PART OF SPEECH TAGGING AND LEMMATIZING OF THE *CORPUS GESPROKEN NEDERLANDS* (SPOKEN DUTCH CORPUS)

Frank Van Eynde

Febuary 2004

Centrum voor Computerlinguïstiek K.U.Leuven

Preface

In addition to digitalized speech, the Corpus Gesproken Nederlands contains various layers of transcription and annotation. This document deals with two of these annotation levels, namely lemmatizing and part of speech tagging.

The first chapter introduces the most important characteristic of these two annotation levels and the notation used. The second chapter presents a complete and detailed overview of the CGN tagset, including the lemmatization guidelines. The third chapter puts the CGN tagset in the broader context of the EAGLES recommendations. The fourth chapter gives a formal summary of the tagset, giving examples[1].

I would like to thank Hans van Halteren, Walter Daelemans, Jakub Zavrel, Ineke Schuurman, Lisanne Teunissen, Richard Piepenbrock, Nelleke Oostdijk, Ard Sprenger, Gosse Bouma, Antal van den Bosch, Ton van der Wouden, Truus Kruyt and Peter-Arno Coppen for their verbal and written comments on the earlier versions of this text.

Frank Van Eynde

Leuven, November 2003

[1] This Protocol was originally written in Dutch and translated into English. The tagset that is described in the protocol was developed for the Spoken Dutch Corpus (Corpus Gesproken Nederlands, CGN). The attributes and values of the tags are in Dutch. An English translation of these terms and tags is given in a list at the end of this document (p. 60). In running text the English terms are used as much as possible. However, where there is a direct reference to an attribute or a value the Dutch term is used. Also see chapter 3 (Comparison with EAGLES, p. 44) for a description of the tags. All examples are in Dutch and have not been translated into English.

1 INTRODUCTION

The first step in the linguistic expansion of the corpus covers the allocation of tags and lemmas to the items in the orthographic transcription, i.e. the words, punctuation marks and special symbols. In this, we endeavour to meet the following requirements:

- The tags must include the information which is traditionally associated with what is commonly understood as 'parsing'; and should therefore correspond with the classification used in grammars designed for general use, such as the *Algemene Nederlandse Spraakkunst* (ANS 1997). This is covered in detail in the section on the tagset, see chapter 2.

- The tagset should tie in with current international standards as much as possible. In practice this means keeping in line with the *EAGLES* standard in particular, see chapter 3.

- Only one tag is allocated to each token. This implies that a maximum of one tag can be allocated and that disambiguation is required. This also implies that at least one tag must be allocated, and that arrangements need to be made for less frequent tokens, such as interrupted and incomprehensible 'words' and dialect or foreign words.

- Only one lemma is allocated to each token. So here, too, we must disambiguate where necessary.

- The tags must provide a suitable basis for the higher linguistic annotation levels, such as syntactic analysis and the identification of multiword combinations (lexicological linking).

- The notation must be clear, compact and easy to read.

Bearing in mind the size of the corpus, we have chosen to automate both tagging and lemmatizing as much as possible. Based on an evaluation carried out in 1999, (Zavrel en Daelemans 1999) and bearing in mind limitations in availability, we have chosen to use MBLEM (Memory-based Lemmatizer, Tilburg) for lemmatizing and the TIMBL combi-tagger for tagging. Characteristic for the combi-tagger is, that, through systematic comparison of the results of a number of taggers working independently, it obtains a result which is more accurate than that of the individual taggers.

In the case of the TIMBL combi-tagger, this is a combination of the results from TnT (Trigram 'n Tags, Saarbrücken), MBT (Memory-based Tagging, Tilburg), a maximum entropy tagger and a Brill tagger. In order to guarantee a higher quality, the results of automatic tagging and lemmatizing are checked manually and corrected where necessary. For the Dutch part of the corpus this was carried out in Nijmegen and for the Flemish part, in Leuven.

1.1 LEMMATIZATION

Lemmatizing involves converting inflected forms back to their base forms; in the case of verbs these are the infinitives. In the case of other word types, this is the stem. If the stem is not an existing word, the base form is identified through an inflected form; this is the case, for example, with intrinsically plural nouns (*hersenen, mazelen*) and with nouns which only occur in the diminutive form (*akkefietje, ootje*). Exactly how lemmatizing is carried out is explained for each separate word type in chapter 2.

Lemmatizing is carried out word for word. The parts of a separable compound verb, as in *ze geeft prachtig verluchte boeken uit,* are therefore each allocated their own lemma. The conversion of *geven* and *uit* to *uitgeven* is carried out at the level of lexicological linking. That is also the level where compound proper nouns such as Den Haag are recognized as a single entity.

1.2 THE TAGSET

Tagging involves allocating lexical and morphosyntactic characteristics to word forms in a context. The relevant characteristics include at least **word type**. The other characteristics are laid down in the tagset.

1.2.1 Tagsets for Dutch

Various tagsets have already been developed for tagging Dutch; an overview can be found in Schuurman (1998) and the appendix of Zavrel and Daelemans (1999).

The majority of these tagsets are characterized by a limited **granularity**, in the sense that they differentiate less than 50 different tags. The INL tagset for example, which is included in the CORRie tagger and others, limits itself to differentiating word types and does therefore not include any additional morphosyntactic characteristics. The KEPER (24 tags), D-Tale (45 tags) and Xerox (49 tags) tagsets make relatively little differentiation. Tagsets with such limited differentiation can suffice for some applications, but for the annotation of the CGN corpus we have chosen a higher degree of granularity, because, for a large part of the corpus, the tags are the only form of linguistic parsing, (syntactic analysis is only done for 10 % of the corpus). Tagsets which do make a large number of differentiations, and in that respect are more attractive for CGN, are those of WOTAN-2 and PAROLE. Besides a higher granularity, an important criterion is the extent to which it is in line with existing **standards**. This is not only important for recognition in an international context, but also for the accessibility of the annotations for the end users. In order to guarantee the first criterion, we have looked for compatibility with the EAGLES recommendations; to guarantee the second criterion we have looked for compatibility with the ANS-97. The latter has two advantages: (1) for additional information we can refer to the relevant sections in the ANS, and (2) as the ANS has been written for a broad public and makes use of familiar concepts and distinctions as much as possible, there is a greater likelihood that the CGN tags can be interpreted by non-linguists. A third selection criterion relates to the accompanying **documentation**. A tagset which restricts itself to listing word types and distinctions used is ill-suited for CGN, because it offers no guarantee that the information will be interpreted unequivocally. In order to be useful, the tagset must also contain instructions and guidelines which enable tagging and the interpretation of the different tags to be carried out uniformly. Furthermore, the need for such documentation increases as the granularity of the tagset increases.

