Vraagzinnen met meerdere WH-elementen	173
4	Nevenschikking	175
4.1	Inleiding	175
4.2	Etcetera	181
4.3	Modificatie van conjuncten	182
4.4	Onverwachte woordvolgorde bij <i>zowel . . . als</i>	182
4.5	Asyndetische nevenschikking, en lijsten	183
4.6	Woordgroepen die met een coordinator beginnen	184
5	Discourse-structuren	185
5.1	Overzicht	185
5.2	Discourse-markeerders: DLINK	186
5.3	Uitbreidingen van de zin: SAT	186
5.4	Uitbreidingen van de zin: TAG	188
5.4.1	Tussenwerpsels	188
5.4.2	Gereduceerde zinnen	189
5.4.3	Aansporingen	190
5.4.4	Inleiding van directe rede	190

5.5	Asyndetische constructies: DP	192
5.5.1	Lijsten	194
5.5.2	Kopteksten	194
5.5.3	Hoe meer, hoe beter	194
6	Dekking, desambiguatie en ongrammaticale invoer	196
6.1	Dekking	196
6.2	Desambiguatie	196
6.3	Ongrammaticale invoer	198
6.3.1	Missende woorden	198
6.3.2	Ten onrechte opgenomen woorden	199
6.3.3	Foutgespelde woorden	201
7	Automatische annotatie	202
7.1	Overzicht additionele attributen	202
7.2	Details	203
A	Appendix	207
A.1	Gebruikte categorieën	208
A.2	Gebruikte dependenterelaties	208

Hoofdstuk 1

Introductie

1.1 Inleiding

In dit document bespreken we de syntactische annotatie in Lassy in termen van grammaticale relaties en syntactische categorieën. Het doel van dit rapport is om te documenteren hoe het Lassy corpus syntactisch geannoteerd is.

De syntactische annotatie bevat de toekenning van woordsoorten en de toekenning van lemma's. Deze annotatielaag is gedocumenteerd in een afzonderlijk document ([Van Eynde, 2005](#)).

Dit document is gebaseerd op CGN SYNTACTISCHE ANNOTATIE ([Hoekstra et al., 2004](#)) maar een aantal typische spreektaalfenomenen worden in dit manual niet behandeld, terwijl andere constructies juist zijn toegevoegd. Om de beknoptheid te bevorderen zijn overwegingen voor of tegen een bepaalde analyse veelal ingekort of verwijderd. Ten slotte is een aantal in D-Coi voorgestelde wijzigingen met betrekking tot de annotaties doorgevoerd in deze tekst. Daarnaast zijn er in vergelijking met de CGN annotatie manual ook een flink aantal constructies toegevoegd.

In het vervolg van dit eerste hoofdstuk vatten we de achterliggende filosofie van de annotatie kort samen, we introduceren de gebruikte illustraties, en we presenteren het gebruikte XML formaat.

Hoofdstuk 1.5 is bedoeld voor gebruikers van het Lassy corpus die ervaring hebben met het CGN corpus. We leggen in dit hoofdstuk uit wat de verschillen zijn met CGN. Daarnaast bevat dit hoofdstuk een handleiding voor het gebruik van onder andere DtSearch voor gebruikers die eerder met TigerSearch hebben gewerkt.

De behandeling van een groot aantal syntactische constructies neemt het grootste deel van dit document in beslag. We bespreken achtereenvolgens de lokale depen-

dentiestructuren (hoofdstuk 2), niet-lokale dependentiestructuren (hoofdstuk 3), nevenschikking (hoofdstuk 4) en discourse-structuren (hoofdstuk 5). Hoofdstuk 6 gaat in op kwesties zoals ongrammaticale invoer, meerduidige invoer, en wat te doen in geval van constructies die niet in dit document besproken worden.

Het Lassy corpus bestaat uit twee delen. Het eerste deel, Lassy Klein, omvat een miljoen woorden. Dit corpus is manueel geverifieerd, en zou dus geen fouten behoren te bevatten. Het tweede deel, Lassy Groot, omvat een half miljard woorden. Dit corpus is automatisch geannoteerd door de Alpino parser (van Noord, 2006), en niet verder gecorrigeerd. Automatische annotatie is nog lang niet perfect, dus Lassy Groot zal ook foute annotaties bevatten. Daarnaast wordt in de automatisch toegekende annotaties wat meer informatie toegevoegd dan in de handmatig geverifieerde annotaties. Dit betreft meestal lexicale informatie zoals het gebruikte subcategorisatie-frame. Deze additionele informatie wordt gerepresenteerd door een aantal extra XML-attributen, en wordt gedocumenteerd in hoofdstuk 7.

1.2 Uitgangspunten van de syntactische annotatie

De basisfilosofie van de annotatie is erop gericht een optimale verhouding tussen input en output te bereiken, in de volgende zin:

Input Aan de inputkant willen we de gehanteerde annotatieschema's zo eenvoudig mogelijk houden, om daarmee de werklust van de annotatie- en correctieprocedures te minimaliseren.

Output Aan de outputkant willen we de gebruikers zo rijk mogelijke annotatie-informatie aanbieden.

Om deze doelen te realiseren, gaan we uit van een theorie-neutraal annotatieniveau in termen van dependentiestructuren. Deze primaire annotatie wordt verrijkt met categorie, POS-tag en lemma informatie.

Gerelateerd aan de wens om rijke output te leveren, is de wens om de annotaties zo aan te leveren, dat ze zich goed lenen voor automatische verwerking. Daarom is gekozen voor een representatie van de annotaties in XML die goed aansluit bij de W3 standaarden voor het werken met XML (zoals XSLT, XPATH en XQUERY). Tools om dependentiestructuren te bewerken worden in een ander document (Kloosterman, 2007) beschreven; we noemen hier de belangrijkste:

DtView Grafische interface om sets van annotaties te bekijken. Annotaties worden als boomstructuren gevisualiseerd.

DtSearch Gereedschap om grote hoeveelheden annotaties te doorzoeken aan de hand van XPATH queries.

DtTrEd Gereedschap voor het handmatig aanpassen van annotaties. Gebaseerd op de TrEd editor (Pajas and Štěpánek, 2008) die oorspronkelijk voor de Praag Treebank werd ontwikkeld.

DtXslt Gereedschap om door middel van stylesheets dependentiestructuren om te zetten in andere formaten. De illustraties in dit document zijn bijvoorbeeld automatisch gegenereerd op basis van dependentiestructuren in de standaard XML notatie.

Gerelateerd aan de wens om rijke output te leveren is ook de beslissing om voor automatisch gemaakte annotaties meer informatie te leveren; deze extra informatie komt aan de orde in hoofdstuk 7.

1.3 Dependentiestructuren

Een dependentiestructuur is een gelabelde gerichte, acyclische graaf. We beschikken over disjuncte verzamelingen CAT (categorielabels) en DEP (dependentielabels) voor de labeling van, respectievelijk, de knopen en de bogen ('edges').

- Knopen: CAT: categorielabels, de vereniging van lexicale labels POS-tag en frasale labels SYNCAT.
- Bogen: DEP: dependentielabels

We onderscheiden *gelede* en *ongelede* dependentiestructuren. Een ongelede dependentiestructuur is simpelweg een knoop met een categorielabel uit POS-tag. Een ongelede dependentiestructuur is bovendien altijd geassocieerd met een woord (token) uit de zin, en het corresponderende lemma. Andersom geldt ook, dat elk token uit de zin terug te vinden is als een ongelede dependentiestructuur.

De manier waarop POS-tag en lemma informatie aan de tokens wordt toegekend, wordt in een ander document beschreven (Van Eynde, 2005).

De elementaire bouwstenen van gelede dependentiestructuren noemen we lokale *dependentiedomeinen*. De moederknoop van een dependentiedomein is gelabeld met een frasaal label uit SYNCAT. De dochters hebben categorielabels uit CAT. De dependentielabels voor de moeder-dochtertakken worden gevormd door een *hoofd*, samen met de *complementen* en de *modificeerders* van dat hoofd.

Hoofd We onderscheiden verschillende types, waarvoor de dependentielabels HD, RHD, WHD, CMP, CRD, DLINK worden gereserveerd. Een hoofd komt per domein normaliter hoogstens één keer voor.

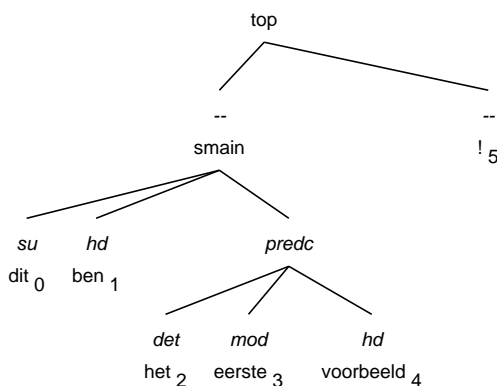
Complementen Het complementatiepatroon bepaalt de interpretatie van het hoofd in termen van thematische structuur. Een complementlabel komt per domein normaliter hoogstens één keer voor. De dependentielabels SUP, SU, OBJ1, OBJ2, VC, POBJ1, PREDC, ME, PC, LD, SE, SVP, OBCOMP worden voor complementen gereserveerd. SVP complementen kunnen soms meerdere keren voorkomen binnen een domein, in geval van werkwoordelijke uitdrukkingen waarbij meerdere delen worden onderscheiden – zie sectie 2.1.13.2 op bladzijde 62.

Modificeerders Modificerende elementen laten het categorielabel van de moederknoop ongemoeid; ze zijn (in principe) weglaatbaar zonder effect op de thematische structuur. Een modificeerderlabel kan binnen een domein meerdere voorkomens hebben. De dependentielabels voor modificeerders zijn MOD, PREDM, DET, APP.

De wortel van een dependentiestructuur heeft altijd categorie TOP. Alle dochters van de wortel hebben relatie --. De eigenlijke analyse van de zin is een van deze dochters. De overige dochters (indien aanwezig) worden voor punctuatie gebruikt, en eventueel ook andere tokens in de zin die niet in de eigenlijke dependentiestructuur zijn opgenomen.

In dit document gebruiken we de volgende visualizatie voor dependentiestructuren:

(1) Dit is het eerste voorbeeld !



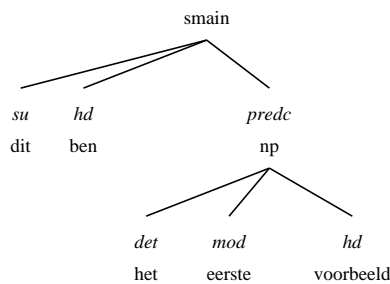
Deze visualizaties zijn vereenvoudigingen van de volledige dependentiestructuren, omdat we in dit document geen pogingen doen om in de voorbeelden juiste

waardes voor POS-tag en lemma te geven, uit angst hiermee inconsistenties met de documentatie op het gebied van POS-tag en lemma te introduceren (Van Eynde, 2005).

Om de beknoptheid te bevorderen van de voorbeelden, introduceren we nog twee vereenvoudigingen in de visualizaties:

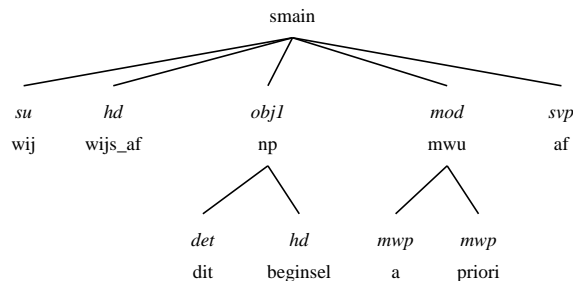
- De stringposities die elke ongelede knoop met zich draagt, om eenduidig te verwijzen naar het woord in de betreffende zin, worden niet getoond. In dit document zijn deze uit de context altijd wel terugvindbaar.
- Ook geven we in dit document vaak niet de eerste verdieping van de dependentiestructuur (de knopen direct onder TOP), indien dit niet ten koste gaat van de duidelijkheid.

Het voorgaande voorbeeld ziet er met deze vereenvoudigingen als volgt uit:



Versteende woordreeksen die zich syntactisch als een enkel woord gedragen (*multi-word-units*) worden gerepresenteerd met een eigen domein met als categorielabel MWU waarbinnen elk token gerepresenteerd wordt met een ongelede dependentiestructuur, steeds met dependentielabel MWP.

(2) Wij wijzen dit beginsel a priori af



Het samennemen van complementatie en modificatie binnen één dependentiedomein leidt tot ‘ondiepe’ annotatiestructuren. Enkele gevolgen:

- een nieuw domein (hiërarchisch niveau) wordt pas geopend als een nieuw hoofd daar aanleiding toe geeft
- complementatie en modificatie zijn *relaties* die frases onderhouden met betrekking tot een hoofd: als er geen complementen of modificeerders zijn, is er ook geen aanleiding tot niet-vertakkende projecties
- dependentiedomeinen zijn, in het standaardgeval, *lexicaal verankerd*. Een dependentiedomein is lexicaal verankerd, als het hoofd een ongelede dependentiestructuur is, of als het hoofd een multi-word-unit is. De eis dat dependentiedomeinen lexicaal verankerd zijn wordt streng toegepast voor de labels HD en CMP. Hoofden van het type RHD of WHD zullen vaak wel geleed zijn (zie hoofdstuk 3 op bladzijde 159). CRD tenslotte is soms afwezig; dit geeft dus aanleiding tot een dependentiedomein zonder hoofd (zie hoofdstuk 4 op bladzijde 175), ook zijn er soms meerdere CRD hoofden.

Deze aannames leiden tot ‘ondiepe’ of ‘platte’ verbale projecties. We onderscheiden op categorielabel niveau *finiete* en *niet-finiete* verbale projecties. De persoonsvorm is hoofd van de finiete zinstypen; de infinitief of het deelwoord van de niet-finiete. Er is in de finiete zin geen tussenliggend VP-niveau.

1.4 XML-formaat

1.4.1 DTD

Het XML-formaat van de annotaties is gespecificeerd als de DTD in de figuren 1.1 en 1.2. Voor de automatisch geannoteerde delen van het corpus is nog een aantal verdere attributen voor *node* beschikbaar, die gedocumenteerd worden in hoofdstuk 7 op bladzijde 202.

De gebruikte attributen *rel*, *cat* en *word* representeren de dependentie-relatie, de categorie, en het woord in de input. De attributen *begin* en *end* representeren de begin- en eindpositie van de betreffende knoop in de zin. De positie voor het eerste woord in de zin is hierbij positie 0. Voor niet lexicale knopen geldt, dat de begin-positie gelijk is aan de meest linkse (laagste) begin-positie van de dochters; de eind-positie is de meest rechtse (hoogste) positie van de dochters. Gecoïndiceerde knopen hebben dezelfde begin- en eind-positie. Elke knoop heeft een unieke waarde voor het attribuut *id*, dat bedoeld is voor interne doeleinden.

De attributen *postag* en *lemma* representeren de POSTAG en LEMMA zoals in een afzonderlijk document gedocumenteerd (Van Eynde, 2005). De attributen *pt*

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!ELEMENT alpino_ds      (node, sentence, comments?)      >
<!ATTLIST alpino_ds
      version            NMTOKEN          #IMPLIED          >
<!ELEMENT node           (node*)          >
<!ATTLIST node
      rel      (hdf|hd|cmp|sup|su|obj1|pobj1|obj2|se|pc|vc|svp|predc
                |ld|me|predm|obcomp|mod|body|det|app|whd|rhd|cnj|crd|
                nucl|sat|tag|dp|top|mwp|dlink|--)          #REQUIRED
      cat      (smain|np|ppart|pp|ssub|inf|cp|du|ap|advp|ti|rel|whrel|
                whsub|conj|whq|oti|ahi|detp|sv1|svan|mwu|top) #IMPLIED

      begin    CDATA          #IMPLIED
      end      CDATA          #IMPLIED
      word     CDATA          #IMPLIED
      index    CDATA          #IMPLIED
      id       CDATA          #IMPLIED

      lemma    CDATA          #IMPLIED
      postag   CDATA          #IMPLIED

      pt       (let|spec|bw|vg|lid|vnw|tw|ww|adj|n|tsw|vz) #IMPLIED
      dial     (dial)          #IMPLIED
      ntype    (soort|eigen)   #IMPLIED
      getal    (getal|ev|mv)   #IMPLIED
      graad    (basis|comp|sup|dim) #IMPLIED
      genus    (genus|zijd|masc|fem|onz) #IMPLIED
      naamval  (stan|nomin|obl|bijz|gen|dat) #IMPLIED
      positie  (prenom|nom|postnom|vrij) #IMPLIED
      buiging  (zonder|met-e|met-s) #IMPLIED
      getal-n  (zonder-n|mv-n) #IMPLIED
      wvorm    (pv|inf|od|vd)  #IMPLIED
      pvtijd   (tgw|verl|conj) #IMPLIED
      pvagr    (ev|mv|met-t)   #IMPLIED
      numtype  (hoofd|rang)    #IMPLIED

```

Figuur 1.1: DTD voor Lassy dependentiestructuren (deel 1/2)

wvtype	(pr pers refl recip bez vb vrag betr excl aanw onbep)	#IMPLIED	
pdtype	(pron adv-pron det grad)	#IMPLIED	
persoon	(persoon 1 2 2v 2b 3 3p 3m 3v 3o)	#IMPLIED	
status	(vol red nadr)	#IMPLIED	
npagr	(agr evon rest evz mv agr3 evmo rest3 evf)	#IMPLIED	
lwtype	(bep onbep)	#IMPLIED	
vztype	(init versm fin)	#IMPLIED	
conjtype	(neven onder)	#IMPLIED	
spectype	(afgebr onverst vreemd deeleigen meta comment achter afk symb enof)	#IMPLIED	
root	CDATA	#IMPLIED	
pos	CDATA	#IMPLIED	
			>
<!ELEMENT	sentence	(#PCDATA)	>
<!ELEMENT	comments	(comment+)	>
<!ELEMENT	comment	(#PCDATA)	>

Figuur 1.2: DTD voor Lassy dependentiestructuren (deel 2/2)

... spectype representeren de afzonderlijke onderdelen van de POSTAG op redundante wijze, zodat ook makkelijk gezocht kan worden naar ondergespecificeerde POSTAG waardes. Dit wordt beschreven in sectie 1.4.3.

De attributen pos en root representeren de door Alpino gebruikte POSTAG en ROOT waardes. Deze worden hier niet afzonderlijk gedocumenteerd, en zijn geen officieel onderdeel van de annotatie.

1.4.2 Kanonieke dependentiestructuren

Zowel Alpino als de DtTrEd editor leveren Lassy dependentiestructuren op in het hierboven geformuleerde XML formaat. Bovendien worden deze XML bestanden door beide tools op een kanonieke wijze geformatteerd. Hierdoor krijgt een gegeven dependentiestructuur altijd een unieke representatie in XML. Het algoritme dat de kanonieke structuur berekent houdt onder andere rekening met:

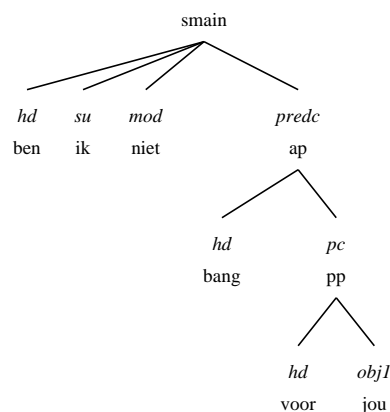
- De volgorde van dochterknopen binnen een domein. Deze volgorde volgt zo veel mogelijk de links-rechts volgorde in de zin
- In geval van gecoïndiceerde knopen bepaalt het algoritme welk voorkomen

van de knoop voluit wordt gespecificeerd; de andere voorkomens bevatten alleen een index attribuut waarmee naar de eerste knoop wordt verwezen. In de kanonieke representatie wordt het meest-linkse voorkomen van een gecoïndiceerde knoop voluit geschreven.

Omdat er een interactie bestaat tussen bovengenoemde twee uitgangspunten is het algoritme vrij ingewikkeld. Details zijn te vinden als commentaar in het python-script `Treebank/canonicalize/canonicalize.py`, en in het bestand `src/dt.pl` van de Alpino distributie. In het Lassy corpus worden de XML bestanden in deze kanonieke vorm opgeleverd.

Een consequentie van deze kanonieke representatie en het vóórkomen van discontinue constituenten is, dat de links-rechts volgorde in de resulterende dependentiestructuur meestal wel, maar niet altijd precies overeen komt met de volgorde in de zin:

(3) bang ben ik niet voor jou



Dit leidt niet tot problemen omdat de links-rechts volgorde in de dependentiestructuur voor elke knoop vastgelegd is door de begin en end attributen.

1.4.3 Representatie van POSTAG en LEMMA

Zoals hierboven al genoemd representeren de XML attributen `postag` en `lemma` de POSTAG en LEMMA annotaties die in een afzonderlijk document zijn gedocumenteerd (Van Eynde, 2005). De representatie van POSTAG waarden is bovendien uitgebreid met extra XML attributen om het gemakkelijk te maken te zoeken naar niet volledig gespecificeerde POSTAG waarden. In deze sectie geven we aan hoe de POSTAG waarden met behulp van XML attributen worden gespecificeerd.

De POSTAG waardes hebben steeds de vorm $POS(val1, val2, \dots)$. Deze volledige POSTAG wordt gerepresenteerd met het XML attribuut `postag`. De POS waarde wordt apart gerepresenteerd als waarde van het XML attribuut `pt`, waarbij de waarde bovendien in kleine letters wordt geschreven. Elk van de waardes $val1, val2, \dots$ wordt vervolgens gerepresenteerd als waarde van een XML attribuut. Als voorbeeld, de POSTAG waarde $VNW(aanw, det, stan, nom, met-e, mv-n)$ wordt in XML met de volgende attribuut-waarde paren gerepresenteerd:

```
postag="VNW(aanw, det, stan, nom, met-e, mv-n) "  
pt="vnw"  
vwtype="aanw"  
pdtype="det"  
naamval="stan"  
positie="nom"  
buiging="met-e"  
getal-n="mv-n"
```

De naam van het betreffende XML attribuut is direct afgeleid uit ([Van Eynde, 2005](#)) en geven we hieronder, wellicht ten overvloede, als omgekeerde sleutel van waarde naar attribuut.

Elke waarde $val1, val2, \dots$ kan deterministisch herleid worden tot het bijbehorende XML attribuut, met behulp van onderstaande tabel. In de meeste gevallen is de mapping van waarde naar attribuut voor elke POS hetzelfde, maar in sommige gevallen moet naar de waarde van POS gekeken worden. In die gevallen staat de POS hieronder genoemd. Voor de waarde `mv` moet bovendien bij voornaamwoorden nog verder gekeken worden of het als lidwoorden gebruikte onbepaalde voornaamwoorden betreft.

1.5 Relatie met CGN syntactische annotatie

1.5.1 Illustraties

De illustraties in dit document zien er anders uit dan die in CGN, maar de verschillen zijn vooral cosmetisch.

De zogenaamde *secondary edges* in de CGN annotaties worden in dit document aangegeven met co-indexatie; daarnaast is de volgorde van woorden in boom niet noodzakelijkerwijs de volgorde van de woorden in de zin. De positie van elk woord in de zin wordt expliciet gemaakt met het XML attribuut *begin* (waarbij het eerste woord op positie 0 begint). Het voordeel van deze twee beslissingen is

Waarde	POS	Attribuut	Waarde	POS	Attribuut
1		persoon	met-s		buiging
2		persoon	met-t		pvagr
2b		persoon	meta		spectype
2v		persoon	mv	N,VNW,LID	getal
3		persoon	mv	VNW(onbep,det,...)	npagr
3m		persoon	mv	WW	pvagr
3o		persoon	mv-n		getal-n
3p		persoon	nadr		status
3v		persoon	neven		conjtype
aanw		vwtype	nom		positie
achter		spectype	nomin		naamval
adv-pron		pdtype	obl		naamval
afgebr		spectype	od		wvorm
afk		spectype	onbep	LID	lwtype
agr		npagr	onbep	VNW	vwtype
basis		graad	onder		conjtype
bep		lwtype	onverst		spectype
betr		vwtype	onz		genus
bez		vwtype	pers		vwtype
bijz		naamval	persoon		persoon
comment		spectype	postnom		positie
comp		graad	pr		vwtype
conj		pvtijd	prenom		positie
dat		naamval	pron		pdtype
deeleigen		spectype	pv		wvorm
det		pdtype	rang		numtype
dial		dial	recip		vwtype
dim		graad	red		status
eigen		ntype	refl		vwtype
enof		spectype	rest		npagr
ev	N	getal	rest3		npagr
ev	WW	pvagr	soort		ntype
evf		npagr	stan		naamval
evmo		npagr	sup		graad
evon		npagr	symb		spectype
evz		npagr	tgw		pvtijd
excl		vwtype	vb		vwtype
fem		genus	vd		wvorm
fin		vztype	verl		pvtijd
gen		naamval	versm		vztype
genus		genus	vol		status
getal		getal	vrag		vwtype
grad		pdtype	vreemd		spectype
hoofd		numtype	vrij		positie
inf		wvorm	zijd		genus
init		vztype	zonder		buiging
masc		genus	zonder-n		getal-n
met-e		buiging			

Figuur 1.3: Codering van kenmerken van POSTAG door XML attributen

dat er niet langer een noodzaak bestaat tot (zogenaamde) bomen met kruisende takken. Formeel gezien zijn beide representaties overigens equivalent.

Net zoals in CGN worden onderdelen van de zin die niet tot de eigenlijke dependentiestructuur worden gerekend (zoals punctuatie) geannoteerd met de speciale - - relatie. Alle knopen worden vervolgens aangehecht aan de root knoop van de dependentiestructuur die categorie *top* heeft. Dit ziet er wat anders uit dan in de CGN visualisaties, maar opnieuw betreft dit louter een cosmetisch verschil.

1.5.2 XML formaat

Het XML formaat waarin de annotaties van Lassy worden opgeslagen verschilt van de XML representatie die in CGN werd gebruikt. In de Lassy representatie worden de syntactische boomstructuren ook in XML als boomstructuren gerepresenteerd. Het grote voordeel hiervan is, dat de XML standaard zoektaal XPATH direct gebruikt kan worden voor het zoeken naar specifieke syntactische constructies in het corpus. Hierdoor hoeft geen specifieke corpusexploitatie software te worden ontwikkeld, maar kan direct worden gebruikgemaakt van bestaande XML toolboxes.

Een ander voordeel van deze beslissing is, dat er keuzemogelijkheden ontstaan met betrekking tot de software die gebruikt kan worden om de annotaties te bewerken. Eerder gebruikten we hiervoor de Thistle editor, maar inmiddels gebruiken we Tred (Pajas and Štěpánek, 2008), de editor die ontwikkeld werd bij de totstandkoming van de Prague Dependency Treebank.

Merk overigens op dat beide formaten automatisch in elkaar zijn over te zetten.

1.5.3 Verschillen in de annotatie zelf

Een belangrijke inhoudelijke uitbreiding van de Lassy annotaties in vergelijking met CGN betreft de in D-Coi voorgestelde annotaties van begrepen complementen in werkwoordclusters. In CGN werd voor werkwoorden normaliter alleen bij een persoonsvorm een subjectrelatie geannoteerd. In Lassy wordt de informatie toegevoegd dat bijvoorbeeld in geval van hulpwerkwoorden of modale werkwoorden het onderwerp van de zin thematisch gezien ook fungeert als het onderwerp van het hoofdwerkwoord. Daarnaast verschilt de annotatiepraktijk van Lassy op sommige detailpunten van de annotaties zoals die in het CGN manual werden voorgeschreven. Deze verschillen bespreken we nu als volgt.

Volledigere specificatie van thematische structuur bij werkwoorden In Lassy wordt de informatie toegevoegd dat bijvoorbeeld in geval van hulpwerk-

woorden of modale werkwoorden het onderwerp van de zin thematisch gezien ook fungeert als het onderwerp van het hoofdwerkwoord. Op vergelijkbare wijze wordt de thematische structuur bij hulpwerkwoorden van de lijdende vorm compleet gemaakt door het onderwerp van het hulpwerkwoord te markeren als lijdend voorwerp. Ook voor AcI werkwoorden en controlewerkwoorden wordt de thematische structuur op deze wijze gecompleteerd. Deze informatie is vrij nauwkeurig automatisch toe te kennen en hierdoor levert dit weinig extra annotatiewerk op. De relevante voorbeelden worden besproken in sectie 2.1.9 op bladzijde 44.

Aanhechting van MOD elementen in het werkwoordelijke domein In CGN werden hiervoor geen duidelijke richtlijnen gespecificeerd. In de Lassy annotatie worden bepalingen van type MOD bij voorkeur aan het hoofdwerkwoord gerelateerd. Bepalingen van type PREDM (die vaak ook betrekking hebben op het onderwerp van de zin) worden bij voorkeur aangehecht op het niveau van de persoonsvorm. De relevante voorbeelden worden besproken in sectie 2.1.14.4 op bladzijde 72.

Partitieven In de Lassy annotaties wordt niet langer gebruik gemaakt van de PART dependency label. Een zin zoals

(4) één van de drie kwam op tijd

wordt geanalyseerd met *één* als hoofd, en de PP *van de drie* fungeert als MOD bij dit hoofd. Een argument voor deze analyse is de overeenstemming in getal met de persoonsvorm die aangeeft dat één hier als hoofd optreedt (zie ook paragraaf 2.5.2.3 op bladzijde 127).

Prt In CGN werd voor de zogenaamde ‘modale partikels’ of ‘oordeelspartikels’ het label PRT gebruikt. Bovendien werden bij elkaar behorende groepjes van zulke partikels samengenomen zonder dat een afhankelijkheidsrelatie tussen de bijwoorden werd verondersteld, en zonder dat deze groepjes een hoofd hadden. In de huidige annotatie fungeert elk van de partikels als MOD bij het hoofd. Het label PRT wordt niet langer gebruikt. Zie ook paragraaf 2.1.14.3 op bladzijde 72.

HD vertakt nooit In de Lassy annotatie wordt strikt de hand gehouden aan de eis dat het hoofd van een woordgroep lexicaal verankerd moet zijn (een dependentiedomein is lexicaal verankerd, als het hoofd een ongelede dependentiestructuur is, of als het hoofd een multi-word-unit is). In CGN werd sporadisch gewerkt met hoofden die zelf verdere structuur bevatten. In Lassy gebeurt dit niet. Zie voor verdere uitleg bladzijde 12.

Sbar complementen als VC Finiete en niet-finiete zinscomplementen krijgen in de Lassy annotatie steeds het label VC. In de CGN annotatie werd hiervoor soms de OBJ1 en soms de VC relatie gebruikt – afhankelijk van de aanwezigheid van een NP die als OBJ1 fungeert. Zie verder sectie 2.1.9 op bladzijde 44.

Werkwoordelijke uitdrukkingen Net als in CGN wordt het label SVP gebruikt voor vaste delen van een werkwoordelijke uitdrukking. In de huidige annotatie wordt zo'n vast deel van een werkwoordelijke uitdrukking vaak niet verder geanalyseerd, maar als multi-word-unit behandeld – tenzij het vaste deel modificatie toestaat. Zie ook sectie 2.1.13.2 op bladzijde 62.

Lijsten In de Lassy annotatie wordt geen onderscheid gemaakt tussen lijsten enerzijds en conjuncties zonder coordinaties anderzijds. Lijsten worden geanalyseerd als een hoofdloze conjunctie. Het label LP en de categorie LIST wordt niet meer gebruikt. Voorbeelden geven we in sectie 4.5 op bladzijde 183.

1.6 XPATH in plaats van TigerSearch

1.6.1 Inleiding

Omdat de XML representatie van Lassy anders is dan de representatie die voorheen in CGN werd gebruikt, moet ook een beroep worden gedaan op andere gereedschappen om die representaties te bekijken en daar in te zoeken. Omdat de representaties gebruik maken van XML zijn de standaard XML oplossingen toepasbaar, zoals met name XSLT, XPATH, XQUERY. In deze sectie geven we enkele voorbeelden voor het zoeken met behulp van XPATH zoekpatronen. XPATH wordt bijvoorbeeld gebruikt in DtSearch en DACT, twee hulpprogramma's die voor Lassy data beschikbaar zijn.

De XPATH zoekpatronen kunnen verwijzen naar hiërarchische structuur, de links-rechts volgorde, en naar categoriale informatie. Een aantal voorbeelden zal verduidelijken hoe dit in zijn werk kan gaan. Voor een volledige beschrijving van XPATH of voor een XPATH tutorial verwijzen we naar het web waar deze in overvloed beschikbaar zijn. De volgende query:

```
//node[@cat="pp"]
```

identificeert alle constituenten (weergegeven als een XML node element) waarvoor geldt dat de categorie gelijk is aan de waarde pp. In de praktijk zal deze

query dus alle zinnen opleveren waarin een PP-constituent voorkomt. In deze query wordt de dubbele slash gebruikt om aan te geven dat we een node element willen vinden, op een willekeurige plek in de XML structuur. Vervolgens worden binnen de vierkante haken de condities genoemd waar deze knoop aan moet voldoen. Deze condities zullen vaak verwijzen naar de attributen van het node element, zoals we in dit voorbeeld deden met het cat attribuut.

Conditie kunnen worden gecombineerd met conjunctie (and), disjunctie (or) en negatie (not). We kunnen bijvoorbeeld op de volgende manier eisen dat de pp aan het begin van de zin staat, door een extra conditie toe te voegen:

```
//node[@cat="pp" and @begin="0"]
```

Gewone haakjes kunnen hierbij de gewenste structuur van de vraag expliciteren, en worden altijd gebruikt voor negatie, zoals in het volgende voorbeeld:

```
//node[(@cat="pp" or @cat="advp") and not(@begin="0")]
```

De binnen de vierkante haken opgesomde condities kunnen ook verwijzen naar de context van het node element, bijvoorbeeld naar de dochters van de geïdentificeerde knoop.

```
//node[@cat="pp" and node[@rel="hd" and not(@pt="vz")]]
```

Deze query geeft alle zinnen waarin een PP voorkomt met een hoofd waarvoor geldt dat het geen POSTAG heeft van de vorm VZ(. .). Dit levert een flink aantal hits, waarbij in het merendeel van de gevallen het hoofd een multi-word-unit is. Omdat een multi-word-unit geen POSTAG waarde heeft, zal er ook geen waarde voor pt zijn waardoor aan de eis wordt voldaan. Om in te zoomen op voorbeelden waarbij het hoofd wel lexicaal is, zou de query als volgt kunnen worden uitgebreid:

```
//node[@cat="pp" and  
      node[@rel="hd" and @word and not(@pt="vz")]]
```

Dit vereist dat het hoofd een word attribuut heeft, ongeacht welke waarde.

Om naar de links-rechts volgorde van knopen te kunnen verwijzen kan gebruik worden gemaakt van de begin en end attributen, en de operatoren <, > en =. Als voorbeeld geven we de volgende query die PP's identificeert waarbij het hoofd voorafgegaan wordt door de obj1 dependent. Voor sommige implementaties van

XPATH is het noodzakelijk om de waarden van het begin en end attribuut expliciet te converteren van string naar getal met behulp van de `number()` functie. Dit doen we hier in de voorbeelden steeds:

```
//node[@cat="pp" and
      node[@rel="hd"]/number(@begin)
      > node[@rel="obj1"]/number(@begin)]
```

De operator `=` kan worden gebruikt om te eisen dat een constituent direct vooraf moet gaan aan een andere constituent. Stel dat we zinnen willen identificeren waarbij een PP direct volgt op de persoonsvorm in een bijzin van type `ssub`. Onze eerste poging levert:

```
//node[@cat="ssub" and
      node[@rel="hd"]/number(@end)
      = node[@cat="pp"]/number(@begin)]
```

Dit geeft inderdaad bijzinnen waarbij de persoonsvorm direct gevolgd wordt door een PP. De query vereist echter ook, dat deze PP een dependent is van de persoonsvorm (in andere woorden, het moet ook een dochterknoop van de `ssub` categorie zijn). Om een willekeurige PP toe te staan kunnen we ook ingebed gebruik maken van de dubbele slash:

```
//node[@cat="ssub" and
      node[@rel="hd"]/number(@end)
      = //node[@cat="pp"]/number(@begin)]
```

Het is ook mogelijk om condities aan hoger gelegen knopen te stellen. De volgende query identificeert PP's die een dependent binnen een hoofdzin zijn:

```
//node[@cat="pp" and ../@cat="smain"]
```

Merk ten slotte het verschil op tussen de notatie `//` en `../`. De eerste notatie refereert aan een willekeurige afstammeling van de topknoop, terwijl de tweede notatie refereert aan een willekeurige afstammeling van de huidige knoop. Voorbeelden:

```
//node[@cat="pp" and //node[@root="op"]]
```

Deze query zal met alle PP knopen in een boom matchen, zolang ergens in die boom een knoop voorkomt met als root-vorm op. De query

```
//node[@cat="pp" and ../node[@root="op"]]
```

matcht met elke PP die een knoop bevat (eventueel diep ingebed) met root-vorm op.

1.6.2 Meer voorbeelden

In deze sectie geven we nog een aantal voorbeelden van queries die wellicht inspirerend kunnen werken. Een aantal van deze queries zijn gebruikt om *verdachte* handmatige annotaties nog eens extra te controleren.

```
//node[node[@getal="mv" and @rel="hd"] and  
      node[@rel="det" and (@root="het" or  
                          @root="een" or  
                          @root="dit" or  
                          @root="dat")]]
```

Deze query vindt meervoudige hoofden die combineren met één van de determiners een, het, dit, of dat.

```
//node[@vwtype="betr" and not(@rel="rhd" or ../@rel="rhd")]
```

De query vindt zinnen waarbij een relatief pronomen niet als rhd dependent wordt gebruikt.

```
//node[@rel="rhd" and @word and  
      not(@vwtype="betr" or  
          @vwtype="vb" or  
          @pt="bw")]
```

De query vindt zinnen met een rhd dependent die niet een verwachte postag waarde heeft.

```
//node[node[@rel="hd" and @pt and not(@pt="ww")] and  
      node[@rel="su"]]
```

Een knoop met een onderwerp die niet een werkwoord als hoofd heeft.


```
//node[@wvorm="pv" and not(../@cat="smain" or
    ../@cat="ssub" or
    ../@cat="sv1" or
    ../@cat="mwu" or
    ../@cat="top" or
    ../@cat="du") ]
```

Een persoonsvorm die niet optreedt als hoofd van een zin.

1.6.3 Valkuilen

Gecoïndiceerde knopen worden in de XML-structuur slechts eenmaal voluit gegeven, en daarna wordt er door middel van het XML-attribuut `index` aangegeven dat de knopen identiek geacht worden te zijn. Dit levert soms beperkingen op bij het gebruik van XML. Als voorbeeld verwijzen we hier vooruit naar de zin je zal wel moe zijn, zoals gegeven in voorbeeld 5.

De query:

```
//node[node[@rel="hd" and @word="zijn"] and
    node[@rel="su" and @word="je"]]
```

zal in dit geval *niet* matchen, omdat het onderwerp van de infinitief zijn hier geen eigen attributen heeft, maar slechts een attribuut `index` dat verwijst naar hetzelfde attribuut voor het onderwerp van de hoofdzin.

De volgende aangepaste query geeft aan hoe je in zulke gevallen soms toch verder kunt:

```
//node[node[@rel="hd" and @word="zijn"] and
    node[@rel="su" and
        (@word="je" or
         @index=//node[@word="je"]/@index)]
```

Hier wordt de query uitgebreid met een alternatieve eis voor het onderwerp: ofwel het onderwerp heeft als woord je, ofwel er is een knoop met dezelfde index als het onderwerp, en die knoop is het woord je.

Dit voorbeeld toont aan dat gecoïndiceerde knopen niet direct problematisch zijn voor XPATH, maar dat ze soms wel leiden tot veel ingewikkeldere queries, en dat je erop bedacht moet zijn dat het voorkomen van gecoïndiceerde knopen kan leiden tot het wellicht onverwacht missen van een aantal relevante gevallen.

- Q1. Vind zinnen waarin het woord 'zag' voorkomt.
- Q2. Vind zinnen waarin het woord 'zag' niet voorkomt.
- Q3. Vind NP waarvan het meest rechtse dochter een zelfstandig naamwoord is.
- Q4. Vind SMAIN met daarin een werkwoord dat onmiddellijk voorafgaat aan een NP, die op zijn beurt onmiddellijk voorafgaat aan een PP
- Q5. Vind de eerste gemeenschappelijke voorouder van NP, PP sequenties (de NP en de PP hoeven niet noodzakelijk siblings te zijn)
- Q6. Vind een NP die het woord 'donker' domineert, waarbij er een tussenliggende PP categorie is
- Q7. Vind een NP die door een SMAIN wordt gedomineerd. Geef alleen de NP terug.

Figuur 1.4: Zeven voorbeeld queries, vergelijkbaar met de queries van Lai en Bird (2004).

1.6.4 De voorbeelden van Lai en Bird

In (Lai and Bird, 2004) worden aan de hand van zeven voorbeelden een aantal bestaande *query languages* voor het zoeken in syntactisch geannoteerde corpora met elkaar vergeleken. We laten hier zien dat de zeven voorbeelden (vertaald naar het Nederlands en de in Lassy gebruikte ontologie) met weinig moeite met behulp van XPATH kunnen worden geformuleerd.

De voorbeelden van Lai en Bird (2004) werden gebruikt om een aantal bestaande *query languages* voor het zoeken in syntactisch geannoteerde corpora met elkaar te vergelijken. In een recent paper (Lai and Bird, 2010) stellen dezelfde auteurs een nieuwe query taal voor, LPATH, die queries kan uitdrukken die in XPATH niet mogelijk zijn. LPATH is een uitbreiding van XPATH op drie fronten:

- de extra *immediate following* as
- de scope operator { . . }
- de node alignment operatoren ^ en \$

Het blijkt dat alle voorbeeld queries die met deze uitbreidingen mogelijk zijn ook in XPATH kunnen worden gedefinieerd, onder de aanname dat de links-rechts

- Q1.** `//node[@word="zag"]`
- Q2.** `not(//node[@word="zag"])`
- Q3.** `//node[@cat="np" and node[@pt="n"]/number(@end)=number(@end)]`
- Q4.** `//node[@cat="smain" and node[@pt="ww"]/number(@end)
= node[@cat="np"
and number(@end)
= following-sibling::node[@cat="pp"]/number(@begin)
]/number(@begin)]`
- Q5.** `//node[./node[@cat="np"]/number(@end) =
./node[@cat="pp"]/number(@begin) and
not(node[./node[@cat="np"]/number(@end) =
./node[@cat="pp"]/number(@begin))]`
- Q6.** `//node[@cat="np" and
./node[@cat="pp" and ./node[@word="donker"]]]`
- Q7.** `//node[@cat="smain"]/node[@cat="np"]`

Figuur 1.5: XPATH code voor de zeven voorbeeld queries, vergelijkbaar met de queries van Lai en Bird (2004).

volgorde van de knopen ook door middel van XML attributen wordt gespecificeerd. Dit is in de LASSY representatie het geval (ook al omdat in LASSY de structuren niet altijd bomen zijn, maar grafen). Een LPATH query die gebruik maakt van de *immediate following* door te eisen dat een knoop x direct vooraf gaat aan y kan in XPATH worden gespecificeerd door te eisen dat het end attribuut van knoop x gelijk is aan het begin attribuut van knoop y . Ook de gegeven voorbeelden die het gebruik van de scope operator en de node alignment motiveren kunnen worden gespecificeerd in XPATH zolang de begin en end attributen de links-rechts volgorde specificeren. Bijvoorbeeld, de LPATH query

```
//node[@cat="smain"]{//node[@cat="np"]$}
```

waarbij een smain knoop wordt gezocht die een rechts-gealigneerde NP knoop bevat, kan in XPATH zo worden gedefinieerd:

```
//node[@cat="smain" and
      number(@end)
      = ../node[@cat="np"]/number(@end)]
```

Daarom trekken we de conclusie dat XPATH voor XML representaties die verrijkt zijn met attributen die de links-rechts volgorde van de constituenten expliciet maken voldoende krachtig is, en er dus geen reden is om te kiezen voor een niet-standaard formalisme zoals LPATH.

1.6.5 Voorbeelden met XPATH 2.0

Het gebruik van XPATH voor het zoeken in dependentiestructuren wordt bekritiseerd in het proefschrift van Gerlof Bouma (([Bouma, 2008](#))) vanwege het ontbreken van de mogelijkheid tot het definiëren van macro's (macro's worden ondersteund in DACT), en het ontbreken van de mogelijkheid om naar specifieke knopen te verwijzen om er later naar terug te kunnen verwijzen. De laatstgenoemde beperking is een beperking van XPATH 1.0. De nieuwere XPATH 2.0 standaard kent gekwantificeerde uitdrukkingen die deze functionaliteit bieden.

Als voorbeeld geven we de volgende query waarbij we zoeken naar een NP die een VP complement bevat, zodanig dat er ergens in de structuur een noun is die voorafgegaan wordt door het hoofd van de NP, maar voorafgaat aan het VP complement. Hierbij gebruiken we de mogelijkheid voor gekwantificeerde expressies die in XPATH 2.0 geïntroduceerd is:

```
//node[@cat="np" and
```

```

    ( some $tussen in //node[@pos="noun"]
satisfies (  $tussen/number(@begin)
             < node[@rel="vc"]/number(@begin) and
             $tussen/number(@end)
             > node[@rel="hd"]/number(@end)
             )
    )
]

```

Overigens lijkt voor het vinden van *vorfeld*-constituenten de extra kracht van XPATH 2.0 niet noodzakelijk. De definitie op bladzijde 76 van (Bouma, 2008) kan in XPATH 1.0 ook als volgt worden gerealiseerd (in dit voorbeeld beperken we de notie *key dependent* tot knopen met de relatie *hd* or *cmp* - dit kan natuurlijk eenvoudig uitgebreid worden):