In view of these criteria, the tagsets of WOTAN-2 and PAROLE would seem to be the most suitable options for the CGN. The problem was, however, that at the start of the CGN project, these were still in the process of being developed. The WOTAN-2 tagset existed only in a preliminary version, which was modified several times during the early stages of the CGN project, and there was no documentation available for the PAROLE tagset. Consequently, during the course of 1998-1999 a new tagset was designed for use in the CGN project. The most important source of inspiration for this were EAGLES, ANS-97 and the provisional versions of WOTAN-2. In view of the similarities between the languages and, more particularly the shared premises, the Stuttgart-Tübingen Tagset for German (STTS) also turned out to be an interesting source. To test the practical use of the tagset, a manual annotation experiment was carried out in the Spring of 1999. The results of this experiment are explained in detail in Zavrel (1999) and have also played a role in defining the CGN tagset.

1.2.2 Setting up the CGN tagset

In the CGN tagset word types are given two sets of morphosyntactic characteristics. The first group consists of **lexical characteristics**, such as the split into coordinate en subordinate conjunctions, or the difference between definite and indefinite articles. The second group consists of characteristics which define morphological variation, such as number in the case of nouns or comparative degrees in the case of adjectives. The **morphological characteristics** include at least those which are not reflected in the lemma after lemmatization. For example, the noun *tafel*s is linked to the lemma *tafel* and consequently the information regarding number must be included as a separate feature. The morphological characteristics included in the CGN tagset are those which encode inflectional variation (number, verb tense, case, etc.), with a number of characteristics added which encode derivations containing word type, such as diminutives in the case of nouns. Whether a characteristic is lexical or morphological sometimes depends on word type. Number, for example, is a morphological characteristic in the case of nouns, but a lexical characteristic in the case of pronouns. Exactly which characteristics are included in the tags is explained for each word type separately in chapter 2. Semantic characteristics are not included. For example, no difference is made between concrete and abstract nouns. Tagging is carried out in the same way as lemmatization: word for word. A fixed expression, such as *te goeder trouw*, is therefore treated as a sequence of preposition, adjective and noun. This method has consequences for setting up the tagset. For example, the nouns and adjectives must allow for the possibility that they may be in the dative form, as such forms occur relatively often in fixed expressions, such as *ter plaatse, van harte, in feite*. A further consequence is the need for a special tag for parts of proper nouns. In *Den Haag*, for example, both words are given the tag *SPEC (deeleigen)*, see 2.12. The identification of the proper noun as a single entity takes place at the level of lexicological linking. For the same reason valency patterns are not included in the tagset. After all, a separable compound verb does not necessarily have the same valency as the verb at its centre: for example, *uitgeven* is transitive, whereas *geven* is ditransitive, and *uitlachen* is transitive, whereas *lachen* is intransitive. The allocation of valency patterns is only worthwhile after lexicological linking. During the allocation of tags we have systematically chosen a morphosyntactic perspective. The numerical values, for example, are interpreted not in conceptual but in morphosyntactic terms; nouns such as *boel* and *aantal* are therefore singular. The choice of the morphosyntactic perspective is also important in the interpretation of word types. A word such as *maandag* for example, is often used as an adverbial, as in *ik heb hem maandag nog gesproken*, but is a noun, as far as word type is concerned and is therefore treated

not as an adverb, but as a noun when tagging.

Dialect words, i.e. those given the symbol '*d' during the orthographical transcription are not given complete tags, but only word type and the TYPE features relevant for that word type, such as NTYPE for nouns and CONJTYPE for conjunctions. The reason for omitting the other features is that they are mainly used to indicate morphological information and it is precisely the morphology in dialects which can significantly vary from standard language. The pronouns in combinations such as *dieën boek* and *dieë vent* for example cannot be described in terms of the characteristics used for standard language. For the same reason no reduction to a base form is carried out when lemmatizing dialect words, but the lemma is set the same as the word form itself.

1.2.3 Ambiguity and underspecification

Ambiguity is a relative concept: a word form is not ambiguous in itself, but only in the context of a differentiation system. The form *snel* for example is *POS-ambigu* if we allocate different word types to the attributive use in *een snel paard* and the adverbial in *dat paard loopt snel*. On the other hand, it is not *POS-ambigu* if we regard these as different uses of the same adjective. The number and types of ambiguities the tagger is confronted with depends on the choices made when setting up the tagset.

When setting up the tagset a difference is made between occasional and systematic ambiguities. The first concerns individual words: the word *bij* for example can be a noun or a preposition. The second applies to groups of words and mainly apply to different possible uses. The fact that the adjective *snel* can be both attributive and adverbial is something many other adjectives also have in common. The tagset is defined in such a way that occasional ambiguities are recognized and solved, while we avoid postulating systematic ambiguities as much as possible. The CGN tagset allows a certain degree of **underspecification**. This can be explained by using the feature *NAAMVAL*.

NAAMVAL = nominatief, oblique, genitief, datief

Examples of these values are:
ik, jij, hij, wij for the *nominatief*, *mij, jou, hem, ons* for *oblique* ,'s *avonds, wiens, elkaars* for the *genitief* and ter *plaatse, in koelen bloede, in der minne* for the *datief*. As we know, the difference between nominative and oblique is only relevant for pronouns: for nouns and adjectives it is systematically neutralized. In order to avoid such cases resulting in systematic ambiguity, we apply an intermediate value ('*standaard'*) which generalizes between nominative and oblique. We do something similar for the genitive and the dative: as in adjectives the genitive and the dative forms are systematically the same form, we have introduced an intermediate generalized value ('*bijzonder*). The partition then looks as follows:

NAAMVAL = standaard (nominatief, oblique), bijzonder (genitief, datief).

In this hierarchy '*standaard'* is used for the forms without a case ending and '*bijzonder'* for forms with one. The level of differentiation is determined per word type. In some cases it is also useful to have a value which generalizes over all the specific values. Nouns for example have a *GENUS* feature with the values '*zijdig'* and '*onzijdig'*. In most cases this differentiation can easily be made, but there is a small number of nouns which can take both gender values (*de/het riool,*

de/het filter); when the context does not permit the allocation of a specific *GENUS* to such nouns, we allow the use of the value '*genus'*.