```

//node[( ancestor::node[@cat="smain"/
           node[@rel="hd"]/number(@begin)
         > node[@rel="hd" or @rel="cmp"]/number(@begin)
or ancestor::node[@cat="smain"/
           node[@rel="hd"]/number(@begin)
         > number(@begin) and @pos
        )
and not ( ancestor::node[ (
           ancestor::node[@cat="smain"/
             node[@rel="hd"]/number(@begin)
           > node[@rel="hd" or @rel="cmp"]/number(@begin)
or ancestor::node[@cat="smain"/
             node[@rel="hd"]/number(@begin)
           > number(@begin) and @pos
         )
        ]
        )
]
]

```

Hoofdstuk 2

Lokale dependentiestructuren

We bespreken achtereenvolgens dependentiedomeinen met werkwoorden, bijvoeglijke naamwoorden, voorzetsels, zelfstandige naamwoorden en bijwoorden als hoofd. Onderschikkende voegwoorden worden besproken in de sectie over werkwoorden. Nevenschikkende voegwoorden komen in een later hoofdstuk (hoofdstuk 4 op bladzijde 175) aan de orde.

Voor de verschillende domeinen gebruiken we onder andere de dependentielabels HD (hoofd) en MOD (bepaling). Deze dependentielabels krijgen dus een ‘context-afhankelijke’ interpretatie: de POS-tag informatie van het hoofd bepaalt de moederknoop van het domein in kwestie (NP, AP, ADVP, PP, of een van de werkwoordelijke projecties van figuur 2.2 op bladzijde 31); het categorielabel van de moederknoop, op zijn beurt, maakt meestal duidelijk of MOD een bijvoeglijke dan wel bijwoordelijke bepaling is.

Ook voor de dependentielabels van complementen maken we van deze contextgevoelige interpretatie gebruik. Zo is OBJ1 het meest prominente argument in werkwoordelijke zowel als naamwoordelijke domeinen (‘lijdend voorwerp’ bij een verbaal hoofd, ‘oorzakelijk voorwerp’ bij een adjectief-hoofd, etc.).

2.1 Het verbale domein

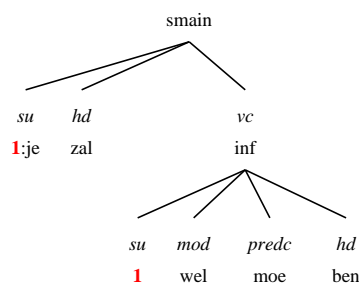
2.1.1 Werkwoordelijk hoofd

We onderscheiden finiete en niet-finiete verbale domeinen. Het werkwoord (finiete dan wel niet-finiete vorm) krijgt in beide gevallen het dependentielabel HD. De combinatie van het dependentielabel en de kenmerk-informatie in het C-label differentieert tussen de finiete en de niet-finiete domeinen.

	dependentielabel	OMSCHRIJVING
<i>hoofd</i>	HD	werkwoordelijk hoofd (finiet of niet-finiet)
<i>complement</i>	SU	subject, onderwerp
	SUP	voorlopig subject
	OBJ1	direct object, lijdend voorwerp
	POBJ1	voorlopig direct object
	SE	verplicht reflexief object
	OBJ2	secundair object (meewerkend, ...)
	PREDC	predicatief complement
	VC	verbaal complement
	PC	voorzetselvoorwerp
	ME	maat (duur, gewicht, ...) complement
	LD	locatief of directioneel complement
	SVP	scheidbaar deel van werkwoord
<i>modificatie</i>	MOD	bijwoordelijke bepaling
	PREDM	bepaling van gesteldheid 'tijdens de handeling'

Figuur 2.1: dependentielabels: verbale domeinen

(5) je zal wel moe zijn



2.1.2 Categorielabels

In Figuur 2.2 geven we de categorielabels die geprojecteerd worden vanuit (finiete dan wel niet-finiete) werkwoordelijke hoofden. Bij de categorielabels voor finiete zinnen (SSUB, SV1, SMAIN) beperken we ons tot zuiver vormelijke onderscheidingen in termen van de positie van de persoonsvorm (respectievelijk: finaal, initieel of op de tweede plaats).

categorielabel	OMSCHRIJVING
SMAIN	declaratieve zin (werkwoord op tweede plaats)
SSUB	bijzin (V finaal)
SV1	werkwoordsinitiële zin (ja/nee-vraag, bevel)
INF	kale infinitiefgroep
TI	infinitiefgroep met <i>te</i>
OTI	infinitiefgroep met <i>om te</i>
PPART	voltooid/passief-deelwoordgroep
PPRES	tegenwoordig-deelwoordgroep

Figuur 2.2: categorielabels: moederknopen van verbale domeinen

2.1.2.1 Declaratieve zin

Hoofd is de persoonsvorm (met label HD). In declaratieve zinnen komt de persoonsvorm normaliter op de tweede plaats.

- (6) het waait
- (7) soms regent het
- (8) dat vind ik niet lekker
- (9) aldus geschiedde

Merk op dat constituentvragen als *wanneer vertrekt de trein naar Brussel* een ander label krijgen: zie sectie 3.3 op bladzijde 168.

In sommige exclamatieven staat het verbogen werkwoord eveneens op de tweede plaats:

- (10) hij heeft me toch een mooie boeken gekocht!
- (11) in die kast staan toch een mooie boeken!

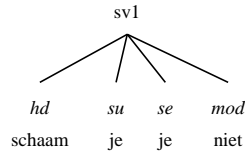
Syntactisch onderscheiden deze uitroepen zich in geen enkel opzicht van mededelingen; in de syntactische annotatie krijgen ze dan ook geen bijzondere behandeling.

2.1.2.2 Werkwoordsinitiële zin

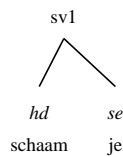
Het label SV1 is voor finiete werkwoordelijke domeinen waarbij de persoonsvorm (HD) op de eerste plaats komt. Het betreft hier vooral polaire vragen (ja/nee-

vragen), directieven (bevelen met een imperatieve vorm van het werkwoord) en ‘voegwoordloze voorwaardelijke bijzinnen’.

(12) schaam je je niet (polaire vraag)



(13) schaam je (bevel)



(14) val dood (bevel)

(15) houd je van vlees, braad je in Croma (voegwoordloze voorwaardelijke bijzin)

Ook topic-drop-zinnen als

(16) doen we

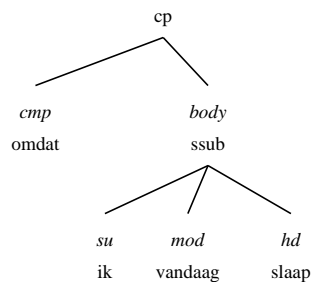
(17) heb ik al gedaan

worden, op grond van de vorm, gecategoriseerd als SV1.

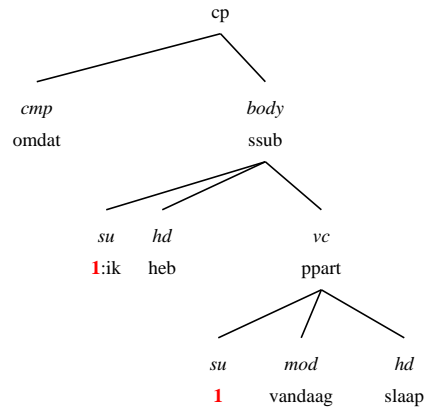
2.1.2.3 Overige categorieën

Finiete bijzinnen krijgen het categorielabel SSUB. Projecties van voltooid deelwoorden krijgen het label PPART, en tegenwoordig-deelwoordgroepen krijgen het label PPRES. De categorie van beknopte bijzinnen met *te* en *om te* wordt besproken in sectie 2.2 op bladzijde 76. Infinitiefgroepen krijgen de categorie INF.

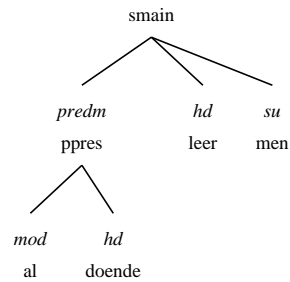
(18) omdat ik vandaag slaap



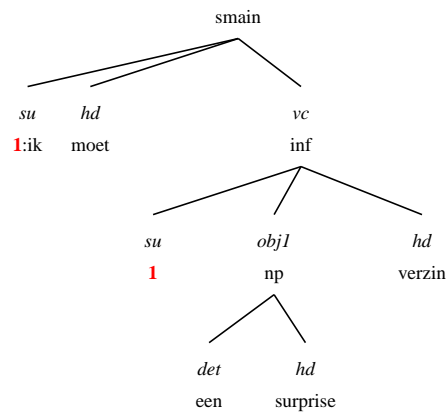
(19) omdat ik vandaag heb geslapen



(20) al doende leert men



(21) Ik moet een surprise verzinnen



2.1.3 Onderwerp, voorlopig onderwerp

De dependentierol SU (onderwerp) komt vaak voor in combinatie met een tempusdragend werkwoordelijk hoofd, i.e., een persoonsvorm.

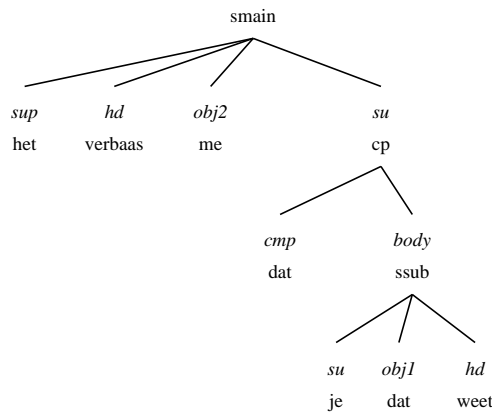
Bij niet-finiete werkwoordelijke hoofden komt de SU rol soms ook voor, in geval

van controle en raising constructies. In die gevallen is de SU gecoïndiceerd.

De dummies ‘het’ en ‘t’ kunnen als voorlopig onderwerp (SUP) fungeren als het inhoudelijke onderwerp niet zijn kanonieke zinsinitiële positie inneemt. De voorlopige onderwerpen zijn gereduceerde, semantisch lege vormen.

(22) **je** verbaast me (SU)

(23) **het** verbaast me **dat je dat weet** (SUP+SU)



2.1.3.1 Plaatsonderwerp

Ingewikkelder zit het met *er*, dat de positie van het onderwerp kan innemen in zinnen zonder onderwerp (Haeseryn et al. 1997:471):

(24) er wordt hard gereden op de snelweg

(25) er is niets te zien

(26) er valt niets te veranderen

(27)

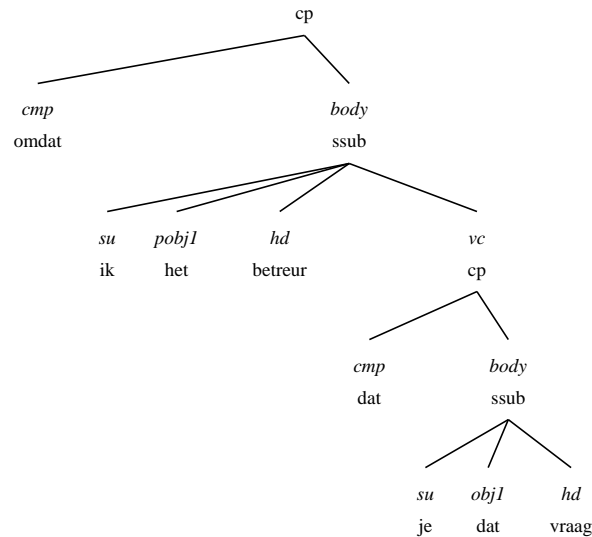
In deze gevallen wordt *er* als MOD geannoteerd.

2.1.4 Direct object, voorlopig direct object

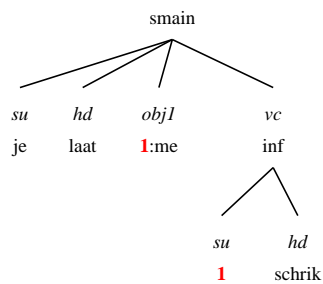
In werkwoordelijke domeinen wordt het dependentielabel OBJ1 toegekend aan het complement dat in het passief in de SU rol verschijnt, en aan het accusatief object in de ‘accusativus cum infinitivo’-constructie (bij causatief *doen*, *laten*, en perceptiewerkwoorden *zien*, *horen*, *voelen*, ...), waar passivering niet mogelijk is. Net als het onderwerp, kan ook de OBJ1-positie door een voorlopig lijdend voorwerp (POBJ1) worden ingenomen.

(28) ik betreur **dit** (OBJ1)

(29) (omdat) ik **het** betreur **dat je dat vraagt** (POBJ1 + VC)



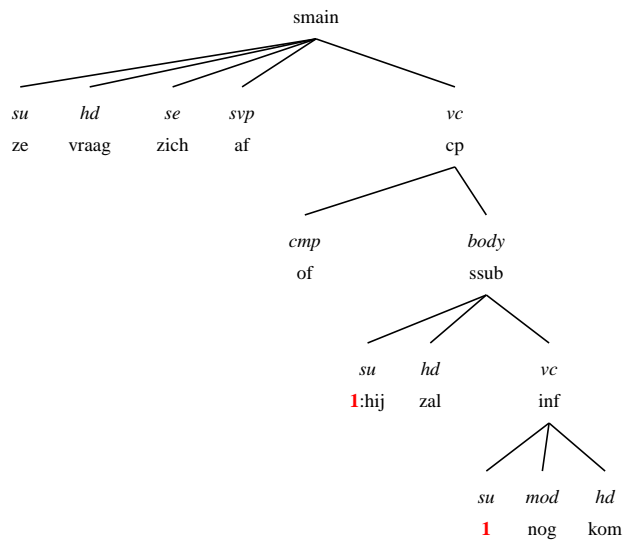
(30) je liet **me** schrikken (OBJ1)



2.1.5 Verplicht reflexief object

Verplicht reflexieve objecten krijgen het label SE. Criterium voor toekenning van het label SE: het is niet mogelijk om, met behoud van betekenis voor het werkwoordelijke hoofd, het SE gelabelde element te vervangen door een volledige NP. Verplichte reflexieve objecten kunnen ook heel gemakkelijk voorafgaan aan het onderwerp. Vergelijk:

(31) ze vroeg (zich | *de jongen) af of hij nog zou komen



- (32) we schaamden (ons | *de jongens)
- (33) hij voelt (zich | *hem) tevreden
- (34) hij wast (zich | zijn voeten) (geen SE)
- (35) in die wijk bevindt zich (SE) de grootste universiteit van het land

OPMERKING. We gebruiken het label SE ook bij niet-verbale hoofden, zoals in

- (36) hij houdt het voor (zich | *de jongens)
- (37) hij is (zich | *de jongens) bewust van de consequenties

2.1.6 Secundair object: meewerkend, belanghebbend, ondervindend voorwerp

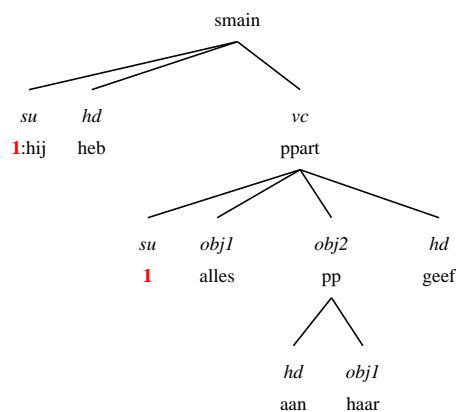
Het is niet altijd gemakkelijk meewerkend, belanghebbend en ondervindend voorwerp van elkaar te onderscheiden.

- (38) hij geeft **haar** alles
- (39) ze schenkt **hem** een borrel in
- (40) het verbaast **me** dat je nog aan het werk bent
- (41) hij rookt **me** te veel

Om de annotatie werkbaar te houden, gebruiken we een overkoepelend label OBJ2 (secundair object), dat dan alleen van direct object (OBJ1) onderscheiden moet worden.

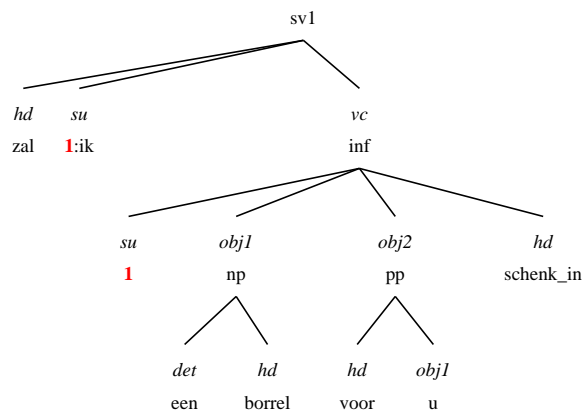
Het OBJ2 label wordt zowel gebruikt voor NP complementen, als ook PP complementen, waarbij het voorzetsel meestal *aan* of *voor* is. Deze analyse wordt gekozen, indien duidelijk is dat de PP een complement is. We beschouwen de PP als complement, indien de PP ook als NP kan worden gerealiseerd met ongeveer dezelfde betekenis. Voorzetselgroepen met *ten bate van*, *ten behoeve van*, *ten gerieve van*, *ten voordele van*, *in het belang van* annoteren we normaal gesproken als MOD.

(42) hij heeft alles **aan haar** gegeven



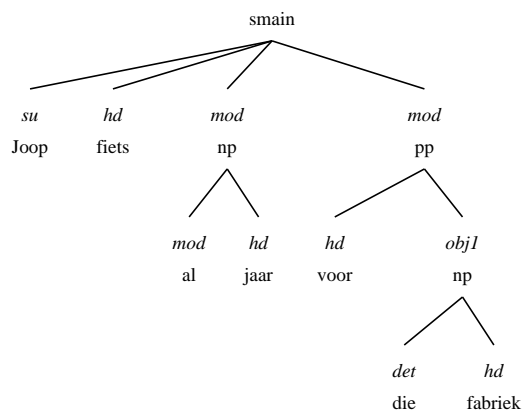
(43) hij heeft haar alles gegeven

(44) zal ik een borrel voor u inschenken ?



(45) zal ik je een borrel inschenken ?

(46) Joop fietst al jaren **voor die fabriek**



(47) *Joop fietst die fabriek al jaren

In de volgende voorbeelden:

(48) hij werd me toch kwaad

(49) hij liep/fietste/schreeuwde/. . . me toch hard

maakt *me* geen deel uit van het complementatiepatroon van de werkwoorden. Ook is er geen agreement met het onderwerp. Deze gevallen worden als MOD geannoteerd.

In:

(50) waar zit 'm dat in

(51) daarin zit 'm de charme

(52) daar zit 'm de kneep

(53) dat is het 'm nou net

(54) daar wringt 'm de schoen

beschouwen we 'm een onderdeel van de werkwoordelijke uitdrukking. Hiervoor wordt dus het label SVP gebruikt, zoals uitgelegd in sectie 2.1.13.2 op bladzijde 62.

2.1.7 Loos *het*

Behalve als voorlopig onderwerp, voorlopig lijdend voorwerp enzovoort kent *het* ook een aantal loze gebruiken, bijvoorbeeld als onderwerp (Haeseryn et al. 1997:1133).

(55) het vriest dat het kraakt

- (56) hoe gaat het met u?
- (57) het ontbreekt hem aan moed
- (58) het werkt prettig op deze kamer
- (59) het is feest

In deze gevallen heeft *het* geen eigen betekenis en geen verwijzende, maar alleen een syntactische functie" (Haeseryn et al. 1997:1133). Deze voorkomens van *het* worden geannoteerd als SU.

Het kan ook loos lijdend voorwerp zijn (Haeseryn et al. 1997:1154). In deze gevallen wordt *het* als OBJ1 aangemerkt:

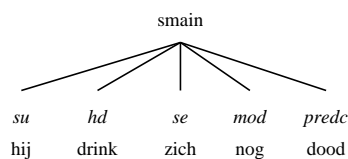
- (60) de presidente had het warm
- (61) samen zullen we het wel rooien
- (62) hij zette het op een lopen
- (63) de veldwachter had het op mij gemunt
- (64) weer een keer dat we het erover hebben

2.1.8 Predicatief complement

Dit is een overkoepelend label voor subject- en object-georiënteerde predicatieve complementen. In de traditionele terminologie (cf. ANS (Haeseryn, 1997)) verschijnt PREDC als:

- naamwoordelijk deel van gezegde bij koppelwerkwoorden
- bepaling van gesteldheid volgens de handeling
- resultatieve werkwoordsbepaling (of 'bepaling van gesteldheid ten gevolge van de handeling').

- (65) ze is **moe**
- (66) ze vindt hem **een schat**
- (67) hij drinkt zich nog **dood**



(68) zij verfdde het kastje **rood**

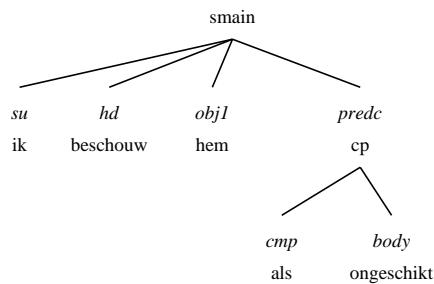
Naast AP en NP kunnen soms ook andere categorieën als PREDC optreden:

(69) De zaak is **van groot belang**

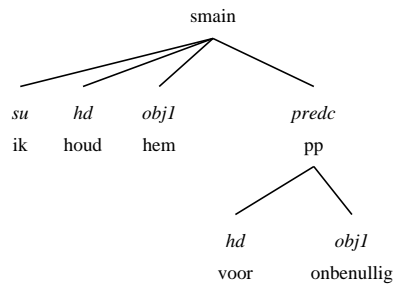
(70) De stand is **3-0**

(71) **Wat je zegt** ben je zelf

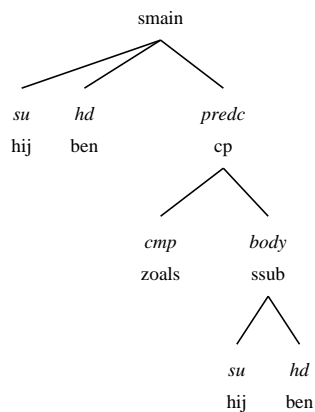
(72) Ik beschouw hem **als ongeschikt**



(73) Ik houd hem **voor onbenullig**



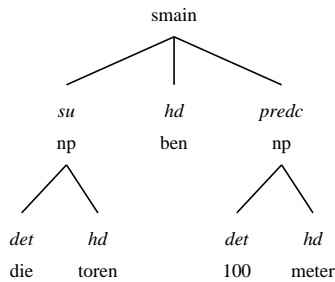
(74) Hij is **zoals hij is**



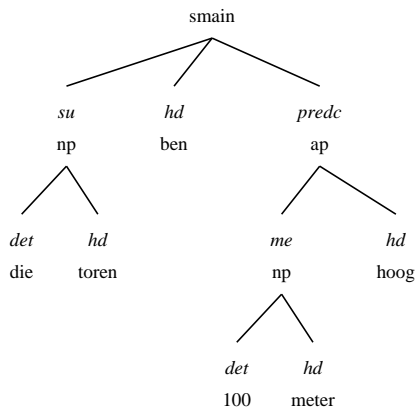
Bij zinnen van het type *hij is dertien* volgen we de traditie, i.c. de ANS (Haeseryn, 1997, blz. 435, blz. 1204): *dertien* wordt beschouwd als een zelfstandig ge-

bruikt hoofdtelwoord dat fungeert als een naamwoordelijk deel van het gezegde, PREDC. Zo ook:

(75) die toren is 100 meter



(76) die toren is 100 meter hoog



(77) het water is 21 graden

(78) het water is 21 graden warm

De traditie (Den Hertog) volgend, noemen we *weg* in *het boek is weg* en *uit de mode* in *die hoed is uit de mode* naamwoordelijk deel van het gezegde (PREDC), en wel omdat het bijwoord en de voorzetselconstituenten de functie van bijvoeglijk naamwoord vervullen.

In de volgende voorbeelden:

(79) wie is deze meneer?

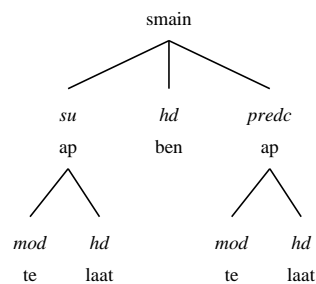
(80) wat wil jij later worden?

(81) wat zijn de kenmerken van de verbrandingsmotor?

wordt de rol van naamwoordelijk deel van gezegde door vragende voornaamwoorden vervuld. Het laatste voorbeeld laat zien dat het werkwoord *zijn* congrueert met

de kenmerken ..., dus die woordgroep moet wel onderwerp zijn. In sommige gevallen is het onderscheid tussen het onderwerp en een naamwoordelijk deel van het gezegde moeilijk te maken:

- (82) Hardlopers zijn doodlopers
- (83) De laatsten zullen de eersten zijn
- (84) Te laat is te laat

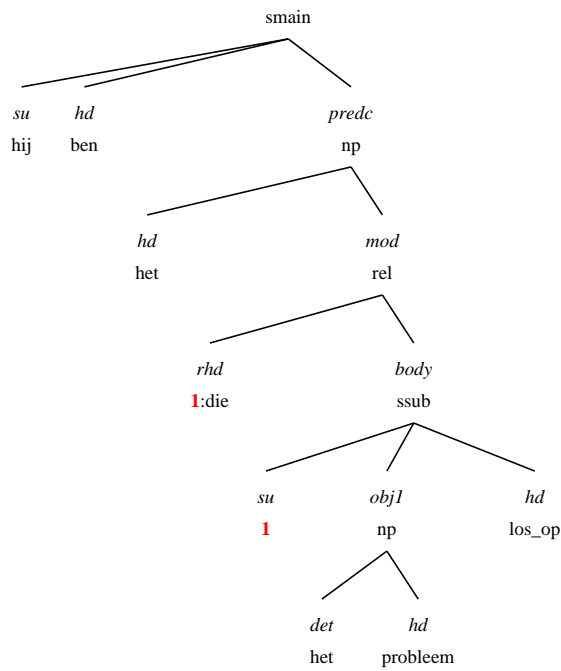


In zulke gevallen kiezen we ervoor om de vooropgeplaatste constituent als onderwerp te beschouwen, omdat dit de niet-gemarkeerde zinsvolgorde is in het Nederlands.

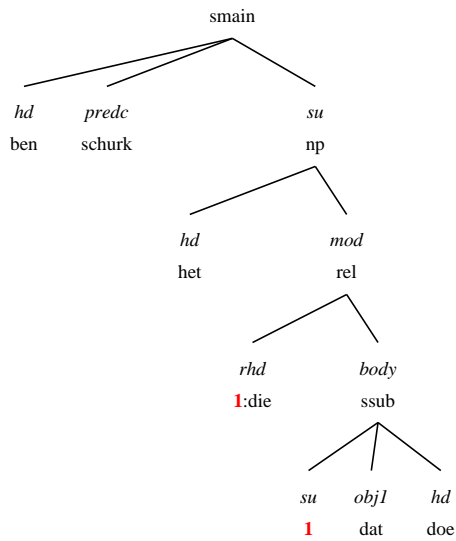
2.1.8.1 Clefts

Bij clefts wordt de relatieve zin aangehecht aan het woordje *het*.

- (85) Hij was het die het probleem oploste

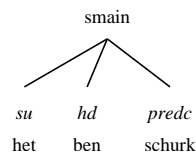


(86) Het zijn schurken die dat doen

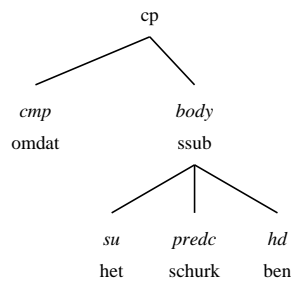


Analoog annoteren we (vergelijk de bijzinsvolgorde van voorbeeld (87) in (88) voor een argument dat *het* hier onderwerp is):

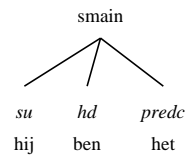
(87) Het zijn schurken



(88) omdat het schurken zijn



(89) Hij was het



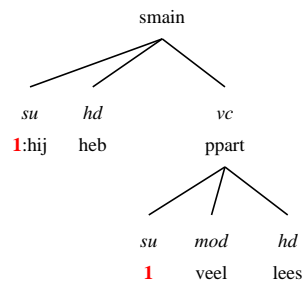
2.1.9 Verbale complementen

Voor verbale en sententiële complementen beschikken we over het dependentielabel VC. De verschillende soorten verbale en sententiële complementen worden hieronder besproken. De dependentierol VC wordt, naast de sententiële complementen, toegekend aan werkwoordgroepen binnen de werkwoordelijke constituent, zoals behandeld in hoofdstuk 18 van de ANS (Haeseryn, 1997, blz. 937–1056). Het gaat om complementen met categorielabels INF, PPART, TI, AHI, OTI (zie de tabellen 2.2 en 2.3).

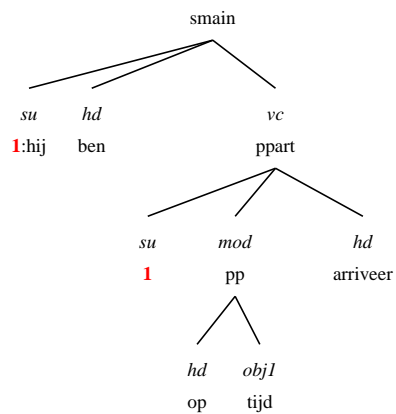
2.1.9.1 Hulpwerkwoorden

Bij hulpwerkwoorden wordt het onderwerp niet alleen geannoteerd bij de persoonsvorm, maar het onderwerp fungeert ook als het onderwerp van het hoofdwerkwoord. In termen van de ANS (Haeseryn, 1997, blz. 966) annoteren we dus hier ook het geïmpliceerd onderwerp. Dit is een uitbreiding van de CGN annotatie. Hulpwerkwoorden voor de passief worden in een volgende sectie behandeld.

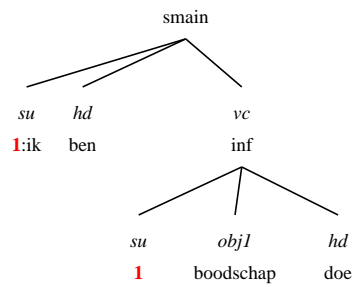
(90) hij heeft veel gelezen



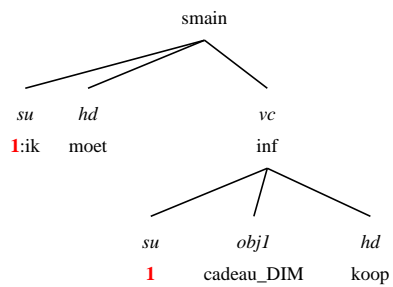
(91) hij is op tijd gearriveerd



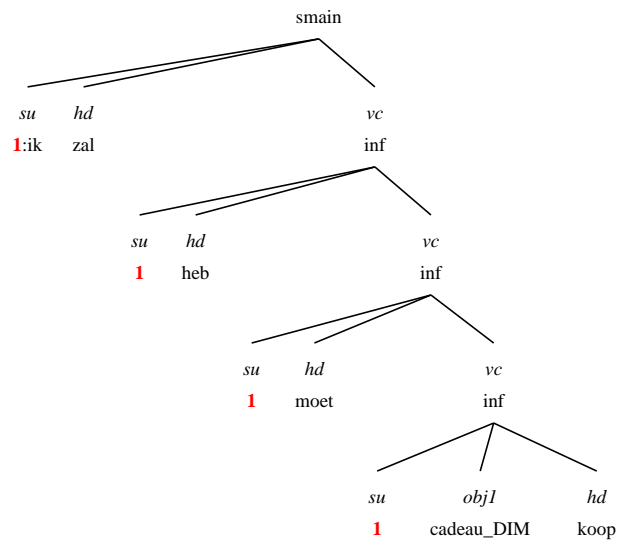
(92) ik ben boodschappen doen



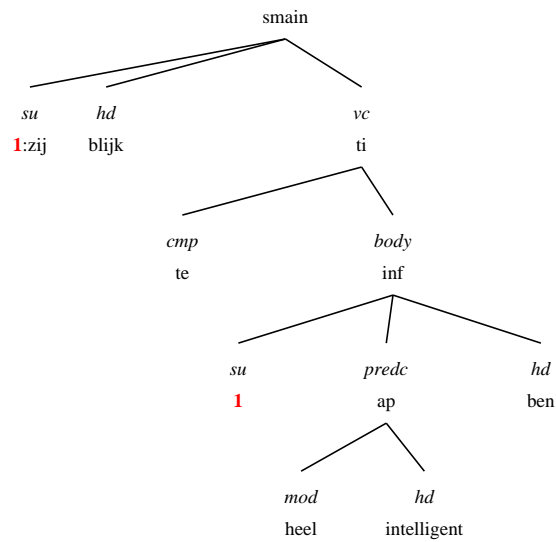
(93) ik moet cadeautjes kopen



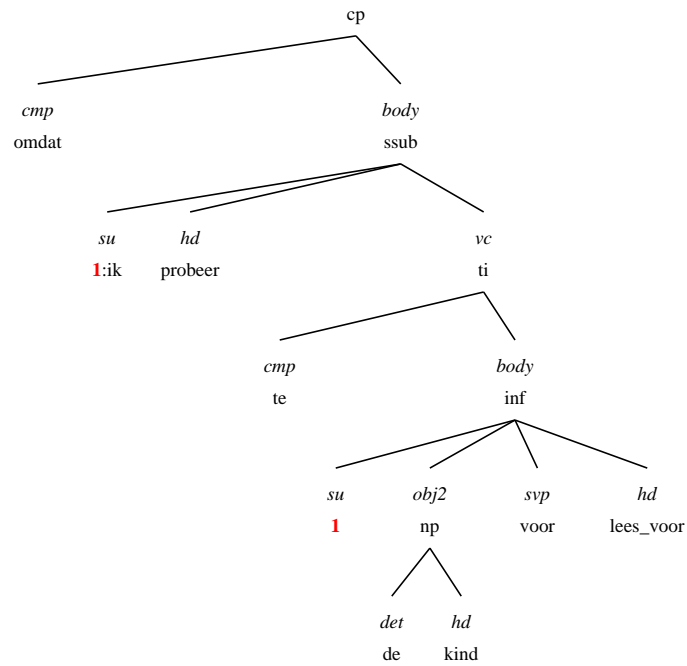
(94) ik zou cadeautjes hebben moeten kopen



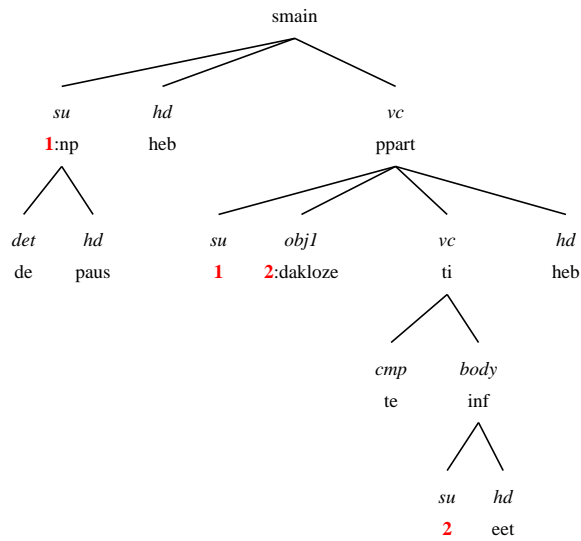
(95) zij bleek heel intelligent te zijn



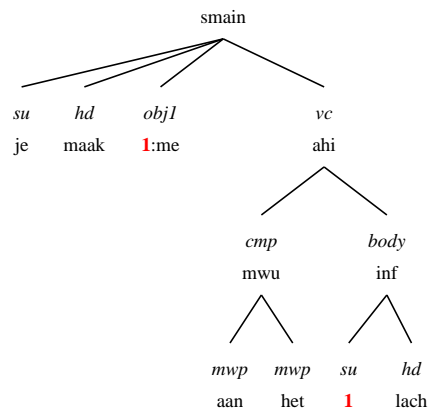
(96) omdat ik de kinderen probeer voor te lezen



(97) de paus heeft daklozen te eten gehad



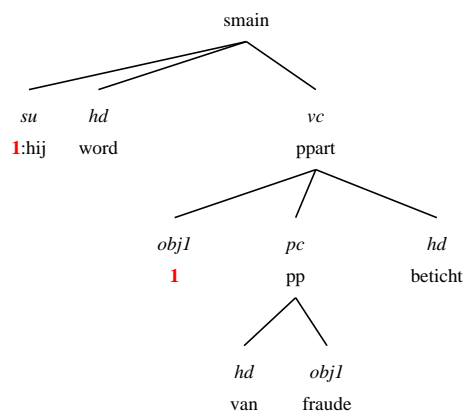
(98) je maakt me aan het lachen



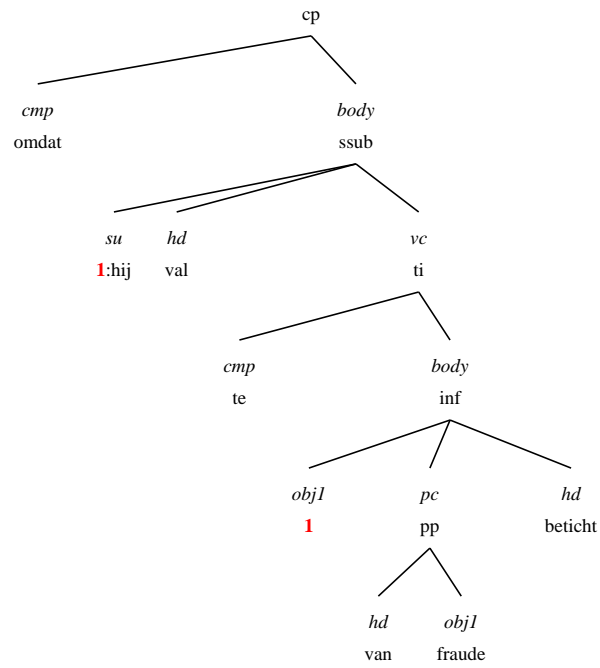
2.1.9.2 Passief

Bij passieve hulpwerkwoorden fungeert het onderwerp van het hulpwerkwoord als lijdend voorwerp (OBJ1) van het hoofdwerkwoord. We geven hier ook de analyse van de verwante constructies met de werkwoorden *krijgen*, *vallen*, *vinden*, *voelen*, *zien*.

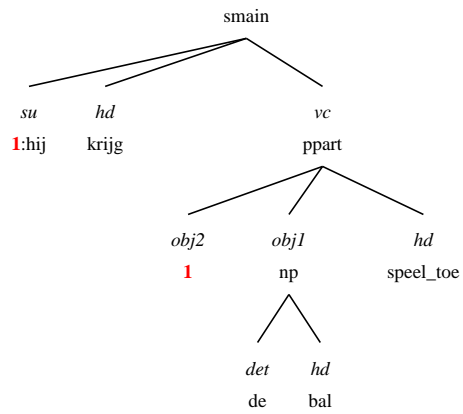
(99) hij wordt van fraude beticht



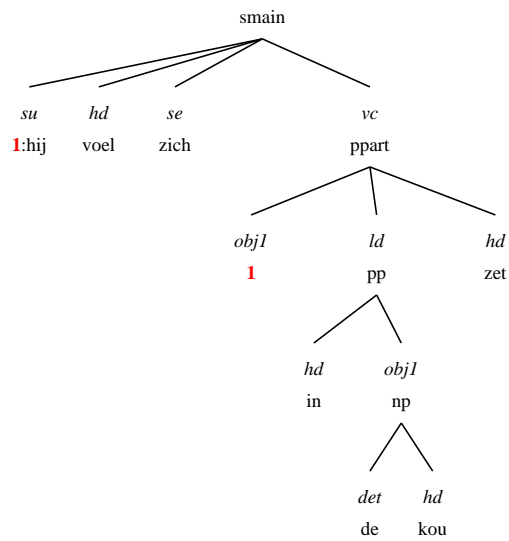
(100) omdat hij van fraude valt te betichten



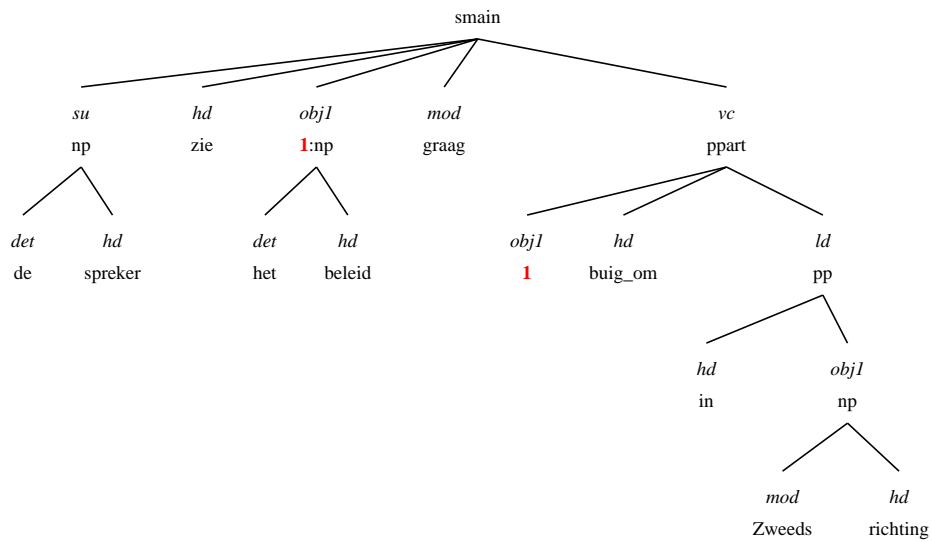
(101) hij krijgt de bal toegespeeld



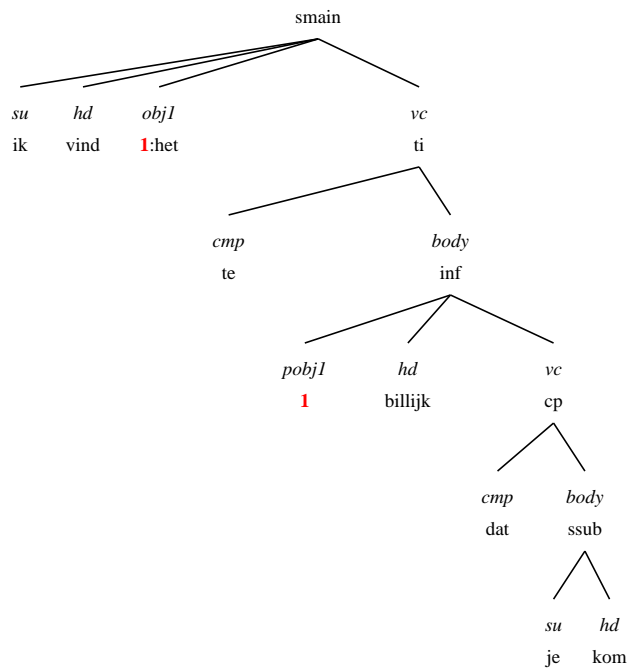
(102) hij voelt zich in de kou gezet



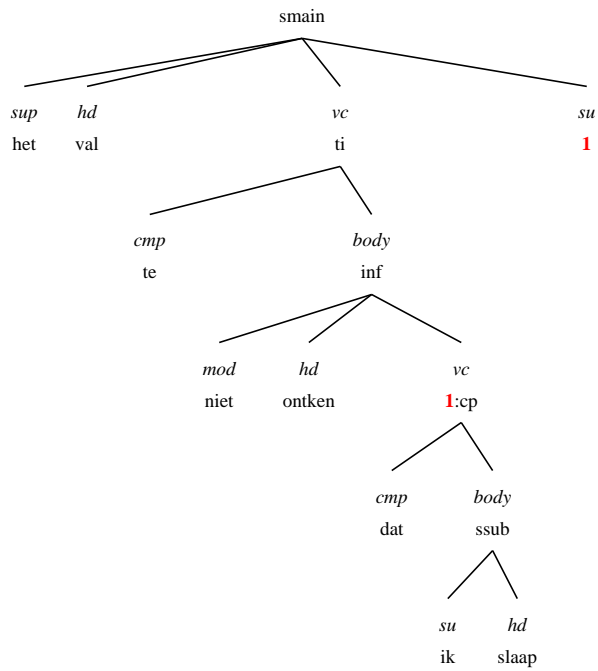
(103) De spreker zag het beleid graag omgebogen in Zweedse richting



(104) ik vind het te billijken dat je komt

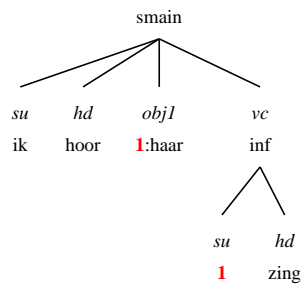


(105) Het valt niet te ontkennen dat ik slaap

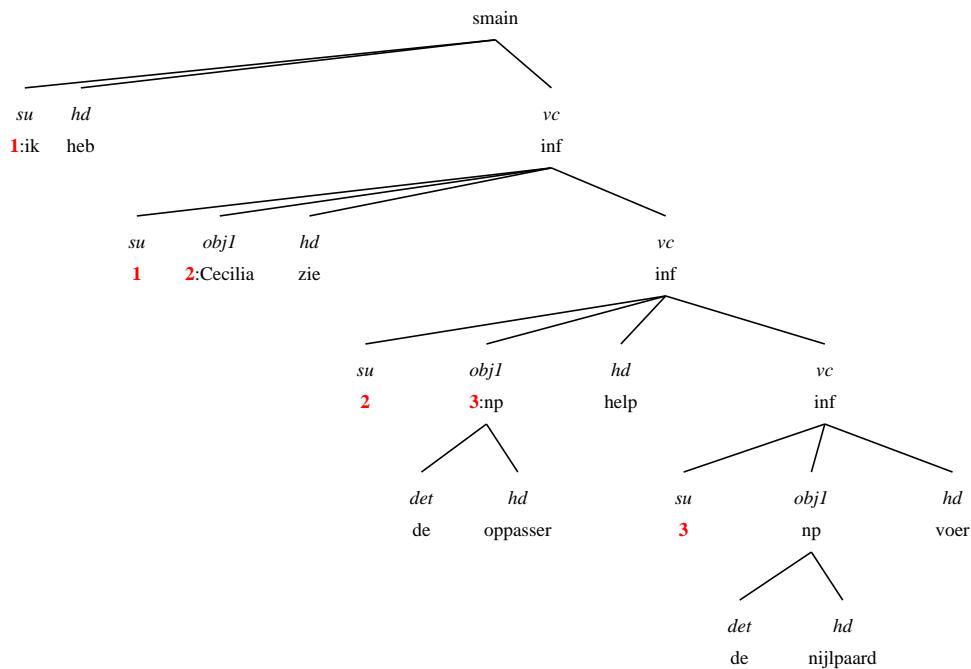


2.1.9.3 AcI

(106) ik hoor haar **zingen**



(107) ik heb Cecilia de oppasser de nijpaarden zien helpen voeren

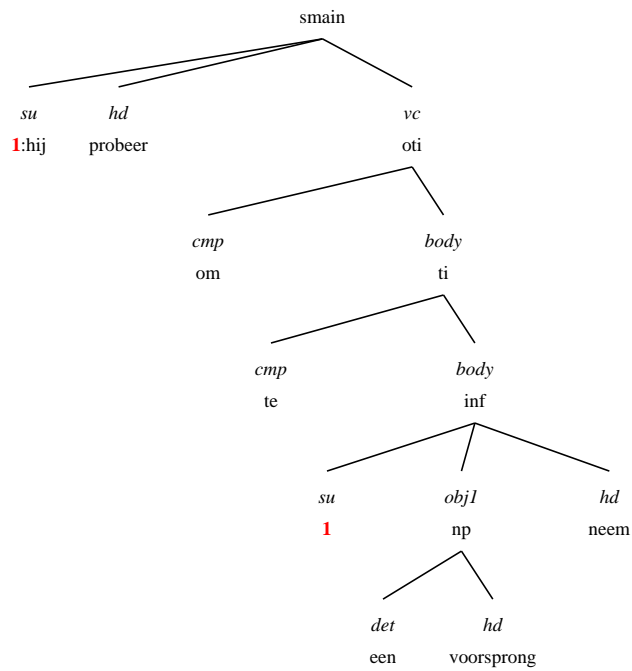


Merk op dat de accusatief in AcI-constructies (*haar* in (106)) wordt aangemerkt als OBJ1 van het matrixwerkwoord *horen* en ook als SU van het ingebedde werkwoord *zingen*.

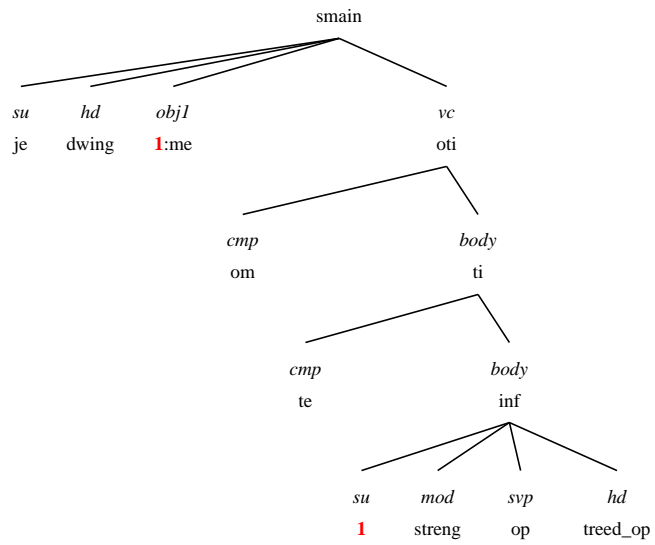
2.1.9.4 Beknopte bijzinnen

De dependentierol VC wordt daarnaast ook toegekend aan de beknopte bijzinnen (categorielabel TI of OTI) waarin geen groepsvorming optreedt, zoals bij subject- en object-controle werkwoorden. Merk hierbij opnieuw op, dat het begrepen, gecontroleerde, onderwerp van de beknopte bijzin wordt geannoteerd (normaliter is dit het onderwerp of het lijdend voorwerp van de hoofdzin).

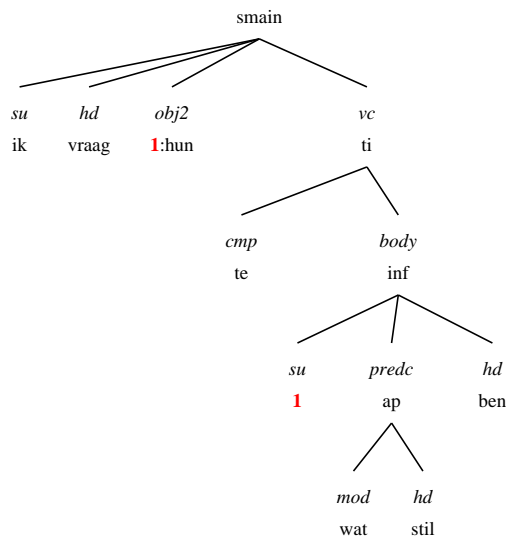
(108) hij probeerde om een voorsprong te nemen



(109) je dwingt me om strenger op te treden



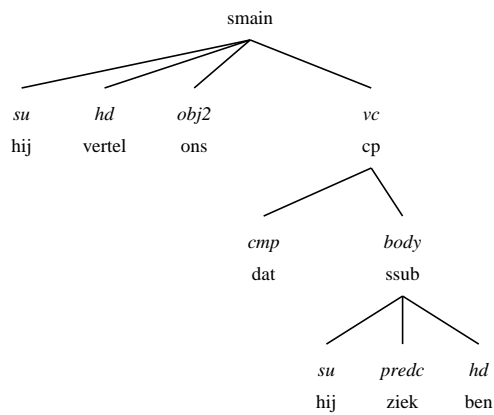
(110) ik vroeg hun wat stiller te zijn



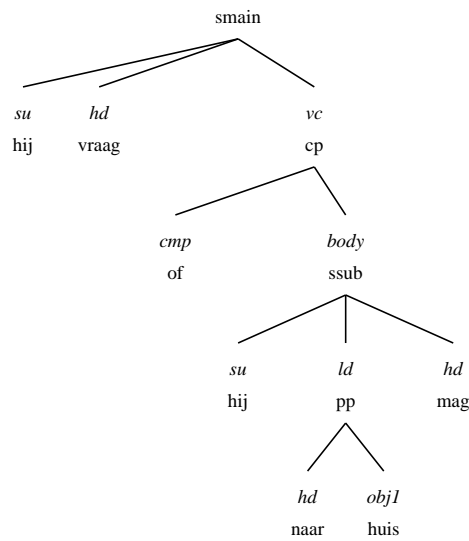
2.1.9.5 Finiete bijzinnen

Daarnaast wordt het label VC ook gebruikt voor zinscomplementen die typisch ingeleid worden met de complementeerders *dat* en *of*. Ook ondergeschikte vraagzinnen treden op als VC. De interne structuur van ondergeschikte vraagzinnen wordt behandeld in hoofdstuk 3 op bladzijde 159.

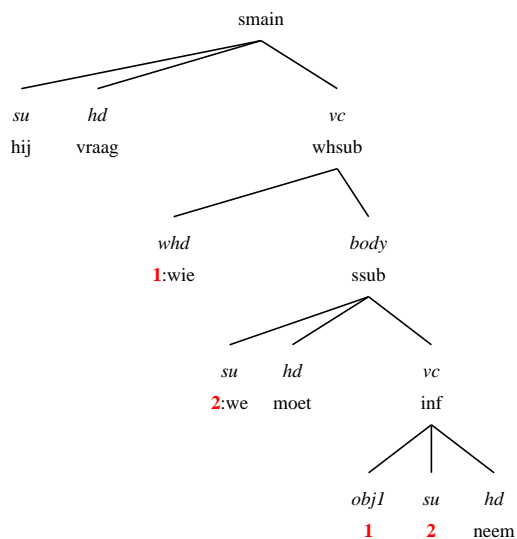