GENUS = genus (zijdig, onzijdig).

Although using underspecification has its advantages, we have used this option sparingly in setting up the tagset. There are two reasons for this: firstly, with excessive use, tags lose their informative value; secondly, if we allow atomic and intermediate values in tags for a single feature, the number of possible tags increases drastically.

1.2.4 The definition of the CGN tagset

Formally speaking, the CGN tagset is a sextuple <*A, W, P, D, I, T*>, where A is a collection of attributes, W of values, P of partitions, D of declarations, I of implications and T of tags. Features are those combinations of attributes and values which meet the restrictions laid down by the partitions. Tags are lists of features which meet the restrictions laid down by the declarations and the implications.

Attributes and values. A feature consists of an attribute and a value. For the former we use capital letters and for the latter we use lower case letters or numbers; the two are separated by an equals sign

GENUS = onzijdig

The values are atomic, i.e. they do not in turn consist of attribute value pairs.

For both the attributes and the values Dutch terms are used. Attributes which are only relevant for a specific word type are given a prefix; *LWTYPE*, for example, is only given to articles and *NTYPE* only to nouns.

Partitions. The possible values are determined for each attribute. This is done in the form of a partition. This can be a simple list but also a hierarchy with intermediate values.

[P03] NTYPE = soortnaam, eigennaam.

[P07] NAAMVAL = standaard (nominatief, oblique), bijzonder
(genitief, datief).

For ease of reference the partitions are numbered. If a feature is relevant for different word types, it is given the same number for the different word types.

Declarations. A tag is a list of lexical and morphosyntactic characteristics. Which characteristics are relevant, and for which word types, is determined in declarations such as

[D08] <POS = werkwoord> $\Rightarrow$ <WVORM>

Characteristics which are only relevant to parts of a category are not associated with the word type as a whole, but with more specific characteristics. In the case of verbs, for example, person, time and mood are relevant for the finite forms but not for the infinitives and the participle.

[D09] <WVORM = persoonsvorm> ⇒ <PVTIJD, PVAGR>

Similarly to partitions, declarations are numbered. If a value is partitioned into one or more subvalues, the subvalues inherit the morphosyntactic characteristics associated with the higher value. If the pronouns have the feature *NAAMVAL*, for example, the determiners also have that feature. More complex forms of inheritance, such as multiple inheritance and default inheritance, are not used.

Tags. Tags are ordered lists of features. They contain a POS feature and all morphosyntactic features which have been declared for the word type in question.

<POS = voegwoord, CONJTYPE = nevenschikkend>

The CGN tags are internally structured and differ in that respect from the monolithic tags used in, for example, the Brown corpus. In addition to the fully written notation, we also use a more compact format, in which the names of the attributes are omitted and the names of the values are abbreviated. This is shown as follows:

[T801] VG (neven)

As with declarations and partitions, tags are also numbered. The first number refers to word type (T8). Tags with an underspecified value, such as *'genus'*, are given a U number instead of a T number.

[U117] N (soort, ev, basis, genus, stan)

Tags intended only for dialect words, are given an R number (for *regional*).

[R5xx] VNW (aanw, det)

Implications. Within a tag there are sometimes dependencies between the values of the features. Diminutive nouns, for example, are always neutral. This can be expressed in terms of an implication such as

<POS = substantief, GRAAD = diminutief> ⇒ <GENUS = onzijdig>

2 The CGN TAGSET

Orthographically transcribed speech is segmented into tokens. These are words, punctuation marks or special symbols.

[P01] TOKENTYPE = woord, speciaal, leesteken.

01. TOKENTYPE. Tokens of the type '*woord'* are the word forms as they occur in the word types; these forms can be the same as the stem but will – in many cases – be inflected. A *POS* (Part of Speech) feature is associated with the words.

[D00] <TOKENTYPE = woord> ⇒ <POS>
[P02] POS = substantief, adjectief, werkwoord, telwoord, voornaamwoord, lidwoord, voorzetsel, voegwoord, bijwoord, tussenwerpsel.

02. POS. We differentiate between four open classes *(noun, adjective, verb, numeral)* and six (more or less) closed classes *(pronoun, article, preposition, conjunction, adverb, interjection).*This split into ten word types is identical to that of ANS-97. There are some differences regarding the classification of specific words: *er* for example is not treated as an adverb but as a pronoun and *veel* and *weinig* are not treated as numerals, but as indefinite pronouns. Such differences are mentioned and justified with each of the different word types. In the following sections we explain how each of the ten word types is differentiated from the other word types (Demarcation), which features are associated with them (Declarations), and which values the features can have (Partitions). All the features are explained with examples, and criteria are given for determining the relevant values. If there are any dependencies between the values of the features (Implications), these are also given for each word type, as are the possible tags. With the latter, a differentiation is made between the most specific combination and the underspecified combinations; the former are given a T number, the latter a U number. Finally, lemmatization instructions are given in each case.

For dialect words a separate tagset has been introduced with a lower granularity (see 2.11). Special tokens and punctuation marks are covered separately in sections 2.12 and 2.13.

2.1 NOUNS

2.1.1 Demarcation

When identifying nouns special care should be taken with the difference between nominally used adjectives (see 2.2), participles and infinitives (see 2.3), numerals (see 2.4) and special tokens (see 2.12). Criteria for differentiating them are given in the sections for the other word types.

2.1.2 Declarations

Nouns are marked according to their type (type name or proper noun), number (singular or plural), and degree (diminutive or not). Case values

(standard, genitive or dative) are only allocated to single nouns, and gender values (masculine/feminine or neutral) to singular nouns without a case suffix.

[D01] <POS = substantief> ⇒ <NTYPE, GETAL, GRAAD>

[D02] <POS = substantief, GETAL = singular> ⇒ <NAAMVAL>

[D03] <POS = substantief, GETAL = singular, NAAMVAL = standaard> ⇒ <GENUS>

NTYPE and *GENUS* are lexical characteristics; *GRAAD*, *GETAL* and *NAAMVAL* are morphological characteristics. The diminutive suffix always precedes the plural suffixes (*hond-je-s*) and the genitive *(Jan-tje-s hond*).

2.1.3 Partitions

[P03] NTYPE = soortnaam, eigennaam.

[P04] GETAL = enkelvoud, meervoud.

[P05] GRAAD = basis, diminutief.

[P06] GENUS = genus (zijdig, onzijdig).

[P07] NAAMVAL = standaard, genitief, datief.