(111) Hij vertelde ons dat hij ziek was



(112) Hij vroeg of hij naar huis mocht



(113) Hij vroeg wie we moesten nemen



2.1.10 Voorzetselvoorwerp

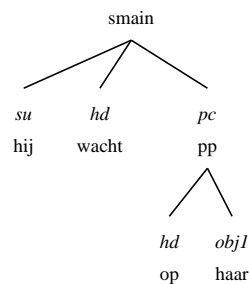
Het label PC wordt gebruikt voor PP-complementen met een ‘vast’ voorzetsel als hoofd. Eigenschappen:

- het ‘vaste’ voorzetsel laat zich niet door een ander voorzetsel vervangen zonder een betekenisverandering te veroorzaken voor het verbale hoofd met het PC-complement;
- het ‘vaste’ voorzetsel draagt niet compositioneel bij tot de betekenisopbouw

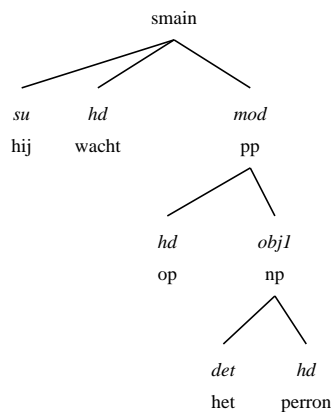
van dit complementatiepatroon (in tegenstelling tot de voorzetsels in modificerders en LD-complementen).

Met andere woorden: de keuze van het voorzetsel ligt vast, maar is niet te voorspellen uit de betekenis van het voorzetsel.

(114) hij wacht **op haar** (PC)



(115) hij wacht **op het perron** (MOD)

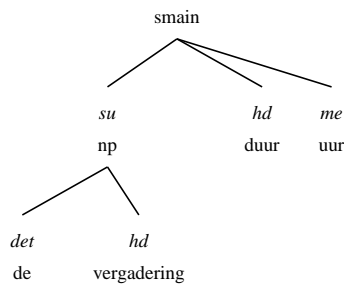


2.1.11 Maatcomplement

Het dependentielabel ME is voor complementen die een maat (gewicht, duur, prijs, ...) van het onderwerp uitdrukken. In tegenstelling tot het lijdend voorwerp (OBJ1) zijn deze ME-complementen niet passiviseerbaar.

(116) het kost **20 gulden**

(117) de vergadering duurt **uren**



Optionele maataanduidingen, als in

(118) we hebben uren gefietst

(119) we hebben vijftig kilometer gewandeld

worden aangemerkt als modificeerder (MOD), in deze gevallen met als categorie NP.

2.1.12 Locatief/directioneel complement

We gebruiken LD als overkoepelend label voor de plaats- en richtingscomplementen.

(120) het ligt **boven**

(121) hij legt het **op tafel**

(122) ze fietste **naar huis**

Het label LD wordt ook toegekend aan nominale complementen in gevallen die vaak als *postposities* worden benoemd:

(123) ik heb hem **het bos** ingestuurd

(124) ik stuur hem **het bos** in

(125) **het bos** zou ik hem nu maar niet in sturen

(126) ik zou hem eigenlijk **het bos** moeten in sturen

In al deze gevallen wordt *in* als SVP van het werkwoord geannoteerd. Deze 'heranalyse' annotatie is gemotiveerd vanwege de observatie dat de postpositie/partikel vaak aan het werkwoord wordt geschreven, en niet noodzakelijkerwijs na de NP volgt. Als derde argument noemen we de mogelijkheid dat de postpositie/partikel in de werkwoordcluster wordt opgenomen.

2.1.13 Scheidbaar deel van samengesteld werkwoord

2.1.13.1 Partikels

Het scheidbare deel (partikel) van een scheidbaar samengesteld werkwoord wordt, indien het los optreedt van het werkwoord, geannoteerd als SVP (ook indien de spelling eigenlijk voorschrijft dat beide delen niet los mogen worden geschreven zoals in voorbeeld (128)).

- (127) ik bel je morgen **terug**
- (128) of je hem morgen **terug** belt
- (129) ik zal je morgen **terug** moeten bellen
- (130) ik probeer je morgen **terug** te bellen
- (131) of je hem morgen **terug** wil bellen

De SVP-rol kan gerealiseerd worden door de categorieën voorzetsel, bijwoord, adjectief en zelfstandig naamwoord.

- (132) of je haar **door** wil laten
- (133) of je het **weg** wil brengen
- (134) of je het **klaar** wil maken
- (135) dat je hier moeilijk **adem** kan halen

Het is te voorzien dat er bij de annotatie problemen zullen zijn wat betreft het onderscheid tussen SVP en PREDC dan wel LD in die gevallen waar

- de semantiek van de combinatie volledig compositioneel is;
- de SVP bij een bepaalde werkwoordelijke stam behalve door een lexicaal element ook gerealiseerd kan worden door een volledige frase.

Voorbeelden:

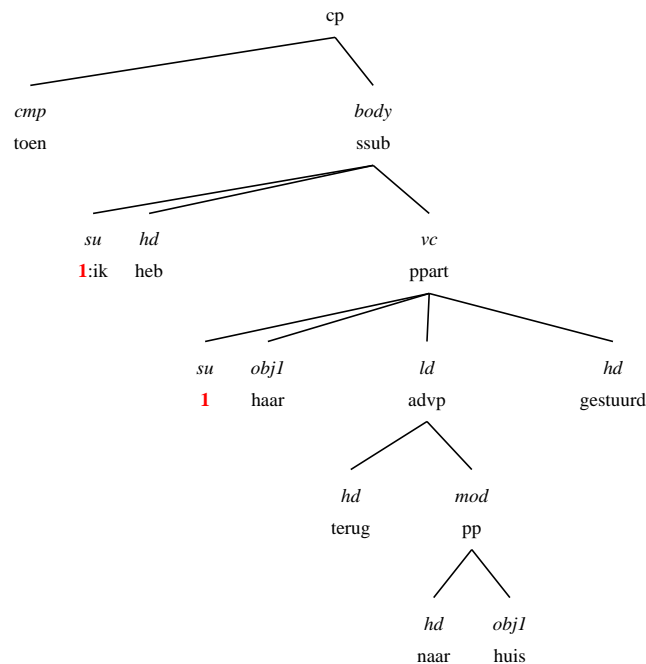
- (136) hij stuurt haar (**weg** | naar huis) (LD)
- (137) de deur viel (**dicht** | in het slot) (PREDC)
- (138) ze steekt haar handen (**omhoog** | in de lucht) (LD)

Hieronder enkele criteria die bij de keuze voor SVP kunnen helpen.

- De combinatie SVP met werkwoordelijk deel kan de input vormen voor morfologische derivatie (bijvoorbeeld: *vrijlating*, *bekendmaking*, *ademhaling*, maar **groenverving*) – voldoende maar niet noodzakelijke voorwaarde.
- De combinatie SVP met werkwoordelijk deel kan bij elkaar blijven binnen de werkwoordelijke cluster (bijvoorbeeld: *(of je haar) wil wegbrengen* naast *weg wil brengen*) – noodzakelijke maar niet voldoende voorwaarde (Vlaamse clusters!).
- De combinatie SVP met werkwoordelijk deel kan niet doorbroken worden door andere bepalingen
- Een partikel kan normaliter niet vooropgeplaatst worden
- Soms kan de passief-test gebruikt worden om aan te tonen dat een zelfstandig naamwoord niet als partikel wordt gebruikt

Het laatste criterium kan worden geïllustreerd met het volgende voorbeeld:

(139) Toen ik haar **terug** naar huis had gestuurd, ... (geen SVP)

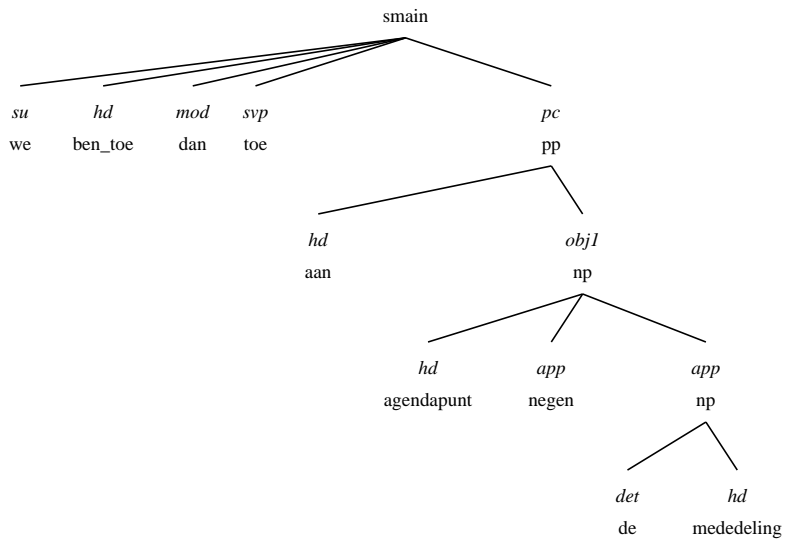


Omdat **terug** in dit geval van het werkwoordelijk deel wordt gescheiden, kan het hier niet de functie SVP toebedeeld krijgen.

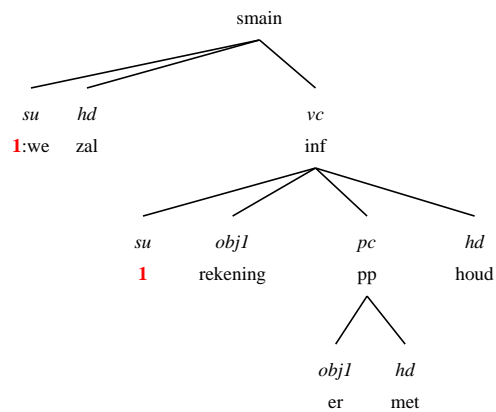
In de volgende voorbeelden zijn respectievelijk *af*, *toe* en *naar voren* geclassificeerd als SVP.

(140) we vroegen ons af of ...

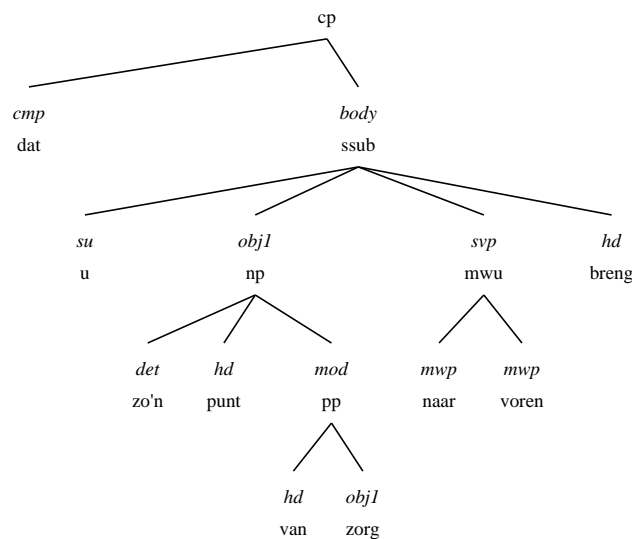
(141) we zijn dan toe aan agendapunt negen de mededelingen



(142) we zullen er rekening mee houden



(143) ik denk dat het belangrijk is dat u zo'n punt van zorg naar voren brengt



Het eerste geval, *af*, is oncontroversieel: het is deel van het werkwoord *zich afvragen* dat geheel andere eigenschappen, en een andere betekenis, heeft dan het (niet-transitieve) grondwoord *vragen*.

De redenering bij *toe* is analoog, hoewel er wellicht geen apart werkwoord *toezijn* (met vast voorzetsel *aan*) in het woordenboek te vinden is.

Naar voren is oorspronkelijk een LD, maar anderzijds wordt voldaan aan het criterium dat de combinatie *naar voren* en werkwoordelijk hoofd *brengt* niet door bijwoordelijke bepalingen doorbroken kan worden – maar dat geldt in het algemeen ook voor richtingsbepalingen. Maar er zijn meer argumenten om *naar voren* hier toch als SVP te benoemen: de letterlijke directionele betekenis is verdwenen, en modificatie is onmogelijk.

Vergelijk:

(144) het punt van zorg dat u (*recht) naar voren brengt (SVP)

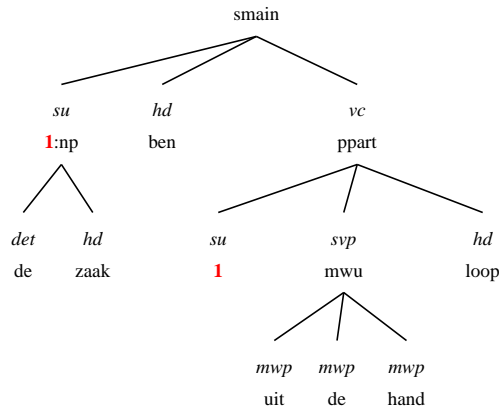
(145) de ballen die de keeper (recht) naar voren schiet (LD)

Het laatste geval, *rekening* is hier geen SVP. De combinatie SVP en werkwoordelijk hoofd wordt doorbroken door *mee*. Daarnaast gedraagt het woord *rekening* zich (in elk geval voor sommige sprekers en schrijvers) als een ‘gewoon’ lijdend voorwerp voor de passief-test (*rekening moet worden gehouden met de volgende punten ...*). Vanwege de passief test, en vanwege de mogelijkheid om *rekening* te modificeren (geen, weinig, (on)voldoende) kiezen we voor de OBJ1 analyse.

2.1.13.2 Idiomatisch deel van werkwoordelijke uitdrukking

In geval van niet-compositionele werkwoordelijke uitdrukkingen worden de idiomatische onderdelen ook als SVP gelabeld.

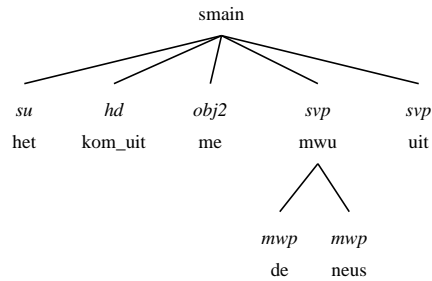
(146) de zaak is **uit de hand** gelopen



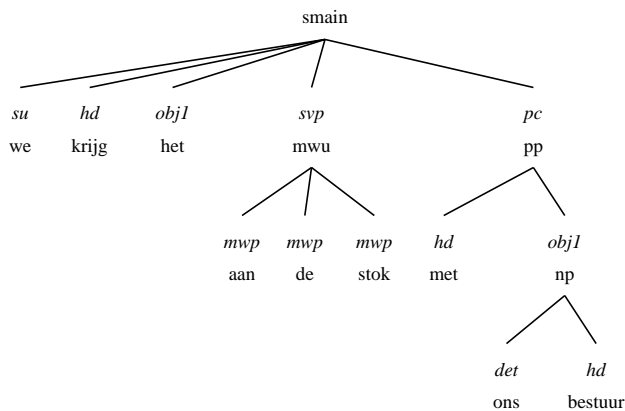
(147) hoe een natuurlijke taal precies **in elkaar** zit

(148) om nieuwe samenwerkingen **tot stand** te brengen

(149) Het komt me de neus uit

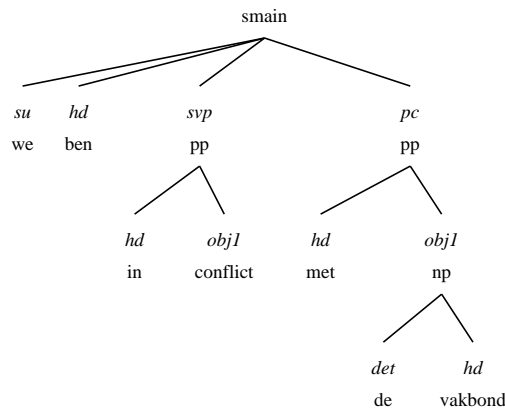


(150) We krijgen het aan de stok met ons bestuur



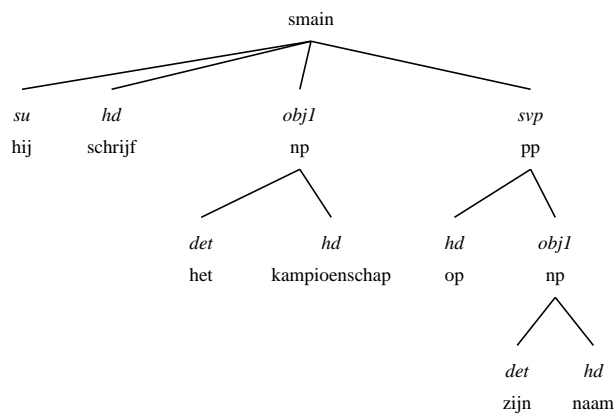
In zulke gevallen is de interne analyse van het als SVP gelabelde zinsdeel normaliter een multi-word-unit (category MWU en alle woorden zijn MWP dochters). Indien echter de onderdelen modificeerbaar zijn, wordt de normale compositionele analyse aangehouden:

(151) We zijn **in conflict** met de vakbond



(152) We zijn **in ernstig conflict** met de vakbond

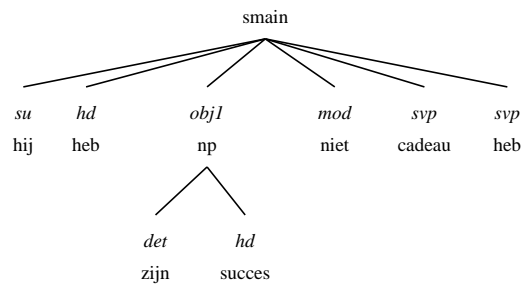
(153) Hij schreef het kampioenschap **op zijn naam**



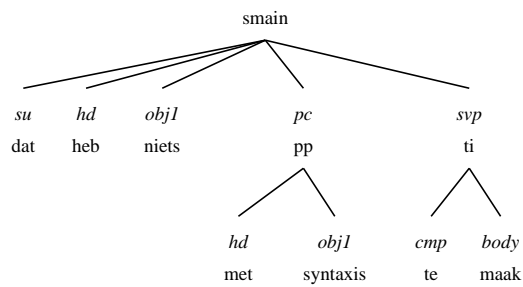
(154) Wij schreven het kampioenschap **op onze naam**

Er zijn ook werkwoordelijke uitdrukkingen die werkwoorden bevatten.

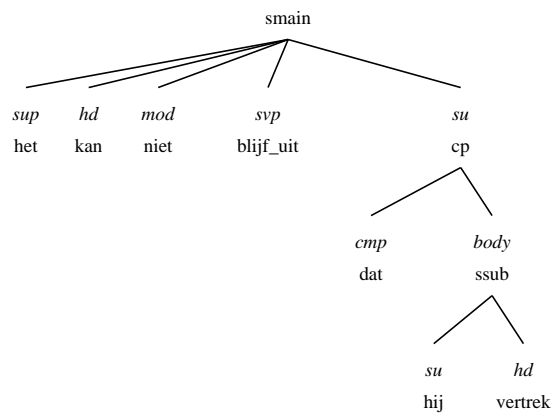
(155) Hij heeft zijn succes niet cadeau gehad



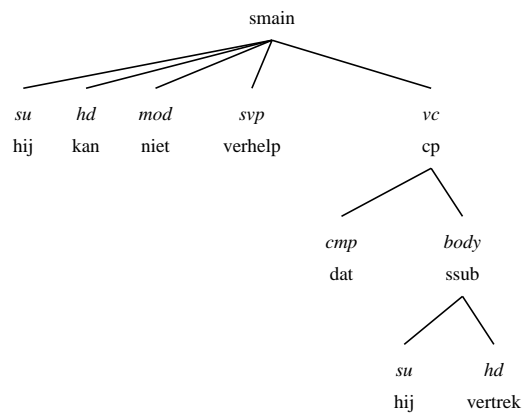
(156) Dat heeft niets met syntaxis te maken



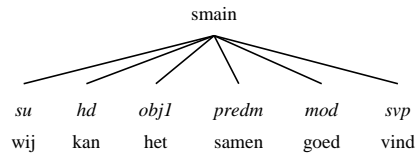
(157) Het kon niet uitblijven dat hij vertrok



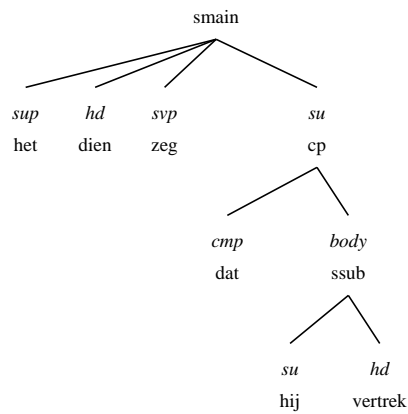
(158) Hij kon niet verhelpen dat hij vertrok



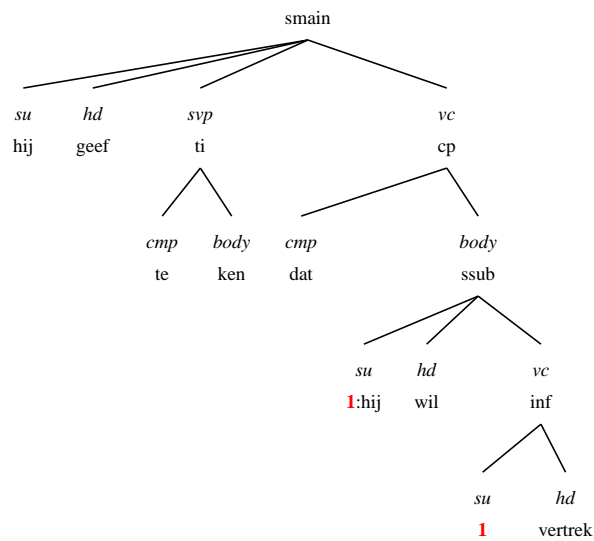
(159) Wij kunnen het samen goed vinden



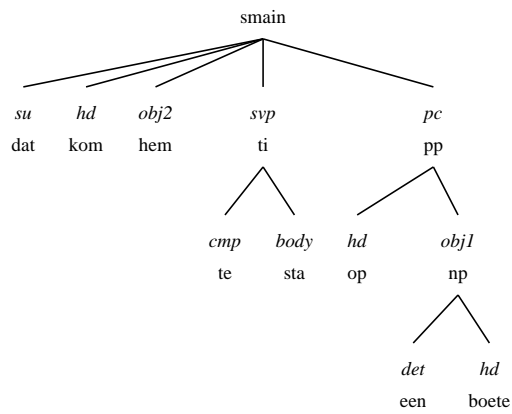
(160) Het dient gezegd dat hij vertrok



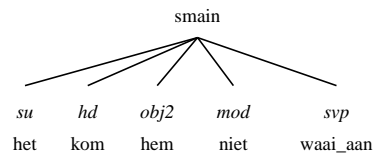
(161) Hij gaf te kennen dat hij wilde vertrekken



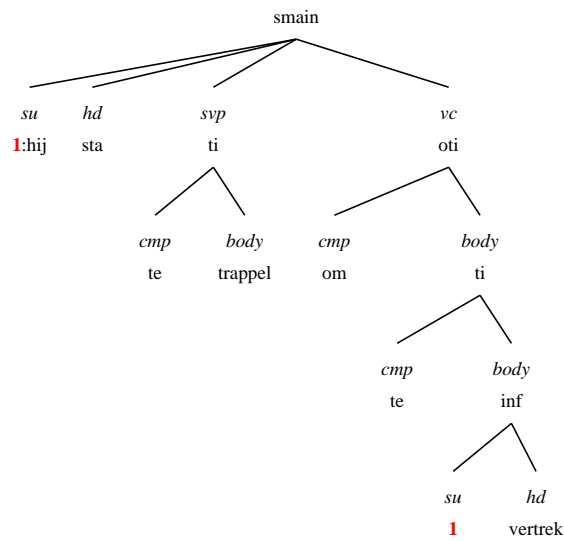
(162) Dat kwam hem te staan op een boete



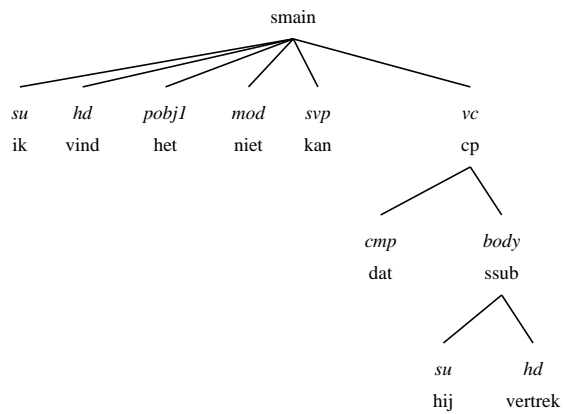
(163) Het komt hem niet aangewaaid



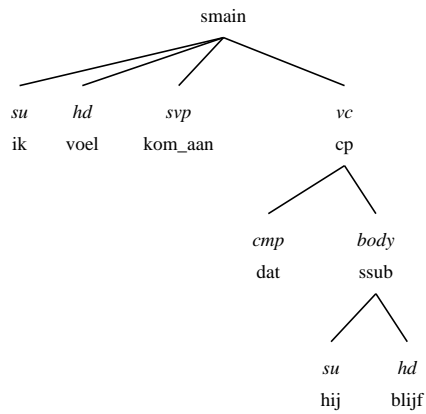
(164) Hij staat te trappelen om te vertrekken



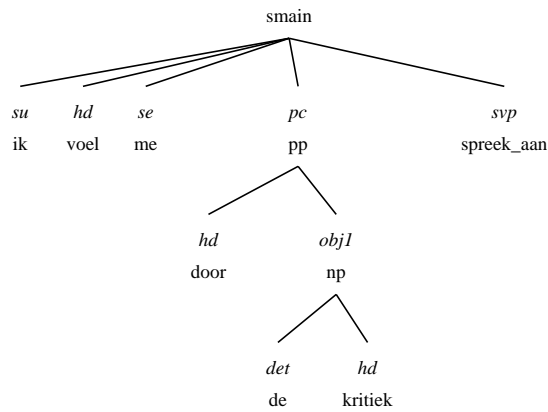
(165) Ik vind het niet kunnen dat hij vertrekt



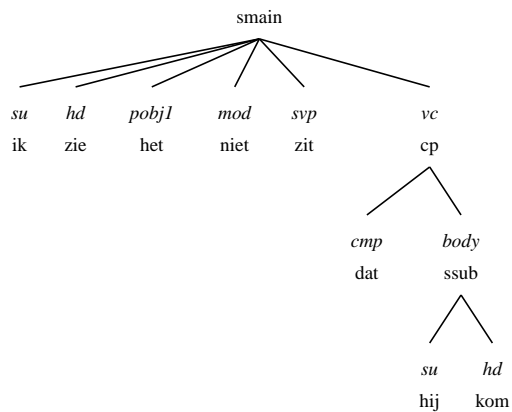
(166) Ik voel aankomen dat hij blijft



(167) Ik voel me niet aangesproken door de kritiek

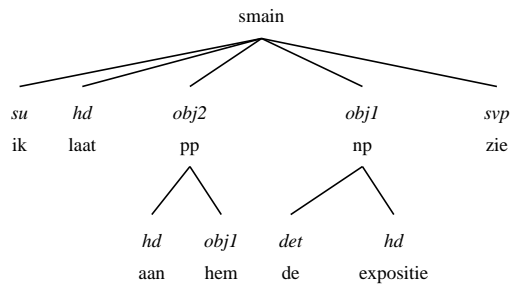


(168) Ik zie het niet zitten dat hij komt

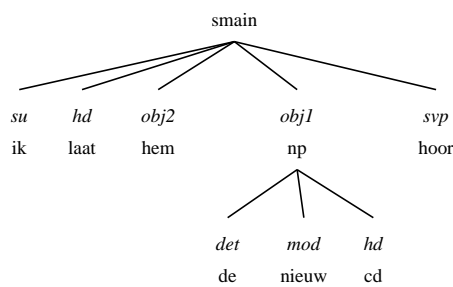


Het werkwoord *laten* staat erom bekend dat het veel in werkwoordelijke uitdrukkingen participeert. Ook gevallen zoals *laten zien* en *laten horen* worden als werkwoordelijke uitdrukking geanalyseerd, onder andere vanwege de mogelijkheid dat een indirect object (OBJ2) in deze gevallen optreedt:

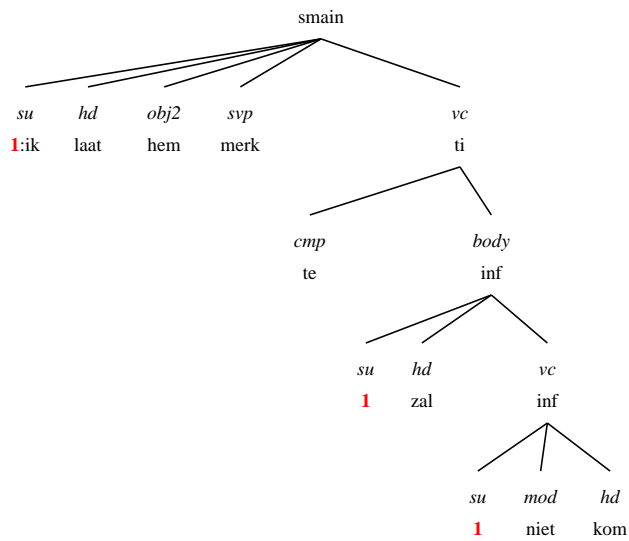
(169) Ik laat aan hem de expositie zien



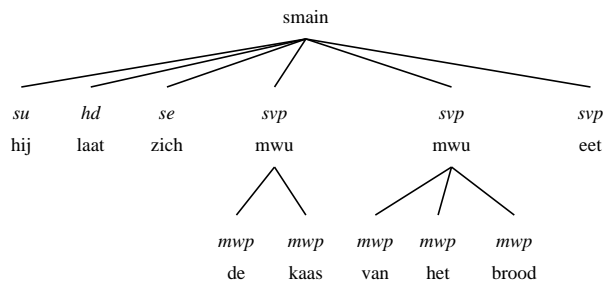
(170) Ik laat hem de nieuwe CD horen



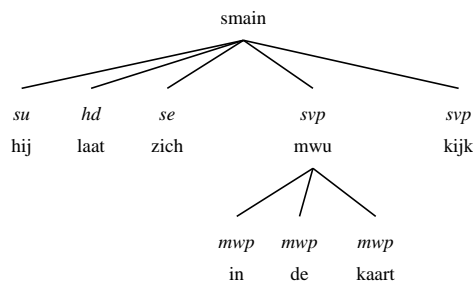
(171) Ik laat hem merken niet te zullen komen



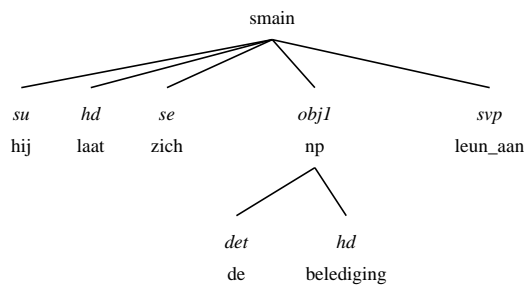
(172) Hij laat zich de kaas van het brood eten



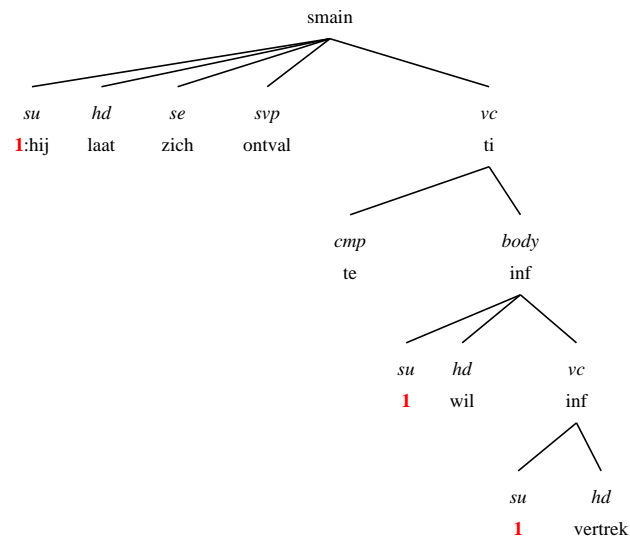
(173) Hij laat zich in de kaart kijken



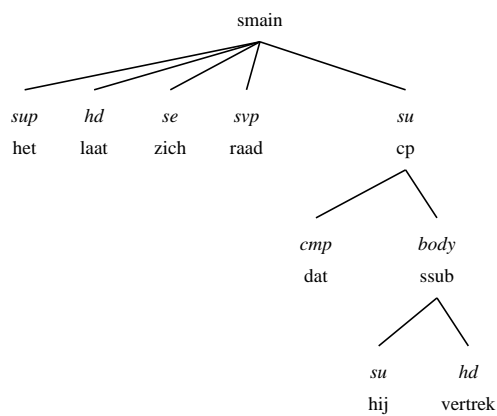
(174) Hij liet zich de beledigingen aanleunen



(175) Hij liet zich ontvallen te willen vertrekken

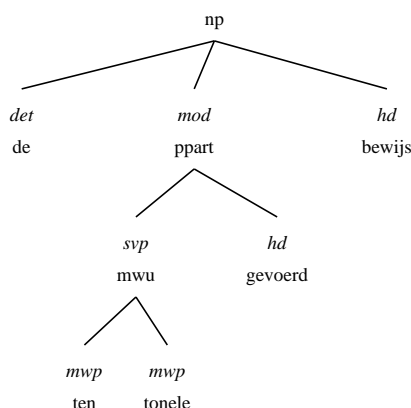


(176) Het laat zich raden dat hij vertrok



Ten slotte komen soms ook in andere domeinen uitdrukkingen voor. Die worden analoog geannoteerd.

(177) De **ten tonele** gevoerde bewijzen



2.1.13.3 Postposities

Postposities in combinatie met een LD complement worden ook als SVP geanalyseerd, zoals beschreven in paragraaf 2.1.12 op bladzijde 57.

2.1.14 Modifierende elementen in het verbale domein

2.1.14.1 MOD en PREDM

Bij de bepalingen binnen het verbale domein maken we een onderscheid tussen PREDM en MOD.

Het dependentielabel PREDM gebruiken we voor wat wel de bepaling van gesteldheid ‘tijdens de handeling’ wordt genoemd. Dit type bepaling van gesteldheid speelt een modifierende rol, en moet dus onderscheiden worden van de bepaling van gesteldheid ‘volgens/ten gevolge van de handeling’ die tot het complementatiepatroon behoort.

- (178) hij kwam **dronken** thuis (‘tijdens de handeling’, PREDM)
- (179) je maakt je nog **vuil** (‘ten gevolge van handeling’, PREDC)
- (180) dat smaakt **lekker** (‘volgens de handeling’, PREDC)

Wat betreft de traditionele bijwoordelijke bepalingen maken we geen verdere semantische onderscheidingen tussen bepalingen van plaats, tijd, doel, gevolg, modaliteit, ... (in de ANS (Haeseryn, 1997) een totaal van 21 soorten): ze krijgen alle het default modificeerderlabel MOD. Dat het om een adverbiale bepaling gaat, blijkt uit het verbale categorielabel van de moederknoop in het lokale dependentiële domein, en/of uit de POS-tag in geval van losse bijwoorden.

2.1.14.2 Door-bepaling in passieve zinnen

De *door*-bepaling in passieve zinnen (“handelend voorwerp”) wordt eveneens benoemd als MOD:

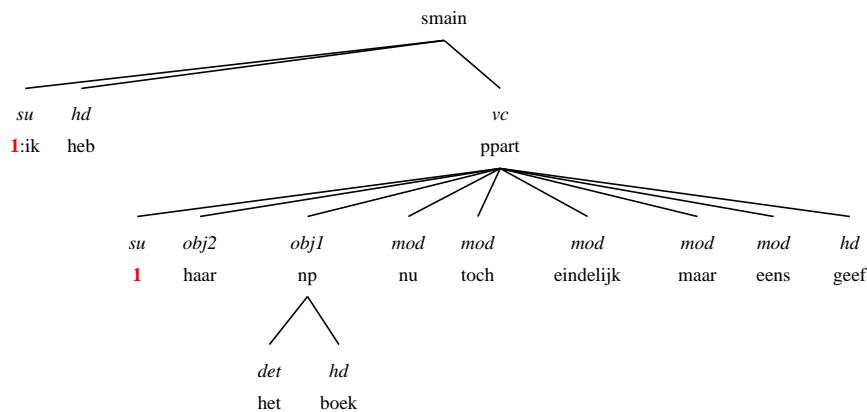
- (181) de winnaar werd gekust door de rondemiss
- (182) de winnaar werd door de rondemiss gekust
- (183) door de rondemiss werd de winnaar gekust
- (184) de overtreders kregen door de politierechter een rijverbod opgelegd

Deze bepaling wordt aangehecht op het niveau van het lexicale werkwoord en niet op het niveau van het hulpwerkwoord. Dit is conform de regel dat modifiers bij voorkeur aan het hoofdwerkwoord worden aangehecht, en niet aan de hulpwerkwoorden.

2.1.14.3 Oordeelspartikels

Speciale vermelding bij de bepalingen binnen het verbale domein verdienen de zogenaamde ‘modale partikels’ of ‘oordeelspartikels’: deze elementen lijken soms een cluster te vormen waarbinnen geen hoofd aan te wijzen is. In tegenstelling tot de oorspronkelijke CGN annotatie, worden deze partikels elk als losse bijwoordelijke bepalingen bij het werkwoord geannoteerd.

- (185) ik heb haar het boek **nu toch eindelijk maar eens** teruggegeven



2.1.14.4 Aanhechting van modifierende elementen in het verbale domein

In geval van hulpwerkwoorden, modale werkwoorden en dergelijke is het vaak lastig te bepalen of modifierende elementen betrekking hebben op het hulpwerkwoord, op het hoofdwerkwoord, of, op de een of andere manier, op de combinatie

van beide. In zulke gevallen kiezen we voor aanhechting van de modificeerders met label MOD aan het hoofdwerkwoord.

Omdat modificeerders met label PREDM vaak ook betrekking hebben op het onderwerp (SU), is de gewoonte ontstaan om deze modificeerders aan het hoogste hulpwerkwoord, meestal de persoonsvorm, aan te hechten. Ter consistentie blijven we deze aanpak volgen.

2.1.15 Kiezen tussen modificeerder en complement

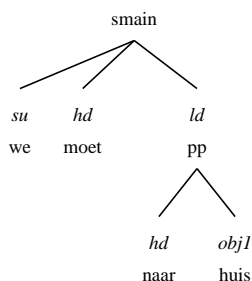
Het is soms lastig te bepalen welke rol een PP speelt bij een werkwoord. De keus is vooral lastig tussen PC, MOD, en LD. Daarnaast kan een PP ook nog voorkomen als PREDM, OBJ2 en PREDC.

2.1.15.1 Argumenten en modifiers bij ‘verzwegen’ werkwoord

Bij zelfstandig optredende modale werkwoorden, lijken de aanwezige zinsdelen te fungeren bij een niet aanwezig hoofdwerkwoord:

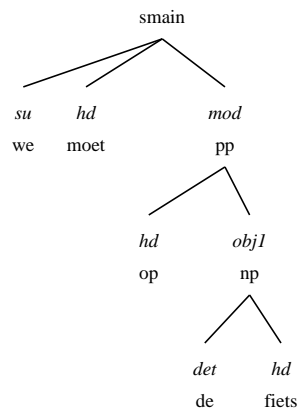
(186) We moeten naar huis gaan

(187) We moeten naar huis



(188) We moeten op de fiets gaan

(189) We moeten op de fiets



In deze gevallen worden de complementen en bepalingen geannoteerd bij het modale hulpwerkwoord. Ze krijgen de rol die ze normaliter bij het hoofdwerkwoord zouden hebben, dus in het eerste geval LD, en in het tweede geval MOD.

2.1.15.2 De rol van PP bij werkwoorden

In de volgende tabel noemen we een aantal frequente werkwoord-voorzetsel combinaties met de gewenste analyse:

werkwoord	voorzetsel	relatie	uitleg
beginnen	met/bij	PC	
denken	van/over	PC	
doen	aan/met	PC	
gaan	aan	PC	aan gang/wandel/bier/dots
gaan	op	PC	op vakantie/excursie/bedevaart...
komen	met	PC	'te berde brengen'
komen	uit	LD	indien letterlijk
komen	uit	PC	<i>er niet uit komen</i>
kijken	naar/richting/...	LD	indien complement
krijgen	van	PC	
maken	van	PC	met OBJ1
praten	over	PC	
vinden	van	PC	'van mening zijn'
werken	aan	PC	
werken	bij, in, voor	MOD	
zoeken	naar, in, door...	LD	indien complement

2.1.15.3 Werkwoorden met *mee*

Het is in het algemeen voorspelbaar dat een eventueel complement van een samengesteld werkwoord dat met *mee* begint een voorzetselgroep met *met* is:

- (190) meelopen met de optocht
- (191) meegaan met de tendens
- (192) twintig minuten laten meekoken met de aardappels
- (193) meekomen met de meester
- (194) meespelen met de Lotto

We benoemen die *met*-bepalingen als PC. Er zijn ook uitzonderingen op deze regel:

- (195) we weigeren mee te werken aan deze schending van de internationale rechtsorde
- (196) ik zou graag willen meedenken over een oplossing

maar ook hier benoemen we de voorzetselbepalingen als PC.

2.1.15.4 Het werkwoord *zijn*

Bij het werkwoord *zijn* is het soms moeilijk uit te maken met welk soort complement we te maken hebben. Voorbeelden:

- (197) het weegt twee kilo ME
- (198) het is zwaar PREDC
- (199) het is twee kilo PREDC
- (200) hij was in Amsterdam LD
- (201) het was in 1993 MOD
- (202) het was in Amsterdam LD/MOD

In voorbeeld (202) kan *in Amsterdam* betekenen ‘*in de stad A’dam*’ of ‘*toen we in A’dam waren | woonden*’. In het laatste geval gaat het om een tijdsbepaling en zou *in Amsterdam* het label MOD moeten krijgen, net als in (201). In (202) is *het was* ambigu tussen ‘*het bevond zich*’ (LD) en ‘*het vond plaats*’ (MOD).

In deze gevallen kiezen we bij twijfel voor het label MOD.

dependentielabel	OMSCHRIJVING
CMP	complementeerder
BODY	romp
categorielabel	OMSCHRIJVING
CP	frase ingeleid door onderschikkend voegwoord
SVAN	<i>van</i> -zin (complement ingeleid door <i>van</i>)
TI	<i>te</i> -infinitief-groep
OTI	<i>om te</i> -infinitief-groep
AHI	<i>aan het/uit</i> -infinitief-groep

Figuur 2.3: Complementeerder-extensies van verbale domeinen

2.2 Extensies van verbale domeinen: complementeerders

2.2.1 Bijzinnen ingeleid met onderschikkend voegwoord

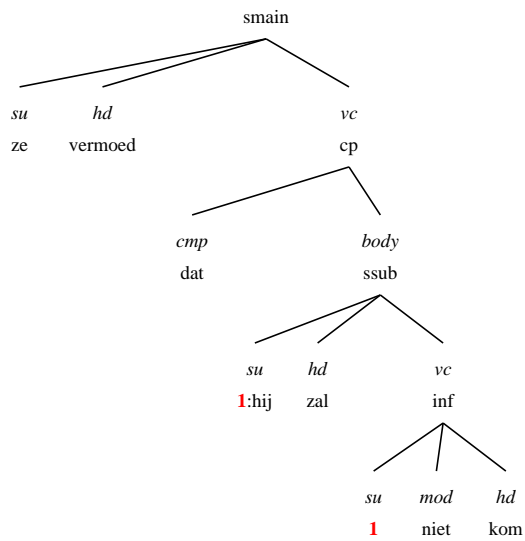
De besproken finiete en niet-finiete verbale projecties kunnen worden uitgebreid met complementeerder-elementen. In Figuur 2.3 geven we de categorielabels die door complementeerders worden geprojecteerd. We gebruiken het dependentielabel CMP voor de complementeerder en BODY voor de romp van de verbale projectie. (Voor de behandeling van extensies van het verbale domein met (vragende, betrekkelijke) voornaamwoorden met labels REL,WHREL,WHQ,WHSUB, verwijzen we naar sectie 3 op bladzijde 159.

CP-frases kunnen een complementsrol of een modificeerder-rol vervullen.

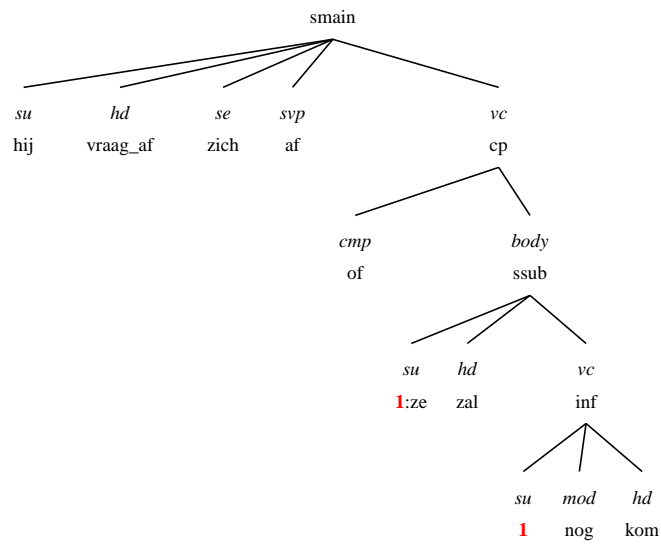
2.2.1.1 Bijzinnen met *dat*, *of*, *om*

De complementsrol betreft vooral bijzinnen ingeleid met de onderschikkende voegwoorden *dat*, *of* en *om*.

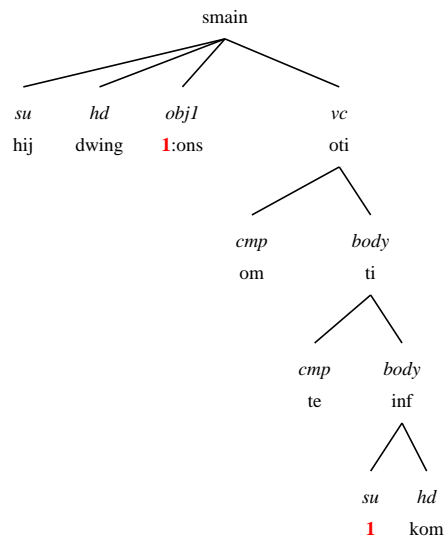
(203) ze vermoedde **dat hij niet zou komen**



(204) hij vroeg zich af **of ze nog zou komen**



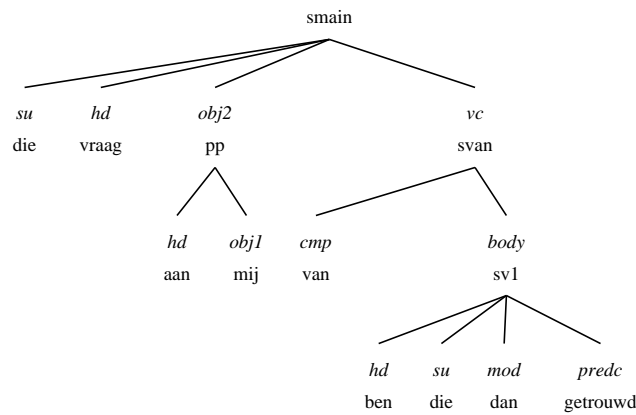
(205) hij dwong ons **om te komen**



2.2.1.2 Bijzinnen met *van*

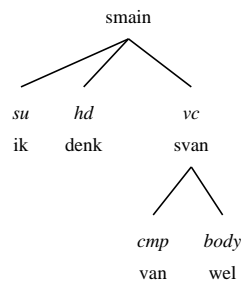
We beschouwen *van* in het voorbeeld hieronder als complementeerder ter inleiding van een complementszin in ‘directe rede’ (SV1 of SMAIN). De bijzin inclusief *van* krijgt categorie SVAN. Ook zinnen zoals (207) krijgen categorie SVAN.

(206) die vroeg aan mij van: is die dan getrouwd?



(207) dan heb ik zoiets van: laat maar...