03. N-TYPE. Nouns written without a capital letter in the orthographic transcription are classed as type names, except where they would be treated as proper nouns in lexicographic practice in general; this applies amongst others to the names of the months (*april*) and the days of the week (*zondag*).

Nouns written with a capital letter are classed as proper nouns, except where there is a case of (1) abbreviations of type names, such as *CD, LP, WC, TV, PC*; (2) compounds of which the core is a type name and which function as a type name (*het Ardennen- offensief, een typisch Randstad-probleem, NAVO-bombardementen*); words such as *Noordzee, Waasland* and *Bondgenotenlaan*, on the other hand, are classed as proper nouns, not least because the addition of a plural ending or indefinite article is marked, compare *een Randstadprobleem* met *? een Noordzee*; (3) nouns which are part of a title, as in *De Naam Van De Roos* and *Man Bijt Hond*; nouns in titles are only tagged as proper nouns, if they are also used as such outside the context of the title; in *Het Verdriet Van België* the first noun is a type name and the second a proper noun. Please note that these are titles (of books, TV and radio programmes, records, films, etc.) and not names of people, places, newspapers, etc. In the person's name *Roos De Hond* and the place name *Den Haag*, the nouns are not given the tag for type names, but *SPEC(deeleigen*). This is a tag given to parts of a proper noun which consists of different words. In combinations such as *Den Haag*, *Freek De Jonge*, *Brusselse Steenweg* and *De Standaard*, the individual words are therefore given the tag *SPEC (deeleigen*), see 2.12. The identification of the proper noun as a single unit occurs at the level of lexicological linking. NB: in combinations such as *het Van Mierlo-effect* the last word is treated as a type name (cf. *Randstad-probleem*) and only *Van* is given the tag *SPEC (deeleigen*).

As this split into type names and proper nouns is lexical orthographic in nature, it is not dependent on specific contexts. A noun such as *Stella* is therefore always tagged as a proper noun, even in contexts where it is not used for an individual,

as in *waar mijn Stella staat* and *drie Stella's a.u.b.*.

Finally, note that only nouns can be given the tag *N (eigen, . . .)*. The adjectives in *de Belgische regering* and *een Italiaanse wijn* are therefore not proper nouns. The same applies to the interjection *AUB* and the foreign *SVP* and *CQ*.

04. GETAL. The plural is marked with a suffix *(-s, -en, ...)* or— as an exception—by suppletion (*zeelui, timmerlieden*). Nouns without a plural suffix are given the value 'singular', even when according to their meaning, they are plural. Type names with a collective meaning such as *boel, aantal, hoop* are therefore given the value '*enkelvoud*'; this also applies to nouns of measure in combinations like *vijf jaar* and *drie liter*. Nouns with a plural suffix are given the value '*meervoud*'. This also applies to pluralia tantum as in *hersenen* and *ingewanden*. With proper nouns, the presence of a plural suffix is not always an indication of a plural. A form such as *Enkhuizen* for example is given the value '*enkelvoud*' because it shows the typical characteristics of a singular noun: in this way it can also be combined with *het*, as in *het/*de Enkhuizen van weleer*, and it demands a singular person form when it serves as a subject, as in *Enkhuizen ligt/*liggen ergens in Nederland.* A form such as *Ardennen* on the other hand is plural, as is apparent in *de/*het Ardennen liggen/*ligt in België.*

05. GRAAD. Diminutive forms are marked with a suffix *(-je, -tje, -pje, -ke, ...)*. Nouns without this suffix are given the value '*basis'*; this value is given to nouns which cannot have a diminutive form, for example *gebergte, vee*. Nouns with a diminutive suffix are given the value '*diminutief*'; this value is also given to nouns which only have a diminutive form such as *ootje, meisje, nippertje*. Typical of diminutive nouns is that they are neutral: in the singular they combine with *het/dat/dit/ons* and not with *de/die/deze/onze*. With proper nouns the presence of a diminutive suffix is not always an indication of a diminutive; a name such as *Nelleke,* for example, combines with typically masculine/feminine determiners *(die/? dat Nelleke toch*) and is therefore not a diminutive.

06. GENUS. In POS tagging we only differentiate between gendered and neutral nouns; the additional differentiation of gendered nouns into masculine and feminine is not made. Gendered nouns are those which in the singular take determiners such as de*/die/deze/onze*, while neutral nouns take determiners such as *het/dat/dit/ons*. As with *GETAL, GENUS* is a morphosyntactic and not a semantic differentiation; *meisje* and *mannetje* are therefore neutral, even when they refer to people. It is also important to realize that the *GENUS* values pertain to word forms and not lemmas; thus *mannetje* is morphosyntactically neutral, even if the corresponding lemma is masculine (*de man).*

A limited number of type names are used as both gendered and neutral.

Where this occurs in combination with a difference in meaning *(de/het bal, de/het blik)*, we differentiate the two through a different gender. If there is no difference in meaning *(de/het riool, de/het filter, de/het soort)*, we regard them as a single lemma and give them the underspecified gender value *('genus'*), where the local context does not enable disambiguation, as in *een filter* (for a list of such nouns, see ANS-97, p. 159); if the local context does allow for disambiguation, as in *de filter*, a specific value is given. In combinations such as *de laatste drie jaar* and *om de vier uur, jaar* and *uur* are given their usual value '*onzijdig'* in spite of the presence of *de*. In this case the pronoun determines not the noun of measure but the numeral.

Gender differentiation is also relevant for proper nouns. Personal names are usually gendered *(de/*het Karel*) and names of cities, countries and languages are usually neutral *(het/*de Brussel van toen, het/*de Frankrijk van de eeuw-*

*wisseling, het/*de Spaans*), see ANS-97, p. 285. Compound proper nouns with a core usually inherit the gender of the type name *(het/*de Waasland, het/*de Zilvermeer, de/*het Noordzee, de/*het Kem- melberg*); an exception is *de Bijlmermeer*. Where there is a free choice of article, choose the generic value; this applies, for example, to brand names such as Linux and Esselte.

As the morphosyntactic gender is systematically neutralized in plural forms and in forms with a case suffix, we only allocate it to singular standard forms.

07. NAAMVAL. The genitive is marked by the suffix *-s* and occurs mainly with nouns which refer to people or times (*Otto's, vaders,'s avonds*); a limited number of nouns takes the suffix *-en (des Heren, des mensen*). The dative is marked by the suffix *-e* and is almost only found in fixed expressions *(ter plaatse, te berde, in der minne);* the dative with the suffix *-en* is very rare, as in *ten voeten uit*. Not every noun ending with *-e* is a dative form: in *aanvrage* and *proeve,* for example, the schwa part is part of the stem, as it is in *bode* and s*meekbede*. Where there is no case suffix the value '*standaard'* is given. As case differentiation with plural nouns is systematically neutralized (*deze dagen, één dezer dagen*), the feature is not used here.