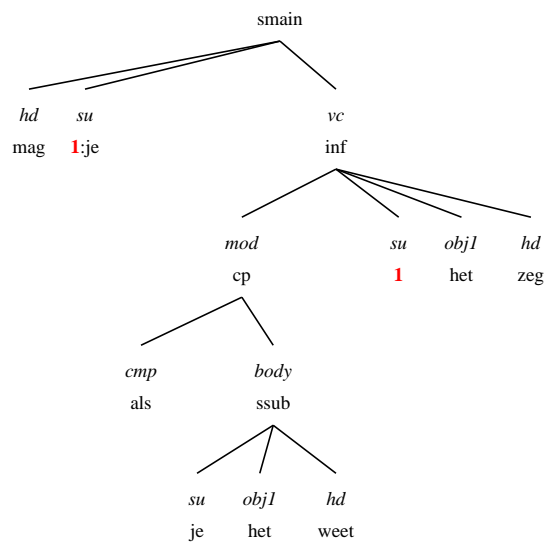
(208) ik denk/hoop/verwacht/vermoed/... van wel



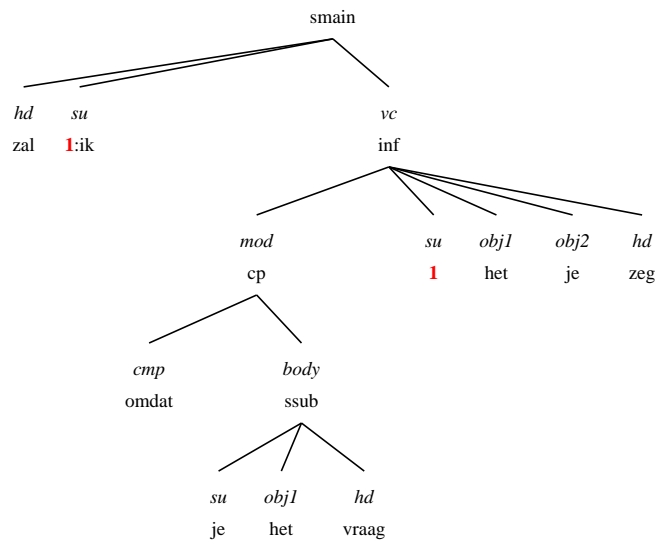
2.2.1.3 Bijzinnen als modificeerder

In de voorbeelden hieronder speelt de CP een modifierende rol. Het argument van de complementeerder kan, zoals hier blijkt, verschillende categorieën hebben.

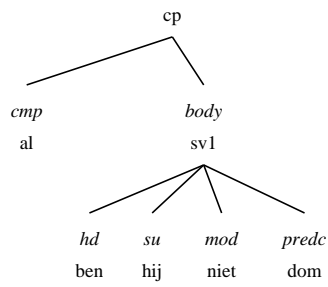
(209) **als je het weet**, mag je het zeggen



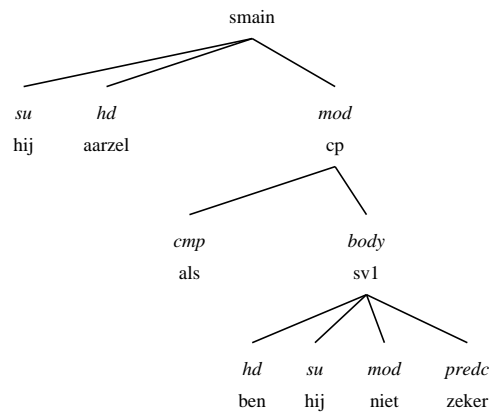
(210) **omdat je het vraagt**, zal ik het je zeggen



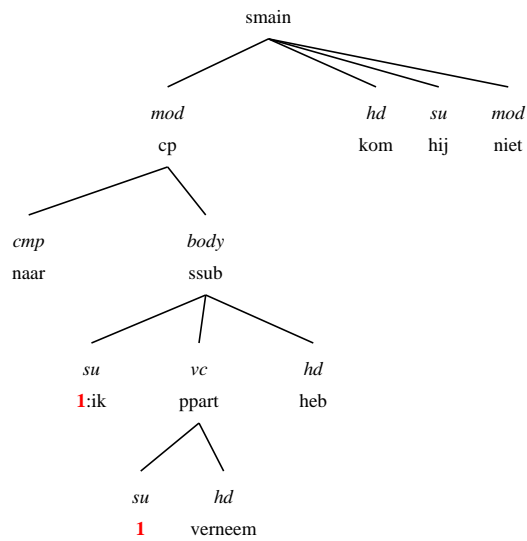
(211) we bieden hem de beurs niet aan, **al is hij niet dom**



(212) Hij aarzelde **als was hij niet zeker**



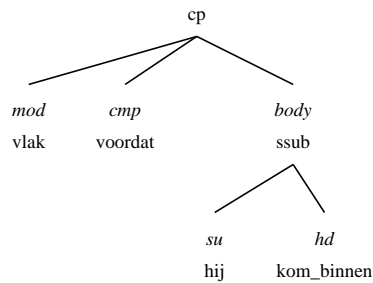
(213) **Naar ik vernomen heb**, komt hij niet



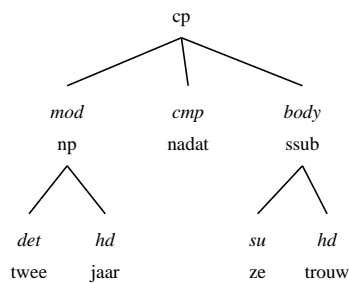
2.2.1.4 Modificatie van complementeerdere zoals *voordat*, *nadat*

Net zoals voorzetsels kunnen complementeerdere soms worden voorafgegaan door een modifierend bijwoord, of een modifierende NP die een maat of duur uitdrukt.

(214) Vlak voordat hij binnenkwam



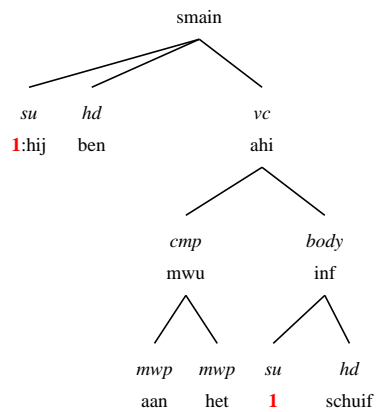
(215) Twee jaar nadat ze trouwde



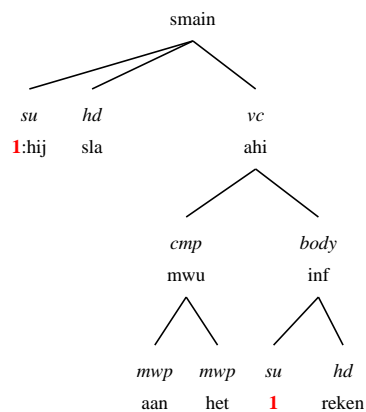
2.2.1.5 Niet-finiete bijzinnen als complement

In het geval van AHI behandelen we de combinatie *aan het* als complexe (meerwoord) complementeerder. De categorie AHI wordt ook gebruikt voor infinitieven die met de complementeerder *uit* combineren. Bij de niet-finiete categorieën OTI, TI en AHI is het vaak zo, dat het onderwerp van het ingebedde werkwoord gecontroleerd wordt door een van de argumenten van het dominerende werkwoord.

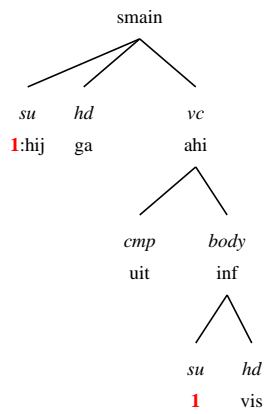
(216) Hij is aan het schuiven



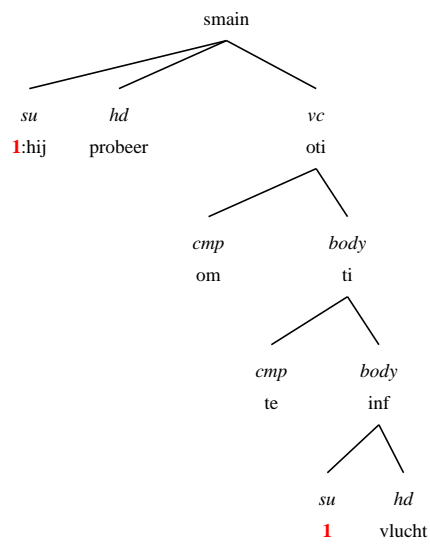
(217) Hij sloeg aan het rekenen



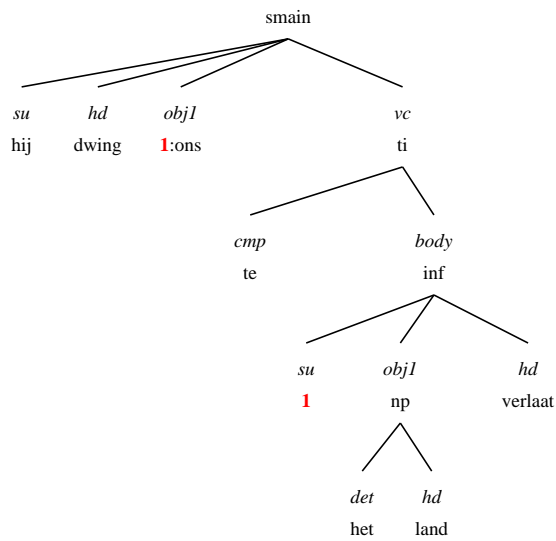
(218) Hij gaat uit vissen



(219) Hij probeerde om te vluchten

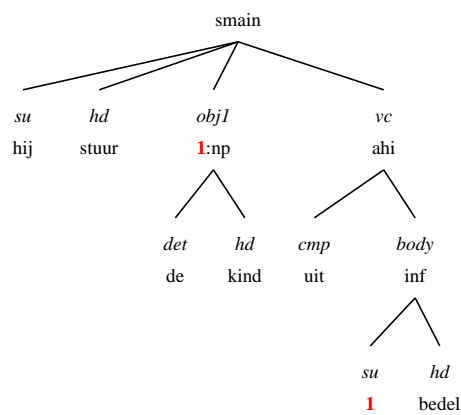


(220) Hij dwong ons het land te verlaten

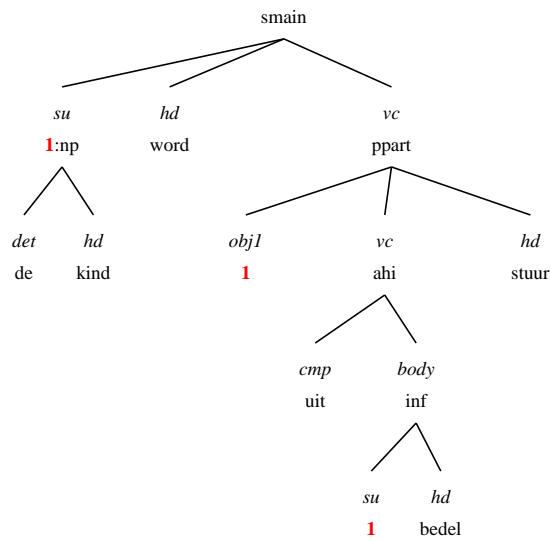


Complementen met de complementizer *uit* zoals in voorbeeld (218) vervullen normaliter de rol van VC.

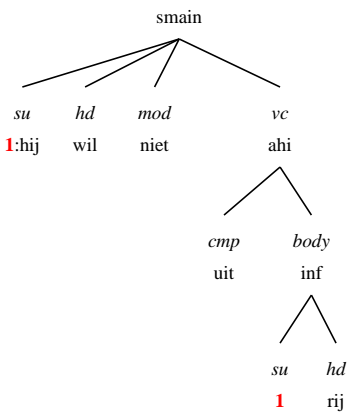
(221) Hij stuurde de kinderen uit bedelen



(222) De kinderen werden uit bedelen gestuurd

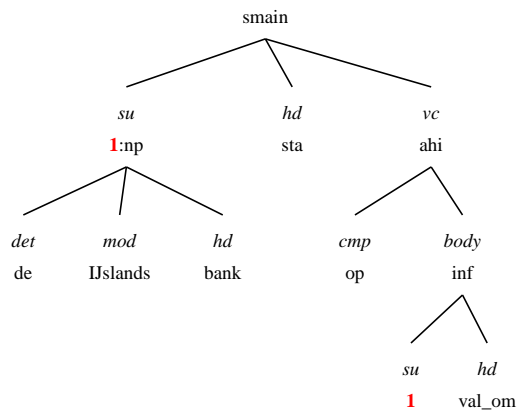


(223) Hij wil niet uit rijden



De ANS (Haeseryn, 1997, blz. 1056–7) suggereert dat ook in de volgende voorbeelden met *op*+infinitief sprake is van een soort van verbaal complement is, en wij volgen deze suggestie:

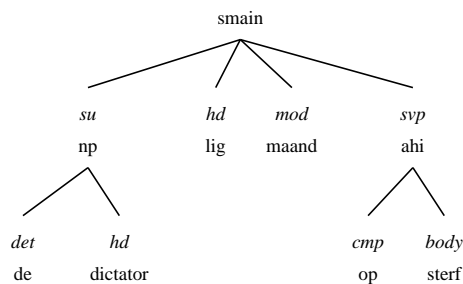
(224) De IJslanse bank stond op omvallen



(225) De dam staat op doorbreken

De uitdrukking *op sterven liggen* wordt als werkwoordelijke uitdrukking geannoteerd:

(226) De dictator lag maanden op sterven



2.2.2 Complexe voegwoorden

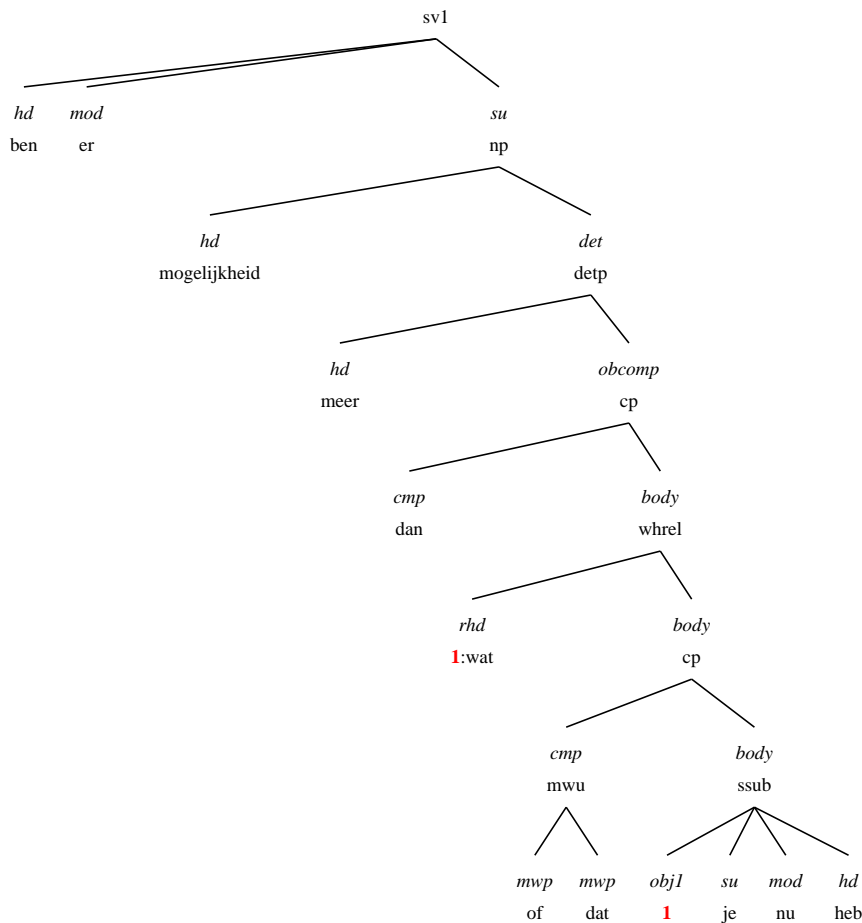
Het is een klassiek gegeven dat de hoofdpositie van ingebedde Nederlandse zinnen soms door meer dan een element tegelijk kan worden vervuld.

(227) zijn er meer mogelijkheden dan *wat of* je nu hebt?

(228) ik zie niet in *wie dat* dat allemaal gaat kunnen realiseren

Varianten met *of* schijnen in België niet of nauwelijks voor te komen, vandaar dat we varianten met drie elementen alleen in Nederland verwachten:

(229) zijn er meer mogelijkheden dan *wat of dat* je nu hebt?



(230) ik zie niet in *wie of dat* dat allemaal gaat kunnen realiseren

De combinaties *of dat* behandelen we als een MWU, die de rol van CMP vervult. Het vraagwoord gedeelte kan echter niet deel zijn van een MWU, omdat in principe ook complexe WH groepen kunnen optreden en er dus sprake van productiviteit is:

(231) Ik vroeg met welke assistent of dat hij komt

(232) Ik vroeg me af welke journalist of dat hij gesproken had

In zulke gevallen kiezen we dus voor een WHD/BODY structuur waar bij de BODY bestaat uit een CP met als CMP de multi-word-unit *of dat*.

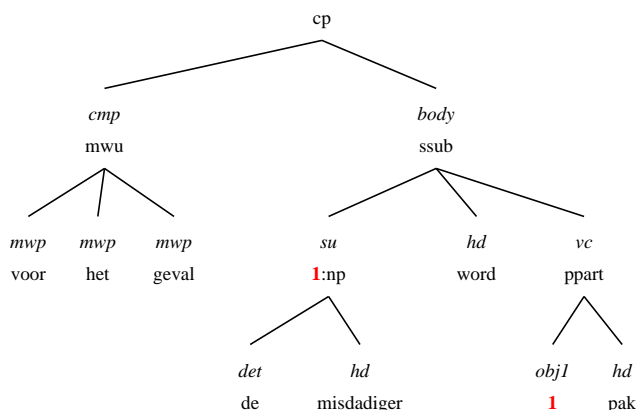
Het komt ook voor dat een voegwoord schijnbaar ontbreekt. In beknopte bijzinnen weten we dat *om* regelmatig weg kan blijven: het label van de bijzin wordt OTI of TI, al naar gelang de aan- of afwezigheid van *om*:

(233) je dwingt me **(om) strenger op te treden** (TI | OTI als VC)

(234) de kans (om) een gouden medaille te winnen (TI | OTI als VC)

Er zijn evenwel ook gevallen waarin het voegwoord *dat* ontbreekt:

(235) voor het geval de misdadiger wordt gepakt



(236) voor het geval ik in de verleiding kwam

De betekenis van deze constructie (*voor*) *het geval* is ongeveer ‘indien’, en *dat* kan alleen wegblijven als *geval* niet gemodificeerd wordt (**voor het zeldzame geval ik in de verleiding kwam*); *het* kan niet vervangen worden door iets anders (**voor een geval ik in de verleiding kwam*, **voor dat geval ik in de verleiding kwam*).

Syntactisch fungeert *voor het geval* hier als een voegwoord (CMP), en zo analyseren we de combinatie ook. Andere enigszins vergelijkbare voegwoorden met een frasale herkomst zijn onder meer *indien*, *voor zover* (*voorzover*) en het nauw verwante *ingeval*.

2.2.3 Of-zinnen

In een zin als

(237) ik zal (hem) eens bellen of hij komt

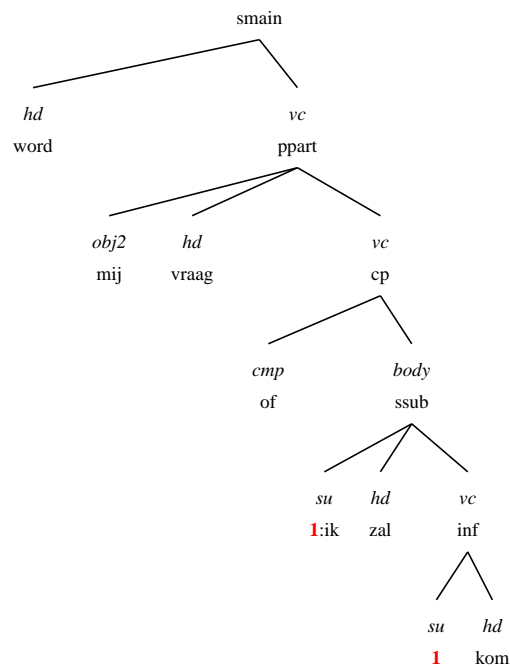
noemen we het *of hij komt*-‘complement’, dat een doel aangeeft, VC, naar analogie met:

(238) ik zal (hem) eens vragen of hij komt

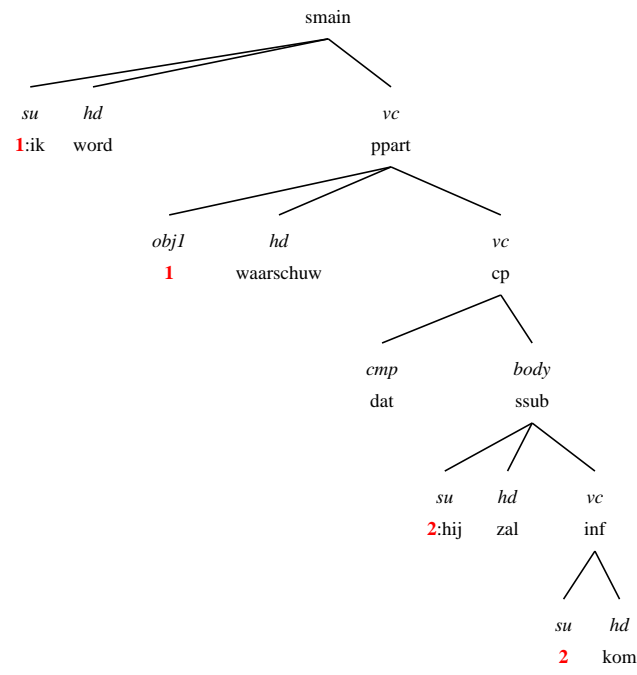
waarin de *of*-zin ook de rol van VC vervult: *of hij komt* kan worden vervangen door een pronominaal lijdend voorwerp (*dat*) en als degene aan wie iets gevraagd wordt, uitgedrukt wordt, dan gebeurt dat met een indirect object.

In een aantal gevallen kan de VC niet vervangen worden door een pronominaal lijdend voorwerp, en lijkt een andere NP de rol van OBJ1 te vervullen, zoals in de voorbeelden (240) en (241). Bij werkwoorden zoals *verzoeken* lijken taalgebruikers beide mogelijkheden tot hun beschikking te hebben. In voorbeeld (242) is *reizigers* niet als onderwerp gebruikt, en blijft het dus OBJ2 bij *verzoeken*. In voorbeeld (243) wordt *reizigers* blijkbaar als onderwerp van het passieve hulpwerkwoord gebruikt, en moet dus wel de rol van OBJ1 bij *verzoeken* spelen.

(239) Mij werd gevraagd of ik zou komen

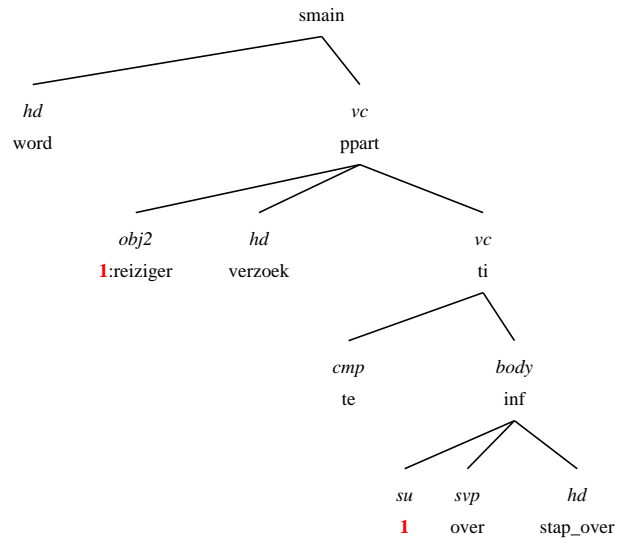


(240) Ik werd gewaarschuwd dat hij zou komen

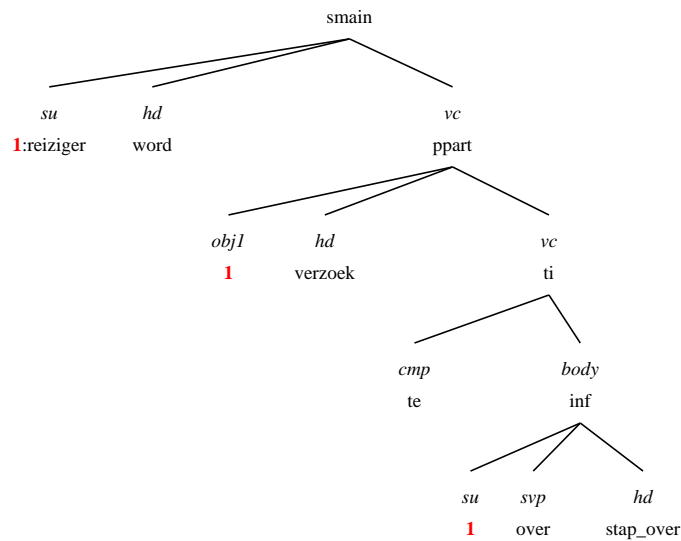


(241) Ik werd getipt dat hij zou komen

(242) Reizigers wordt verzocht over te stappen



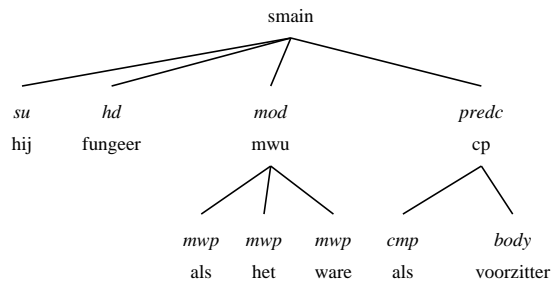
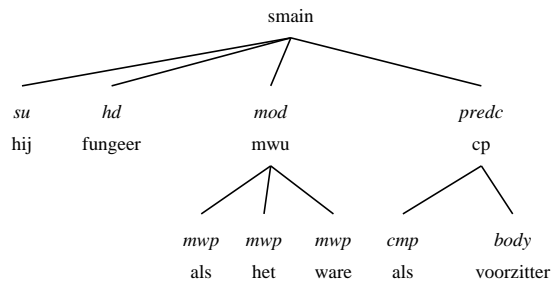
(243) Reizigers worden verzocht over te stappen



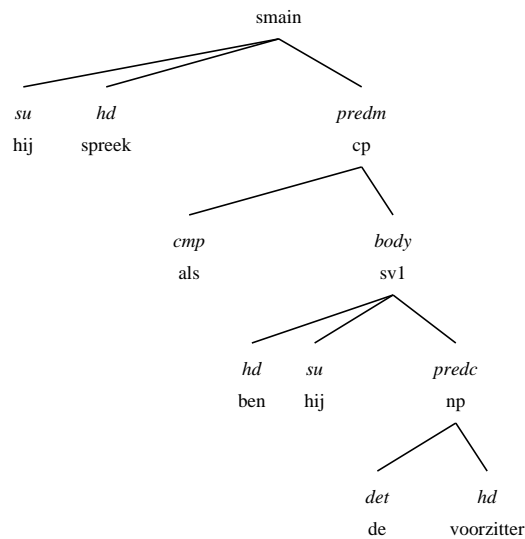
2.2.4 Als het ware

De frase *als het ware* in zinnen zoals in (244) wordt geanalyseerd als een MWU die fungeert als MOD. Zinnen met *als ware/was* zoals in voorbeeld (245) worden geanalyseerd zoals voorbeeld (212).

(244) Hij fungeerde als het ware als voorzitter



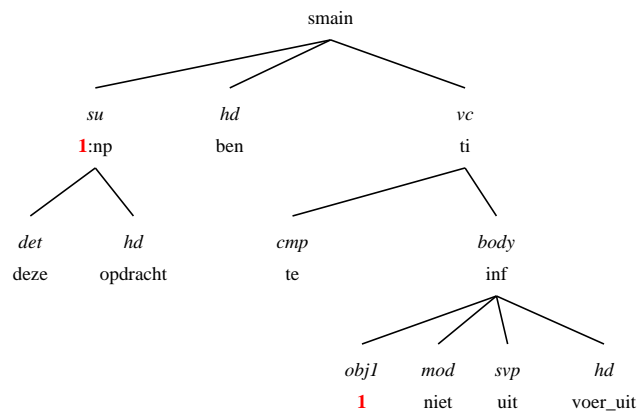
(245) Hij sprak als ware/was hij de voorzitter



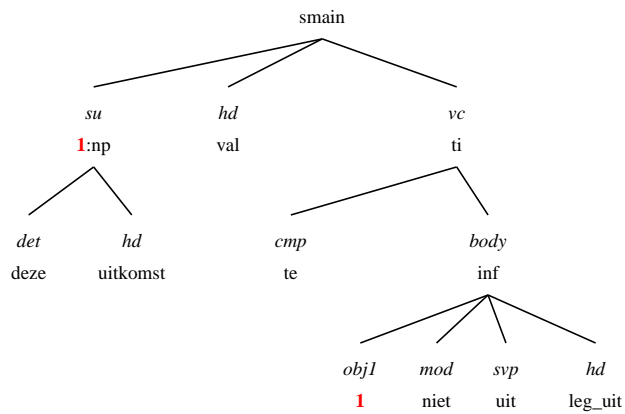
2.2.5 Dat is goed te doen

Het predikaat (TI) van zinnen als *dat is (goed, nauwelijks, ...) te doen (doenbaar/doenlijk)* en *dat is (goed, nauwelijks, ...) te zien (zichtbaar)* wordt geanalyseerd als een VC met de interne opbouw *goed: MOD, te: CMP, rest: BODY*. Deze constructie wordt dus geanalyseerd als een passief constructie met een te-infinitief:

(246) Deze opdracht was niet uit te voeren



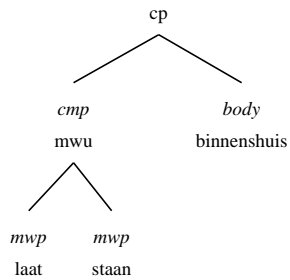
(247) Deze uitkomst valt niet uit te leggen



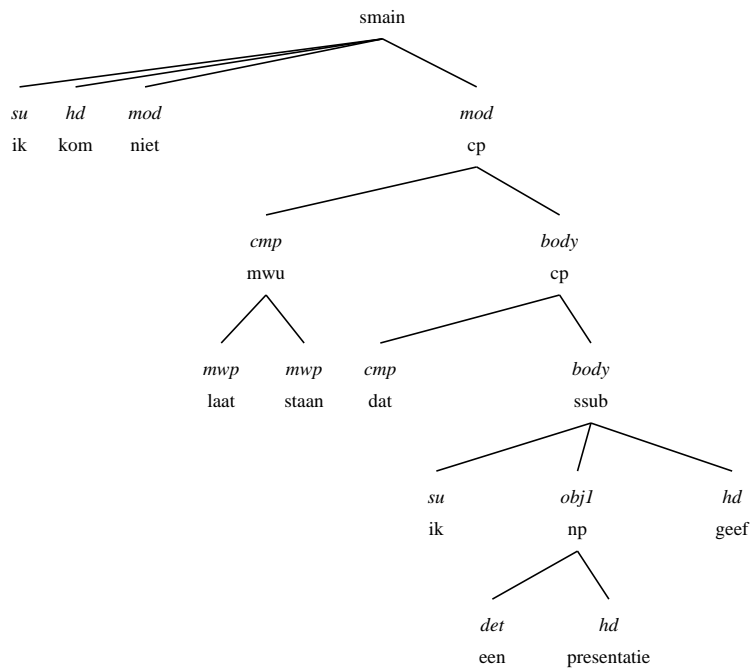
2.2.6 Laat staan

Laat staan fungeert doorgaans als voegwoord. De combinatie wordt als multiword-unit geanalyseerd.

(248) Laat staan binnenshuis



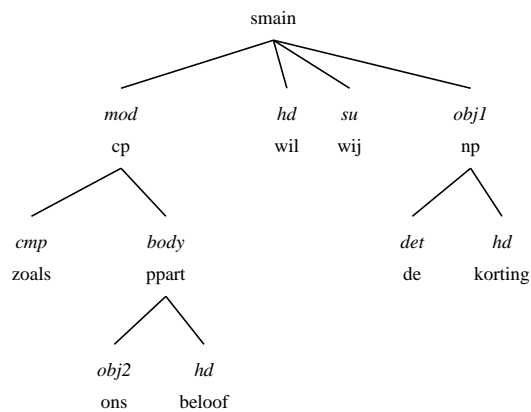
(249) Ik kom niet laat staan dat ik een presentatie geef



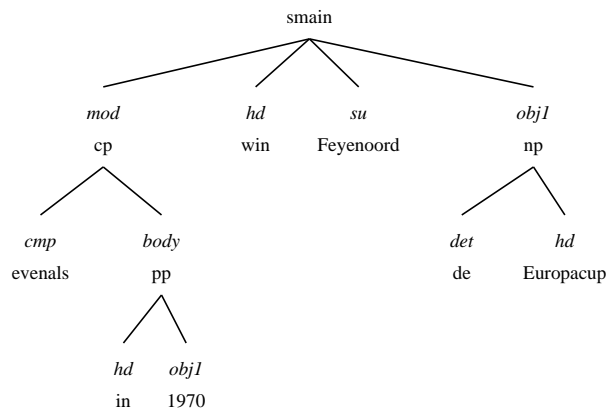
2.2.7 Onderschikkende voegwoorden zonder verbaal vervolg

Het is niet altijd zo dat een onderschikkend voegwoord noodzakelijkerwijs combineert met een bijzin. Complementeerdere combineren ook met andere woordgroepen, en ook in die gevallen gebruiken we de CMP BODY analyse.

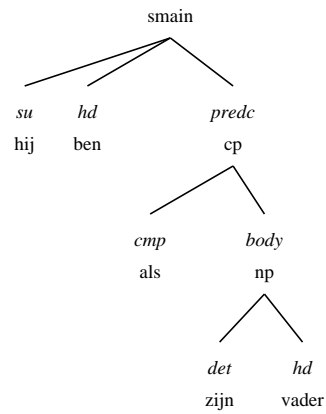
(250) **Zoals ons beloofd** willen wij de korting



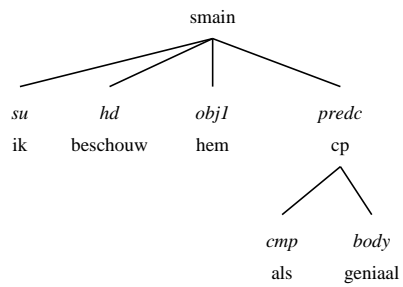
(251) **Evenals in 1970** won Feyenoord de Europracup



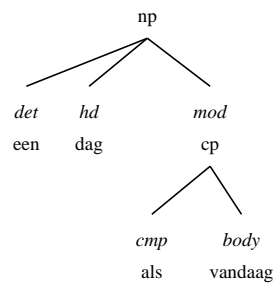
(252) Hij is **als zijn vader**



(253) Ik beschouw hem **als geniaal**



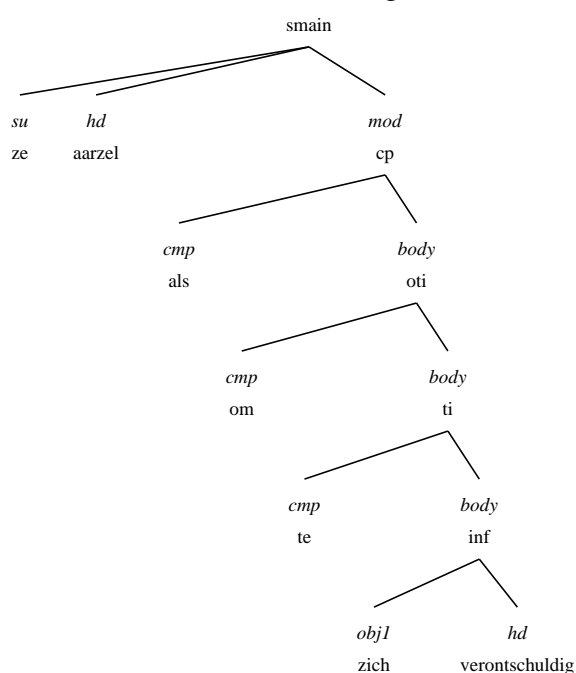
(254) Een dag **als vandaag**



	dependentielabel	OMSCHRIJVING
<i>hoofd</i>	HD	hoofd van AP
<i>compl</i>	OBJ1 SE POBJ1 PC VC OBCOMP ME	oorzakelijk voorwerp verplicht reflexief complement voorlopig oorzakelijk voorwerp voorzetselvoorwerp verbaal complement vergelijkingscomplement maatcomplement
<i>mod</i>	MOD	bepaling

Figuur 2.4: dependentielabels: adjectiefdomein

(255) Ze aarzelde als om zich te verontschuldigen



2.3 Het adjectiefdomein

In Figuur 2.4 geven we de dependentielabels voor adjectiefgroepen. Het categorielabel voor de moederknoop is AP.

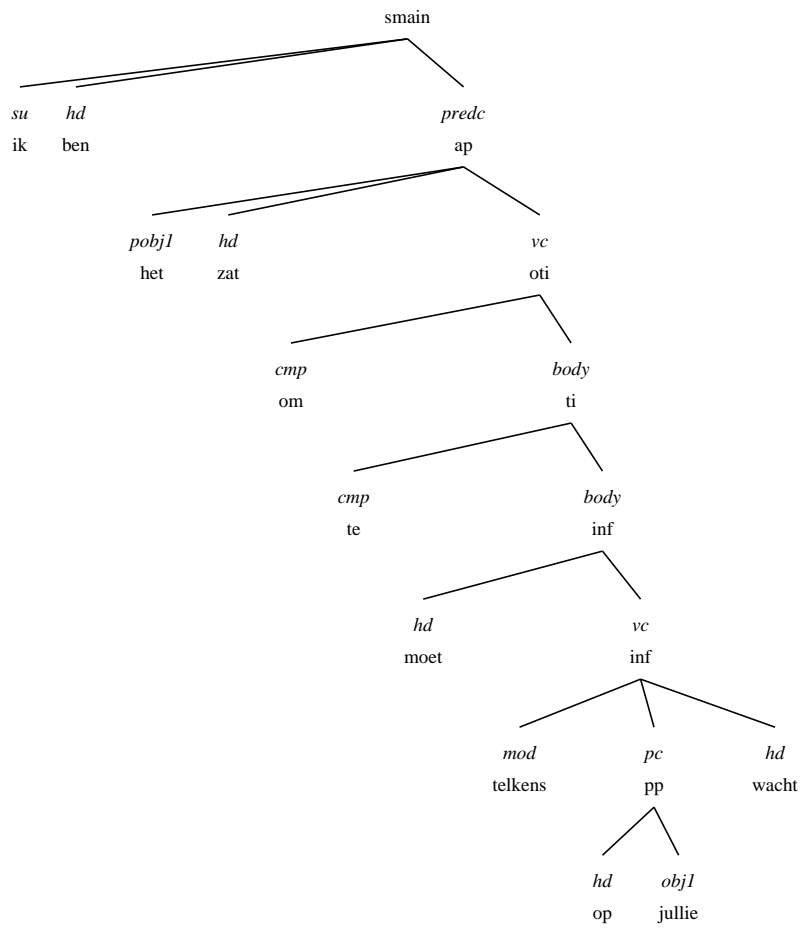
2.3.1 PC, VC, OBJ1

In de adjectiefgroep treffen we de dependentielabels PC en VC aan voor complementen die we ook in het verbale domein hebben ontmoet.

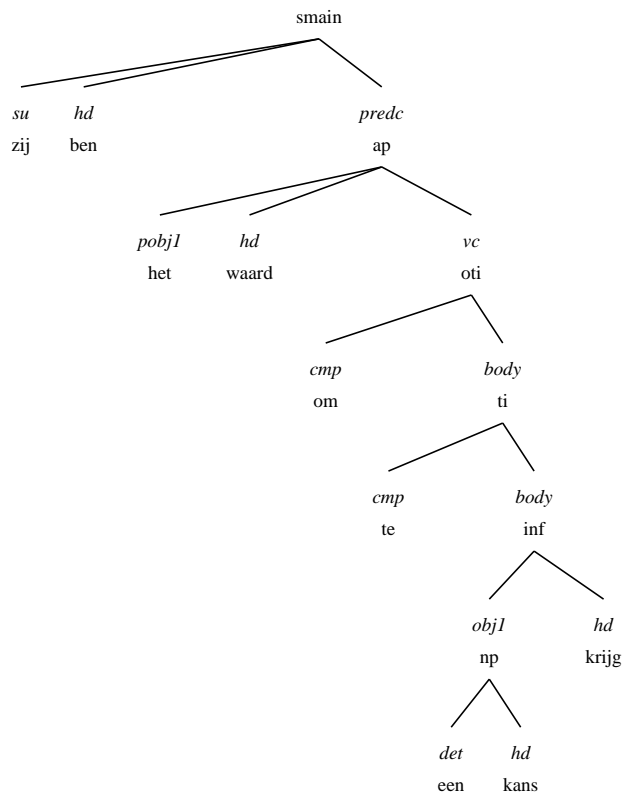
- (256) bereid **tot samenwerking** (PC)
- (257) zij is **op wraak** belust (PC)
- (258) zij is bereid **(om) hem te helpen** (VC)
- (259) zij is bang **dat hij zou opvallen** (VC)
- (260) ik ben nieuwsgierig **wanneer hij komt** (VC)

Het label OBJ1 is voor het naamwoordelijk complement dat traditioneel ‘oorzaikelijk voorwerp’ wordt genoemd. Merk op dat, zoals in het verbale domein, een voorlopige realisering (POBJ1) mogelijk is in combinatie met een (beknopte) bijzin.

- (261) zij is **het gezeur** moe (OBJ1)
- (262) zij is **het** zat (OBJ1)
- (263) wij zijn **het** eens (OBJ1)
- (264) wij zijn **dat** eens (OBJ1)
- (265) ik ben **het** zat **om telkens op jullie te moeten wachten**



- (266) ik ben het zat dat jullie niet luisteren (POBJ1 en VC)
- (267) Zij is de moeite waard (OBJ1)
- (268) Zij is het waard om een kans te krijgen (POBJ1 en VC)



2.3.2 Reflexief complement

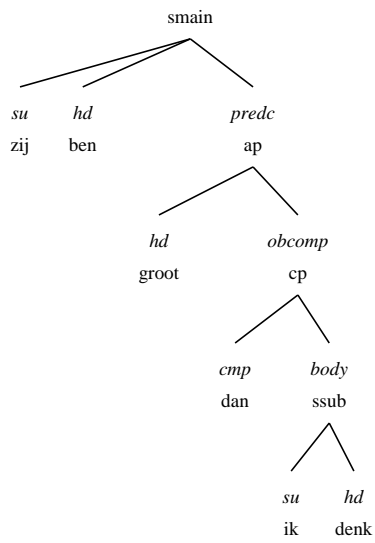
Het label SE wordt gebruikt voor verplicht reflexieve complementen van adjectieven:

(269) hij is (zich | *de jongens) bewust van de consequenties

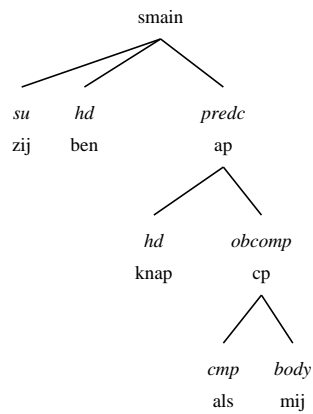
2.3.3 Comparatieven

Het label OBCOMP is voor het complement van comparatieven. Een OBCOMP-complement kan ook voorkomen in het gezelschap van andere complementen, zoals in het derde voorbeeld:

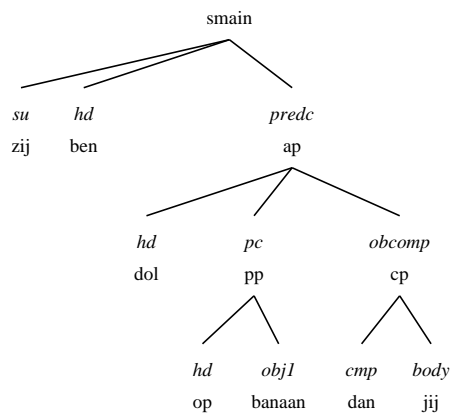
(270) zij is groter dan ik dacht



(271) zij is knapper als mij



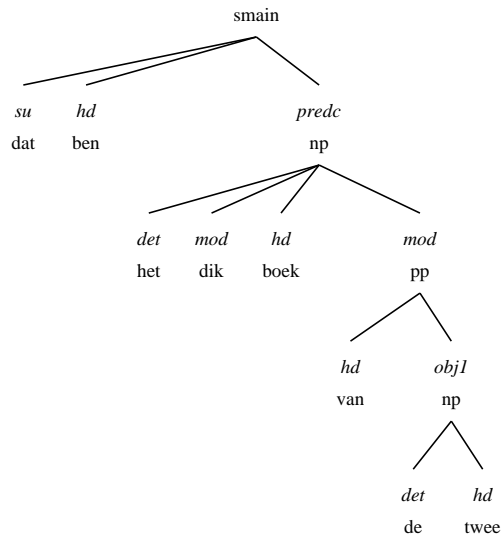
(272) zij is doller op bananen dan jij



Comparatieven en superlatieven kunnen soms ook vergezeld worden door een

(optionele) partitief. Deze partitief benoemen we als een modificeerder van het zelfstandig naamwoord.

(273) dat is het dikkere boek van de twee



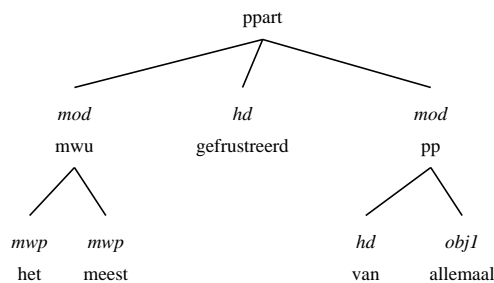
(274) dat is het dikste boek van de twee

(275) dat is het dikke boek van de twee

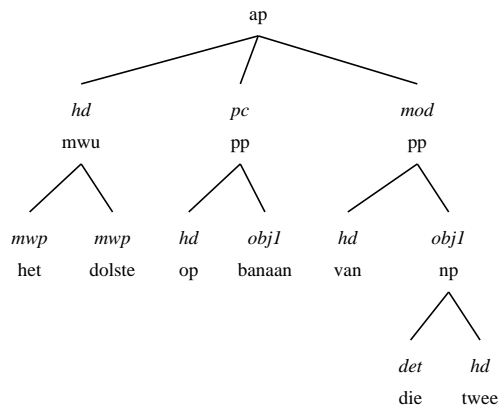
2.3.4 Superlatieven

De annotatie van superlatiefconstructies kent geen bijzonderheden. We volstaan hier met de opmerking dat we het element *het* in de voorbeelden hieronder een puur grammaticale functie toedichten (voorbeelden waarbij geen sprake is van een verzwegen, begrepen zelfstandig naamwoord). We maken daarom een MWU van *het* met de superlatief.

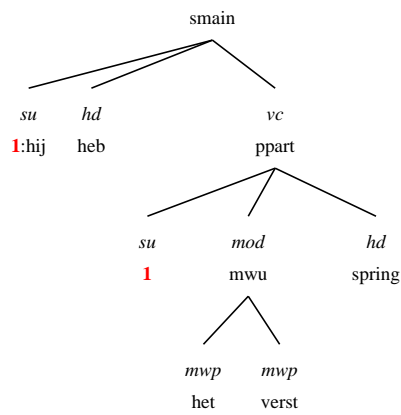
(276) het meest(e) gefrustreerd van allemaal



(277) het dolst(e) op bananen van die twee



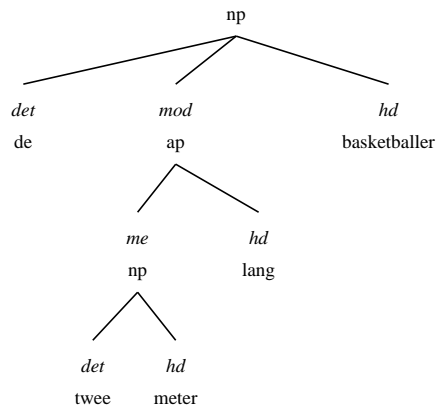
(278) Hij heeft het verst(e) gesprongen



2.3.5 Maatcomplement

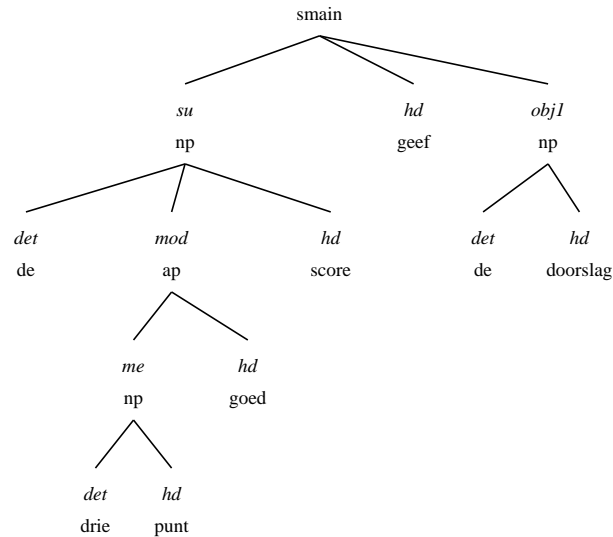
We gebruiken het label ME voor maatcomplementen voor adjectieven zoals *lang*, *diep*, *hoog*, en voor maatcomplementen die optreden bij comparatieven:

(279) De twee meter lange basketballer



(280) De drie jaar oude peuter

(281) De drie punten betere score gaf de doorslag



2.3.6 Andere complementen

Naast de genoemde complementen komen in het adjectief-domein ook andere complementen voor die we eerder bij het werkwoordelijk domein tegenkwamen, bijvoorbeeld in geval van deverbale adjectieven. Enkele voorbeelden:

(282) Hij is *ons veel geld* schuldig (OBJ2 en OBJ1)

(283) De *ons goed* gezinde ambtenaren (OBJ2 en PREDC)

(284) De *mij* loyale medewerkers (OBJ2)

2.3.7 Modificatie

Het dependentielabel voor bepalingen binnen de adjectivische constituent is MOD. De bepalingen hebben verschillende syntactische realiseringen. Hieronder enkele voorbeelden.

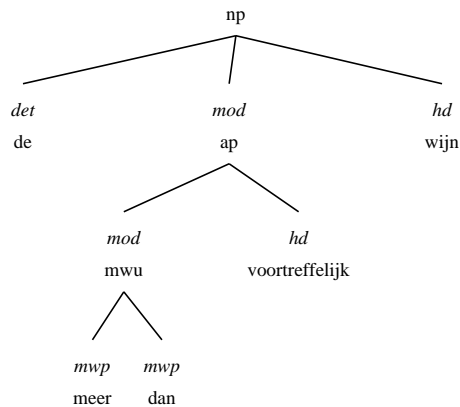
(285) **heel erg** druk (AP als MOD)

(286) **nogal** druk (ADV als MOD)

(287) vers **van de pers** (PP als MOD)

(288) **het meest** gefrustreerd (MWU als MOD)

(289) de **meer dan** voortreffelijke wijn (MWU als MOD)



(290) zijn **niet minder dan** sensationele debuut

(291) een **verre van** geslaagd optreden (MWU als MOD)

In de voorbeelden (289) en (290) is geen sprake van een OBCOMP omdat hier het bijvoeglijk naamwoord hoofd van de bepaling moet zijn. Dit blijkt uit de inflectie, en ook uit de onmogelijkheid om de hypothetische OBCOMP weg te laten of naar rechts te verplaatsen:

(292) *de meer wijn

(293) *de meer wijn dan voortreffelijke

(294) *zijn niet minder debuut

(295) *zijn niet minder debuut dan sensationele

In geval het modificeerders betreft van categorie NP die alleen met comparatieven gecombineerd kunnen worden, dan worden deze als maatcomplement ME geannoteerd:

(296) **een stuk** leuker (NP als ME)

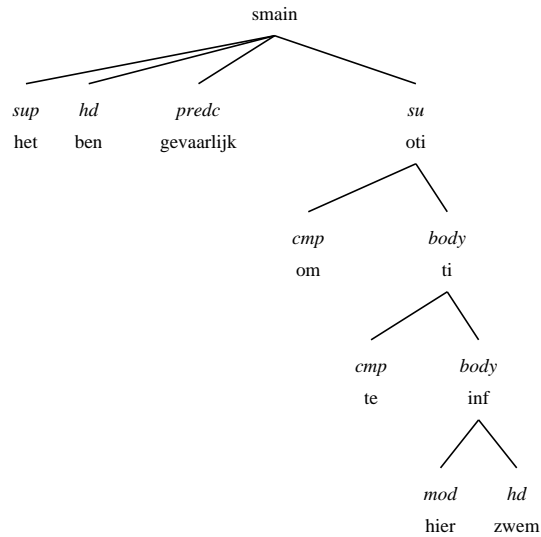
(297) ***een stuk** leuk

2.3.8 Het is ADJ om/dat ...

Een aantal adjectieven kan gecombineerd worden met een bijzin waarbij een copula aanwezig moet zijn:

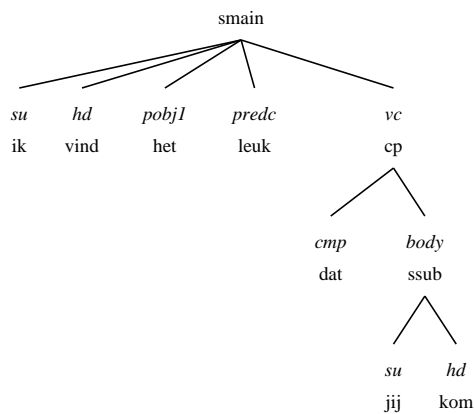
(298) Het is leuk dat je komt

(299) Het is gevaarlijk om hier te zwemmen



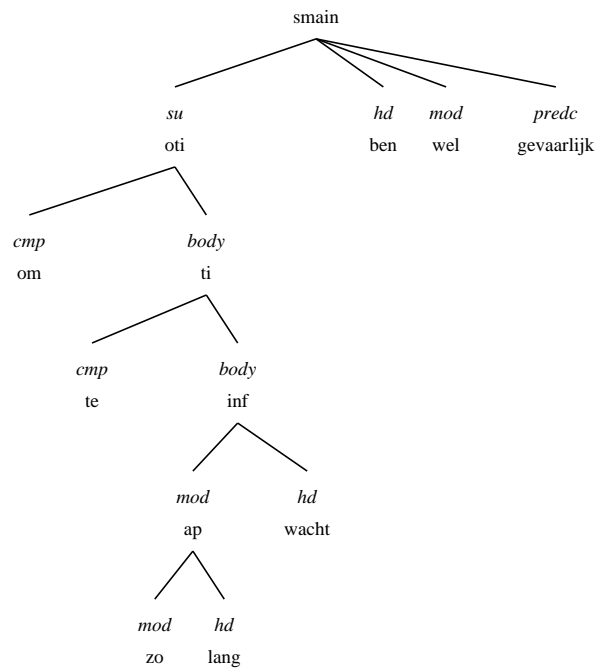
(300) Het is raadselachtig waarom hij niet kwam

(301) Ik vind het leuk dat je komt



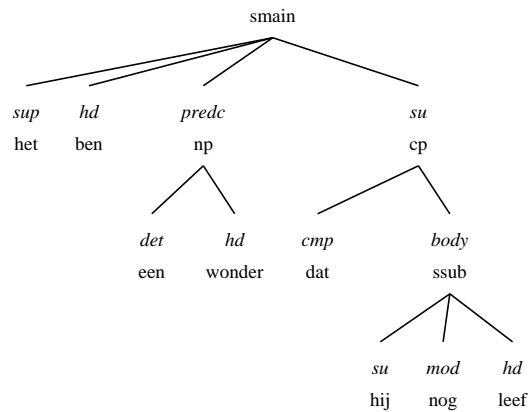
In deze gevallen vormt het adjectief geen constituent met de bijzin, in contrast met de eerdere voorbeelden 258, 259 en 260 op bladzijde 97. De bijzin fungeert als onderwerp van de zin, en *het* annoteren we als voorlopig onderwerp; bij werkwoorden zoals *vinden* is dat respectievelijk POBJ1 en VC. Merk op dat het woordje *het* niet in alle gevallen verplicht is:

(302) Om zo lang te wachten is wel gevaarlijk

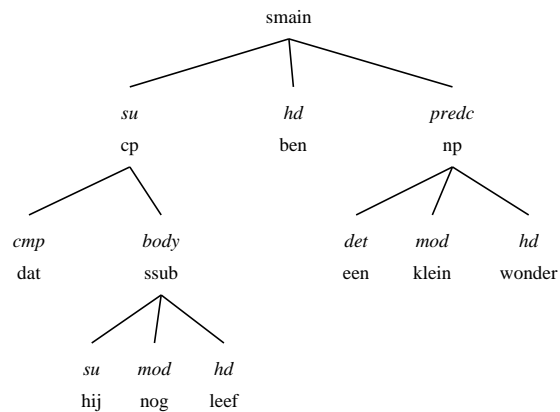


Een aantal zelfstandige naamwoorden kan op dezelfde manier gebruikt worden, en deze gevallen worden analoog geannoteerd. Deze constructie moet dus onderscheiden worden van voorbeelden zoals (305).

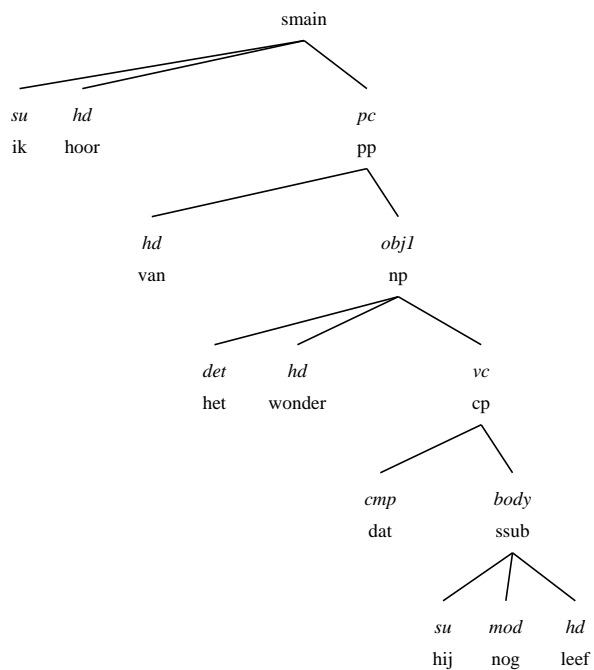
(303) Het is een wonder dat hij nog leeft



(304) Dat hij nog leeft is een klein wonder



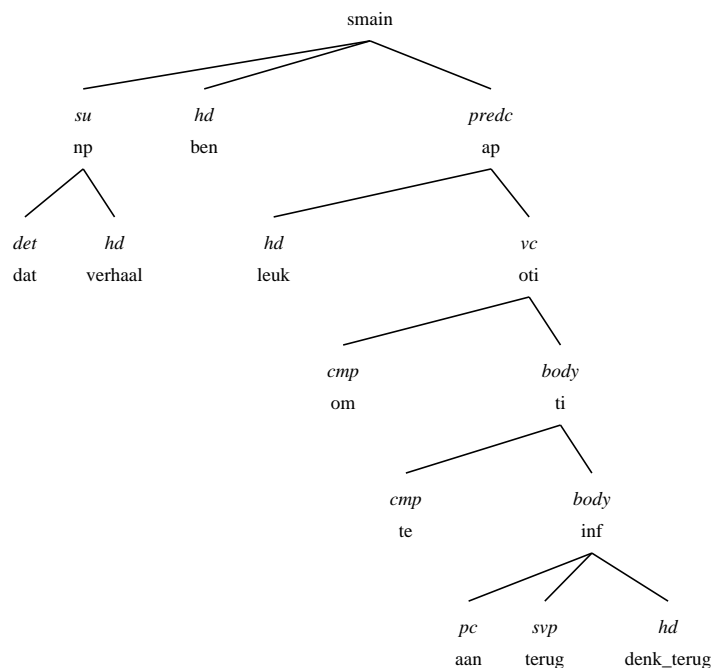
(305) Ik hoorde van het wonder dat hij nog leeft



2.3.9 Leuk om aan terug te denken

Complementen die de vorm nemen van een finiete bijzin (CP of WHSUB) of een beknopte bijzin (OTI of TI) worden geannoteerd als VC. In het algemeen zijn er twee soorten adjectieven die met een VC complement combineren. De volgende voorbeelden geven houvast om te bepalen of de combinatie adjectief met bijzin als complement moet worden beschouwd.

(306) dat verhaal is leuk **om te horen**



- (307) dat verhaal is leuk **om aan terug te denken**
- (308) het is leuk om dat verhaal te horen
- (309) het is leuk dat je het verhaal vertelt
- (310) *leuk om het verhaal te horen vertrok hij
- (311) *leuk dat alles goed afgelopen was vertrok hij
- (312) ik ben blij **om je te zien**
- (313) ik ben blij **dat je er bent**
- (314) *ik ben blij **om aan terug te denken**
- (315) *het is blij om je te zien
- (316) *het is blij dat je er bent
- (317) blij om alles in goede handen achter te kunnen laten vertrok hij
- (318) blij dat alles goed afgelopen was vertrok hij

Bijvoeglijke naamwoorden zoals *leuk* kunnen met een OTI als VC combineren, waarbij een complement in de OTI ontbreekt - zelfs de NP van een PP kan hier ontbreken zoals in voorbeeld (307). Deze klasse bijvoeglijke naamwoorden laat ook de SUP-SU constructie met copula toe (308), terwijl zonder syntactisch verband een complement niet mogelijk is (310).

Adjectieven zoals *blij* aan de andere kant kunnen altijd met een volledige VC optreden, *preposition stranding* is niet mogelijk zoals voorbeeld (314) aantoont. De

	dependentielabel	OMSCHRIJVING
<i>hoofd</i>	HD	hoofd van PP
	HDF	afsluitend element van circumpositie
<i>comp</i>	OBJ1	object van prepositie
	VC	verbaal complement van prepositie
	SE	verplicht reflexief object
	POBJ1	voorlopig object
	PREDC	predicerend complement
	LD	locatief/directioneel complement
<i>mod</i>	MOD	modificeerders binnen PP

Figuur 2.5: dependentielabels: prepositionele domeinen

SUP-SU constructie met copula is nu juist niet mogelijk, terwijl zonder syntactisch verband de VC nog steeds mogelijk is.

Omdat de mogelijkheid van de constructie in (306) en (307) lexicaal bepaald is, kiezen we hier voor de analyse waarbij de OTI een complement is, VC dus.

2.3.10 's Morgens vroeg

's Morgens vroeg is een ADVP met *vroeg* als MOD, en *'s morgens* als HD.

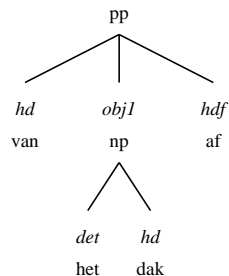
2.4 Het prepositionele domein

In Figuur 2.5 geven we de dependentielabels voor voorzetselgroepen. Het categorielabel voor de moederknoop is PP.

2.4.1 Prepositionele hoofden

Het voorzetsel is het hoofd van de PP. We onderscheiden twee hoofdlabels: HD en HDF. Het label HD wordt gebruikt voor de groepen waar er maar één hoofdkandidaat is. Het label HDF gebruiken we voor het afsluitende element van circumpositieconstructies (zie ANS (Haeseryn, 1997, blz. 526 e.v.) voor een opsomming). Het label HDF komt dus alleen voor in combinatie met HD. Door de combinatie HD+HDF krijgen de circumpositieconstructies een niet-geneste annotatie.

(319) van het dak af



(320) tot nu toe

(321) van hier uit

2.4.2 Reflexief object van een prepositie

In enkele gevallen is het object van een voorzetsel verplicht reflexief. In dat geval gebruiken we het label SE. Voorbeeld:

(322) hij houdt het voor (zich | *de jongens)

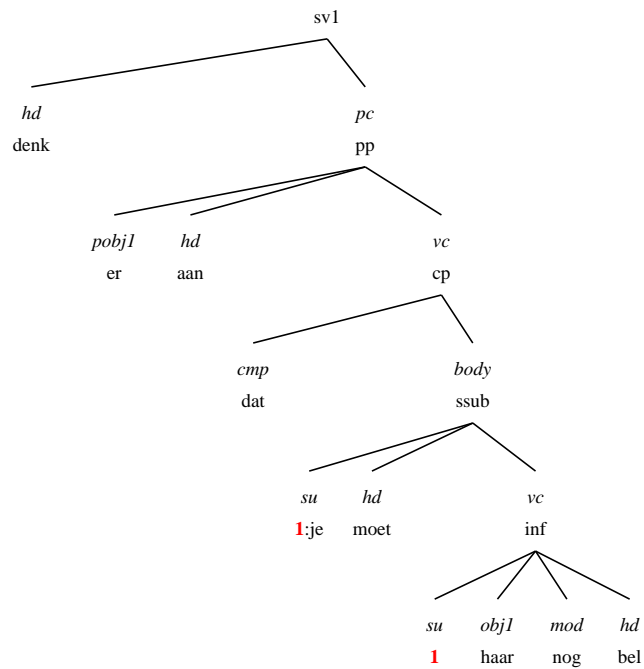
2.4.3 (Voorlopig) object van een prepositie

We maken een onderscheid tussen OBJ1 en VC, voor het 'echte' object van een prepositie, en POBJ1 voor *er* als 'voorlopig' object. Voor sententiële complementen gebruiken we het label VC, analoog met de analyse van complementen van werkwoorden.

Het label POBJ1 kan alleen voorkomen als ook OBJ1 of VC in het betreffende domein is toegekend. Vergelijk:

(323) je moet **er** niet aan denken (OBJ1)

(324) denk *(**er**) aan **dat je haar nog moet bellen** (POBJ1+VC)

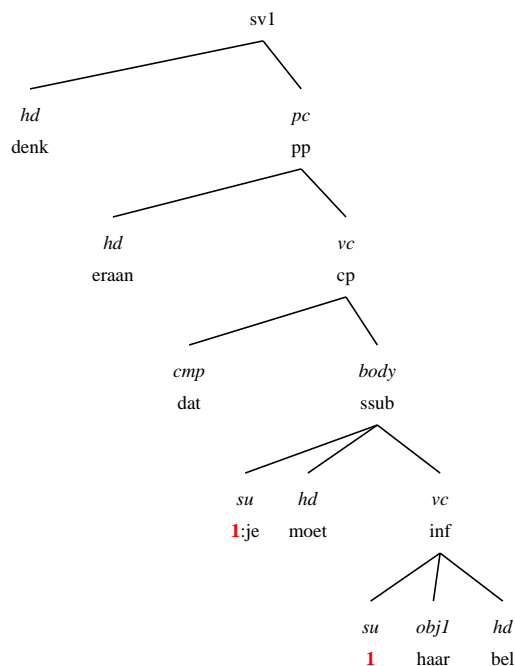


(325) denk *(er) aan **haar vanavond nog te bellen** (POBJ1+VC)

(326) ik heb *(er) niet aan gedacht **wie we moeten bellen** (POBJ1+VC)

De Nederlandse orthografie schrijft evenwel voor dat voornaamwoordelijke bijwoorden dikwijls aan elkaar geschreven moeten worden (*eraan* in plaats van *er aan*). In die gevallen beschouwen we *eraan* als HD.

(327) denk eraan dat je haar moet bellen

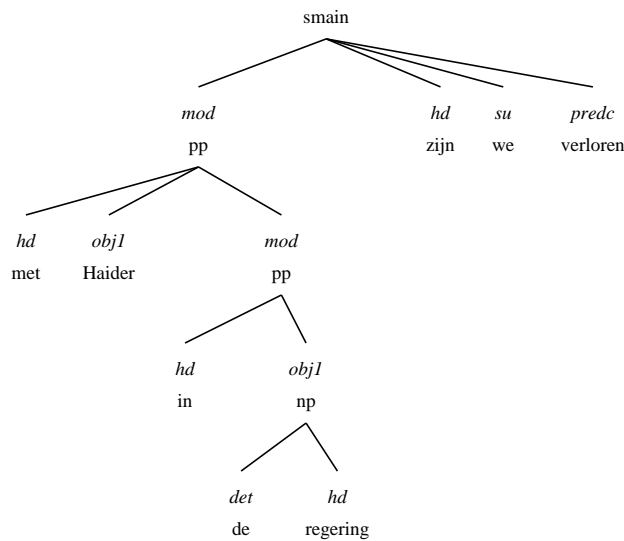


Het label OBJ1 heeft verschillende syntactische realisaties: NP, PP, ADVP, AP; de dependentie VC is gereserveerd voor TI, OTI, CP en WHSUB.

2.4.4 Absolute *met*-constructies

De absolute *met*-constructie is een PP-constituent die de rol van bijwoordelijke bepaling (MOD) of bepaling van gesteldheid (PREDM) vervult. Behalve het prepositioneel object treedt in deze constructie nog een tweede complement op met een MOD of PREDC functie.

(328) met Haider **in de regering** zijn we verloren (MOD)



(329) met de helft van de ploeg **dronken** zijn we verloren (PREDC)

(330) zonder wapens **in huis** zijn we verloren (MOD)

Er kunnen eventueel meerdere MOD elementen optreden:

(331) met hem **zonder ervaring ongekleed dronken achter het stuur** konden problemen niet uitblijven

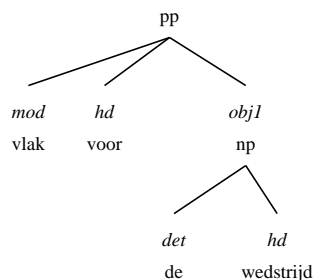
Onduidelijk is hier het van CGN geërfde onderscheid tussen MOD en PREDC. We hanteren de heuristiek dat woordgroepen met categorie AP en TI als PREDC optreden, en andere woordgroepen als MOD. Het label LD dat in CGN wel werd gebruikt, wordt niet gebruikt ten faveure van het label MOD.

2.4.5 Modifierende elementen

Zoals in het verbale domein, gebruiken we een enkel dependentielabel (MOD) voor de bepalingen binnen de voorzetselconstituent.

(332) **pal** achter het voorzetsel

(333) **vlak** voor de wedstrijd

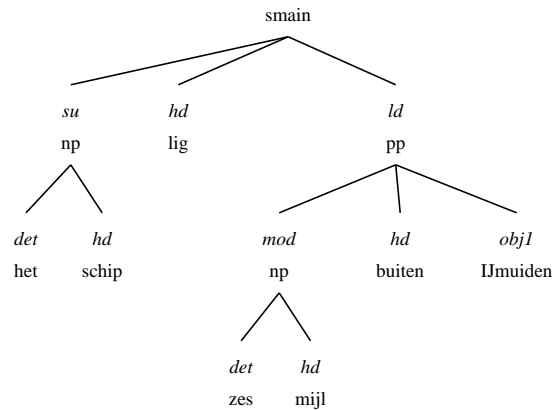


Bij constructies zoals de volgende is het vaak lastig de juiste analyse te bepalen:

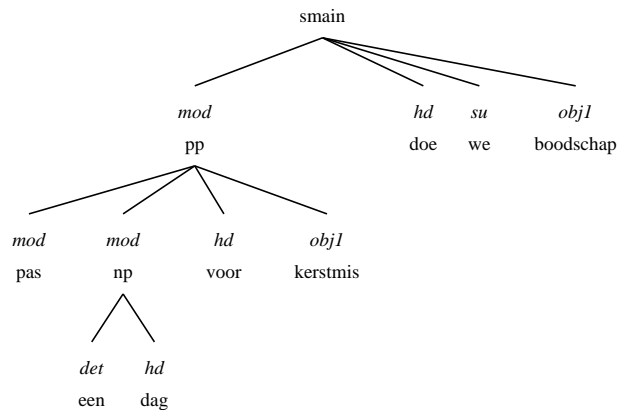
(334) **een uur** voor de wedstrijd

(335) **tientallen meters** onder de zeespiegel

(336) Het schip ligt **zes mijl** buiten IJmuiden



(337) **Pas een dag** voor kerstmis deden we boodschappen



Is bij deze voorbeelden de vetgedrukte NP het hoofd, waarbij de PP als modificeerder optreedt, of is de NP een modificeerder bij de PP? De ANS (Haeseryn, 1997, blz. 923) kiest voor de tweede mogelijkheid, en wij volgen deze analyse.

Het bijwoord *thuis* lijkt soms een PP te modificeren:

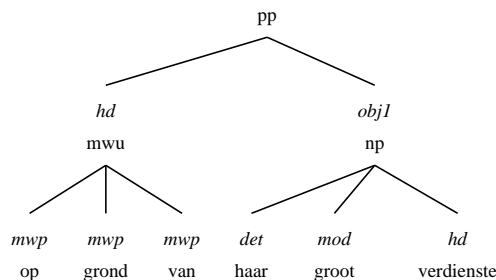
(338) bij hem **thuis**

Een andere mogelijkheid is, dat het bijwoord de NP modificeert (vergelijk met *de mensen thuis*). We kiezen voor de laatste mogelijkheid omdat er verder geen gevallen lijken te bestaan waarbij een bijwoord rechts van de PP als modificeerder van die PP kan optreden.

2.4.6 Voorzetseluitdrukkingen

Met de ANS (Haeseryn, 1997, blz. 510–11) beschouwen we voorzetseluitdrukkingen als lexicaal gefixeerde groepen (ingeleid door een voorzetsel) die in hun geheel de functie van een voorzetsel vervullen. Voorbeelden zijn *op grond van* in (339) en *aan de hand van* in (340). Ze worden daarom geannoteerd als multiword-units.

(339) *op grond van* haar grote verdiensten werd zij in de adelstand verheven



(340) *aan de hand van* deze plattegrond zul je je wel kunnen oriënteren

Groepen als deze kunnen vaak ook een letterlijke betekenis hebben; het zijn dan geen voorzetseluitdrukkingen:

(341) we bouwen hier *op grond van* de gemeente Wymbritseradeel

(342) het kindje liep *aan de hand van* zijn moeder

Je kunt ook nog voornaamwoordelijke bijwoorden krijgen bij voorzetseluitdrukkingen: dat leidt dan tot constructies als *aan de hand daarvan*. Die analyseren we ook als MWU.

2.4.7 Als, zoals

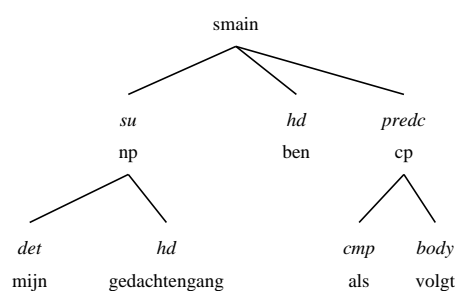
Het woordje *als* kan soms optreden als voorzetsel, en soms als complementeerder. Dit verschil valt soms aan te tonen door de naamvalstoekenning (*als ik* of *als mij*), maar het onderscheid is moeizaam. We kiezen ervoor om *als* altijd als complementeerder te analyseren, waarbij *als* als CMP optreedt en het complement als BODY.

De versteende uitdrukking *als volgt* wordt als MWU geannoteerd.

(343) de redenering luidt als volgt

dependentielabel	OMSCHRIJVING
HD	naamwoordelijke kern
DET	determinator
VC	verbaal complement
MOD	voor- of nabepaling
APP	appositie, bijstelling

Figuur 2.6: dependentielabels: nominale domeinen



Het woord *zoals* is ook altijd een complementeerder.

2.5 Het nominale domein

De globale opbouw van de naamwoordelijke constituent (categorielabel van de moederknoop: NP) kan aan de hand van het volgende schema (zie ANS, (Hae-seryn, 1997, hoofdstuk 14)) inzichtelijk worden gemaakt:

NAAMWOORDELIJKE CONSTITUENT			
Determinator	Voorbepalingen	Kern	Complementen/Nabepalingen

In Figuur 2.6 geven we de dependentielabels binnen het nominale domein. Over de vraag wat het hoofd is van een naamwoordelijke constituent bestaan verschillende meningen, afhankelijk van de interpretatie van het begrip ‘hoofd’. De voorgestelde annotatie stelt zich in deze kwestie theorie-neutraal op.

2.5.1 De nominale kern

We hanteren het label HD voor de nominale kern van de NP.

(344) een lekkere **wijn**

- (345) koel **bier**
 (346) **iets** moois
 (347) die **rode**

2.5.2 De determinator

De ANS (Haeseryn, 1997) onderscheidt verschillende soorten determinator-elementen, waarvoor een aantal volgorde restricties gelden, die gehoorzamen aan de volgende systematiek:

DETERMINERENDE ELEMENTEN			
POSITIE 1	POSITIE 2	POSITIE 3	POSITIE 4
kwantiteits-aanduidende elementen	determinerende elementen	adjectivische elementen	kwantiteits-aanduidende elementen

De verschillende determinator-elementen hebben in deze visie van rechts naar links een toenemend bereik: [de [laatste [twee [opties]]]]. In het geval van *de twee laatste opties* maakt *laatste* volgens de ANS (Haeseryn, 1997) geen deel uit van de determinator, maar fungeert het als bepaling bij *opties*.

Wat ook de verdiensten zijn van dit schema voor determinatoren, het is incompatibel met de aanname dat een NP direct de categorie van de naamwoordelijke kern projecteert. Wij houden het er dan ook op dat zowel in het geval van *de twee laatste opties* als in het geval van *de laatste twee opties*, de de determinator is en dat *laatste* en *twee* beide fungeren als bepaling bij *opties*. Wij baseren ons hierbij op data als de volgende:

- (348) (veel | vele) journalisten belaagden het jonge paar
 (349) de (vele | *veel) journalisten belaagden het jonge paar

waarin het maar de vraag is of *vele* in (349) als determinator-element kan worden opgevat, gegeven dat het hier, anders dan in (348), verplicht verbogen is. Verder is *de vele journalisten* equivalent met *de journalisten*, *die (vele | met velen) waren*, maar *(veel | vele) journalisten* niet met *journalisten(,) die (vele | met velen) waren*.

Kortom, we kiezen voor:

- (350) de vele journalisten DET MOD HD
 (351) veel journalisten DET HD

2.5.2.1 Complexe determinatoren

Determinatoren kunnen natuurlijk ook gemodificeerd worden:

(352) **precies drie** boeken

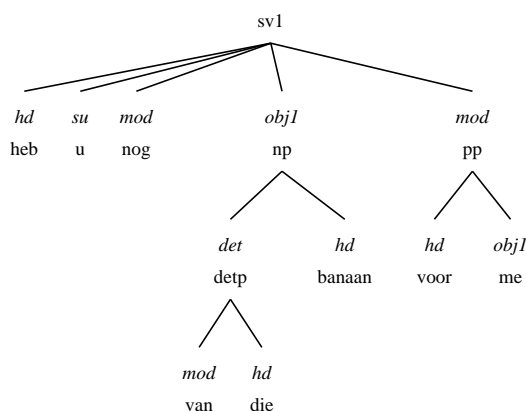
(353) **juist die** kolibrie

(354) **al mijn** vrienden

Daarvoor introduceren we een nieuw categorielabel, DETP (determinator-frase)

Eenzelfde structuur kennen we toe aan de complexe determinatoren *van die*, *van dat*, *van zulke* enz., waarin het paradigmaloze *van* een modifier-rol krijgt toegerekend:

(355) Hebt u nog **van die** bananen voor me?



(356) Je kunt daar **van dat** heerlijke gebak kopen

(357) Ze kan daar **van zulke** diepe dingen over zeggen

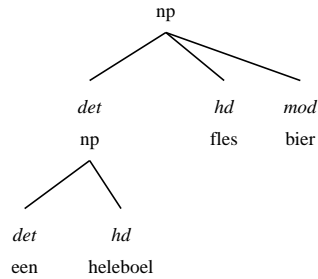
De schijnbaar analoge uitdrukking *van alles* in zinnen als

(358) in de Winkel van Sinkel is **van alles** te koop

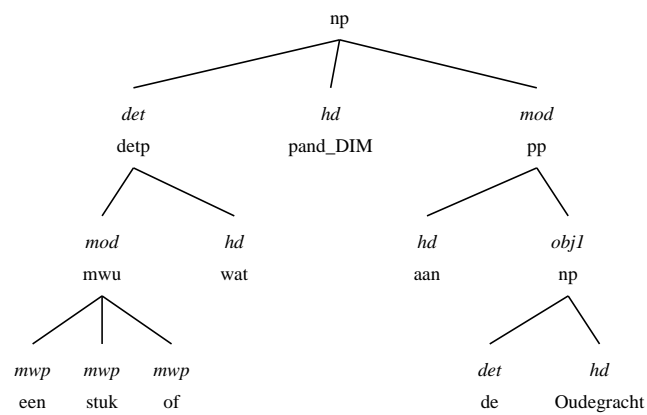
wordt niet nader geanalyseerd: die wordt, als vaste uitdrukking, aangemerkt als een MWU.

Woordcombinaties als *een heleboel* (*flessen bier*) en *een stuk of wat* (*pandjes aan de Oudegracht*) fungeren eveneens als complexe determinator, en hetzelfde geldt voor de discontinue constructie *een N of Telwoord*, als in *een boek of zes*. We annoteren de constructie als een NP met *boek* als HD en een complexe determinatorfrase DETP *een ... of zes*, waarin *zes* als achteropgeplaatst HD fungeert, en *een ... of* een MWU is die een MOD-rol vervult.

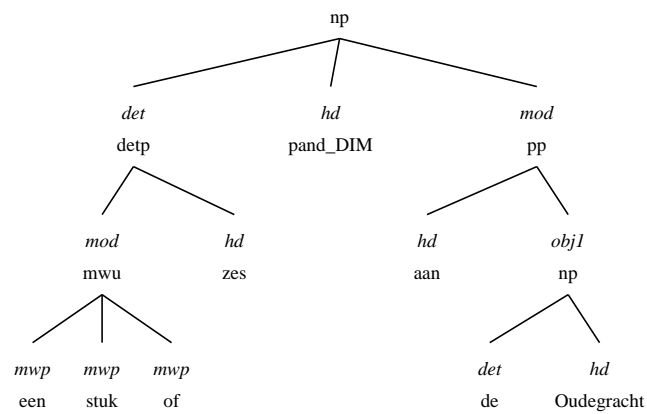
(359) een heleboel flessen bier



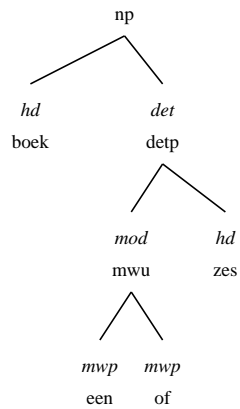
(360) een stuk of wat pandjes aan de Oudegracht



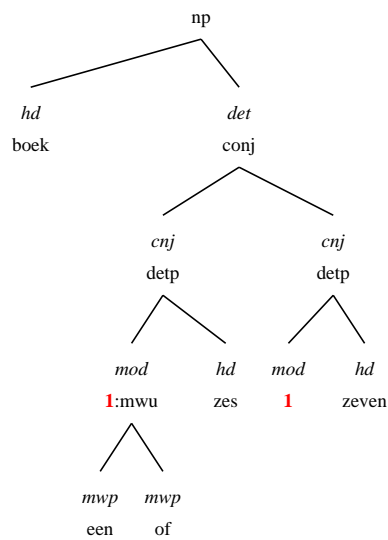
(361) een stuk of zes pandjes aan de Oudegracht



(362) een boek of zes

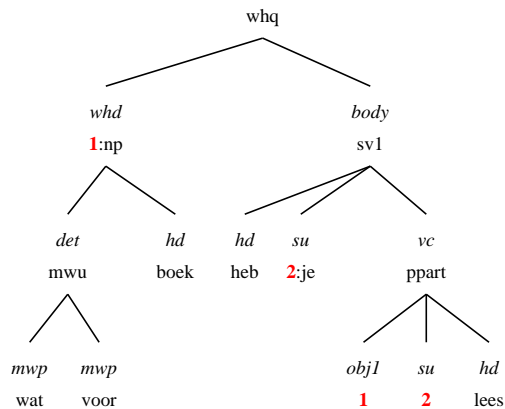


(363) een boek of zes zeven

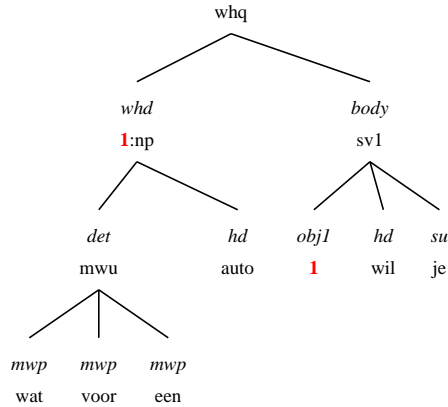


Wat voor (een) is een vragende NP-determinator:

(364) wat voor boeken heb je gelezen?



(365) wat voor een auto wil je?

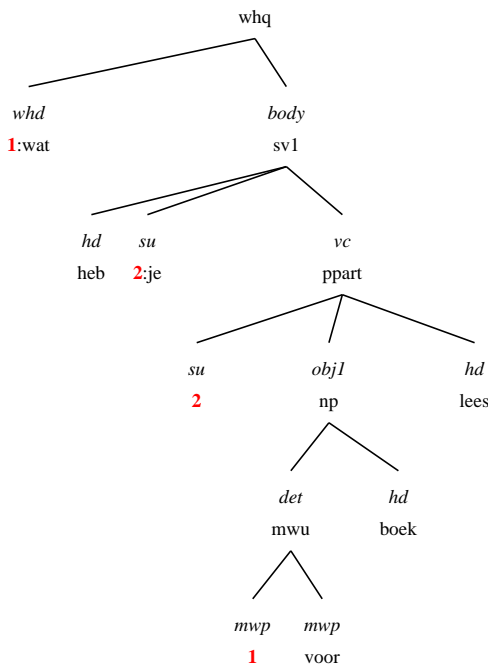


Deze woordgroepen wordt geannoteerd als een MWU.

Zoals andere vraagwoorden staat dit *wat* (*voor*) niet altijd op de plaats waar we het zouden verwachten, maar we benoemen het desondanks als een determinator van het nominale hoofd en storen ons niet aan de discontinue afhankelijkheden.

Wat uit de *wat voor* (*een*)-constructie kan ook “vanuit deze constructie verplaatst worden” en helemaal vooraan de zin staan, los van de NP waarin het thuishoort:

(366) wat heb je voor boeken gelezen?



(367) wat zeg je dat je voor een auto wilt?

We vinden dezelfde constructie ook met sommige onbepaalde voornaamwoorden, met name *iemand* en *iets* (*wat voor iemand is dat eigenlijk?*) en met genominaliseerde adjectieven (*wat voor moois heb je nu weer gekocht?*). In het laatste geval fungeert *wat voor* als hoofd van de NP *wat voor moois* (vgl. *iets moois* in (2.5.1)). Vergelijk verder de combinatie *wat een* die we in exclamatieven aantreffen, ook met meervoudige zelfstandige naamwoorden, die normaliter onverenigbaar zijn met het enkelvoudige lidwoord *een*:

(368) wat een feest was dat zeg!

(369) wat een boeken heeft die man gelezen!

Ook hier is ‘extractie’ van het WH-element mogelijk:

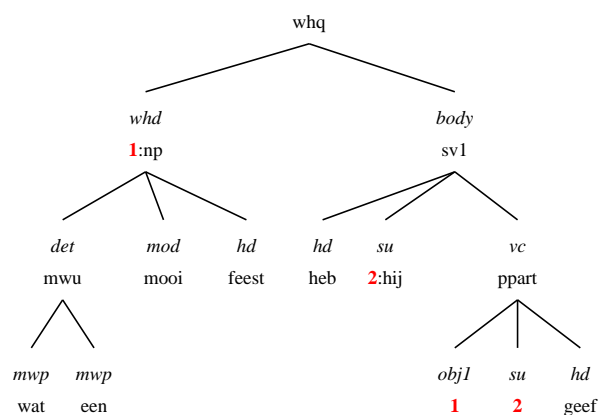
(370) wat heeft die man een boeken gelezen!

Ook in deze gevallen krijgt de complexe determinator het label MWU.

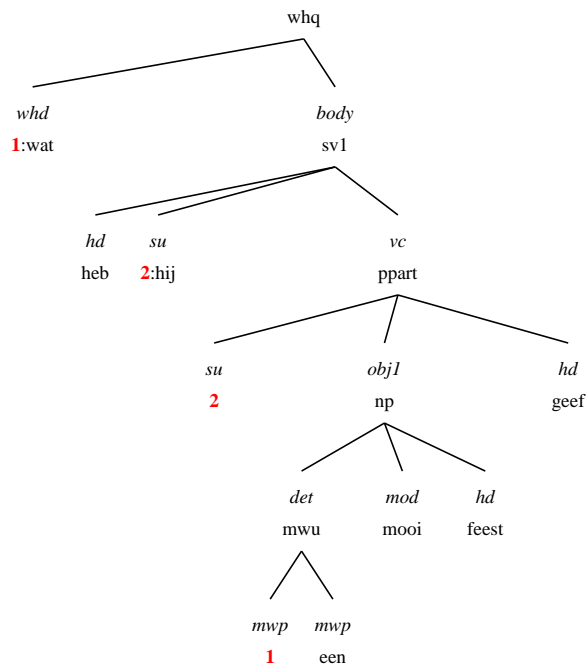
Ook al levert de constructie met “wat een” geen vraagzinnen op, toch wordt deze constructie geanalyseerd als een vraagzin (dus met whq of whsub) omdat de constructie syntactisch gezien op dezelfde manier is opgebouwd, en vooropplaatsing ook verplicht is. Dit gaat ook op voor de combinatie waarbij een exclamatief gebruik van “wat” wordt gecombineerd met een adjectief.

We krijgen dus:

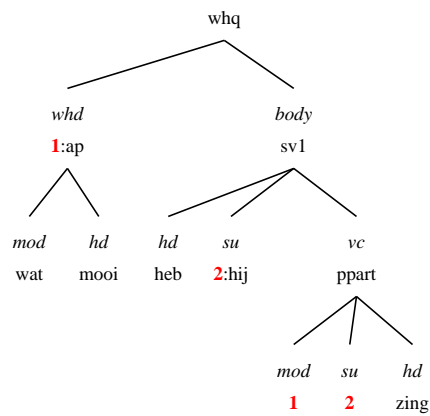
(371) Wat een mooi feest heeft hij gegeven!



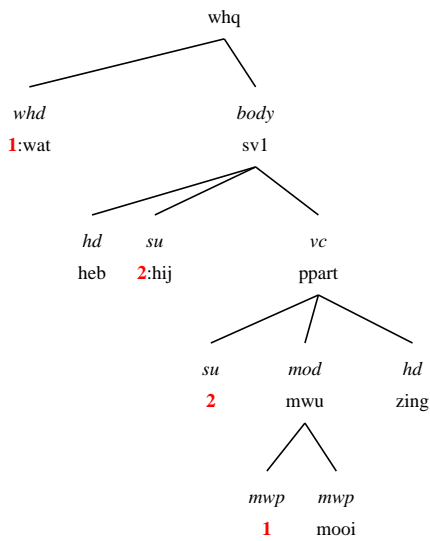
(372) Wat heeft hij een mooi feest gegeven!



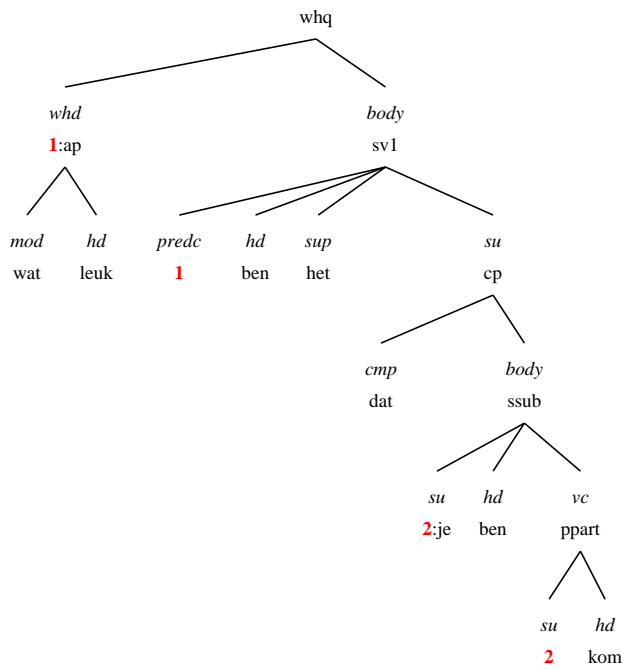
(373) Wat mooi heeft hij gezonden!



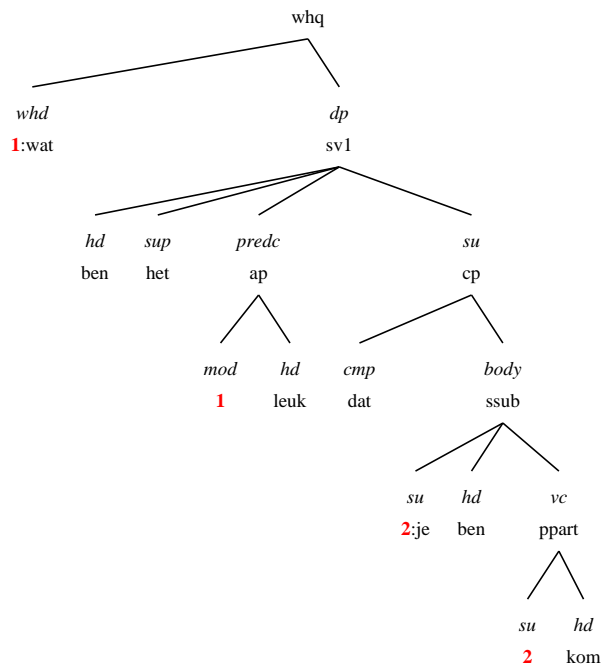
(374) Wat heeft hij mooi gezongen!



(375) Wat leuk is het dat je bent gekomen!



(376) Wat is het leuk dat je bent gekomen!

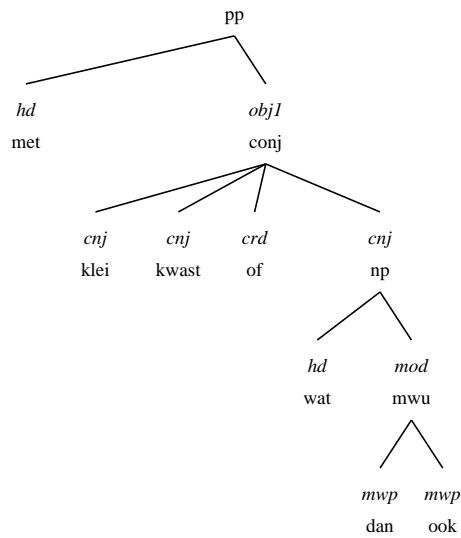


Opmerking. *De constructie met gespleten 'wat... voor (een)' en 'wat... een' wordt niet door Alpino herkend.*

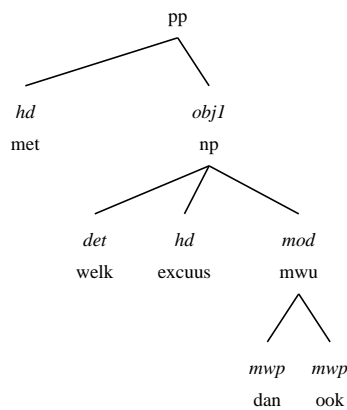
2.5.2.2 Wat dan ook

WH-constituenten kunnen worden gevolgd door het MWU *dan ook*. Deze wordt geannoteerd als MOD. Ook indien het WH aspect wordt geïntroduceerd in een lidwoord wordt de MWU *dan ook* op het niveau van de NP aangehecht.

(377) Met klei, kwast of wat dan ook

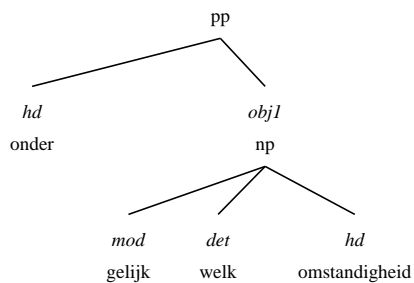


(378) Met welke excuses dan ook



Een vergelijkbare constructie komt voor met *gelijk*, *om het even*. Deze worden ook als MOD aangehecht op het niveau van de NP.

(379) onder gelijk welke omstandigheden



2.5.2.3 De partitieve *van*-constructie

In afwijking van de ANS (Haeseryn, 1997, blz. 835) beschouwen we de nominale kern van de NP in een partitieve *van*-frase niet als het hoofd van de volledige partitieve constructie. In een zin als

(380) één van de aanwezigen nam het woord

is er immers sprake van getalsovereenkomst tussen *nam* en *één*, niet tussen *nam* en *aanwezigen*.

2.5.3 Complementen binnen de nominale constituent

Het dependentielabel VC en hun syntactische realisaties zijn reeds besproken bij de behandeling van het verbale domein.

(381) het eeuwige gewacht bij de auto **op m'n vader** (PC)

(382) z'n heilige geloof **in de medemens** (PC)

(383) twijfels **aan de oprechtheid** (PC)

(384) onze hoop **(om) de trein nog te kunnen halen** (VC)

(385) de verwachting **dat ze er nog wel zou zijn** (VC)

(386) de vraag **of alle moeite voor niets was geweest** (VC)

(387) de vraag **met wie we naar het feest konden** (VC)

Maar ook niet-deverbale nomina kunnen een complement met een vast voorzetsel krijgen. Ook die worden benoemd als PC:

(388) de kans **op winst**

(389) het gebrek **aan mensen** bij de politie

(390) gelegenheid **tot bloot zwemmen**

Opmerking. *PP-bepalingen bij zelfstandig naamwoorden worden door Alpino altijd als MOD geannoteerd, en nooit als PC.*

Neemt het complement een verbale vorm aan (wat niet altijd kan), dan wordt dat benoemd als VC:

(391) de kans **(om) te winnen**

(392) gelegenheid **(om) bloot te zwemmen**

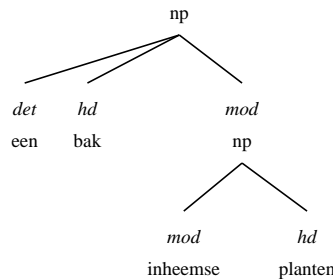
- (393) de hoop **dat hij zal komen**
- (394) de onduidelijkheid **of hij komt**
- (395) de onzekerheid **met wie hij komt**

In Vlaanderen zal men hier, afhankelijk van het nomen, soms *van* in plaats van *om* aantreffen.

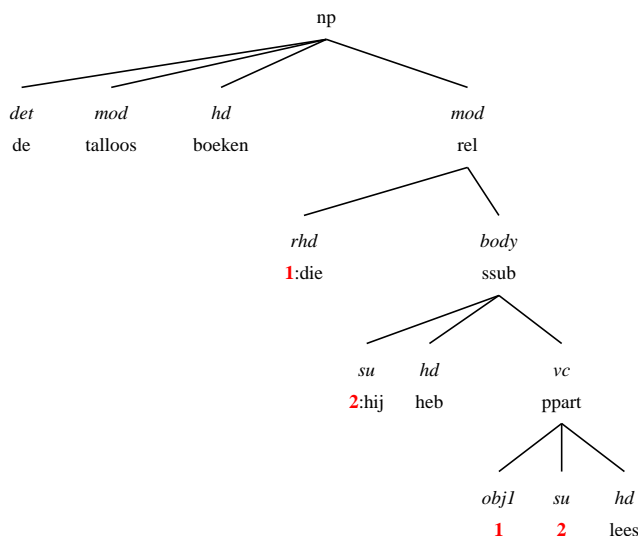
2.5.4 Modifierende elementen

Als modificeerders kunnen tal van syntactische categorieën fungeren. Hieronder een paar voorbeelden.

- (396) een bak **inheemse planten** (NP)



- (397) **heerlijk helder** Heineken (adjectieven)
- (398) de **talloze** boeken **die hij heeft gelezen** (adjectief, REL)

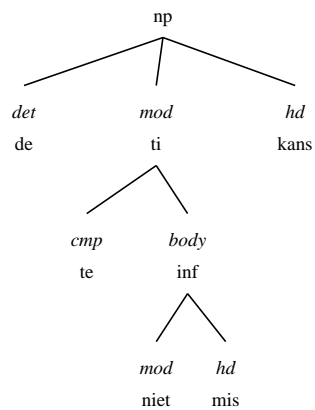


Met de ANS ([Haeseryn, 1997](#), blz. 844) beschouwen we de gecursiveerde zinsdelen in de volgende voorbeelden als modificatoren in het nominale domein:

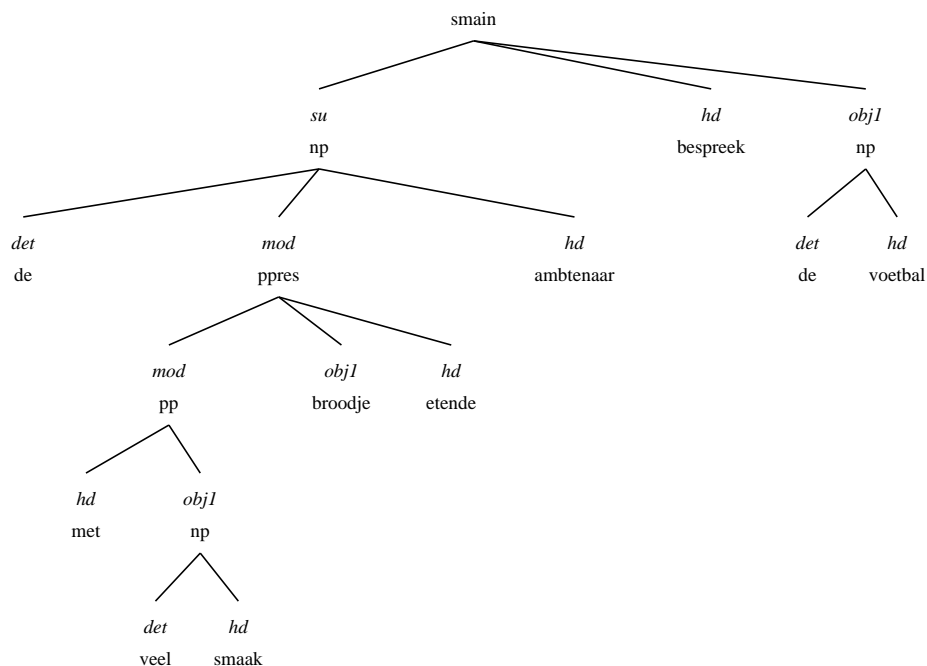
- (399) **flink** wat
- (400) dezelfde **ongeveer**
- (401) **in het geheel** niets
- (402) **om het even** wie
- (403) **eender** wie
- (404) **gelijk** wie
- (405) **al** wie (dat hoorde wou bij hem zijn)

De modificerende bepaling tussen lidwoord en nominale kern kan ook verbaal materiaal bevatten, dat dan ook zo behandeld wordt:

- (406) De **niet te missen** kans



- (407) De **met veel smaak broodjes etende** ambtenaren bespraken de voetbal



Nabepalingen (MOD) bij nominale constituenten kunnen allerlei vormen aannemen; hun aantal is theoretisch onbeperkt (cf. (416)):

- (408) het feest *deze middag* (NP)
- (409) de plek *des onheils* (NP)
- (410) wie *uwer* (NP)
- (411) die herrie *hier (ook altijd)* (ADVP)
- (412) koekjes *voor bij de koffie* (PP)
- (413) kennis *van de natuur* (PP)
- (414) jongens *zonder verantwoordelijkheidsbesef* (PP)
- (415) een dag *als vandaag*
- (416) de tocht (*te voet*) (*van Karel*) (*vorig jaar september*) (*naar Rome*) (PP)
- (417) een kind *om te zoenen* (OTI)
- (418) iets *te eten* (kan ik je niet aanbieden, maar we hebben drank genoeg in huis voor een heel peloton) (TI)
- (419) het dorpje *dat wij allemaal zo goed kennen* (relative bijzin)
- (420) God *almachtig* (A)
- (421) zaterdag *aanstaande* (A)
- (422) iets *heel raars* (AP)

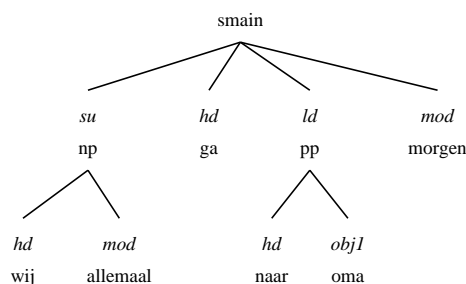
- (423) wij *drieën* (volgens de ANS (Haeseryn, 1997) een telwoord)
 (424) wij *drietjes*
 (425) wij *allemaal*
 (426) wij *als redacteuren van de ANS*

Met name voorzetselbepalingen hoeven niet direct naast de nominale kern te staan, maar dat geldt niet alleen daarvoor:

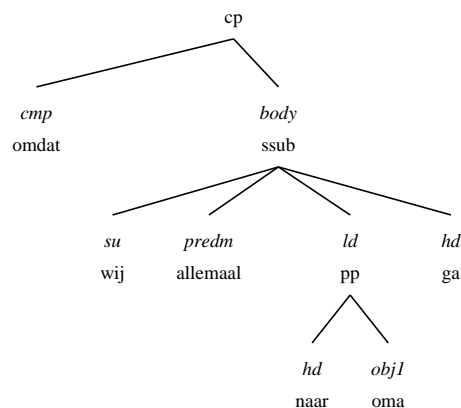
- (427) Jan heeft een boek gekocht *over reptielen*
 (428) ik vind dat je een kind hebt *om door een ringetje te halen*

In (425) en (426) fungeren *allemaal* en *als redacteuren van de ANS* als NP-modificeerders. Dit is echter alleen het geval als er sprake is van adjacentie en de NP de eerste positie in een hoofdzin inneemt. In alle andere gevallen fungeren deze constituenten als PREDM op zinsniveau. Er zijn twee redenen waarom het geval adjacentie + NP-op-de-eerste-positie-in-de-hoofdzin in dezen een aparte behandeling krijgt. De eerste is dat de constituent die 'normaal' als PREDM op zinsniveau fungeert in dit geval wel deel moet uitmaken van de NP, omdat er in de hoofdzin immers maar één constituent voor de persoonsvorm kan staan. De tweede reden is dat een dergelijke NP bijvoorbeeld geen gereduceerd pronomen als kern toestaat (*wij* versus **we* – vgl. ook *arme jij* versus **arme je*), een beperking die alleen geldt in (431), maar niet in (432–435).

- (429) *wij allemaal* gaan naar oma morgen



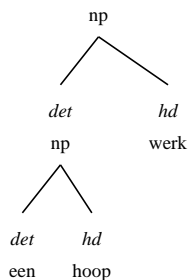
- (430) omdat *wij allemaal* naar oma gaan



- (431) **we allemaal* gaan naar oma morgen
- (432) we gaan *allemaal* naar oma morgen (PREDM)
- (433) omdat we *allemaal* naar oma gaan morgen (PREDM)
- (434) omdat we morgen naar oma gaan *met z'n allen* (PREDM)
- (435) omdat ze dat *als redacteurs van de ANS* belangrijk vinden (PREDM)

In een aantal constructies van het type (396) is een zelfstandig naamwoord zo vaak opgetreden als kwantiteitsaanduiding dat de afgesloten combinatie functioneert als een (complexe) determinator:

- (436) een hoop werk