2.1.4 Implications

[I01] <POS = substantief, GRAAD = diminutief, GETAL = enkelvoud> ⇒ <NAAMVAL ≠ datief>

[I02] <POS = substantief, GRAAD = diminutief, GETAL = enkelvoud, NAAMVAL = standaard> ⇒ <GENUS = onzijdig>

Diminutive nouns do not have a dative form, perhaps because they already end in a schwa and the addition of the dative suffix would therefore make no difference.

Diminutive nouns are always neutral. Note that this refers to the gender of the word form (morpho-syntactic gender) and not the gender of the lemma (lexical gender); the latter could, after all, be gendered, even where the former is neutral.

2.1.5 The tags

The maximum number of specific tags for types names is eight.

[T101]	N (soort, ev, basis, zijd, stan)	die stoel, elke avond, deze muziek, de filter
[T102]	N (soort, ev, basis, onz, stan)	het kind, ons huis, dat brood, dit land, het filter
[T103]	N (soort, ev, dim, onz, stan)	dit stoeltje, ons huisje, op't nippertje
[T104]	N (soort, ev, basis, gen)	's avonds, de heer des huizes, des mensen
[T105]	N (soort, ev, dim, gen)	vadertjes pijp
[T106]	N (soort, ev, basis, dat)	ter plaatse, heden ten dage, te berde brengen
[T107]	N (soort, mv, basis)	stoelen, kinderen, hersenen
[T108]	N (sort, MV, dim)	stoeltjes, huisjes, hersentjes

The number is the same for proper nouns.

[T109]	N(eigen,ev,basis,zijd,stan)	de Noordzee, de Kemmelberg
[T110]	N(eigen,ev,basis,onz,stan)	het Hageland, het Albertkanaal, het Latijn
[T111]	N (eigen, ev, dim, onz, stan)	het slimme Kareltje
[T112]	N (eigen, ev, basis, gen)	des Heren, Hagelands trots, de Aa's bovenloop
[T113]	N (eigen, ev, dim, gen)	Kareltjes fiets
[T114]	N (eigen, ev, basis, dat)	wat den Here toekomt
[T115]	N (eigen, mv, basis)	de Ardennen, de Middeleeuwen, de Kempen
[T116]	N (eigen, mv, dim)	de Maatjes (een natuurreservaat in de Kempen)

The use of underspecified values is relevant for *GENUS*. In cases where the context does not allow for a choice between two gender values, the generic value is given, both for type names and proper nouns.

[U117]	N(soort,ev,basis,genus,stan)	een riool, geen filter
[U118]	N(eigen,ev,basis,genus,stan)	Linux, Esselte

2.1.6 Lemmatization

For lemmatizing type names we use a form without affixes, i.e. the base form, in the singular, without a case suffix. If, however, the noun only has a plural form (*mazelen, hersenen*), that plural form is used as the lemma. The same applies to nouns which only have a diminutive form (*meisje, ootje, akkefietje*); such nouns mainly occur in fixed expressions (*een robbertje vechten, zijn hachje redden, in het ootje nemen, op het nippertje*). Nouns which only occur in the genitive or dative form are given the inflected form as lemma; a rare example is the dative form in *ten behoeve van*.

The same principles apply to lemmatizing proper nouns: they are reduced to a singular basic form *((een) Fiatje* $\Rightarrow$ *Fiat, (drie) Stella's* $\Rightarrow$ *Stella)*, except where the form without the affix does not exist.

This applies, for example, to forms such as *Ardennen, Antillen, Marollen, Enkhuizen, Middeleeuwen* and to forms such as *Nelleke, Sneeuwwitje, Doornroosje*. A small number of nouns have two basic forms, for example *aanvraag/aanvrage* and *proef/proeve*; in that case we also differentiate between the two lemmas. For forms such as *proeven* and *aanvragen* the choice is determined by the context: if they occur in a context where the singular form has a schwa, the lemma of the plural form also has a schwa; otherwise we choose the lemma without a schwa.

Abbreviations such as *TV* and *CD*, and type names which, because of their use in a title, are written with a capital letter, as the nouns in *De Naam van de Roos*, are given a lemma without a capital letter.

2.2 ADJECTIVES

2.2.1 Demarcation

With adjectives we include not only adjectives used prenominally and predicatively, but also those used nominally and adverbially. How we differentiate between nouns and adverbs respectively is explained in 2.2.3. The difference between participle and adjective is explained in the section about verbs, see 2.3.1.

2.2.2 Declarations

All adjectives have the features *POSITIE, GRAAD* and *BUIGING.* Those used nominally are also marked for *GETAL* and — if they are singular and inflected — for *NAAMVAL*. The latter feature is also given to inflected prenominal adjectives.

[D04] <POS = adjectief> ⇒ <POSITIE, GRAAD, BUIGING >

[D05] <POSITIE = nominaal> ⇒ <GETAL-N>

[D06] <POS = adjectief, POSITIE = nominaal, INFLECTION = met-e, GETAL-N = zonder-n> ⇒ <NAAMVAL>

[D07] <POS = adjectief, POSITIE = prenominaal, INFLECTION = met-e> ⇒ <NAAMVAL>

With the exception of *POSITIE* these are morphological characteristics. The degree suffix always precedes the inflectional suffix, which in turn can be followed by a number or case suffix (*de oud-er-e-n*).

2.2.3 Partitions

[P08] POSITIE = prenominaal, nominaal, postnominaal, vrij.

[P05] GRAAD = basis, comparatief, superlatief, diminutief.

[P09] BUIGING = zonder, met-e, met-s.

[P10] GETAL-N = zonder-n, meervoud-n.

[P07] NAAMVAL = standaard, bijzonder.