Frases als (436) zijn in principe ambigu: de *hoop* kan, als de nominale kern iets is dat zich op laat hopen, ook letterlijk ('berg', 'bult') bedoeld zijn. De context zal in het algemeen helpen desambigueren. Indien de context niet helpt, kiezen we bij voorkeur voor de 'normale' compositionele analyse.

2.5.5 Bijstellingen: APP

Bijstellingen worden met het label APP (appositie) geannoteerd.

De ANS (Haeseryn, 1997, blz. 846) verstaat onder bijstelling “elke naamwoordelijke constituent die als aanvullende informatie achter een andere naamwoordelijke constituent geplaatst wordt”. Daaraan wordt nog toegevoegd dat die nominale constituent van verbale origine kan zijn (437) en dat er allerlei zinsdelen in geïncorporeerd kunnen zijn (437, 438):

- (437) de dagelijkse activiteit van een auteur, *stukjes schrijven*, wordt niet altijd gewaardeerd
- (438) die man daar, *tot voor kort gymnasiumleraar hier ter stede*, is onlangs tot hoogleraar benoemd

Niet altijd staat de bijstelling onmiddellijk na het bijbehorende zelfstandige naamwoord:

- (439) we hebben Pollini ontmoet, de beste pianist ter wereld

Het Nederlands kent nog een aantal constructies waarbij een nomen een soort nabepaling heeft, een naam of een telwoord. Denk daarbij aan de volgende types (naar Paardekooper):

- (440) hoofdstuk *vier*
- (441) de maand *maart*
- (442) de stad *Groningen*
- (443) de afstand *Mechelen-Brussel*
- (444) oom *Wim*

Ook voor al dit soort gevallen gebruiken we het (recursieve) label APP.

In sommige gevallen leidt dit tot ambiguïteit, die moet worden opgelost door de voor de hand liggende lezing te kiezen.

- (445) een bundel gedichten MOD
- (446) de gedichtenbundel ‘ Ik kan weer horen ’ (APP)

2.5.6 Deverbale nominalisaties

In de sfeer van de deverbale nominalisatie kunnen we een aantal gevallen onderscheiden:

1. morfologische derivatie

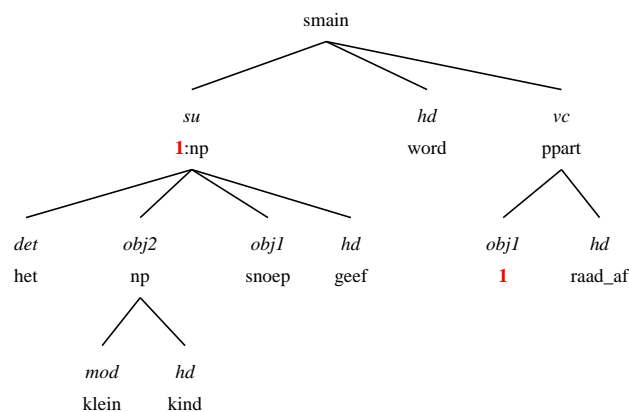
afleiding op *-ing* (*vernedering, verdediging*)
 afleiding op *-atie* (*diskwalificatie, hospitalisatie*)
 afleiding met *ge-* (*getreuzel, gedachte*)
 conversie van werkwoordelijke stam (*hoop, vlucht*)

2. de constructie *het* + infinitief, al dan niet met frasale conversie (zie (a) en (b) hieronder)
3. frasale conversie van een infinitiefgroep

Voorbeelden:

(447) de **verdediging** van de stad (type 1)

(448) het kleine kinderen snoep **geven** (wordt afgeraden) (type 2a)



(449) het **geven** van snoep aan kleine kinderen (type 2b)

(450) cognac **aanbieden** aan tante Fien (vindt hij zonde) (type 3)

In alle gevallen zal het hoofd hier een categorie NP projecteren. De verschillen tussen de types is meestal duidelijk door het POS-tag label.

Wat betreft de NP-complementen van een werkwoord (SU, OBJ1, PC, OBJ2) merken we op dat deze als PP worden gerealiseerd in het nominale domein. Ze zijn dan altijd weglaatbaar, en worden als MOD gelabeld.

(451) het getreuzel *van Piet*

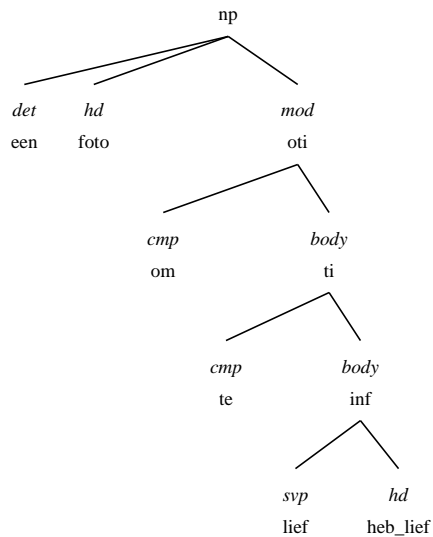
(452) het zingen *van het volkslied*

(453) de gift *door Han van een tientje aan Jantje Beton*

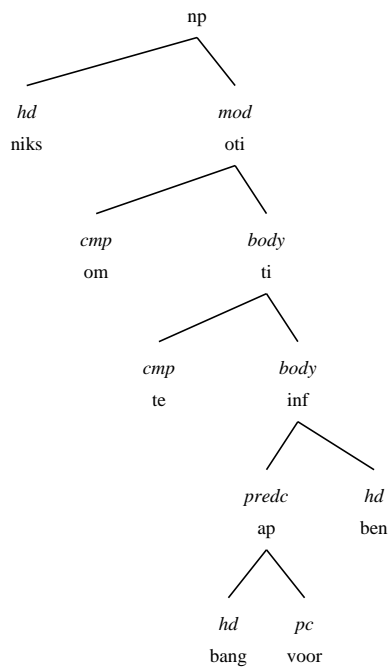
2.5.7 VP modificatie met missend argument

In deze gevallen wordt een zelfstandig naamwoord gemodificeerd door een OTI waarbij een argument lijkt te missen, zelfs *preposition stranding* komt voor.

(454) Een foto om lief te hebben



(455) Niks om bang voor te zijn



2.5.8 Vocatieven

Vocatieven of aanspreekvormen kunnen verschillende vormen aannemen:

- (456) ha, *Kees*
- (457) *bakker*, mag ik een brood?
- (458) die rommel, *Jan*, die moet je toch eens opruimen
- (459) goedemorgen *Nijmegen*
- (460) dat is het juist, *jongen*
- (461) geef maar hier, *liefste*
- (462) maar *honneponnetje*, we hoeven toch niks?
- (463) kijk uit, *hufter*

In voorbeelden als deze zijn ze nooit echt deel van de zin: we benoemen ze dan ook als TAG.

Ook sommige ‘volle’ NPs (met modificatoren en determinatoren) kunnen als aanspreekvorm gebruikt worden; de determinator, voor zover aanwezig, zal meestal *mijn* zijn:

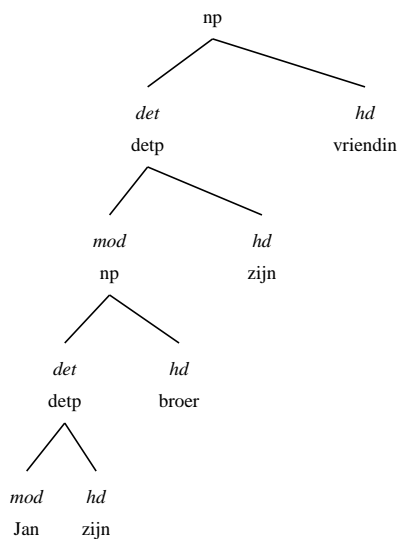
- (464) kom maar hier, *m'n jongen*
- (465) ik zal altijd bij je blijven, *mijn allerliefste lieveling*

Ook deze vocatieven benoemen we als TAG.

2.5.9 Jan zijn fiets

Jan zijn vormt een complexe determinator met *zijn* als HD en *Jan* als MOD. Zo ook met *die mensen hun geld* enz.

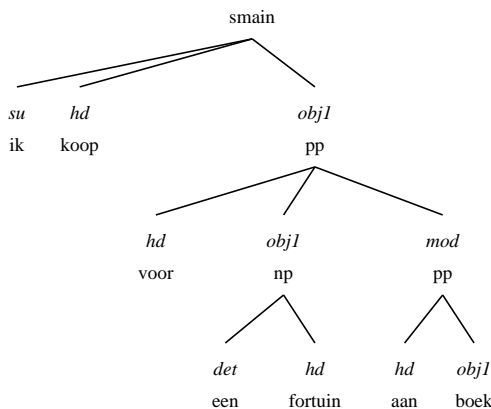
- (466) Jan zijn broer zijn vriendin



2.5.10 Voor NP aan NP

Een lastig geval zijn woordgroepen zoals de volgende die zich als NP gedragen, maar opgebouwd lijken te zijn als een PP. Wij kiezen ervoor om als categorielabel PP te gebruiken. De interne structuur van de PP is drieledig en reflecteert hiermee de uitzonderlijke status van deze constructie.

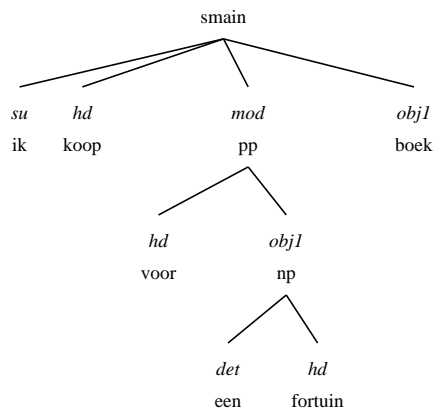
(467) Ik koop voor een fortuin aan boeken



(468) We bestelden voor twaalf euro aan ijsjes

Indien *aan* niet optreedt, wordt aangenomen dat de *voor*-PP een modificeerder van het werkwoord is:

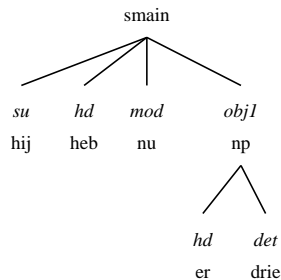
(469) Ik koop voor een fortuin aan boeken



2.5.11 Kwantitatief *er*

We kiezen ervoor om geen nieuw label te introduceren, maar om *er + telwoord* te analyseren als een gewone NP, waarbij *er* het hoofd is, en het telwoord als DET fungeert. Merk hierbij op dat het telwoord en *er* niet adjacent hoeven zijn.

(470) Hij heeft er nu drie



Opmerking. *Deze constructie wordt niet door Alpino herkend.*

2.5.12 Problemen hebben/krijgen met

(471) we hebben/krijgen problemen met de koeling

(472) daar heb/krijg je problemen mee

(473) problemen met de koeling krijg je dan

De voorzetselbepaling is een PC bij het werkwoord in de eerste twee gevallen. In het derde geval kan de PP alleen bij *problemen* horen. In dat geval annoteren we de PP als MOD bij *problemen*.

Er zijn nogal wat vergelijkbare gevallen waarbij het onduidelijk is of een PP bij een naamwoord hoort of bij een werkwoord, of allebei. In de volgende gevallen

annoteren we normaliter de PP als PC bij het werkwoord, behalve in gevallen als (473). Voor motivatie zie (Bouma, 2004). Deze constructie tref je vooral aan met de werkwoorden *hebben, zijn, bestaan, doen, geven, krijgen, leggen, lopen, maken, nemen, stellen*.

2.5.13 Projectie van voornaamwoorden

Sommige voornaamwoorden zijn nogal 'multi-categoriaal': ze kunnen fungeren als hoofd van verschillende soorten constituenten. Een voorbeeld is *veel*:

(474) heel veel mensen (DETP)

(475) hij eet nogal veel (NP)

(476) je zeurt te veel (AP)

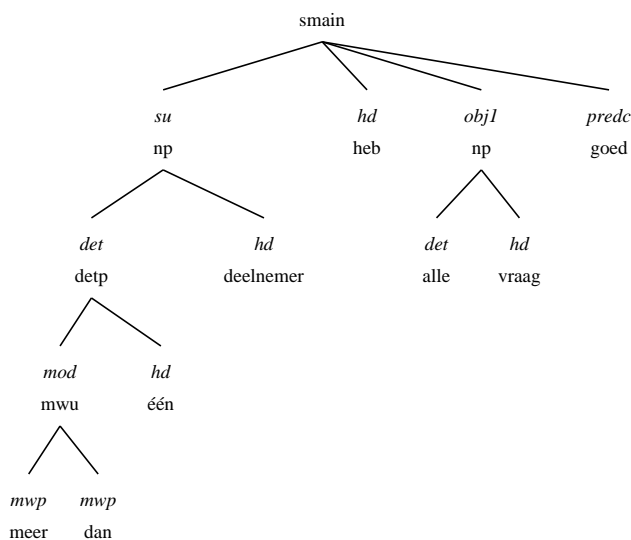
2.5.14 Iets van, rond de, meer dan + telwoord

Voorbeelden:

(477) (iets van | rond de | over de) vijfhonderd kinderen doen d'raan mee

(478) (zo'n | een | nog geen) vijfhonderd kinderen

(479) meer dan één deelnemer(*s) had(*den) alle vragen goed



In (479) is goed te zien dat *meer dan* een modificeerderrol vervult, aangezien er agreement is tussen *één* en het zelfstandig naamwoord, niet tussen *meer* en het zelfstandig naamwoord, en ook het werkwoord de enkelvoudvorm heeft. Zowel

	dependentielabel	OMSCHRIJVING
<i>hoofd</i>	HD	hoofd van ADVP
<i>compl</i>	OBCOMP ME	vergelijkingscomplement maatcomplement
<i>mod</i>	MOD	bepaling

Figuur 2.7: dependentielabels: adverbiaal domein

meer dan als *iets van* en *rond de* modificeren dus het telwoord. *Iets van*, *rond de*, *meer dan* worden hier als multi-word-unit geannoteerd. Zo ook voor *niet meer dan*, *minder dan*, en *niet minder dan*. In minder formeel taalgebruik wordt ook *als* gebruikt in plaats van *dan*.

Merk op dat de combinatie *meer dan* op vergelijkbare wijze ook adjectieven kan modificeren, zie sectie 2.3.7 op bladzijde 103.

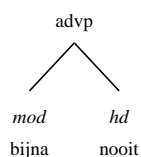
2.6 Het domein van de bijwoorden

2.6.1 Complementen en bepalingen bij bijwoorden

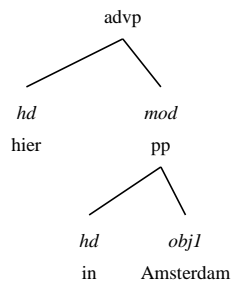
In Figuur 2.7 geven we de dependentielabels voor adverbiale groepen. Het categorielabel voor de moederknoop is ADVP.

Over het adverbiale domein valt niet zo heel veel te vertellen. Hier volgen nog enkele voorbeelden.

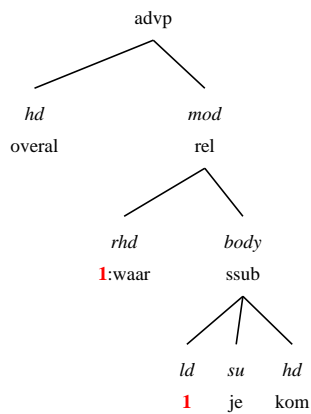
(480) bijna nooit



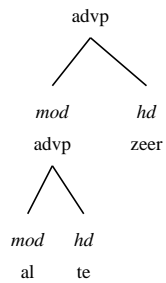
(481) hier in Amsterdam



(482) overal waar je komt



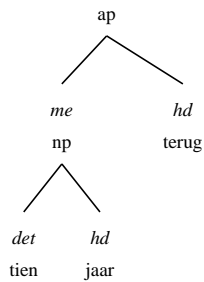
(483) al te zeer



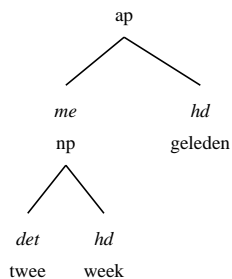
2.6.2 Terug, geleden

De volgende groep woorden worden geanalyseerd als bijvoeglijke naamwoorden omdat ze als PREDC kunnen optreden. Ze combineren verplicht met een ME complement:

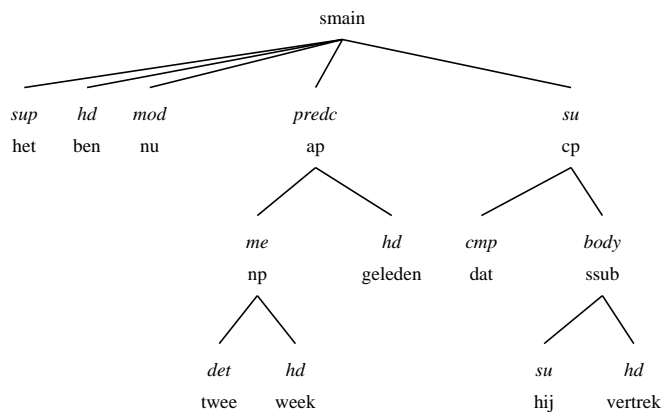
(484) tien jaar terug



(485) twee weken geleden



(486) Het is nu twee weken geleden dat hij vertrok



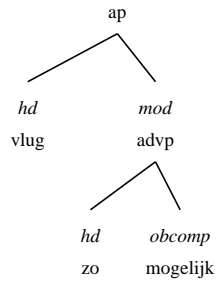
2.6.3 OBCOMP bij bijwoorden

Net als adjectieven kunnen bijwoorden combineren met een vergelijkingscomplement. Een voorbeeld is de vergelijkingsconstructie *even | zo... als...*, waar de *als*-frase fungeert als OBCOMP van het bijwoordelijke hoofd *even | zo*. De constructie als geheel fungeert normaliter als modificeerder van een bijvoeglijk naamwoord. Ook de constructie *zo... mogelijk* wordt op vergelijkbare manier geanalyseerd.

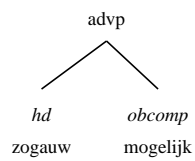
(487) even groot als ik

(488) zo gek als een deur

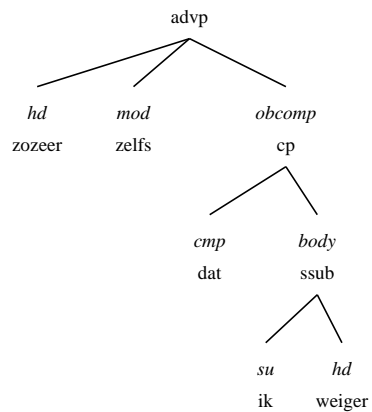
(489) zo vlug mogelijk



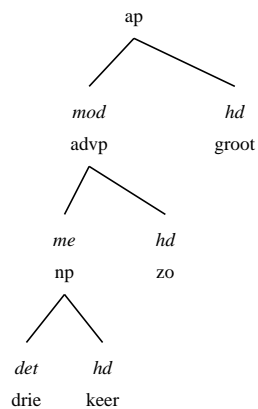
(490) zogauw mogelijk



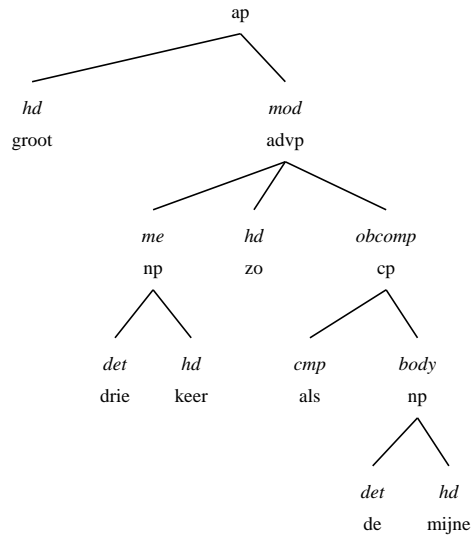
(491) zozeer zelfs dat ik weiger



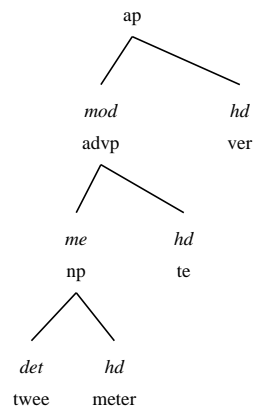
(492) drie keer zo groot



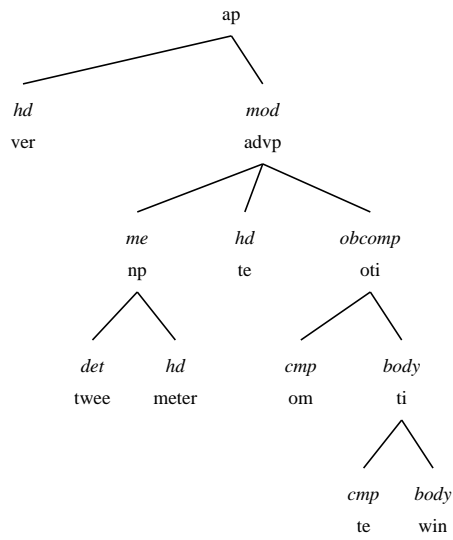
(493) drie keer zo groot als de mijne



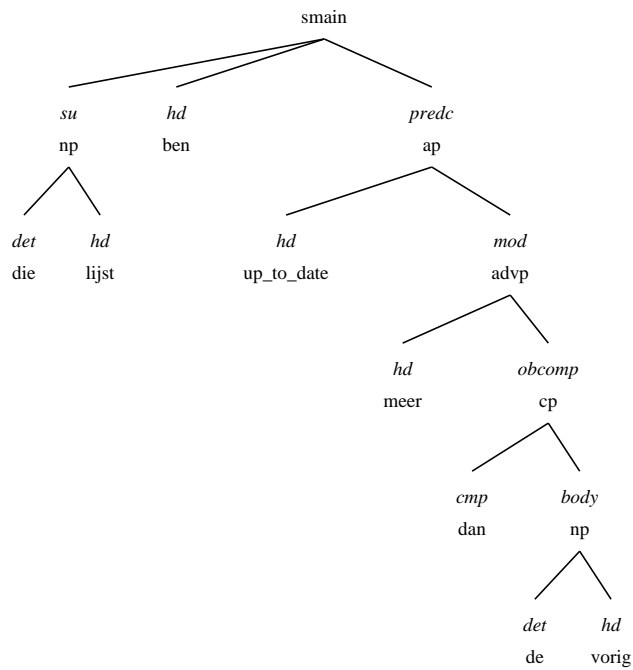
(494) twee meter te ver



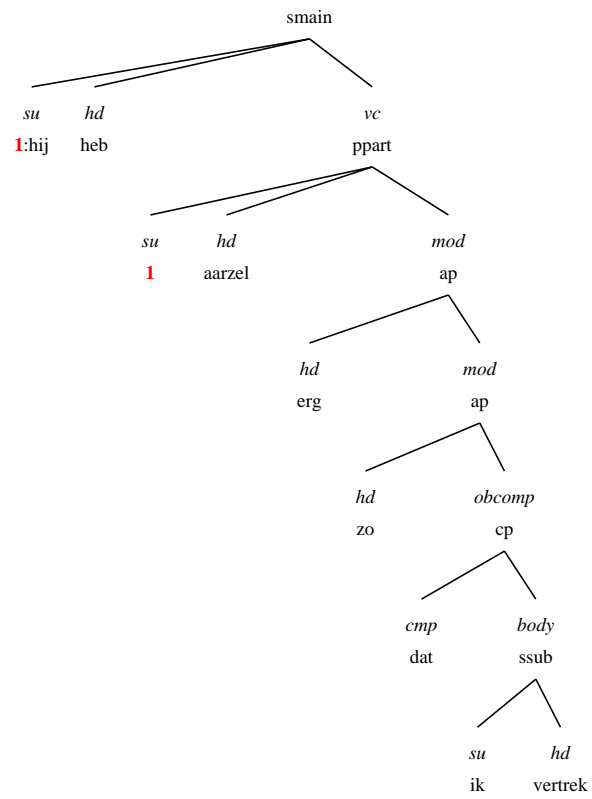
(495) twee meter te ver om te winnen



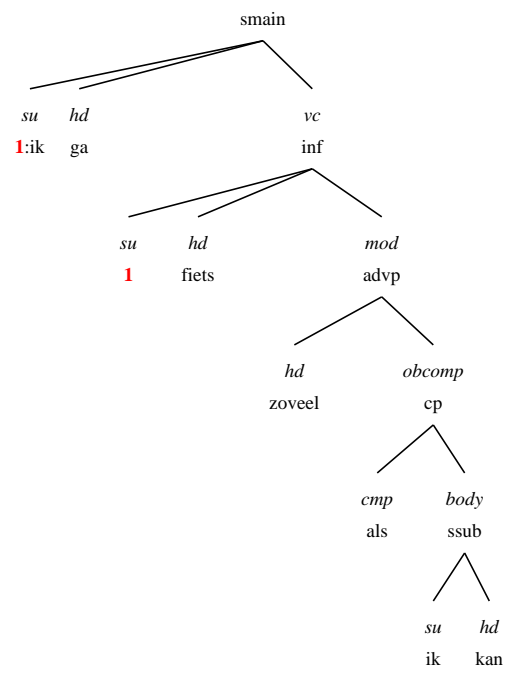
(496) die lijst is meer up-to-date dan de vorige



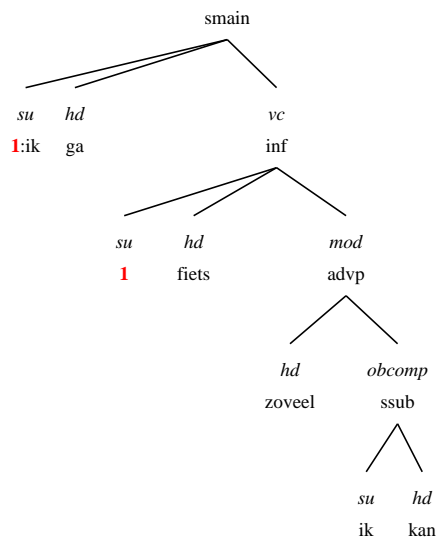
(497) hij heeft zo erg gearzeld dat ik vertrek



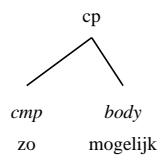
(498) ik ga fietsen zoveel als ik kan



(499) ik ga fietsen zoveel ik kan

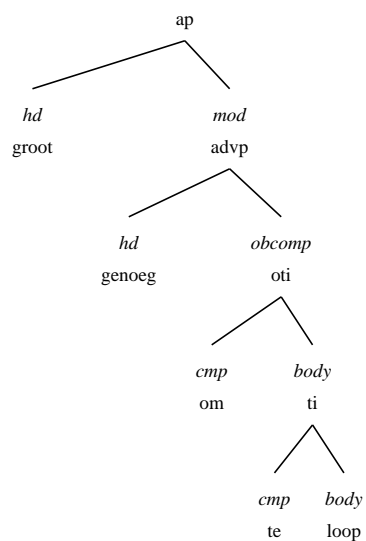


De combinatie *zo mogelijk* wordt geanalyseerd naar analogie met *indien mogelijk*:



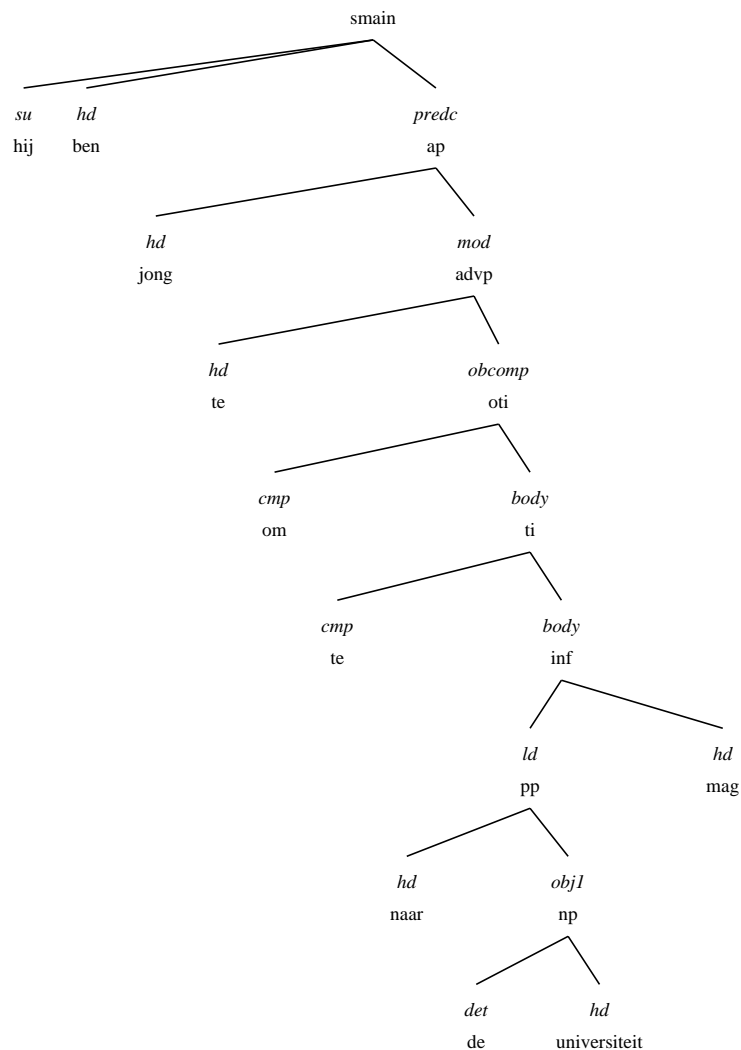
Het bijwoord *genoeg* kan als modificeerder van bijvoeglijke naamwoorden ook een OBCOMP complement nemen:

(500) groot genoeg om te lopen



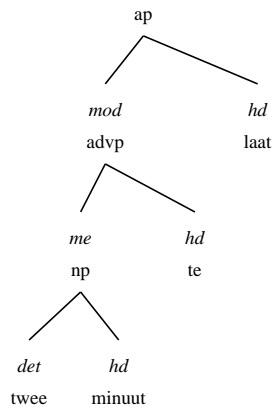
Het bijwoord *te* kan als versterker van een bijvoeglijk naamwoord ook een OB-COMP krijgen, met als categorie OTI:

(501) hij is te jong om naar de universiteit te mogen

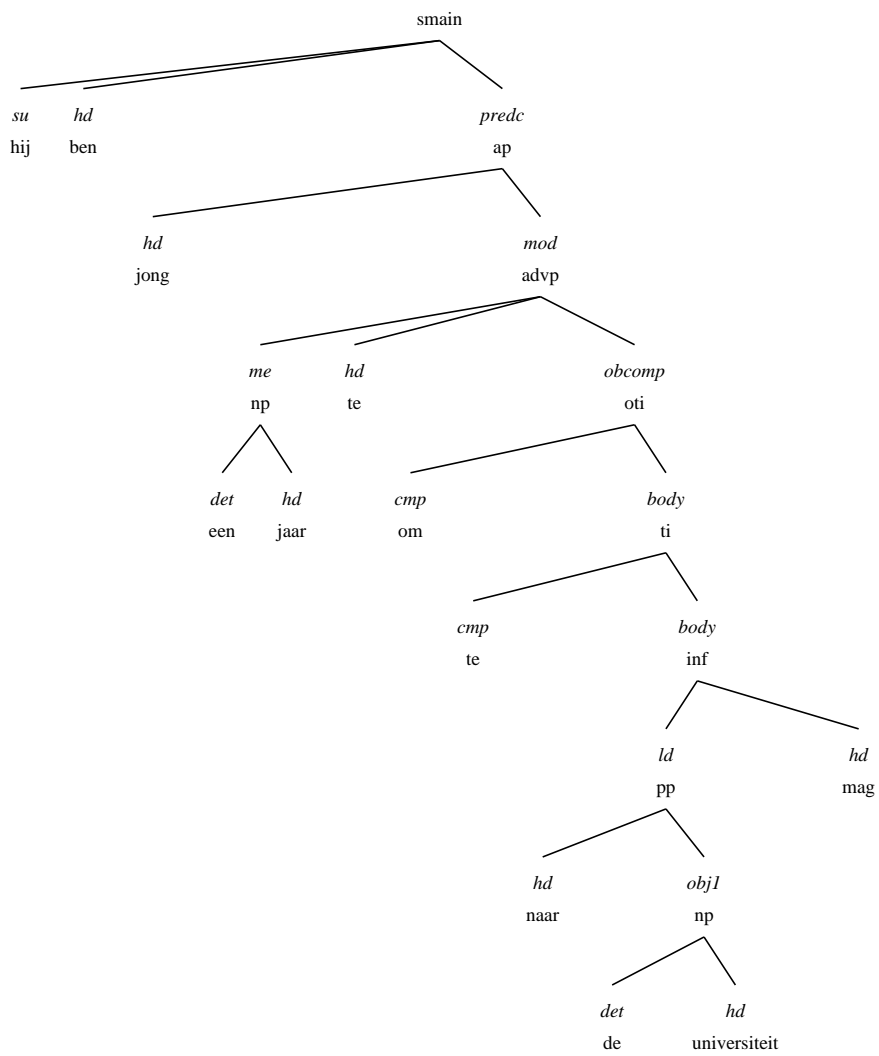


Het bijwoord *te* kan als versterker van een bijvoeglijk naamwoord ook zo'n maat-complement krijgen, eventueel in combinatie met een OBCOMP:

(502) twee minuten te laat



(503) Hij is een jaar te jong om naar de universiteit te mogen

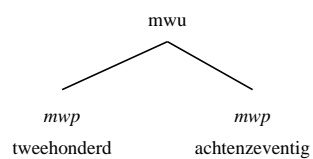


2.7 Named Entities

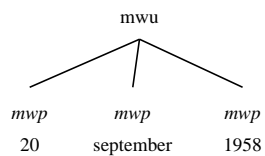
2.7.1 Namen, tijdsuitdrukkingen, data, adressen

Er zijn constructietypes waar de syntaxis niet veel mee kan, omdat niet uit te maken is hoe de dependentieverhoudingen liggen. Voorbeelden zijn getallen, data, complexe eigennamen, adressen.

(504) tweehonderd achtenzeventig

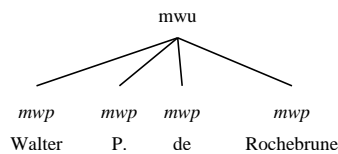


(505) 20 september 1958; 20 september; september 1958



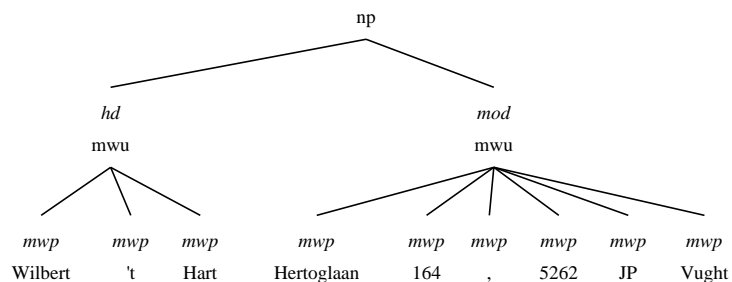
(506) half drie; tien over drie; drie uur; kwart voor twaalf; tien uur vijftien

(507) Walter P. de Rochebrune

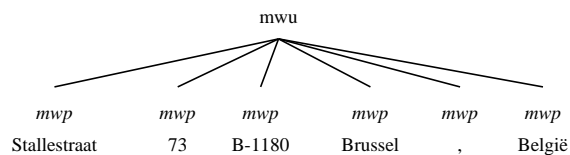


(508) Zo'n **flo 1500,-** kostte de verbouwing

(509) Wilbert 't Hart , Hertoglaan 164 , 5262 JP Vught



(510) Stallestraat 73 B-1180 Brussel , België



We kiezen in dit geval voor een pragmatische, agnostische aanpak: we labelen de knoop MWU. Merk op dat in de laatste twee voorbeelden ook de comma onderdeel uitmaakt van de multi-word-unit. We proberen zoveel mogelijk multi-word-units te vermijden die niet een opeenvolgende reeks woorden van de zin bevat.

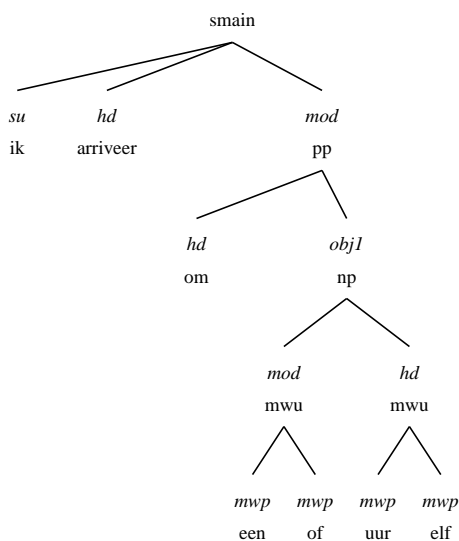
2.7.2 Gemodificeerde multi-word-units

In het volgende voorbeeld:

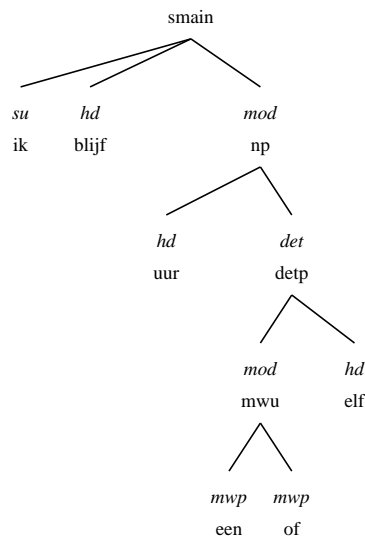
(511) Een uur of elf

zijn er twee mogelijkheden, afhankelijk of een tijdstip of een tijdsduur is bedoeld:

(512) Ik arriveer om een uur of elf



(513) Ik blijf een uur of elf



Zie ook sectie 2.5.2.1 op bladzijde 118.

2.8 Geschreven taal fenomenen

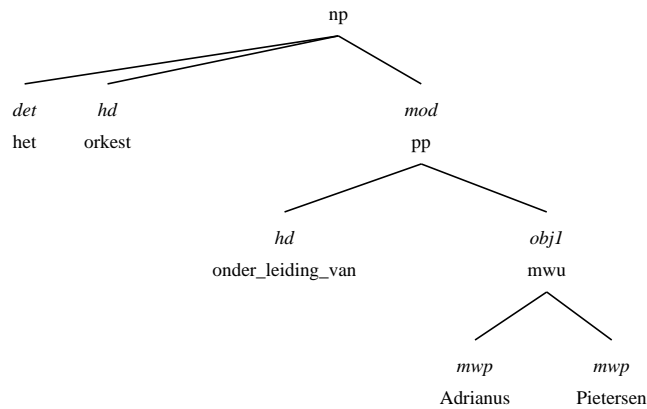
In deze sectie komt een aantal kwesties voorbij die met name voor geschreven taal relevant zijn, en daarom niet of nauwelijks bij de CGN annotatie een rol speelden.

2.8.1 Afkortingen en dergelijke

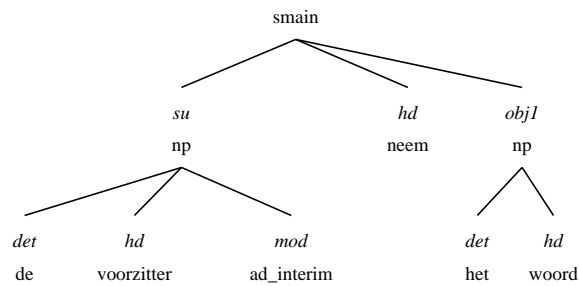
Voor de annotatie van afkortingen waarbij één token meerdere woorden representeert, zoals in de volgende voorbeelden, zijn meestal geen speciale maatregelen noodzakelijk. Indien de afkorting als één woord wordt geschreven, zijn we wel gedwongen om ook in de dependentiestructuur slechts één knoop voor de afkorting te reserveren, ook al gedraagt de afkorting zich syntactisch gezien als een reeks woorden.

Op welke manier de lemma informatie van de afkorting gerepresenteerd wordt, valt buiten het bestek van dit document en behoort tot het domein van de POS-tagging en lemmatizing manual (Van Eynde, 2005).

(514) Het orkest o.l.v. Adrianus Pietersen

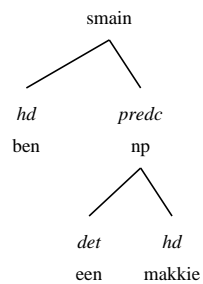


(515) De voorzitter a.i. nam het woord

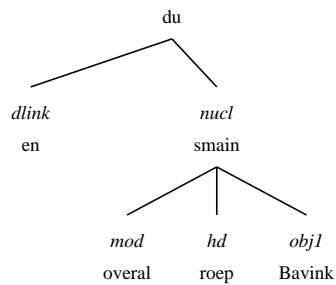


Lastiger zijn de gevallen waarbij de samengenomen woorden geen syntactische eenheid (volgens deze richtlijnen) zijn. In deze gevallen is er één woord, dat eigenlijk twee woorden met twee verschillende rollen bevat. Indien mogelijk kiezen we er dan voor om de rol van hoofd wel aan te geven; de andere rol wordt dan niet expliciet gerepresenteerd.

(516) Da's een makkie

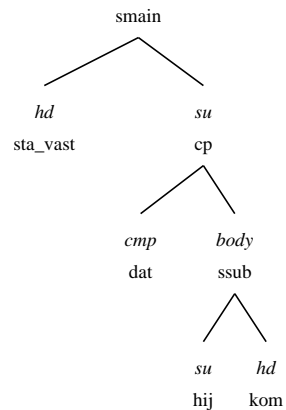


(517) En overal roept-i Bavink



Enigszins verwant met het voorgaande geval waarbij spaties lijken te ontbreken is de volgende constructie, waarbij we opnieuw de annotatie van HD voorrang geven:

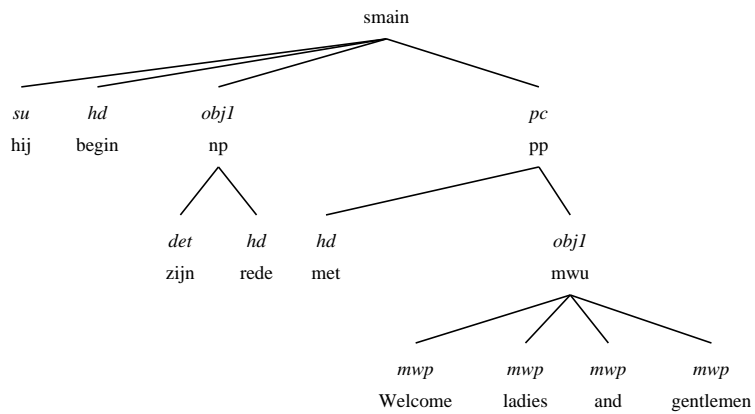
(518) Vaststaat dat hij komt



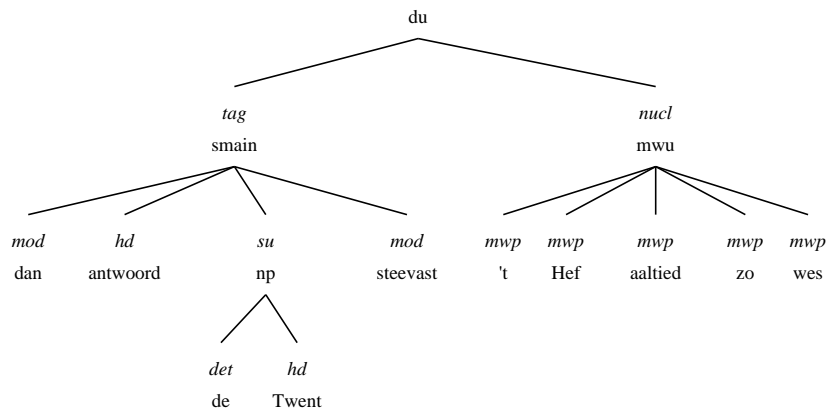
2.8.2 Niet-talig en niet-Nederlands materiaal

We doen geen poging om niet-talig materiaal te analyseren, of om fragmenten uit andere talen van een syntactische analyse te voorzien. In zulke gevallen maken we een MWU van het niet-geanalyseerde deel. Het deel zelf wordt zo mogelijk wel in de analyse opgenomen.

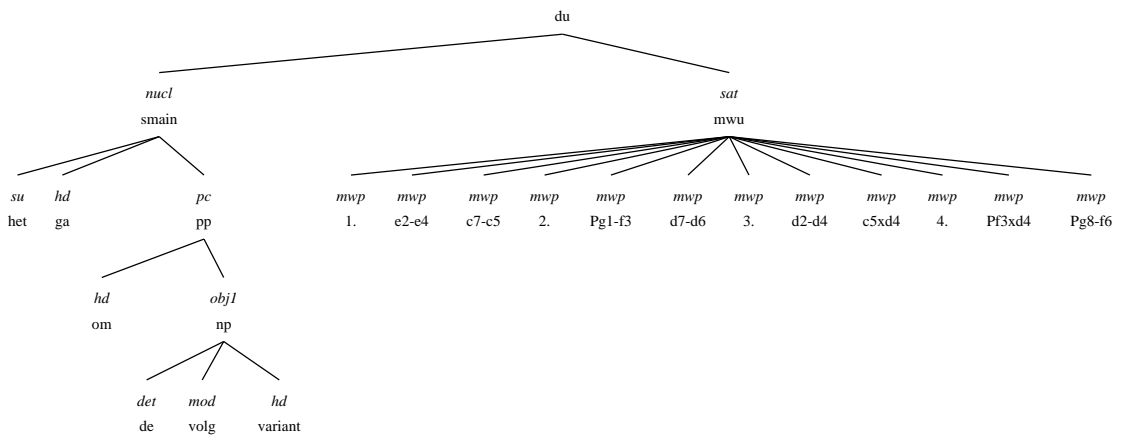
(519) Hij begon zijn rede met: "Welcome ladies and gentlemen"



(520) Vraag nooit waarom , want dan antwoordt de Twent steevast : 't Hef aaltied zo wes .



(521) Het gaat om de volgende variant: 1. e2-e4 c7-c5 2. Pg1-f3 d7-d6 3. d2-d4 c5xd4 4. Pf3xd4 Pg8-f6 ...

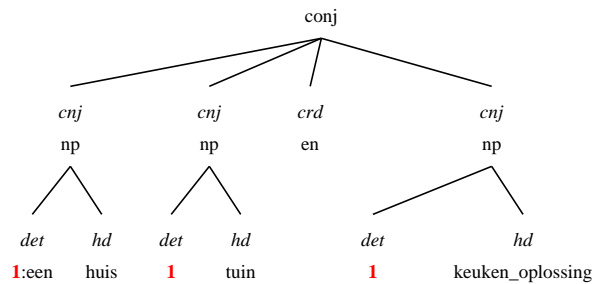


2.8.3 huis-, tuin- en keukenproblemen

Voorbeelden zoals:

(522) de voor- en nadelen

(523) een huis-, tuin- en keukenoplossing



worden geannoteerd als conjunctie, en bijvoorbeeld niet als MWU, omdat in het algemene geval ook andere woorden in of voor de conjuncten kunnen optreden:

(524) Met de linker- en de rechtervoet

(525) Met zijn linker- en met zijn rechtervoet

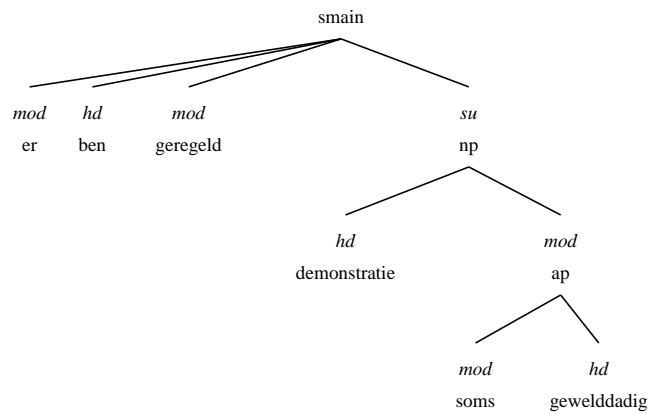
(526) Met zijn linker- en vooral met zijn rechtervoet

2.8.4 Materiaal tussen haakjes

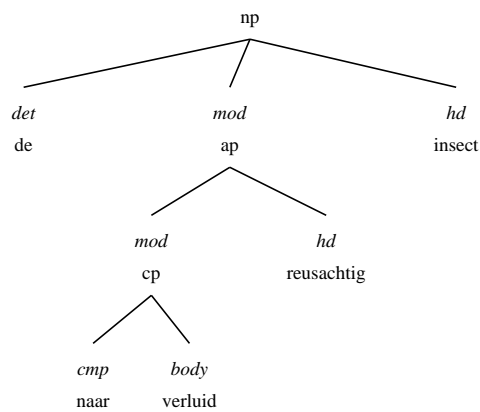
Het is lastig om hiervoor preciese regels op te stellen. In geval het materiaal tussen haakjes een woordgroep vormt, die ook weggelaten kan worden zonder de syntactische structuur van de hele zin te veranderen, dan wordt het tussen haakjes geplaatste zinsdeel meestal als MOD geannoteerd.

In plaats van haakjes kunnen ook gedachtenstreepjes of de schuine streep worden gebruikt.

(527) Er zijn geregeld demonstraties (soms gewelddadig)

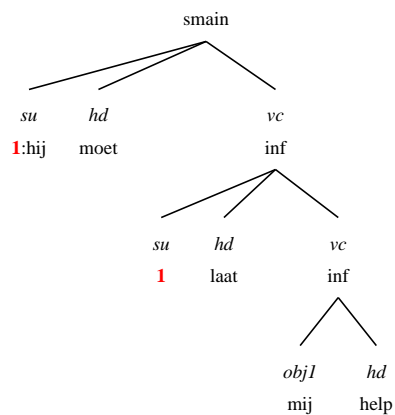


(528) De - naar verluidt - reusachtige insecten

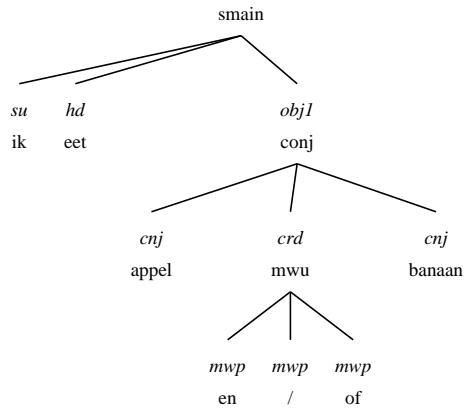


Soms verandert de structuur bij weglating van het tussen haakjes geplaatste materiaal wel degelijk, en lijkt de bedoeling van de spreker om twee verschillende zinnen tegelijk te uiten. Dan kiezen we meestal voor de annotatie van de zin alsof de haakjes er niet zijn:

(529) Hij moet mij (laten) helpen



(530) ik eet appels en / of bananen



In voorbeeld (530) annoteren we *en / of* als een MWU. In de meeste gevallen zal de preprocessor overigens *en/of* als enkel token beschouwen.

Hoofdstuk 3

Niet-lokale dependenties

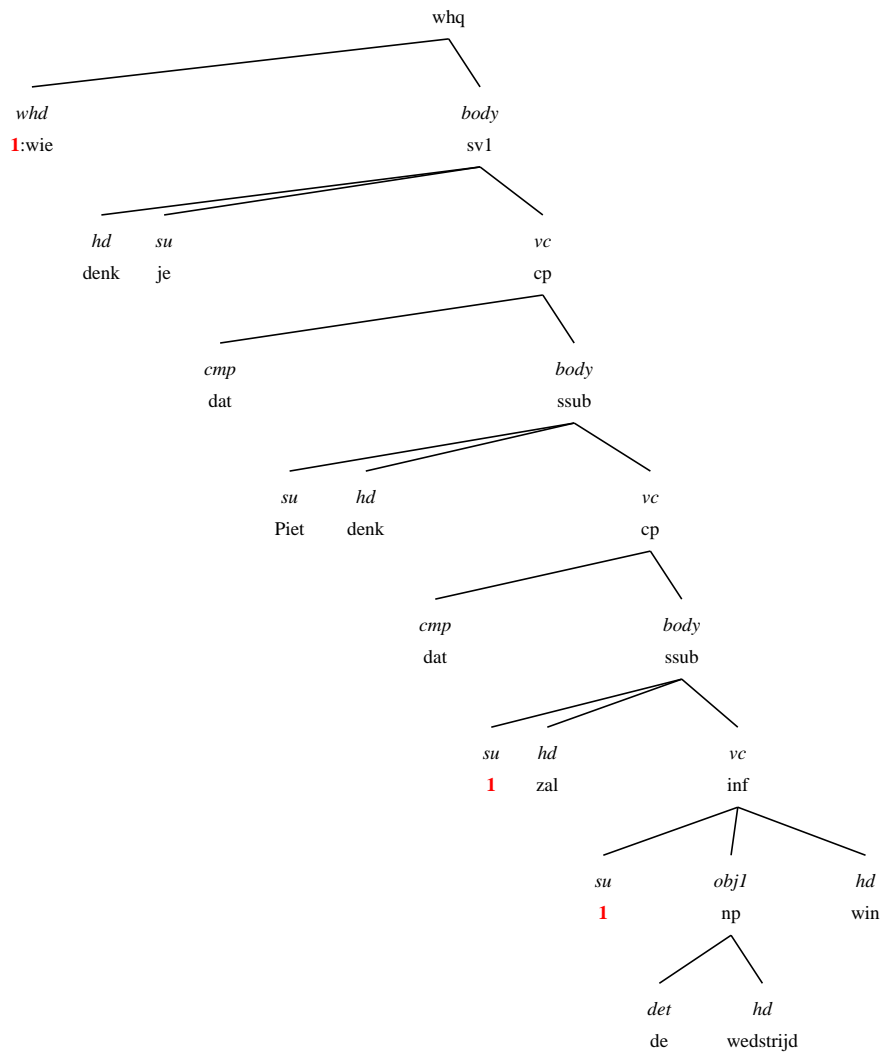
3.1 Introductie

In deze sectie behandelen we de dependentie- en categorielabels in niet-lokale dependenties, zoals die zich voordoen bij:

- relatieve zinnen (met expliciet dan wel met ingesloten antecedent, cf. ANS (Haeseryn, 1997, blz. 329 ev);
- constituentvragen, op hoofdzins- en op bijzinsniveau

De elementen die deze configuraties inleiden (constituenten met een vragend of betrekkelijk voornaamwoord) zijn bepalend voor het categorielabel van de moederknoop (relatieve zin vs vraagzin, ...); dat bestempelt ze tot hoofden. Tegelijk willen we aangeven welke rol deze inleidende elementen vervullen in een dependentiedomein ergens in het vervolgstuk; het relevante dependentiedomein kan arbitrair diep in dat vervolgdeel zijn ingebed. Dit geven we aan door middel van coindexatie, zoals in dit voorbeeld:

(531) Wie denk je dat Piet dacht dat de wedstrijd zou winnen?



Hier fungeert *wie* als hoofd van de vraagzin, en tegelijkertijd vervult het de rol van onderwerp bij de ingebedde persoonsvorm *zou* (en dus ook bij het ingebedde hoofdwerkwoord *winnen*).

In Figuur 3.1 geven we een overzicht van de gebruikte labels.

categorielabel	OMSCHRIJVING
REL	relatieve zin
WHREL	relatieve zin met ingesloten antecedent
WHQ	constituentvraag: hoofdzin
WHSUB	constituentvraag: ondergeschikte zin
dependentielabel	OMSCHRIJVING
WHD	hoofd van WHQ, WHSUB
RHD	hoofd van REL, WHREL
BODY	romp

Figuur 3.1: Labels voor niet-lokale dependenties

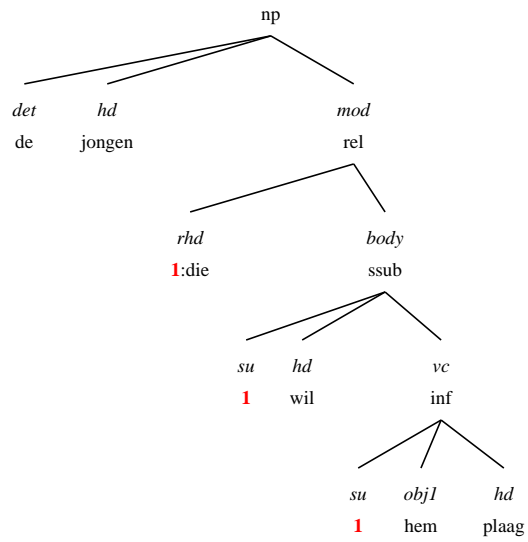
3.2 Relatieve zinnen

3.2.1 Relatieve zinnen als modificeerders bij zelfstandig naamwoorden

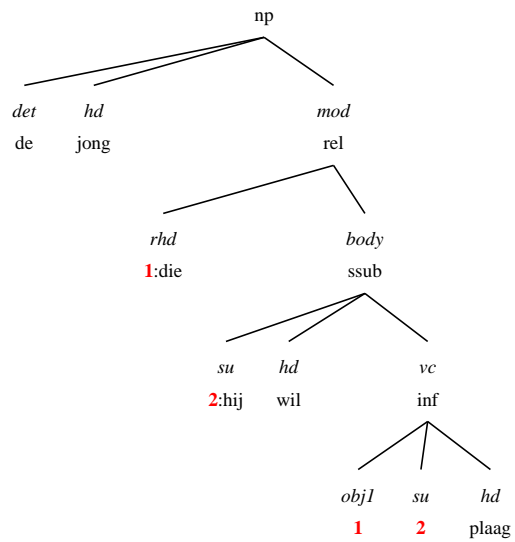