08. POSITIE. Adjectives occur in various types and positions: prenominal (*een mooie tuin*), nominal (*het/de mooie*), postnominal (*iets moois*) and non-nominal or free; under in the latter we include both the predicative *(dit is mooi, de tuin mooi maken*) and the adverbial *(hij praat zo mooi).*

When used prenominally, adjectives precede the noun they determine. In this position both forms with an inflectional *-e* and forms without an inflected ending (*kleine* vs. *klein*) occur. Prenominal use also includes elliptical use, as in the second conjunction in *Hij heeft een wit bord en ik een groen*; the antecedent does not necessarily have to be placed in the same sentence: *Ik heb gisteren een witte telefoon gekocht. Hij past beter in het interieur dan die groene.* An important argument for treating these adjectives as prenominal is that they show the same variation of forms with and without inflected endings in elliptic use as in prenominal use. In nominal and free use we find more or no variation

respectively.

Nominally (or independently) used adjectives are not treated as nouns but as adjectives. This is necessary among other reasons because of the existence of comparative and superlative forms *(de ouderen, de rijksten*), compatibility with adverbial degree indicators (*de zeer rijken)* and the fact that their plural is formed differently from nouns, see *GETAL-N*. When they refer to people, adjectival nouns always end in a schwa and they can take an *-n* in the plural: *de arme (n)* and *de blinde (n)* are therefore adjectives, but *de liberaal* and *de conservatief* are nouns. When they do not refer to people, the adjectival nouns can also occur without an inflected ending, as in *iets in het groen schilderen*.

Postnominal use is marked as stylistic and is rare (*kindeke klein*), except when the adjective has its own determiners (*alle rivieren bevaarbaar in de winter*), or when it is a determiner for a quantifier *(niets bijzonders, iets groters*). In that case the adjective is usually given an inflectional *-s*. Postnominal use only occurs when the adjective is a post-determiner for a noun; in combinations such as *drie maand lang* and *twee meter breed* the adjective is not a post-determiner for the noun, but the adjective is the core of an adjectival group and the NP is a determiner for the adjective. In that case the *POSITIE* value of the adjective is dependent on the broader context: in *een twee meter brede stoep* the adjective is prenominal and in *die stoep is twee meter breed* it is free. In addition to predicative use, non-nominal or free use also includes adverbial use. Adjectives used adverbially are therefore not treated as adverbs but as adjectives, as they are in ANS-97, CELEX, RBN, WOTAN-2 and the Geramn STTS-95. To differentiate adverbs from adjectives used adverbially we apply the following criterion: if the word in question is also used with the same meaning in prenominal positions, it is not an adverb but an adjective. So *vrij* in *de vogels vrij laten rondvliegen* is an adjective in a free position, because in this context it has the same meaning as the prenominal adjective in *een vrije vogel*, while it is a case of an adverb in *dat is vrij hoog*, see 2.9. As there are no morphological differences between adjectives used predicatively or adverbially, and differentiating is only possible on the basis of a complete syntactic analysis, it is not done during POS tagging.

Not all adjectives occur in each of the four positions. Some are only used prenominally (*houten, ijzeren*) and others are only used freely (jammer, beu). For such adjectives lexical specifications can help determine the *POSITIE* value. For adjectives which occur in all four positions, as in *mooi*, the contextually relevant value must be given. Forms such as *een-/het-/de-/dat-/diezelfde* are prenominal when they define a nominal core, as in *hetzelfde paard*, and nominal when they themselves are the core of a nominal group, as in *het zijn altijd dezelfden die niet opdagen*.

05. GRAAD. The comparative is marked by the suffix *-er*, the superlative by *-st* and the diminutive by *-jes*, as in *(zachtjes, fijntjes, warmpjes).* The reason why the diminutive is placed on a par with the comparative and the superlative is twofold: (1) they are in complementary distribution (a form cannot simultaneously be comparative and diminutive); (2) they combine with the same types of adjectives (degreeable adjectives). Where there is no degree suffix the value '*base'* is given. This also applies to adjectives which cannot take a degree indication (*houten, elektrisch*). This value is also given to articulated adjectives in which the degree suffix is not at the end of the word, as in *verderaf, dichterbij* and *hoogstgelegen, laatstleden*.

09. BUIGING . Most adjectives have three forms (*klein, kleine, (iets) kleins*). The form of the suffix is evidence that the last form differs from the genitive; the inflected ending *iets zaligs* is a minimum pair versus the genitive in *zaliger gedachtenis*. Adjectives of which the stem ends in a *sis* sound (*grijs, boos, theoretisch*) cannot take an *-s* and therefore never have the value '*met-s'*. Something simlilar applies to adjectives which cannot take a schwa (*beige, timide, bescheiden, houten, lila, kaki*); they never have the value '*met-e'*. The schwa in *beige, timide, onderste* is part of the stem. The same applies to the schwa in postnominal position *(1 juli) aanstaande*; postnominal adjectives never have an inflectional *-e.*

10. GETAL-N. When used nominally, adjectives can take a plural suffix *(de blinden, de zieken, de rijken).* The fact that these forms are definitely adjectives and not nouns has been explained above (see *POSITIE*); we can now add to the arguments already given the fact that forming the plural with nominal adjectives differs in two aspects from nouns. Firstly, the suffix used is still *-(e)n*, while for nouns which end in a schwa, *-s* is also used (*dames, bodes, gewoontes, (on)voldoendes*). Secondly, the form with the plural sufix can only refer to people; this restriction does not apply to nouns.

This has two consequences: (1) nominally used adjectives without *-n* are not necessarily singular (*de grootste is/zijn al verkocht*); this is why '*meervoud-n'* is not the opposite of '*enkelvoud-n'* but simply of '*zonder-n';* (2) plural forms with *-n* are only adjectives when they refer to people; mathematical terms such as *kromme(n), variabele(n*) and *gemiddelde(n)* are therefore nouns, and not independently used adjectives; the same applies to superlative forms such as *uiterste(n), groteske(n)* and *vereiste(n*).

In view of these differences, in the notation we also differentiate between the feature used to distinguish count for nouns (*GETAL*) and the feature used to distinguish count for adjectives (*GETAL-N).* A minimum pair is the substantive *ouders* versus the nominally used adjective *ouderen*. An additional reason for differentiating between *GETAL* and *GETAL-N* is that some words, more specifically the nominally used possessive prounouns, are marked for both features; *de zijnen*, for example, has the value '*enkelvoud'* for *GETAL* and the value '*meervoud-n'* for *GETAL- N*, see 2.5.

07. NAAMVAL. The case differentiation is only made for (pre)nominally used adjectives with an inflectional *-e*. Where there is a case suffix (*-er or -en*) they are given the value '*bijzonder'*. We do not distinguish between genitive and dative, because this does not correspond with the different forms: the *-er* form is a genitive in *zaliger gedachtenis* and a dative in *te goeder trouw*, and the *-en* form is a genitive in *des Allerhoogsten* and a dative in *in koelen bloede*. In the absence of a case suffix the value '*standaard'* is given.

2.2.4 Implications

[I03] <POS = adjectief, GRAAD = superlatief> $\Rightarrow$ <POSITIE $\neq$ postnominaal>
[I04] <POS = adjectief, GRAAD = diminutief> $\Rightarrow$ <POSITIE = vrij>
[I05] <BUIGING = met-s> $\Rightarrow$ <POSITIE = postnominaal>
[I06] <BUIGING = met-e> $\Rightarrow$ <POSITIE = (pre)nominaal>

The first two implications concern relationships between *GRAAD* and *POSITIE*:

the superlative forms cannot be used postnominally and the diminutive forms only occur in predicative or adverbial positions. The diminutive form in *de kleintjes* is therefore not an adjective, but the plural form of the noun *kleintje*; the same is the case in *de oudjes*.The last two implications link *INFLECTION* and *POSITIE*. The *-s* forms only occur in the postnominal positions; we do not regard the *s-* forms in *van jongs af aan, er is nieuws, lekkers krijgen* as inflected forms of the adjective, but as nouns with separate lemmas. The *-e* form only occurs in nominal and prenominal positions; the form *hele* in *een hele mooie tuin* is treated as prenominal and not as free. Please note that these two implications are valid for all word types with a feature, and therefore also for participles and determiners.

2.2.5 The tags

The interaction of the declarations with the implications produces a maximum of nine specific combinations for prenominal adjectives: three for the base forms, three for comparatives and three for superlatives.

[T201]	ADJ(prenom,basis,zonder)	een mooi huis, een houten pot
[T202]	ADJ(prenom,basis,met-e,stan)	mooie huizen, een grote pot
[T203]	ADJ(prenom,basis,met-e,bijz)	zaliger gedachtenis, van goeden huize
[T204]	ADJ(prenom,comp,zonder)	een mooier huis
[T205]	ADJ(prenom,comp,met-e,stan)	mooiere huizen, een grotere pot
[T206]	ADJ(prenom,comp,met-e,bijz)	van beteren huize
[T207]	ADJ(prenom,sup,zonder)	een alleraardigst mens
[T208]	ADJ(prenom,sup,met-e,stan)	de mooiste keuken, het grootste paard
[T209]	ADJ(prenom,sup,met-e,bijz)	bester kwaliteit

For nominal adjectives the count differentiation also plays a role, resulting in more combinations:

[T210]	ADJ(nom,basis,zonder,zonder-n)	in het groot, in het bijzonder, het groen
[T211]	ADJ(nom,basis,zonder,mv-n)	de timiden, dezelfden
[T212]	ADJ(nom,basis,met-e,zonder-n,stan)	het leuke is dat ..., geef mij maar een grote met tartaar
[T213]	ADJ(nom,basis,met-e,zonder-n,bijz)	hosanna in den hogen
[T214]	ADJ(nom,basis,met-e,mv-n)	de rijken
[T215]	ADJ(nom,comp,zonder,zonder-n)	
[T216]	ADJ(nom,comp,met-e,zonder-n,stan)	een betere
[T217]	ADJ(nom,comp,met-e,zonder-n,bijz)	
[T218]	ADJ(nom,comp,met-e,mv-n)	de ouderen
[T219]	ADJ(nom,sup,zonder,zonder-n)	op z’n best, om ter snelst
[T220]	ADJ(nom,sup,met-e,zonder-n,stan)	het leukste is dat, het langste blijven
[T221]	ADJ(nom,sup,met-e,zonder-n,bijz)	des Allerhoogsten
[T222]	ADJ(nom,sup,met-e,mv-n)	de slimsten

With the postnominal adjectives case and number are not relevant and the variation in degree is restricted to two values (*base* and *comparative*).

[T223]	ADJ(postnom,basis,zonder)	rivieren bevaarbaar in de winter
[T224]	ADJ(postnom,basis,met-s)	iets moois
[T225]	ADJ(postnom,comp,zonder)	een getal groter dan drie
[T226]	ADJ(postnom,comp,met-s)	iets gekkers kon ik niet bedenken

Finally, case and number are also irrelevant in free use and the value of the inflection feature is invariable.

[T227]	ADJ(vrij,basis,zonder)	die stok is lang, lang slapen
[T228]	ADJ(vrij,comp,zonder)	deze stok is langer, langer slapen
[T229]	ADJ(vrij,sup,zonder)	welke stok is het langst, het langst slapen, de verst afgelegen dorpen
[T230]	ADJ(vrij,dim,zonder)	het is hier stilletjes, stilletjes wegsluipen

2.2.6 Lemmatization

To specify the lemma we use a form without affixes i.e. the base form, without inflection, count or case suffix. If the adjective does not have a base form, as in the comparative in *eerdere pogingen*, we take the comparative as the lemma. We also do this when the comparative form is autonomous; the lemma of the comparatives in *verder denk ik dat dit niet klopt* and *het verdere verloop van de procedure* is *verder*, and not *ver*; the comparatives in *vroeger was alles beter vroegere pogingen*, *later zal je dat wel begrijpen* and *latere successen* are similar. Typical of the autonomous comparative forms is that they cannot be combined with the determiner *dan* and they cannot take an adverbial degree determiner, such as *veel* in *zij springt veel verder dan ik* and *het was later/vroeger dan ik dacht.* The lemma is therefore the same as the base forms *ver, laat, vroeg*. When stripping the inflected ending you should bear in mind that not every adjective which ends in a schwa is inflected; in *beige, timide, morbide, onderste,* for example, the schwa is part of the stem (and thus the lemma), as forms such as *beig, timid* and *onderst* do not exist. Adjectives which only occur with a case suffix, such as *arren* in *in arren moede*, are also used as a lemma.