De primaire dependentierelaties worden in de betreffende bijzin als volgt gerealiseerd. Het inleidend element is ofwel een betrekkelijk voornaamwoord op POS-tag niveau, ofwel een constituent die een betrekkelijk voornaamwoord bevat. Het daarop volgende rompdeel is een SSUB. De volledige relatieve zin REL vervult de functie van MOD in het nominale domein. We maken in de annotatie geen onderscheid tussen beperkende en uitbreidende relatieve zinnen.

Hieronder enkele voorbeelden van secundaire dependentierelaties die het RHD-element onderhoudt binnen SSUB.

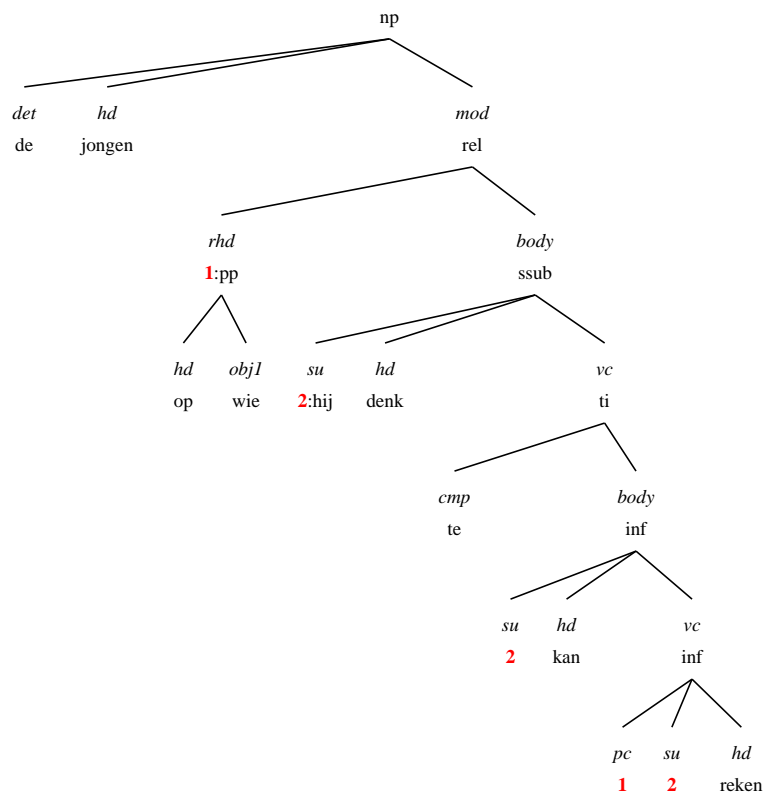
(532) de jongen **die** hem wil plagen



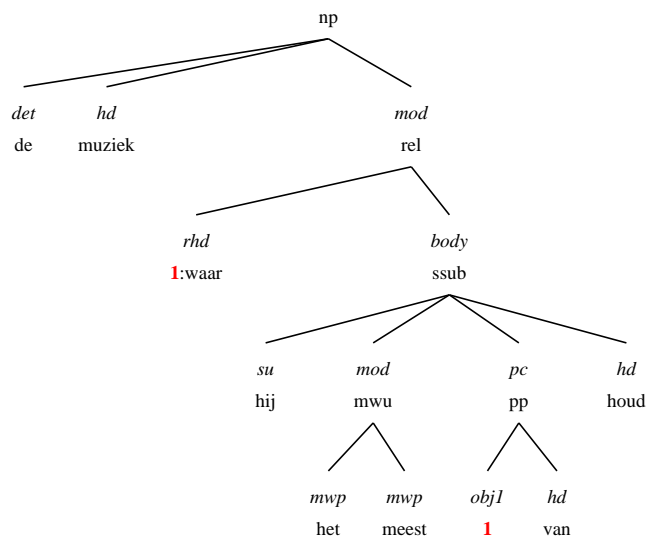
(533) de jongen **die** hij wil plagen



(534) de jongen **op wie** hij dacht te kunnen rekenen



(535) de muziek **waar** hij het meest van houdt

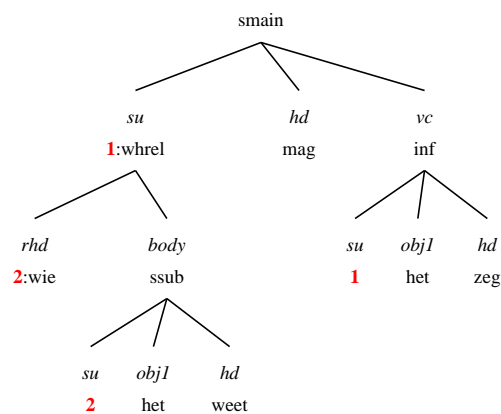


3.2.2 Relatieve zin met ingesloten antecedent

Anders dan de relatieve zin, die een modifierende functie bij een zelfstandig naamwoord heeft, vervult de betreffende bijzin met ingesloten antecedent de functie van een naamwoordelijke constituent.

De voorbeelden hieronder illustreren mogelijke secundaire dependentierelaties binnen SSUB.

(536) **wie** het weet, mag het zeggen



(537) **wat** je zegt, ben je zelf (OBJ1 bij 'zegt')

(538) **wie** de meeste antwoorden goed heeft, geef ik de prijs (OBJ2)

De RHD-elementen in een WHREL zijn op POS-tag niveau verankerd: een grotere XP-constituent waarin het betreffende voornaamwoord is ingebed, is als in-leider van WHREL uitgesloten (een uitzondering lijkt te worden gevormd).

In de voorbeelden die volgen maakt het voorzetsel dus geen deel uit van WHREL; WHREL speelt de rol van OBJ1 bij dat voorzetsel.

(539) nog een aanvulling bij **wat ik je gisteren vertelde** (OBJ1 bij 'bij')

(540) geef het aan **wie het verdient heeft** (OBJ1 bij 'aan')

Het verschil tussen WHSUB en WHREL in termen van de traditionele grammatica: WHSUB is voor afhankelijke vraagzinnen; een WHSUB hangt dan ook af van een werkwoord, bijv. *weten* in:

(541) ik weet niet wat hem bezielt

WHREL is voor relatieve bijzinnen met ingesloten antecedent, hetgeen inhoudt dat het relativum vervangen moet kunnen worden door een aanwijzend voornaamwoord + relativum. In het bovenstaande zinnetje kan dat niet:

(542) *ik weet niet datgene wat hem bezielt

In de zin

(543) ik geloof nog steeds in wat ik gezegd heb

kan dat wel:

(544) ik geloof nog steeds in datgene wat ik gezegd heb

Een WHREL fungeert als, en kan dus vervangen worden door, een 'echte' NP:

(545) ik geloof nog steeds in jou/de quantumtheorie

Een WHSUB kan alleen vervangen worden door een voornaamwoord in de derde persoon, oftewel een 'zinsvervangende' NP, vergelijk

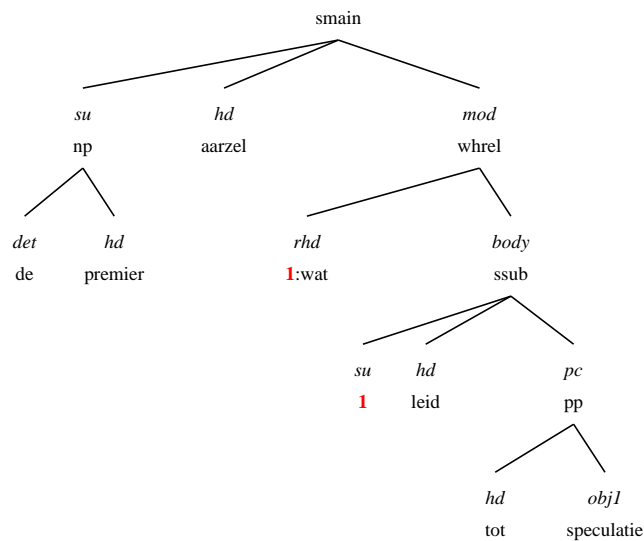
(546) ik weet het niet

(547) ik weet *jou/*de quantumtheorie niet

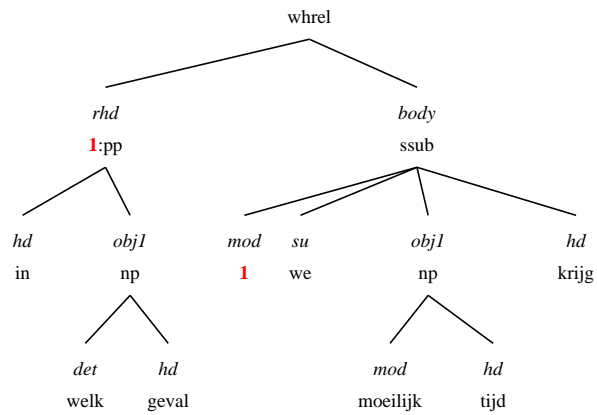
3.2.3 Relatieve zin als modificeerder van werkwoorden

Een relatieve zin met ingesloten antecedent treedt soms op als modificeerder in het werkwoordelijke domein. Ook in die gevallen gebruiken we label REL en dochters RHD en BODY. In die gevallen lijkt de beperking dat de RHD niet complex kan zijn, te vervallen.

(548) De president aarzelde , wat leidde tot speculaties

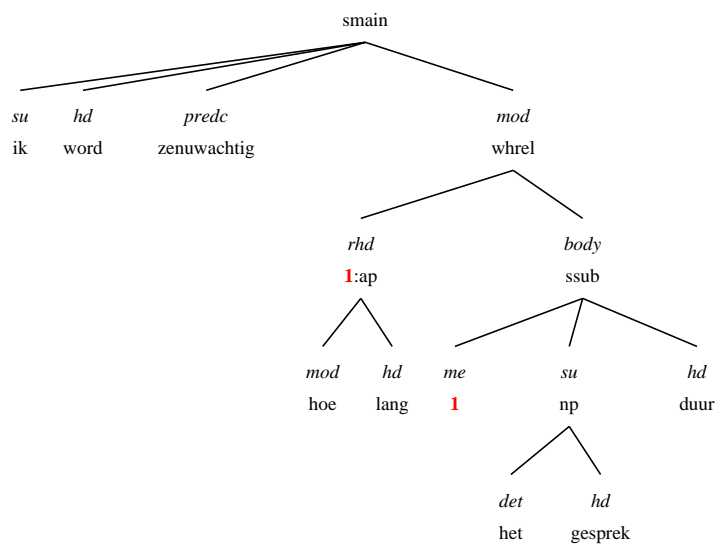


(549) De president sloot sancties niet uit , in welk geval we moeilijke tijden krijgen

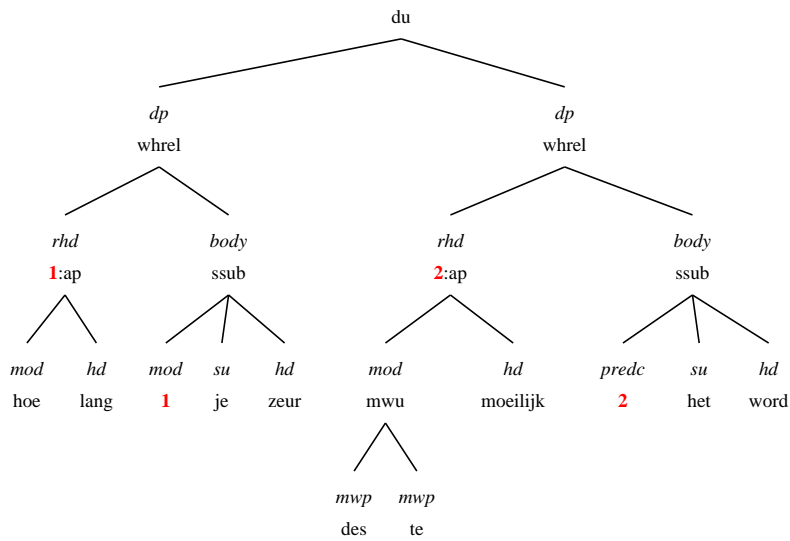


Ook de constructie *hoe* met een adjectief wordt op deze manier geannoteerd; naar analogie hiermee worden zinnen met *des te* geannoteerd. De combinatie *des te* wordt als MWU geannoteerd.

(550) Ik werd zenuwachtiger hoe langer het gesprek duurde



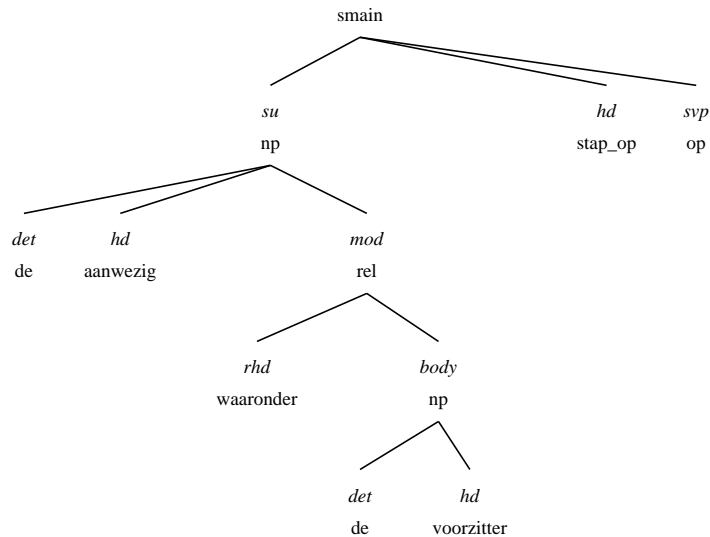
(551) Hoe langer je zeurt , des te moeilijker het wordt



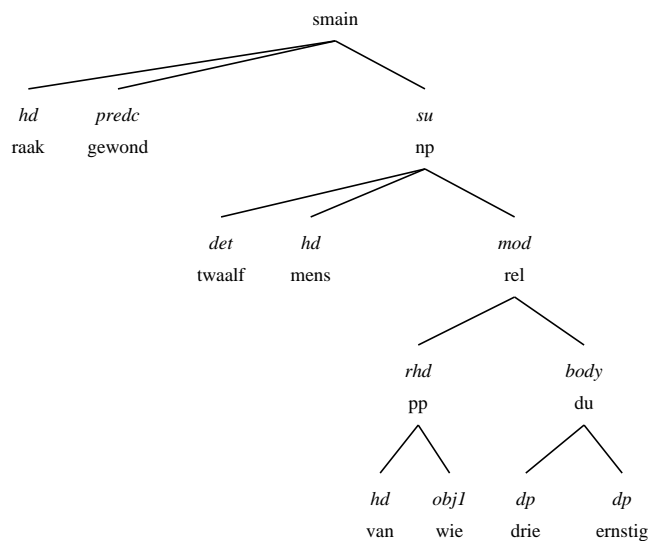
3.2.4 Verkorte relatieve zinnen

Niet altijd leidt een betrekkelijk voornaamwoord een betrekkelijke bijzin, soms volgt een andere categorie zoals NP:

(552) De aanwezigen , waaronder de voorzitter , stapten op



(553) Twaalf mensen raakten gewond, van wie drie ernstig



3.3 Constituentvraag: hoofd- en bijzin

Constituentvragen op het hoofdzins- en bijzinsniveau hebben dezelfde inleiders: vragende voornaamwoorden, dan wel grotere XP-constituenten waarin deze voornaamwoorden (of de DET ‘welk(e)’) zijn ingebed. De twee typen constituentvragen verschillen in de syntactische realisering van het rompdeel: een werkwoords-initiële SV1 in het geval van WHQ, en een SSUB in het geval van WHSUB.

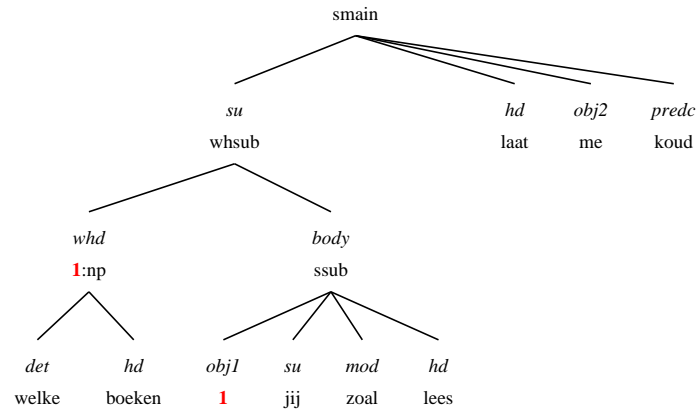
Zoals in het geval van de relatieve zinnen, vervult het WHD-element een secundaire dependentiefunctie binnen de BODY-component. We geven hieronder enkele voorbeelden.

- (554) WHQ: **wat** wil hij zeggen? (OBJ1 bij ‘zeggen’)
- (555) WHSUB: (ik vermoed) **wat** hij wil zeggen (OBJ1 bij ‘zeggen’)
- (556) WHQ: **hoeveel** denk je dat hij weegt? (ME bij ‘weegt’)
- (557) WHSUB: (ik vroeg) **hoeveel** je denkt dat hij weegt (ME bij ‘weegt’)

OPMERKING. In eerdere versies van dit manual werd de zin *Welke boeken jij zoal leest laat me koud* geanalyseerd met een WHREL. Dit is onterecht zoals de volgende zinnnetjes aantonen.

- (558) Het laat me koud dat jij komt
- (559) Dat jij komt laat me koud
- (560) Het laat me koud welke boeken jij leest

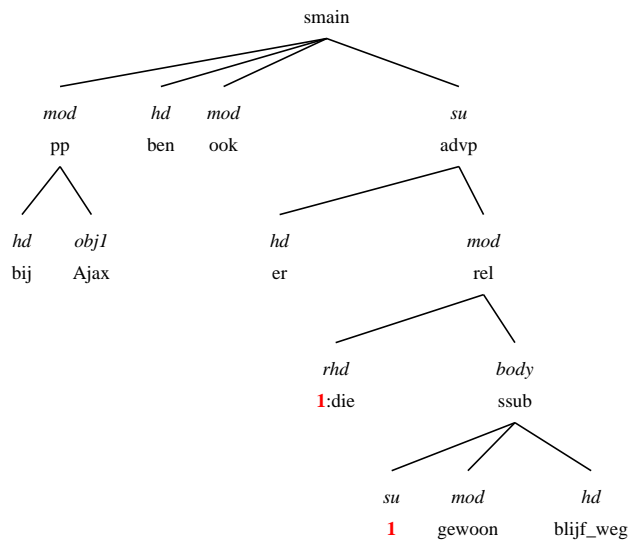
(561) Welke boeken jij zoal leest laat me koud



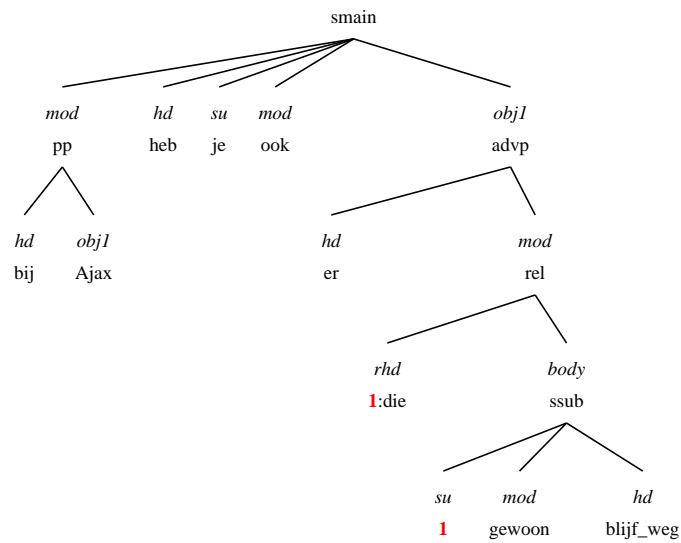
3.4 Er zijn er/Je hebt er

Hier treedt verplicht *er* met een relatieve zin op, die we annoteren als MOD bij *er*:

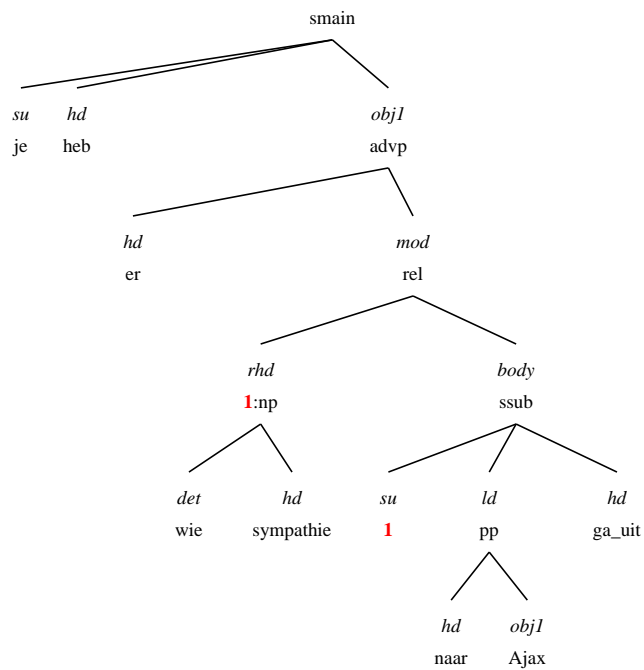
(562) bij Ajax zijn er ook die gewoon wegblijven



(563) bij Ajax heb je er ook die gewoon wegblijven



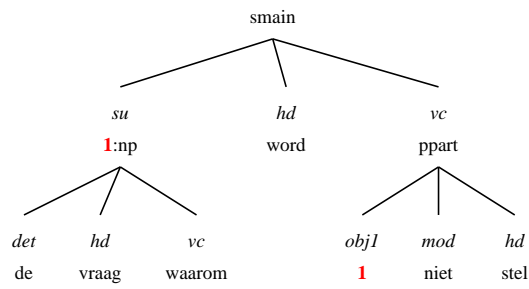
(564) je hebt er wier sympathie naar Ajax uitgaat



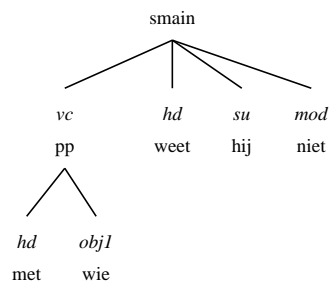
3.5 Verkorte vraagzinnen

In sommige gevallen wordt een vraagzin op bijzinniveau afgekort tot de constituent die normaliter de rol van WHD zou spelen:

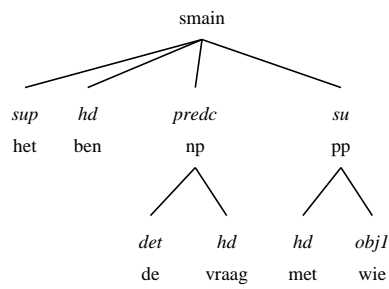
(565) de vraag waarom werd niet gesteld



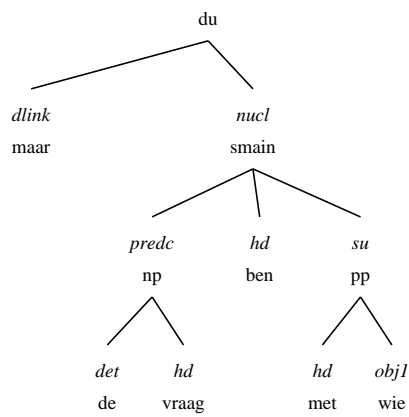
(566) Met wie wist hij niet



(567) Het is de vraag met wie



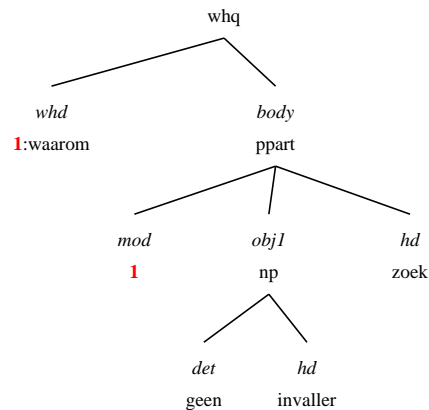
(568) Maar de vraag is met wie



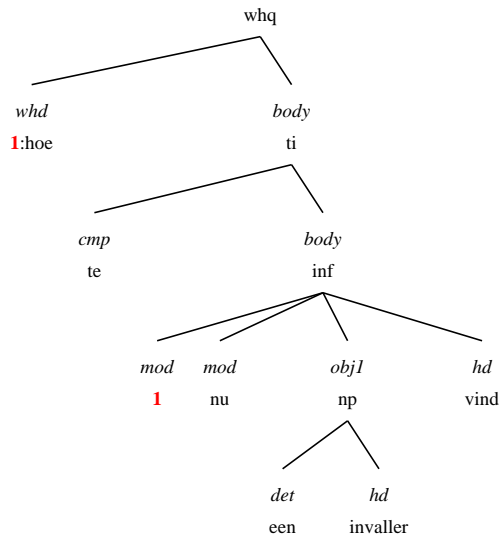
3.6 Vraagzinnen zonder persoonsvorm

Een vraagzin hoeft niet altijd de normale hoofdzinsvolgorde met persoonsvorm te hebben. We annoteren wat er staat:

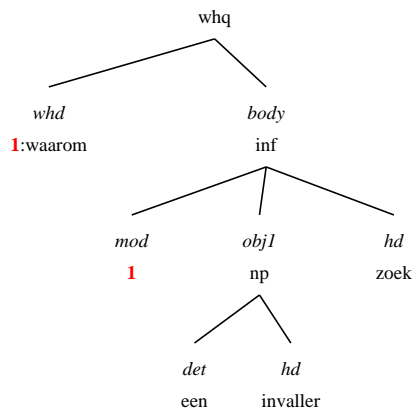
(569) Waarom geen invaller gezocht?



(570) Hoe nu een invaller te vinden?



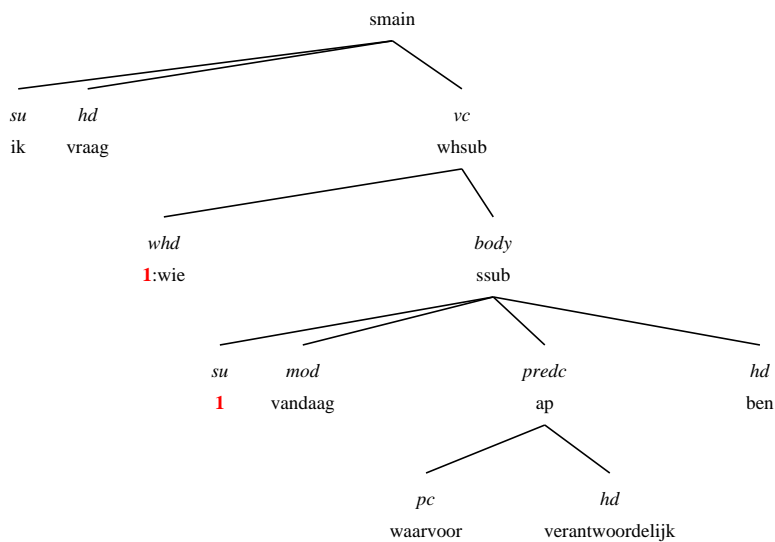
(571) Waarom een invaller zoeken?



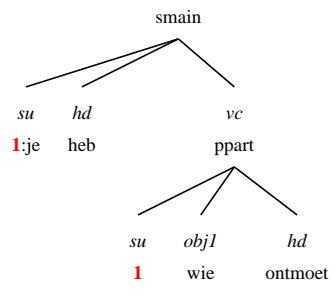
3.7 Vraagzinnen met meerdere WH-elementen

Een vraagzin kan soms meerdere vraagwoorden bevatten. In deze gevallen wordt alleen de WH-groep waarmee de vraag begint als WHD geannoteerd. De andere groepen worden geannoteerd zonder aandacht te besteden aan het vraagwoord. Op vergelijkbare wijze worden vraagwoordconstituenten die niet vooropgeplaatst zijn ook niet als WHD geannoteerd:

(572) Ik vroeg wie vandaag waarvoor verantwoordelijk is



(573) Je hebt wíe ontmoet?



Hoofdstuk 4

Nevenschikking

4.1 Inleiding

De dependentie- en categorielabels voor de annotatie van nevenschikkingen zijn te vinden in Figuur 4.1. Schematisch ziet het structurele patroon voor een nevenschikking eruit als in (4.1).

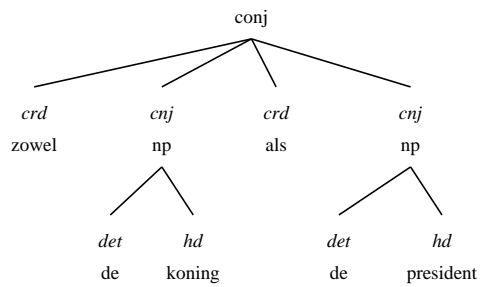
- In een nevenschikking worden de leden van de nevenschikking met dependentielabel CNJ verbonden door nevenschikkende voegwoorden of reeksvormers met dependentielabel CRD. De moederknoop krijgt altijd het categorielabel CONJ.

(574) Zowel de koning als de president

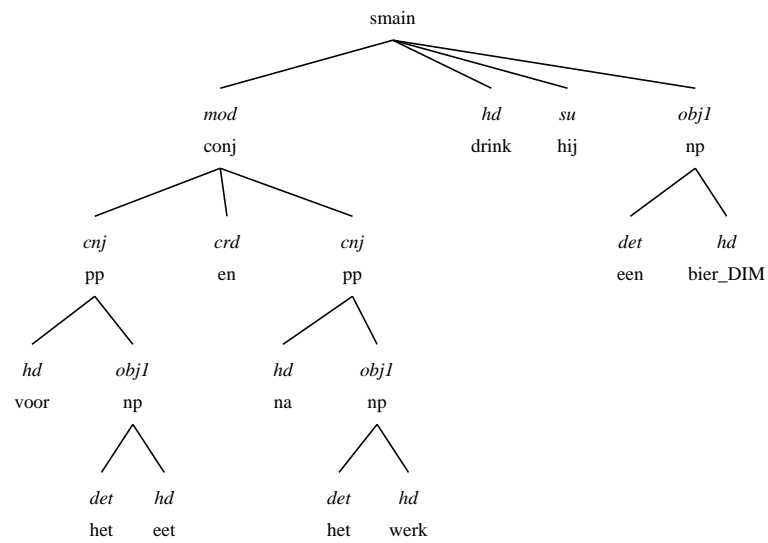
dependentielabel	OMSCHRIJVING
CRD	nevenschikker (voegwoord of reeksvormer)
CNJ	lid van nevenschikking

categorielabel	OMSCHRIJVING
CONJ	nevenschikking (polymorf)

Figuur 4.1: Labels voor nevenschikking

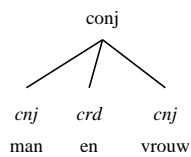


(575) Voor het eten en na het werk drinkt hij een biertje

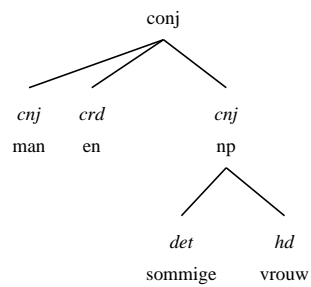


- Leden van een nevenschikking kunnen geled of ongeled zijn.

(576) Mannen en vrouwen



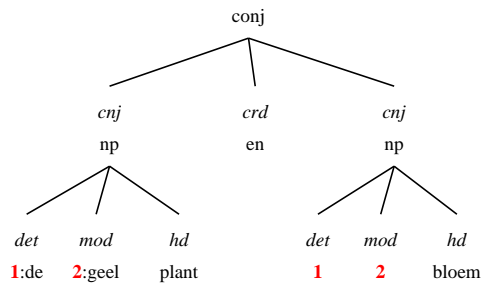
(577) Mannen en sommige vrouwen



- Wel is het zo, dat we blijven uitgaan van de regel dat hoofden normaliter niet splitsen.

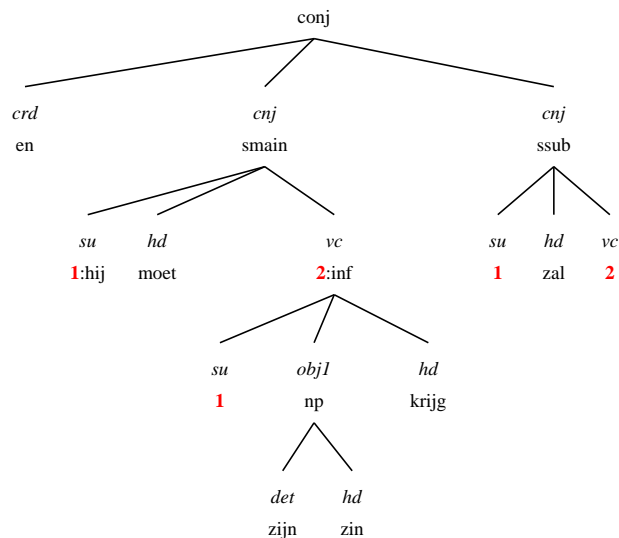
(578) De gele planten en bloemen

Dit voorbeeld is meerduidig. *Gele* kan betrekking hebben op alleen *planten*. In dat geval annoteren we zoals voorbeeld (577). Indien de meer voor de hand liggende lezing wordt gekozen waarbij *Gele* ook betrekking heeft op *bloemen*, dan wordt geen extra domein geopend, maar worden door middel van secundaire dependentierelaties de relaties vastgelegd:

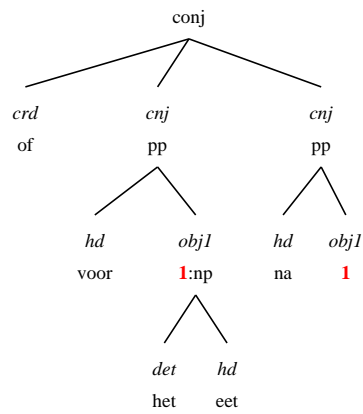


Deze aanpak wordt dus steeds gevolgd indien het hoofd van een projectie of een niet-maximale projectie gecoördineerd wordt, zoals in de volgende voorbeelden:

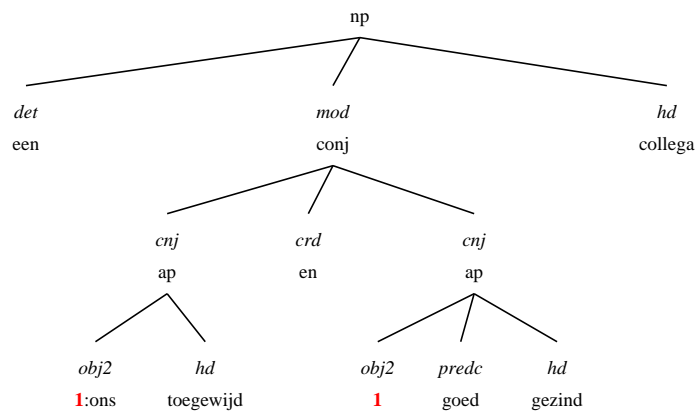
(579) Hij moet en zal zijn zin krijgen



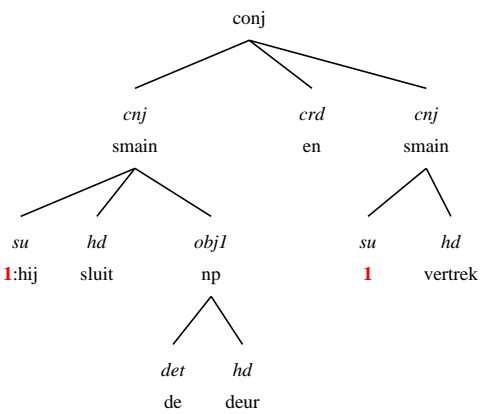
(580) Voor of na het eten



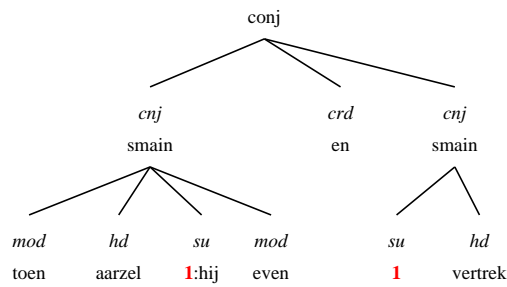
(581) Een ons toegewijde en goed gezinde collega



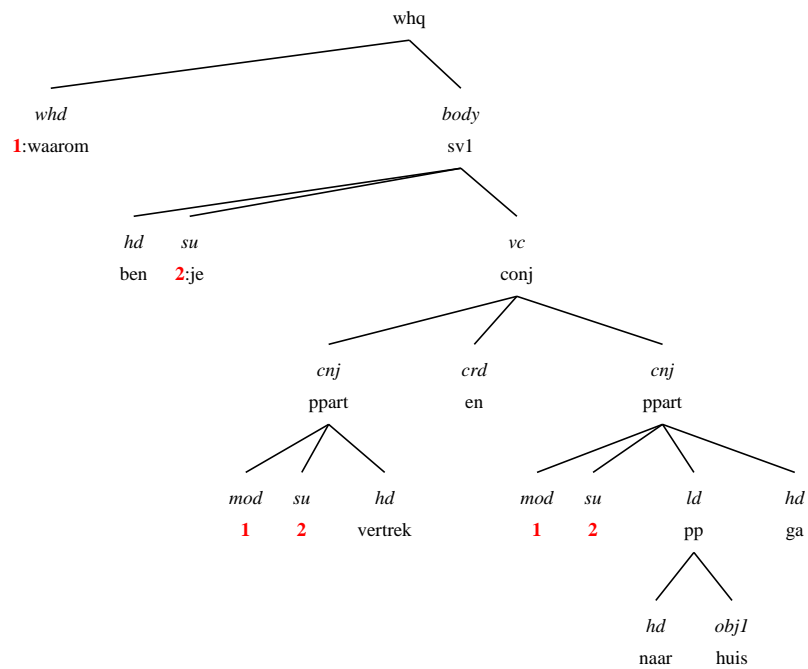
(582) Hij sloot de deur en vertrok



(583) Toen aarzelde hij even en vertrok

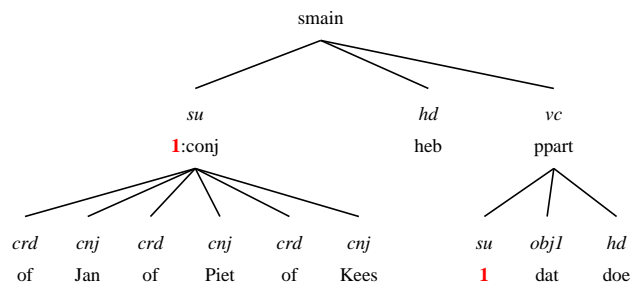


(584) Waarom ben je vertrokken en naar huis gegaan?

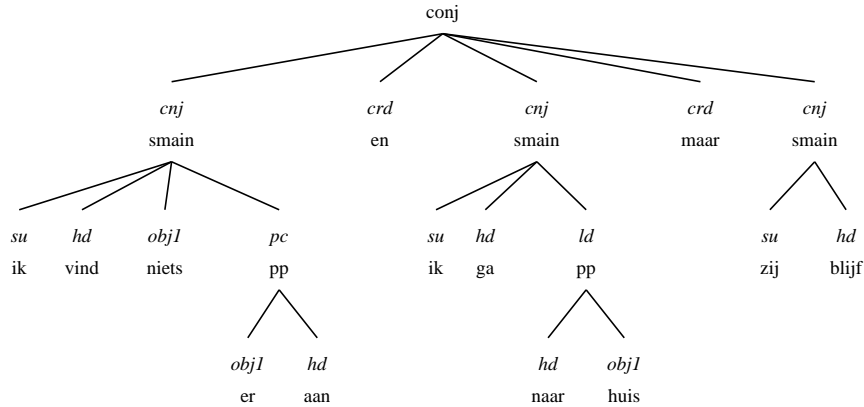


- In geval er meerdere coordinatoren zijn, kan men soms denken aan een coordinatie waarbij de conjuncten zelf weer uit een coordinatie bestaan. Wij kiezen bij voorkeur voor een platte analyse, met meerdere CRD dochters:

(585) Of Jan of Piet of Kees heeft dat gedaan



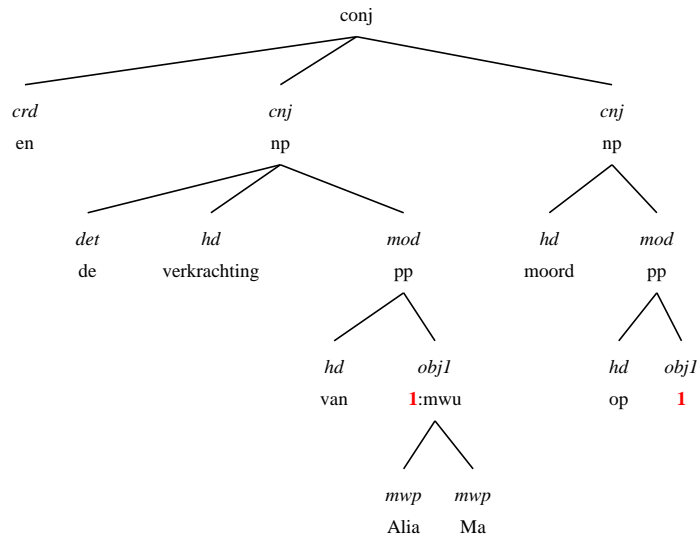
(586) Ik vond er niets aan en ik ging naar huis maar zij bleef



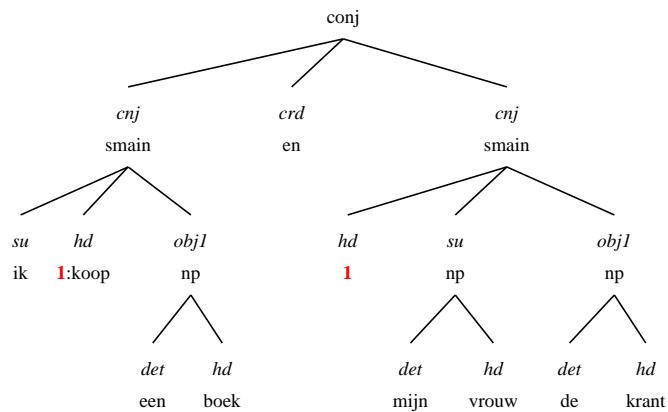
- In het geval van nevenschikking waarbij duidelijk sprake is van samentrekking (ANS, (Haeseryn, 1997, hoofdstuk 27)) vullen we de restantleden aan door middel van secundaire dependentierelaties met betrekking tot de volledige leden.

Voorbeelden van samentrekking:

(587) De verkrachting van en moord op Alia Ma



(588) Ik kocht een boek en mijn vrouw de krant

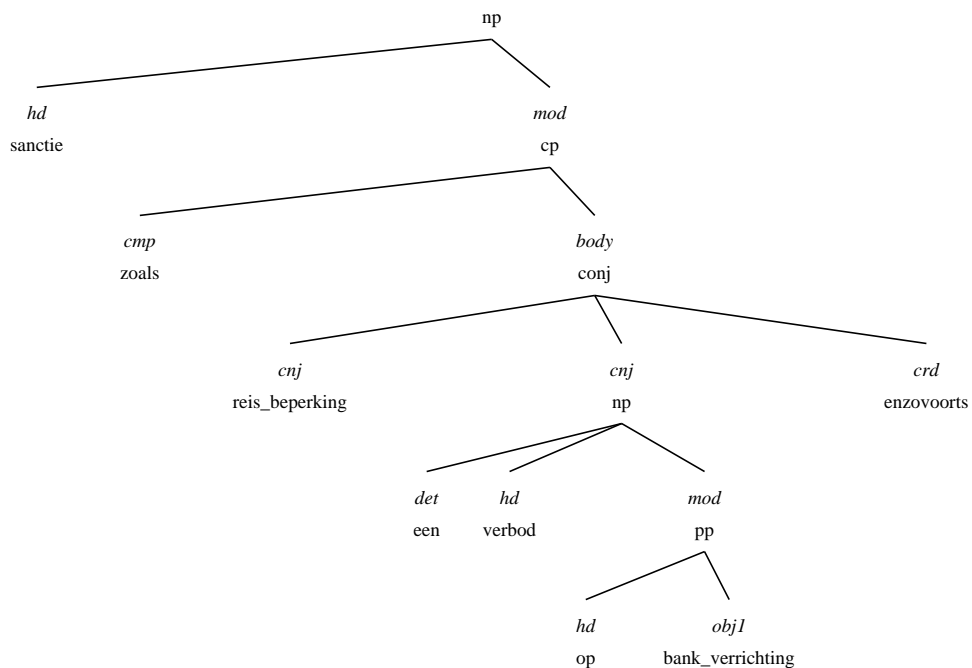


Opmerking. *De meeste gevallen van nevenschikking met samentrekking worden niet door Alpino herkend.*

4.2 Etcetera

Een coordinatie kan afgesloten worden met *enzovoorts* of *etcetera*. Deze woordgroep lijkt zowel de rol van de laatste coordinator te spelen, als de laatste conjunct te bevatten. Wij kennen de relatie CRD toe aan deze elementen.

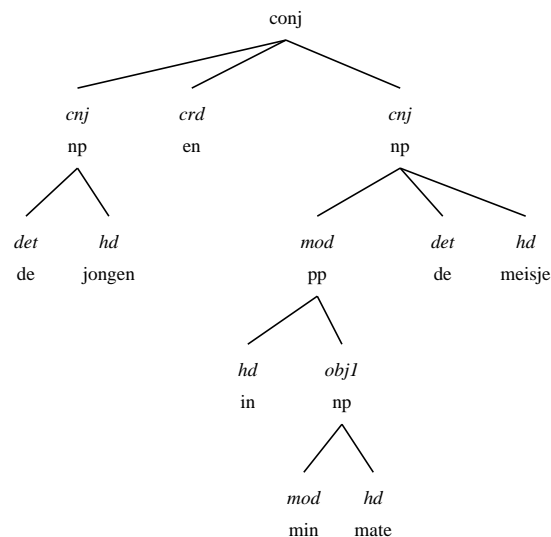
(589) sancties zoals reisbeperkingen , een verbod op bankverrichtingen , enzovoorts



4.3 Modificatie van conjuncten

Conjuncten in coordinaties kunnen vaak gemodificeerd worden:

(590) De jongens en in mindere mate de meisjes

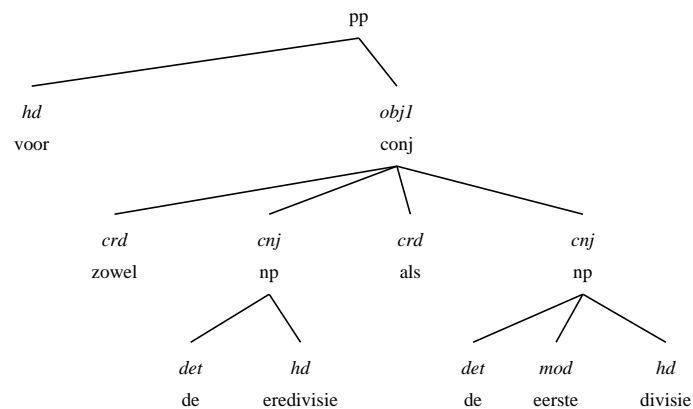


In deze gevallen hechten we de modifier aan het hoofd van de conjunct, en er wordt dus geen apart niveau ingevoerd.

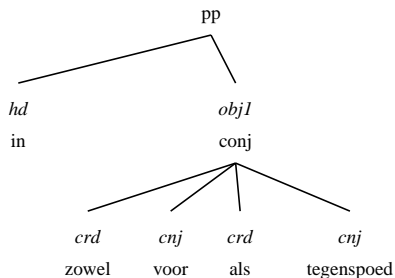
4.4 Onverwachte woordvolgorde bij *zowel ... als*

In voorbeelden als de volgende lijkt het woordje *zowel* thematisch gezien ten onrechte links van het voorzetsel op te treden:

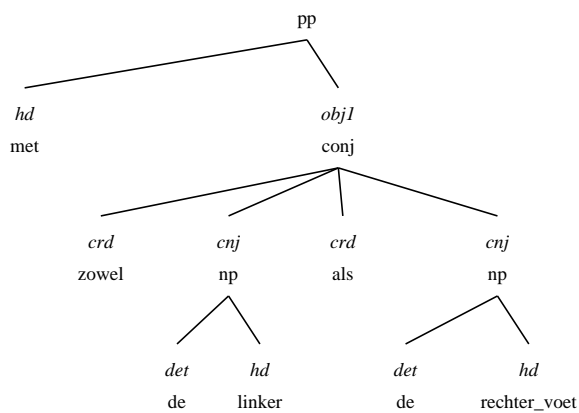
(591) *zowel* voor de eredivisie als de eerste divisie



(592) Zowel in voor- als tegenspoed



(593) Zowel met de linker- als de rechterspoed



Deze gevallen worden dus geannoteerd als de corresponderende zin *in zowel voor- als tegenspoed*.

4.5 Asyndetische nevenschikking, en lijsten

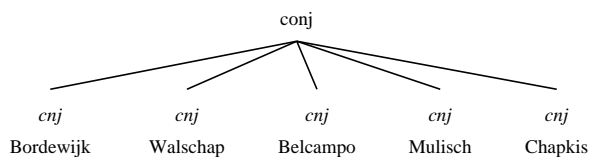
Asyndetische nevenschikkingen, zoals

(594) zijn er verder nog vragen, opmerkingen?

passen niet goed in de hoofd-gebaseerde aanpak van de annotatie. We beschouwen dit dan maar als een coordinatie zonder coordinator.

Dezelfde analyse gebruiken we voor lijsten:

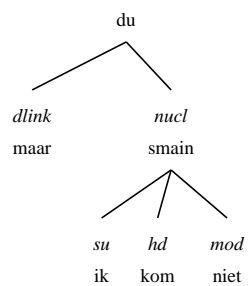
(595) Bordwijk, Walschap, Belcampo, Mulisch, Chapkis



4.6 Woordgroepen die met een coordinator beginnen

Uitingen kunnen soms beginnen met een coordinator zoals *en*, *maar*. In deze gevallen wordt de DLINK dependentierelatie uit hoofdstuk 5.2 op bladzijde 186 gebruikt.

(596) Maar ik kom niet



Hoofdstuk 5

Discourse-structuren

5.1 Overzicht

De annotatie is gericht op het aanbrengen van zinsinterne dependentierelaties. Daarbij worden we ook geconfronteerd met dependenties die behoren tot het domein van de discourse. Het betreft enerzijds uitbreidingen van de kernzin, anderzijds asyndetische constructies waarbinnen de samenhang alleen in een wijder discourse-verband vast te stellen valt. Figuur 5.1 geeft een overzicht van de dependentielabels en categorielabels die in deze sectie aan bod komen.

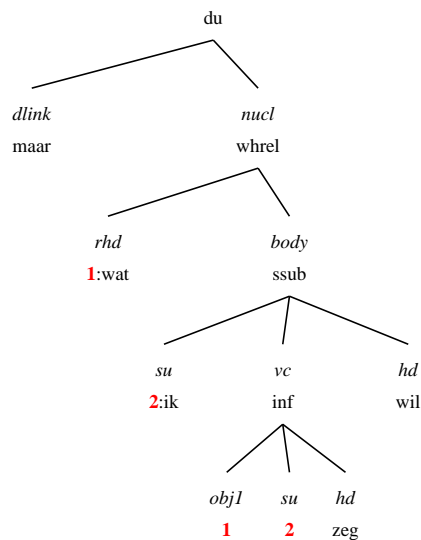
dependentielabel	OMSCHRIJVING
NUCL	kernzin
DLINK	discourse-link
SAT	satelliet: aan- of uitloop
TAG	aanhangsel, tussenvoegsel
DP	discourse-part
categorielabel	OMSCHRIJVING
DU	discourse-unit

Figuur 5.1: Labels voor discourse structuren

5.2 Discourse-markeerders: DLINK

In samenhangende tekst komt het vaak voor dat een spreker een zin begint of eindigt met iets dat op een voegwoord lijkt, maar geen syntactisch verband heeft:

(597) maar wat ik zeggen wou ...



(598) dus je bent zeker wel blij dat je niet naar Brussel bent gegaan?

(599) is dat nou gewoon alleen een broek of ...

Deze elementen krijgen het dependentielabel DLINK. De kernzin krijgt het dependentielabel NUCL. De moederknoop krijgt het categorielabel DU (discourse unit).

Merk op dat woorden als *dus* verschillende analyses krijgen afhankelijk van de zinspositie waarin ze voorkomen:

(600) Jan gaat dus naar huis (na de persoonsvorm: MOD)

(601) dus gaat Jan naar huis (eerste zinspositie: MOD)

(602) dus Jan gaat naar huis (buiten zinsverband: DLINK)

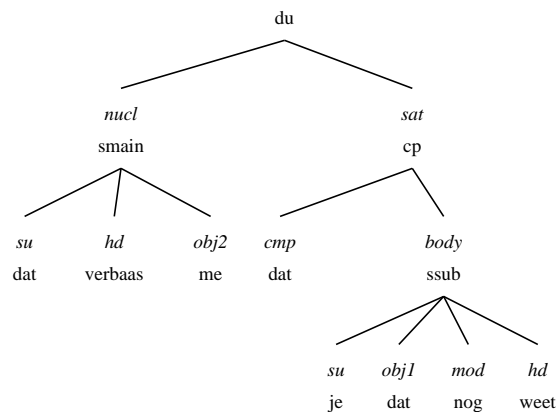
5.3 Uitbreidingen van de zin: SAT

Onder de bedoelde uitbreidingen vallen ‘aanloop’- en ‘uitloop’-elementen die een anaforische relatie onderhouden met verwijzwoorden in de kernzin. Zie ANS,

(Haeseryn, 1997, hoofdstuk 21.8 en 21.9). De aanloop/uitloop-constructies kunnen met leestekens gemarkeerd zijn waardoor ze van extrapositie te onderscheiden zijn: vergelijk (603) en (604). Maar ook zonder die intonatie-informatie is (604) als NUCL+SAT te identificeren: immers, het element *dat* in de kernzin kwalificeert niet als dummy-SUP.

(603) **het** verbaast me **dat je dat nog weet** (SUP + SU)

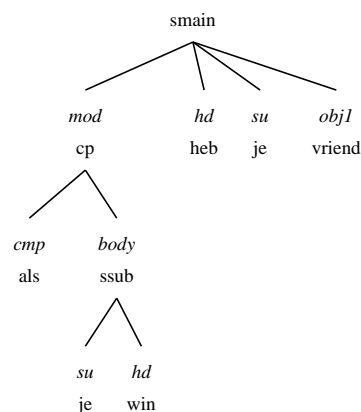
(604) dat verbaast me , **dat je dat nog weet** (NUCL + SAT)



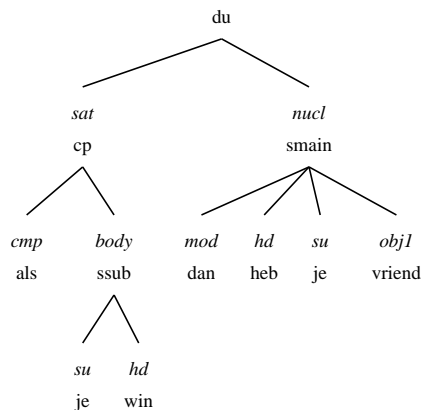
We gebruiken het dependentielabel SAT (satelliet) zowel voor de aanloop- als voor de uitloopelementen.

Ook *als-dan*-constructies worden op deze manier behandeld: vergelijk (605) en (606). *Dan* fungeert hier als het verwijzwoord in de kernzin waar de *als*-zin een relatie mee onderhoudt.

(605) **als je wint** heb je vrienden (eerste zinspositie: MOD)



(606) **als je wint** dan heb je vrienden (buiten zinsverband: SAT)



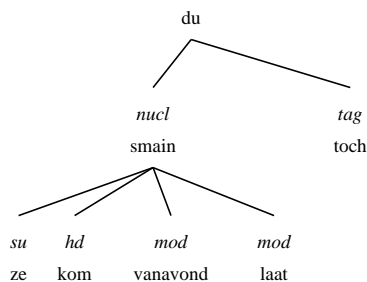
5.4 Uitbreidingen van de zin: TAG

Een tweede type uitbreiding van een NUCL-component wordt gevormd door een heterogene groep van elementen die, op een andere manier dan de hierboven beschreven DLINK en SAT, buiten de kernzin staan. We geven deze uitbreidingen het dependentielabel TAG.

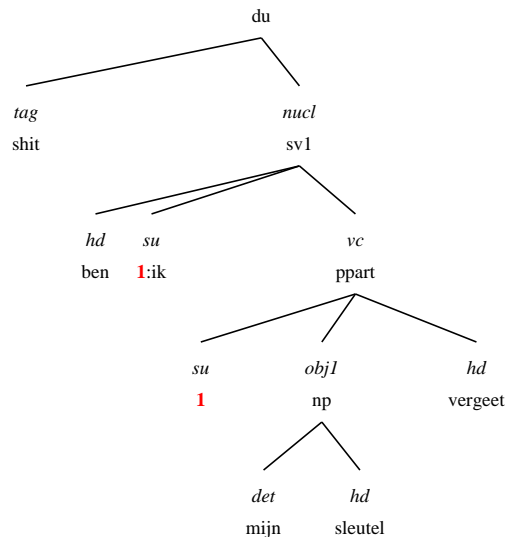
5.4.1 Tussenwerpsels

Een eerste groep van TAG-uitbreidingen zijn elementen die op POS-tag niveau als tussenwerpsels zijn benoemd, dan wel die een tussenwerpsel-achtige functie vervullen, ook al hebben ze een andere POS-tag waarde gekregen (bijvoorbeeld, zelfstandig naamwoord, werkwoord, ...). Het functielabel TAG maakt in die gevallen duidelijk dat de elementen niet hun 'normale' rol vervullen van bijvoorbeeld (MOD binnen de zin), zelfstandig naamwoord (hoofd van een NP), etc.

- (607) **hè**, wat zitten we hier toch weer gezellig (tussenwerpsel)
- (608) 't was gisteren weer ouderwets gezellig, **hè** (tussenwerpsel)
- (609) je gelooft me niet, **wel**? (bijwoord)
- (610) ze kwam vanavond later, **toch**? (bijwoord)



- (611) **verrek**, da's waar ook (werkwoord)
 (612) **shit**, ben ik mijn sleutels vergeten (zelfstandig naamwoord)



- (613) **God**, wat was ik misselijk (eigenaam)
 (614) **Jan**, doe jij de deur even dicht? (eigenaam)

5.4.2 Gereduceerde zinnen

Een tweede groep van TAG-elementen wordt gevormd door (gereduceerde) zinnen die als inleider, aanhangsel of tussenvoegsel gerealiseerd worden. Ook hier maakt het TAG-label duidelijk dat het niet gaat om imperatieven of vragen in de gebruikelijke zin.

- (615) dat is lang geleden, **zeg**
 (616) **zeg eens**, was je nu nog van plan iets te vertellen
 (617) **kom**, ik ga er maar weer eens vandoor
 (618) **kijk**, daar heb je haar ook weer
 (619) **moet je horen**, ik ben hier niet voor niks naartoe gekomen
 (620) dan heb ik zoiets van: laat maar, **weet je**
 (621) hij is daar nog niet klaar voor, **denk ik**
 (622) hij is daar, **denk ik**, nog niet klaar voor

5.4.3 Aansporingen

We gebruiken de combinatie NUCL+TAG ook voor aansporingen met een vocatief en een kern die een imperatiefzin is.

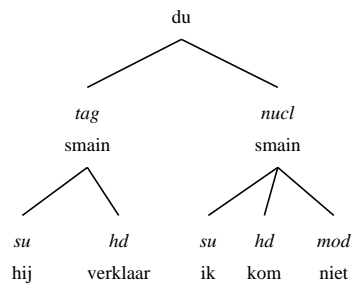
(623) hou je mond, **jij** (TAG)

(624) **jongens**, opgepast (TAG)

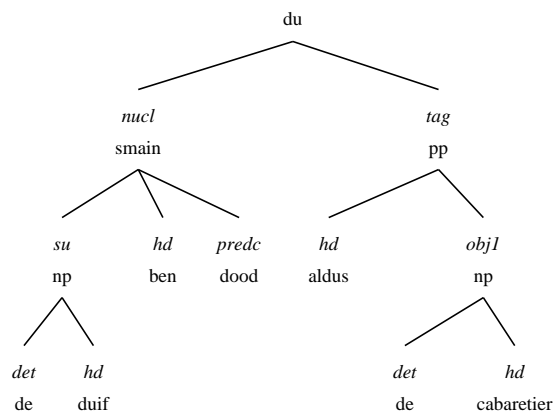
5.4.4 Inleiding van directe rede

Een ander geval waarbij de NUCL+TAG structuur wordt gebruikt betreft inleiders van aanhalingen in de directe rede. Of hierbij de directe rede wel of niet met aanhalingstekens begeleid wordt, is niet van belang.

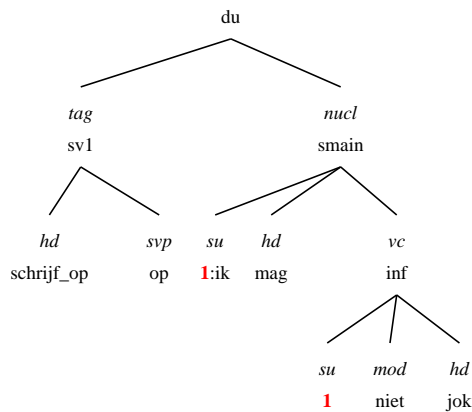
(625) Hij verklaarde : “Ik kom niet”



(626) “De duif is dood”, aldus de cabaretier

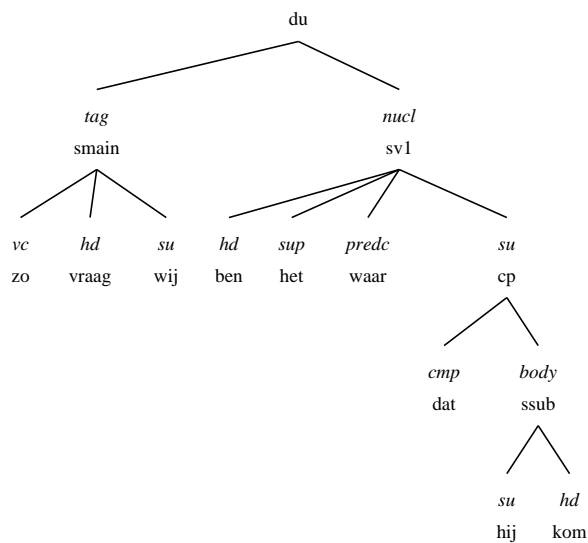


(627) Schrijf op: “Ik mag niet jokken”.

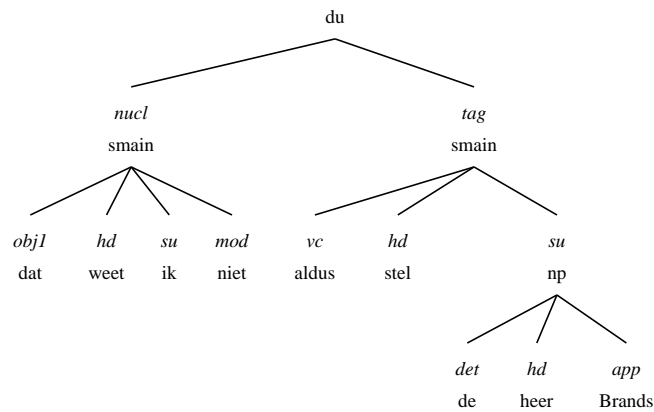


In deze context merken we op, dat werkwoorden die typisch zo'n directe rede inluiden, vaak beginnen met de bijwoorden *zo* of *aldus* die hier lijken op te treden als plaatsvervanger van een sententieel complement (VC). In deze gevallen wordt het bijwoord als VC geannoteerd:

(628) Is het waar, zo vroegen wij, dat hij komt?



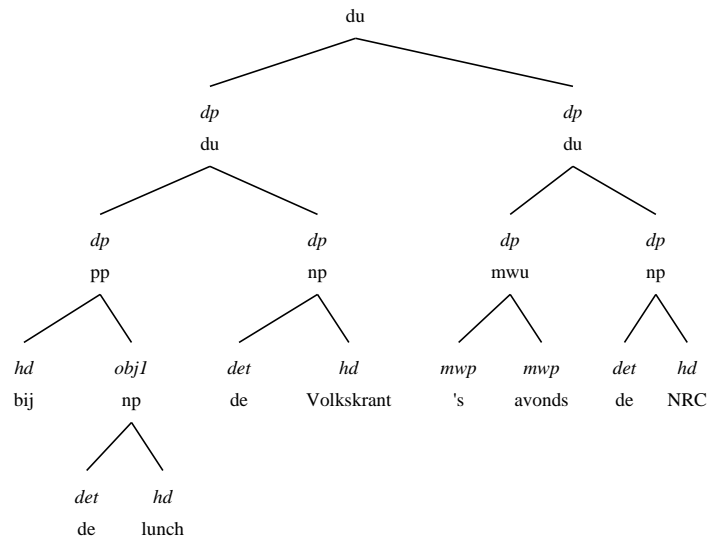
(629) Dat wist ik niet, aldus stelt de heer Brands



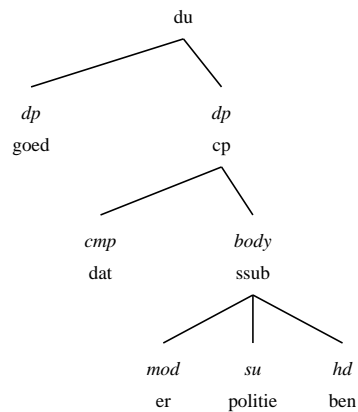
5.5 Asyndetische constructies: DP

Met asyndetische constructies doelen we hier op reeksen van woordgroepen waar-tussen geen expliciete dependentierelaties zijn uitgedrukt (vergelijk ook sectie 4.5 op bladzijde 183 over asyndetische nevenschikking), maar die als elliptische zin-nen kunnen worden opgevat. Bijvoorbeeld:

(630) (welke kranten lees jij?) bij de lunch de Volkskrant; 's avonds de NRC

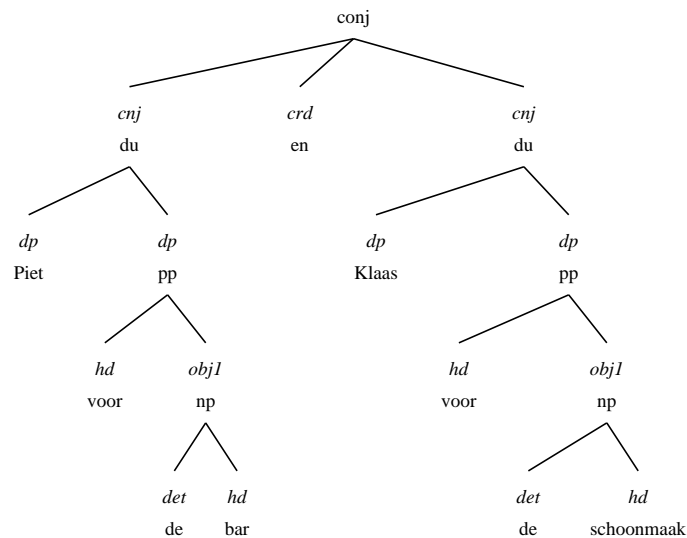


(631) goed dat er politie is



(632) (Wordt er gevoetbald?) In Amsterdam wel

(633) (Wie is er verantwoordelijk?) Piet voor de bar en Klaas voor de schoonmaak



We geven de coherentie in dergelijke reeksen weer door de frases onder een DU-knoop samen te voegen. We zien er evenwel van af om expliciete dependentierelaties te reconstrueren: we beschouwen dit als een inferentie-taak, niet als een taak van de basisannotatie zelf. De dochters van DU krijgen in deze gevallen een uniform dependentielabel DP ('discourse-part').

Het laatste voorbeeld toont bovendien aan dat een dergelijke discourse unit niet altijd de topknoop van de uiting vormt.

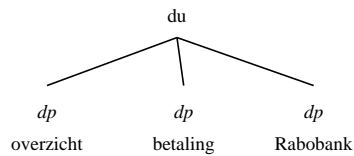
5.5.1 Lijsten

Voor lijsten hebben we de keuze tussen een hoofdloze coordinatie aan de ene kant, en een asyndetische constructie zoals in de vorige paragraaf behandeld aan de andere kant. Bij lijsten van namen en dergelijke kiezen we voor een coordinatie zonder CRD.

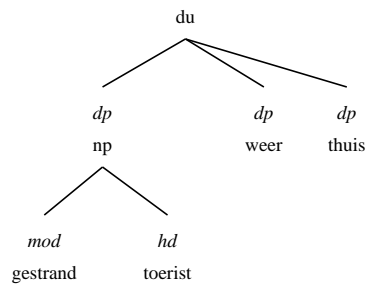
5.5.2 Kopteksten

In schrijftaal komen vaak ook kopjes en dergelijke voor, die ook met behulp van de DP labels worden geannoteerd, er wordt dus geen poging gedaan de uitgebreide versie van de kopjes te herconstructuren.

(634) Overzicht betalingen RABObank

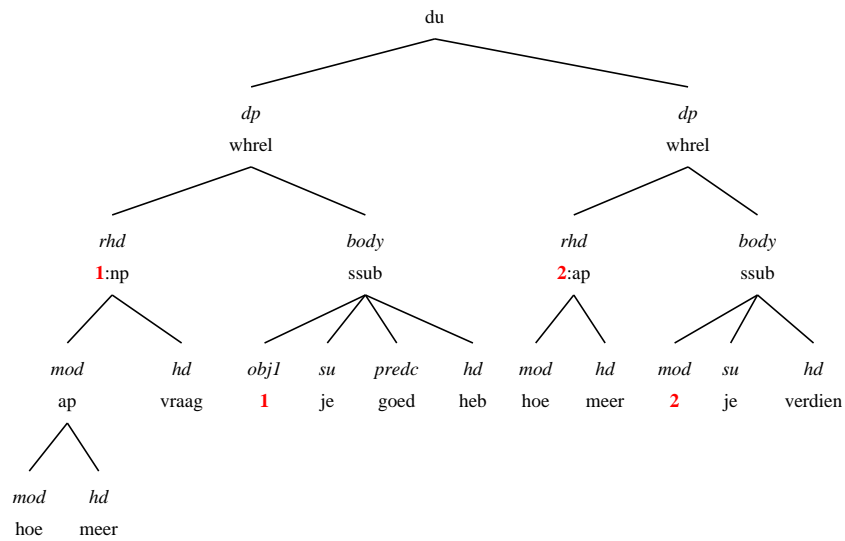


(635) Gestrande toeristen weer thuis



5.5.3 Hoe meer, hoe beter

(636) hoe meer vragen je goed hebt, hoe meer je verdient



Zowel het elliptische *hoe meer hoe beter* als de 'complete' gevallen met twee werkwoorden (*hoe meer vragen je goed hebt*, *hoe meer je verdient*) analyseren we als een asyndetische DU (DP DP), bestaande uit twee WHRELS.

Hoofdstuk 6

Dekking, desambiguatie en ongrammaticale invoer

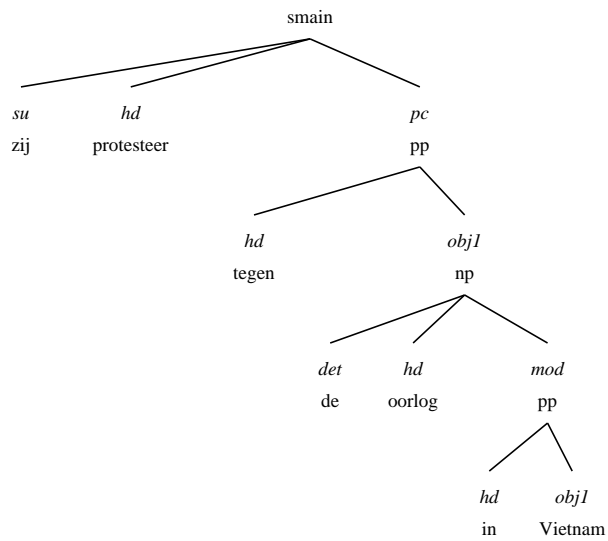
6.1 Dekking

Hoewel in dit document een fors aantal constructies van het Nederlands wordt beschreven, is het vanzelfsprekend onmogelijk om de complete Nederlandse grammatica hier uitputtend te behandelen. Niet behandelde fenomenen dienen zo veel mogelijk in de geest van dit document te worden geanalyseerd. De annotatoren documenteren dit soort gevallen, alsmede de genomen beslissingen. Daarnaast proberen de annotatoren natuurlijk consistent te zijn met eerdere annotatiebeslissingen. Het verdient dus aanbeveling voor de annotator om in geval van twijfel te zoeken naar eerdere voorkomens van de constructie. Ten slotte kan de Alpino parser gebruikt worden om te zien of de door Alpino gebouwde analyse toepasbaar is. Deze methode moet natuurlijk met zorgvuldigheid worden toegepast.

6.2 Desambiguatie

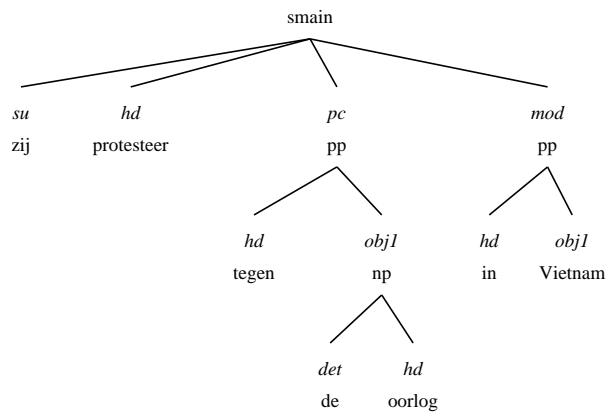
Daarnaast is de moeilijkste taak van de annotatoren om in geval er meerdere mogelijke annotaties zijn (bijvoorbeeld in geval van PP-ambiguiteit), de meest voor de hand liggende annotatie te kiezen. Hierbij kan gebruik worden gemaakt van de context, of van andere kennis. Dit kan ertoe leiden dat dezelfde zin op andere plaatsen in het corpus op andere wijze wordt geannoteerd. Bijvoorbeeld:

(637) Zij protesteren tegen de oorlog in Vietnam



Dit is de voor de hand liggende analyse in geval de context geen uitsluitel geeft. Echter, indien bovenstaande zin in een context voorkomt, kan de keus anders worden:

(638) De anti-Golfoorlog-demonstranten arriveerden vanochtend in Saigon. Zij protesteren tegen de oorlog in Vietnam.



Soms is het eigenlijk niet goed mogelijk om tussen verschillende aanhechtingen te kiezen. Indien de keus is tussen de aanhechting aan een hulpwerkwoord of modaal werkwoord aan de ene kant, of de aanhechting aan het hoofdwerkwoord, dan wordt het hoofdwerkwoord gekozen. In alle andere gevallen hebben we in onduidelijke gevallen een voorkeur voor de hoogste aanhechting.

6.3 Ongrammaticale invoer

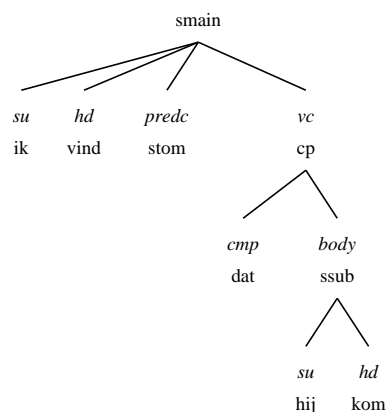
Ten slotte besteden we hier enige aandacht aan ongrammaticale invoer. Hierbij gaan we uit van het principe dat we zoveel mogelijk annoteren wat er staat, en ons dus niet overgeven aan de vraag wat de schrijver mogelijk had willen schrijven.

Zinnen kunnen op vele manieren ongrammaticaal zijn; we bespreken hier een aantal veelvuldig voorkomende gevallen, met hun annotaties:

6.3.1 Missende woorden

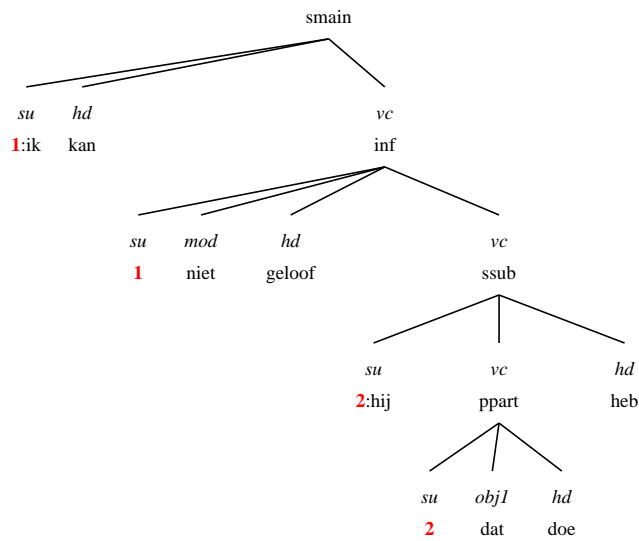
Missende woorden. Indien dit woord (bij aanwezigheid) tot een ongelede dependentiestructuur zou geleid hebben, laten we het bij afwezigheid natuurlijk gewoon weg:

(639) Ik vind stom dat hij komt

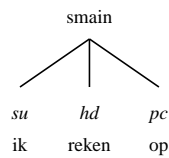


Indien het missende woord hoofd zou zijn van een gelede structuur, wordt de keus lastiger. Ook in zulke gevallen proberen we zoveel mogelijk te annoteren wat er staat; zoals gebruikelijk vermijden we indien mogelijk niet vertakkende dependentiedomeinen, en hoofdloze dependentiedomeinen.

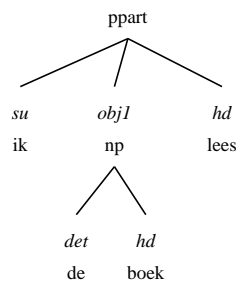
(640) Ik kon niet geloven hij dat gedaan had



(641) Ik reken op



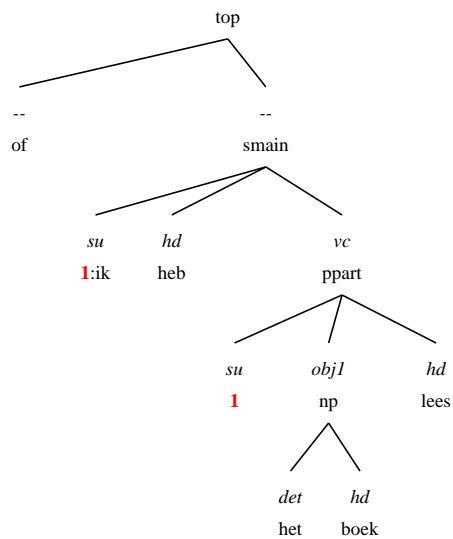
(642) Ik de boeken gelezen



6.3.2 Ten onrechte opgenomen woorden

Een woord dat ten onrechte in de zin lijkt te zijn opgenomen, wordt niet geannoteerd en (dus) direct aan de topknoop gehangen met de speciale - - relatie.

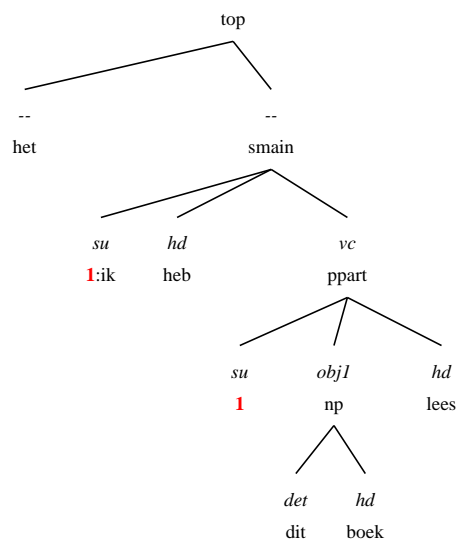
(643) Ik heb of het boek gelezen



- (644) En toch bleef zijn denkwerk op hetzelfde niveau *is* gehandhaafd
- (645) Tussen 1917 en 1930 is er uitgebreid over deze restauratie *en* gediscussieerd
- (646) Hoe noemt men de *van* fijne , droge schilfertjes tussen de hoofdharen ?

We nemen hierbij de gewoonte over uit de CGN praktijk om in geval er twee op elkaar volgende woorden optreden waarvan onduidelijk is, welk van beide moet worden genegeerd, we de eerste negeren:

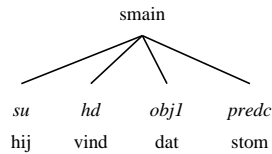
- (647) Ik heb het dit boek gelezen



6.3.3 Foutgespelde woorden

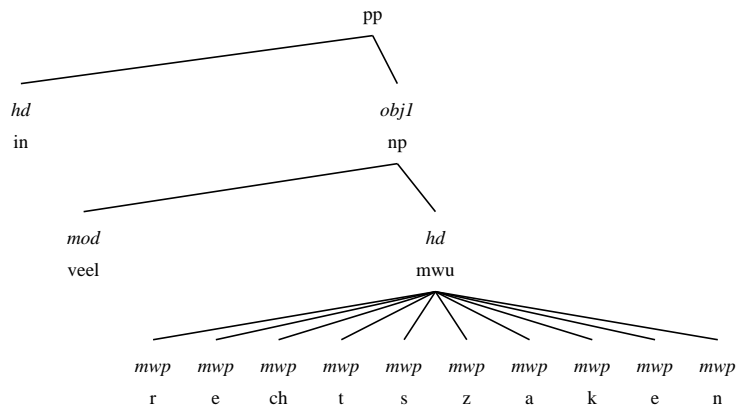
In geval van foutgespelde woorden ten slotte kiezen we voor een syntactische analyse alsof het woord juist was gespeld. We stellen ons dus niet prescriptief op:

(648) Hij vind dat stom

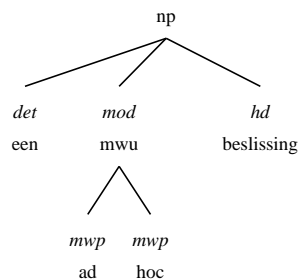


Ditzelfde geldt ook voor gevallen waarbij extra spaties worden gebruikt; indien te weinig spaties worden gebruikt moet vaak een creatieve oplossing worden gekozen.

(649) Ze zijn verwickeld in vele r e c h t s z a k e n



(650) een ad hoc beslissing



De overige gevallen van onverwachte invoer zijn te divers om in duidelijke regels te vatten.

Hoofdstuk 7

Automatische annotatie

7.1 Overzicht additionele attributen

Ook de automatisch door Alpino geannoteerde zinnen in Lassy Groot zien er zoveel mogelijk uit volgens de in dit document gespecificeerde richtlijnen. Omdat automatische ontleding niet altijd honderd procent correct kan verlopen, moet rekening gehouden worden met de mogelijkheid dat er ook fouten in dit deel van het corpus optreden. De hoop is, dat Lassy Groot (een grote treebank, maar wel met fouten) voor sommige toepassingen bruikbaar is dan Lassy Klein (een handmatig geverifieerde treebank), omdat Lassy Groot vele malen groter is.

Een ander verschil tussen Lassy Groot en Lassy Klein is, dat in Lassy Groot een aantal extra attributen in de XML code beschikbaar is. Deze attributen worden gebruikt om extra informatie waarover Alpino beschikte (bijvoorbeeld informatie uit het Alpino woordenboek) aan de gebruikers van Lassy Groot beschikbaar te stellen.

De volgende tabel noemt de veelgebruikte extra attributen:

attribuut	OMSCHRIJVING
case	naamval van nomina (nom,acc,dat,..)
comparative	type complement van comparatives (dan,als,dat,..)
def	definitieiteit van nomina (def,indef,..)
frame	volledige Alpino POS-tag
gen	gender van nomina (de,het,..)
infl	inflectie van adjectieven, determiners
lcat	zoals het attribuut cat, maar voor lexicale knopen
neclass	classificatie van namen (org, loc, per, misc)
num	getalinformatie van nomina (sg,pl,..)
per	persoonsinformatie van nomina (fir,thi,..)
pos	vereenvoudige POS-tag
refl	reflexiviteit van pronomina (refl)
root	stam van betreffende woord
sense	van root en sc afgeleide betekenisrepresentatie
sc	subcategorizatie frame
special	vergaarbak voor andere speciale features
tense	geeft aan dat persoonsvorm verleden tijd is
wh	geeft vraagwoordinformatie voor pronomina (ywh,nwh,rwh)

7.2 Details

De verschillende attributen bespreken we nu heel in het kort als volgt. Het voert te ver om hier een uitputtende documentatie van de features te presenteren, omdat dit gedetailleerde kennis van de Alpino parser bij de lezers vooronderstelt.

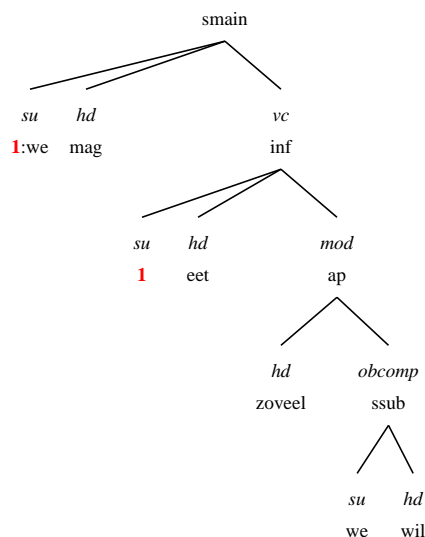
case Dit attribuut geeft de naamvalsinformatie van persoonlijke voornaamwoorden weer. Waardes die gebruikt worden zijn `nom` (nominatief), `acc` (accusatief), `gen` (genitief) en `dat` (datief). Merk op, dat de naamvalsinformatie alleen wordt gegeven indien deze lexicaal wordt toegekend. Dus wel voor pronomina, maar niet voor de meeste zelfstandig naamwoorden. De waardes `nom_acc`, `dat_acc` en `both` worden gebruikt voor woorden die op meerdere manieren gebruikt kunnen worden.

Relatieve voornaamwoorden gebruiken het attribuut `case` om aan te geven of ze normaliter wel of niet met een voorzetsel kunnen combineren. De verschillende waardes zijn `obl`, `no_obl` en `both`.

comparative Dit attribuut wordt voor comparatieven gebruikt, en geeft aan wat voor type OBCOMP complement geselecteerd moet worden. De waardes

bevatten dan, als, dat, om, e_als, en de waardes geven aan welk woord geacht wordt als hoofd van de OBCOMP te fungeren. De waarde e_als wordt gebruikt voor zoveel in zinnen van het type:

(651) We mogen eten zoveel we willen



def Dit attribuut geeft voor lidwoorden de definitetheid aan. Waardes zijn def, indef, both.

frame Dit attribuut geeft de volledige lexicale categorie die Alpino gebruikt.

gen Dit attribuut geeft de gender informatie, dat wil zeggen het onderscheid tussen neuter en nonneuter voor nomina. Gebruikte waardes zijn de, het, both.

infl Dit attribuut geeft de inflectieinformatie van werkwoorden en bijvoeglijke naamwoorden. Voor werkwoorden zijn de gebruikte waardes als volgt: sg, sg1, sg3, sg_heeft, sg_hebt, modal_u, modal_inv, pl, imp(sg1), imp(modal_u) subjunctive, psp, inf, inf(e), inf(no_e), inf_ipp.

Voor adjectieven gebruiken we de waardes e, no_e, er, both, pred, anders, ge_e, ge_no_, ge_both, stof, postn_no_e, postn_e, postn_both, meer, ere, ste, st, het_st, end, ende, sten, eren

Voor determiners gebruiken we de waards de, het, een, pron, geen, pl_num, alle, enkele, wat, welke, elke, der, pl_indef, al, enkel, zulke, welk, den, des, sg, sg_num, pl, both

Voor getallen worden de waardes pl_num, sg_num, both gebruikt

lcat De waarden die het attribuut `lcat` kan aannemen zijn in principe hetzelfde als de categorielabels. De waarde van `lcat` voor een lexicale knoop kan begrepen worden als de waarde van de categorielabel van dezelfde knoop indien het woord complementen zou hebben.

In de treebank is het enigszins omslachtig om te zoeken naar NPs. Een NP kan een knoop met categorie NP zijn. Indien de NP echter lexicaal is, wordt geen extra niveau geïntroduceerd, en dus is er ook geen categorie NP. Een query naar categorie NP zou dus niet matchen op de dependentiestructuur van de volgende zin:

(652) Ik slaap

Met het feature `lcat` is het mogelijk ook zulke lexicale NPs te identificeren. De gebruikte waarden van `lcat` komen overeen met de waarden voor het attribuut `cat`.

neclass Namen hebben het attribuut `neclass` om aan te geven wat voor type naam het betreft. De waarden zijn `LOC`, `PER`, `ORG`, `MISC`, `year` voor locaties, persoonsnamen, organisaties, namen van een ander type, en jaartallen.

num Dit attribuut onderscheidt enkelvoudige en meervoudige nomina. De waarden zijn `sg`, `pl`, `both`, `meas`, `bare_meas`, `pl_num`, `sg_num`.

per Persoonsinformatie voor nomina. De waarden zijn `u`, `both`, `fir`, `inv`, `je`, `thi`, `u_thi`.

pos Pos is de vereenvoudigde POS-tag die binnen Alpino traditioneel wordt gebruikt. De gebruikte waarden zijn `noun`, `name`, `punct`, `det`, `prep`, `verb`, `adj`, `adv`, `vg`, `num`, `pron`, `comp`, `part`, `pp`, `fixed`, `tag`, `comparative`, `prefix`, `etc`, `max`. De speciale waarde `--` wordt gebruikt voor niet geanalyseerde zinnen. De speciale waarde `UNKNOWN` indiceert ook een probleem van de parser met de betreffende zin.

root Root geeft de stam van het betreffende woord.

sc Dit feature geeft het subcategorisatie-frame voor, met name, werkwoorden en adjectieven.

sense Dit attribuut combineert informatie uit het root attribuut en de subcategorisatie informatie. De resulterende waarde van `sense` geeft aan in welke lezing het betreffende woord is gebruikt. De `sense` waarde incorporeert de root waarde van het attribuut en alle volgens het subcategorisatie-frame

verplichte lexicale materiaal waarmee dit gebruik van het woord combineert. Voorbeelden: *besta-uit*, *behoor-tot*, *zich-bevind*, *trouw-met*, *win-van*, *neem_deel-aan*, *vergelijkbaar-met*, *krijg-met-maak*, *op-de-markt-breng*, *in-de-gaten-houd*, *in-verband-breng-met*.

special vergaarbak voor andere speciale features.

tense de enige waarde is *past* om de verleden tijd van de persoonsvorm aan te geven.

wh geeft vraagwoordinformatie voor pronomina. De gebruikte waardes zijn *ywh*, *nwh*, *rwh*, *wh*, *rel*.

Het is wellicht goed om duidelijk te maken dat de extra XML attributen de *lexicale* kennis representeert. Bijvoorbeeld:

(653) De Clintons komen niet

In dit voorbeeld is *Clintons* volgens het woordenboek een noun waarvan de getalinformatie *num=both* is, dat wil zeggen dat het woordenboek zowel enkelvoud als meervoud toestaat. Uit de context is het uiteindelijk voor de parser duidelijk dat de NP als geheel meervoudig moet zijn. Het werkwoord bevat het attribuut *infl=pl*. Toch bevat de XML alleen *num=both*, dus de a-priori lexicale informatie, en niet de door de parser afgeleide informatie.

Bijlage A

Appendix

A.1 Gebruikte categorieën

categorielabel	OMSCHRIJVING
AP	bijvoeglijk-naamwoordgroep
ADVP	woordgroep met bijwoord als hoofd
AHI	<i>aan het</i> -infinitief-groep
CONJ	nevenschikking
CP	frase ingeleid door onderschikkend voegwoord
DETP	woordgroep met lidwoord als hoofd
DU	discourse unit
INF	kale infinitiefgroep
NP	naamwoordelijke constituent
OTI	<i>om te</i> -infinitief-groep
PPART	voltooid/passief-deelwoordgroep
PP	voorzetselconstituent
PPRES	tegenwoordig-deelwoordgroep
REL	relatieve zin
SMAIN	declaratieve zin (werkwoord op tweede plaats)
SSUB	bijzin (V finaal)
SVAN	<i>van</i> -zin
SV1	werkwoordsinitiële zin (ja/nee-vraag, bevel)
TI	<i>te</i> -infinitief-groep
WHREL	relatieve zin met ingesloten antecedent
WHSUB	constituentvraag: ondergeschikte zin
WHQ	constituentvraag: hoofdzin

A.2 Gebruikte dependentierelaties

dependentielabel	OMSCHRIJVING
APP	appositie, bijstelling
BODY	romp (bij complementizer))
CMP	complementizer
CNJ	lid van nevenschikking
CRD	nevenschikker (als hoofd van conjunctie)
DET	determinator
DLINK	discourse-link
DP	discourse-part
HD	hoofd
HDF	afsluitend element van circumpositie
LD	locatief of directioneel complement
ME	maat (duur, gewicht, ...) complement
MOD	bijwoordelijke bepaling
MWP	deel van een multi-word-unit
NUCL	kernzin
OBCOMP	vergelijkingscomplement
OBJ1	direct object, lijdend voorwerp
OBJ2	secundair object (meewerkend, belanghebbend, ondervindend)
PC	voorzetselvoorwerp
POBJ1	voorlopig direct object
PREDC	predicatief complement
PREDM	bepaling van gesteldheid 'tijdens de handeling'
RHD	hoofd van een relatieve zin
SAT	satelliet; aan- of uitloop
SE	verplicht reflexief object
SU	subject, onderwerp
SUP	voorlopig subject
SVP	scheidbaar deel van werkwoord
TAG	aanhangsel, tussenvoegsel
VC	verbaal complement
WHD	hoofd van een vraagzin

References

- [Bouma2008] Bouma, Gerlof. 2008. *Starting a Sentence in Dutch*. Ph.D. thesis, University of Groningen. 27, 28
- [Bouma2004] Bouma, Gosse. 2004. Treebank evidence for the analysis of PP-fronting. In S. Kubler, J. Nivre, E. Hinrichs, and H. Wunsch, editors, *Third Workshop on Treebanks and Linguistic Theories*, pages 15–26, Seminar für Sprachwissenschaft, Tübingen. 139
- [Haeseryn1997] Haeseryn, Walter, et al. (eds.). 1997. *Algemene Nederlandse Spraakkunst*. Martinus Nijhoff and Wolters Plantijn. 2e, geheel herz. dr. 39, 40, 44, 71, 85, 109, 114, 115, 116, 117, 127, 128, 131, 133, 159, 180, 187
- [Hoekstra et al.2004] Hoekstra, Heleen, Michael Moortgat, Machteld Schoupe, Ineke Schuurman, and Ton van der Wouden, 2004. *CGN Syntactische Annotatie*. 7
- [Kloosterman2007] Kloosterman, Geert, 2007. *An overview of the Alpino Treebank Tools*. 8
- [Lai and Bird2004] Lai, Catherine and Steven Bird. 2004. Querying and updating treebanks: A critical survey and requirements analysis. In *In Proceedings of the Australasian Language Technology Workshop*, pages 139–146. 25
- [Lai and Bird2010] Lai, Catherine and Steven Bird. 2010. Querying linguistic trees. *Journal Logic Language Information*, 19:53–73. 25
- [van Noord2006] van Noord, Gertjan. 2006. **At Last Parsing Is Now Operational**. In *TALN 2006 Verbum Ex Machina, Actes De La 13e Conference sur Le Traitement Automatique des Langues naturelles*, pages 20–42, Leuven. 8
- [Pajas and Štěpánek2008] Pajas, Petr and Jan Štěpánek. 2008. Recent advances in a feature-rich framework for treebank annotation. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 673–680, Manchester, UK, August. Coling 2008 Organizing Committee. 9, 18
- [Van Eynde2005] Van Eynde, Frank, 2005. *Part Of Speech Tagging En Lemmatisering Van Het D-Coi Corpus*. 7, 9, 11, 12, 15, 16, 152