2.3 VERBS

2.3.1 Demarcation

Compared to other word types, there are hardly any problems with demarcation as far as the finite forms are concerned; it is obvious that *vliegen* in *we vliegen* is a verb, whereas it is a noun in *wat een vervelende vliegen.* We treat infinitives as verbs, even when they occur in nominal positions (*het vallen van de bladeren, het polsstokspringen*). There are a few infinitive forms which have a noun homonym such as *leven*. Unlike the infinitive, the substantive has a plural form (*levens*) and a diminutive form (*wat een leventje*) and is usually used with an

indefinite article (*een leven*). We also treat participles as verbs, except where they clearly show adjectival characteristics. This is the case with forms (1) with a non-verbal stem, such as *getand, geaderd, gevleugeld*; (2) with a typical adjectival prefix (*on-gehoord, on-deugend, aarts- ingewikkeld*); (3) with a degree suffix, mainly a comparative or superlative (*opgewekter, spannendste*), see ANS-97, p. 388-389; (4) with an inflectional *-s (iets uitdagends, iets gezochts), see* ANS-97, p. 412; (5) with a case suffix, mainly a dative (*te gepasten tijde, te bestemder plaatse, met voorbedachten rade*), see ANS-97, p. 413. These criteria are easy to apply, because they are related to the form itself, irrespective of the context in which it occurs. For the other forms the choice of word type is determined by the context. Here, too, the default is to choose a verbal analysis. However, we opt for an adjectival analysis when the participle (6) is combined with a degree determiner, such as *zo opgewekt, zeer beperkte voorraad, te ingewikkeld, heel spannende film, hoogst opwindende lingerie, ...*; (7) has a valency which is not the same as that of the corresponding verb; compare for example the verbal use of *bedacht* in *dat heb je zeker zelf bedacht* and *dat is bedacht (door ...)* with the adjectival use in *daar was ze niet op bedacht*. In the first two examples it is a form of the transitive verb *iets bedenken*, but in the latter there is no verb with the corresponding valency: **op iets bedenken*; instead it is a case of the adjectival *op iets bedacht zijn*; (8) with the combination 'imperfect participle + noun' it is not possible to paraphrase as 'noun + rel.pron. + finite verb' and still retain the lexical meaning. The participle in *een spannende film* is adjectival, because it cannot be paraphrased as *een film die spant*, while the participle in *een doodlopende straat* is a verb, as it can be paraphrased as *een straat die doodloopt*; (9) the combination 'past participle + *zijn'* must be first in subordinate clauses; compare the adjectival participle in * *dat hij na de lange strijd was uitgeput* with the verbal in *dat hij toen al was vertrokken*.

2.3.2 Declarations

The verbs are associated with different features, depending on whether they are used as a finite or non-finite form.

[D08] <POS = werkwoord> ⇒ <WVORM>
[D09] <WVORM = persoonsvorm> ⇒ <PVTIJD, PVAGR>
[D10] <WVORM = buigbaar> ⇒ <POSITIE, BUIGING>
[D05] <POSITIE = nominaal> ⇒ <GETAL-N>

The finite forms therefore have three features as do the non-finite (or non-finite) forms, except in nominal use; then they have four.

2.3.3 Partitions

[P11] WVORM = persoonsvorm, buigbaar (infinitief, onvdw, voltdw).
[P12] PVTIJD = tegenwoordig, verleden, conjunctief.
[P13] PVAGR = enkelvoud, meervoud, met-t.
[P08] POSITIE = pronominaal, nominaal, vrij.
[P09] BUIGING = zonder, met-e.
[P10] GETAL-N = zonder-n, meervoud-n.

11. WVORM. The non-finite verb forms are the infinitive and the participles. The infinitives have the suffix *-en* or in some cases *-n* (*zijn*,

gaan, slaan, staan, doen, zien) and the imperfect participles are formed by adding *-d* of *-de* to the infinitive. The past participles have the suffix *-d, -t* or *-en* and in many cases also the prefix *ge-*. Differentiating between past and passive participle is not a job for the tagger.

12. PVTIJD. The present tense is not morphologically marked, the past tense has the suffix *-de* or *-te* (suppletion in the case of the irregular verbs) and the conjunctive has the suffix *-e* (*moge, leve, kome*), except when the stem ends in a vowel (*het zij zo, het ga je goed).* Some verbs have more than one base form for the present tense (*kun/kan, zul/zal* and *ben/is/wees*). This phenomenon also occurs in the past tense *(joeg/jaagde, wou/wilde*) and in the conjunctive (*zij, weze, ware*). The imperative is not differentiated from the present tense.

13. PVAGR. Where there is no *AGR* suffix the value *'enkelvoud'* is given, because such forms are always singular; the *'persoon'* value is variable: first person (*ik kom, ik kwam, ik speelde, ik maakte*), second person (*kom je, je mag, je kwam, je speelde, je maakte, kom*) or third person (*hij mag, ze kwam, het speelde, hij maakte, leve de koning*). In combination with the *PVTIJD* distinction this produces three forms for most verbs, one for the present tense and the imperative (*kom, leef*), one for the past tense (*kwam, leefde*) and one for the conjunctive (*kome, leve).* If a verb has more than one form for a particular *PVTIJD* value, then of course more combinations are possible with *PVAGR. Kun* is not the only singular form of the present tense of *kunnen*, but also *kan*; similarly for *ben* and *is* in the case of *zijn*. The finite forms with an *-(e)n* suffix are always plural and are therefore given the value *'meervoudl'* ; here, too, the *persoon* value is variable (*wij/jullie/zij komen, wij/jullie/zij kwamen, wij/jullie/zij zijn)*. There are two forms for most verbs in combination with the *PVTIJD* distinction: one for the present tense (*komen, leven, maken)* and one for the past tense (*kwamen, leefden, maakten*). The imperative and conjunctive forms do not take this suffix.

The forms with a *-t* suffix can be both singular and plural, and are given the value *'met-t';* they may be the second person *(je komt, je bent, gij waart, gaat u zitten, geeft acht)* or the third person (*hij komt*). Most verbs have two forms with this suffix: one for the present tense and the imperative (*komt, geeft, gaat*) and one for the past tense (*kwaamt, gaaft*); the last two only exist for irregular verbs in combination with *gij* as the subject, and therefore occur mainly in Flemish. As we would expect, here too *zijn* has greater variation: instead of one form for the present tense and the imperative it has three (*bent, zijt, weest*). The conjunctive forms do not take this suffix. The value *'met-t'* is only allocated where *-t* is an *AGR* suffix is; if it is part of the stem, as in *hij zit*, the value 'enkelvoud' is given.

08. POSITION. As with the adjectives, the participles occur in different positions: prenominal (*een fraai versierde kerstboom, een slapend lid*), nominal (*de gedupeerde, het geschrevene, de wachtenden*) and free (*we zijn opgelicht, wat is hier gaande, achternagestaard door de menigte reed hij langzaam weg, hij liep luid lachend weg).* There are no separate tags for postnominal use, because the participles in that position do not express any variation and are therefore systematically the same as freely used participles (*een boom versierd met slingers, alle burgers residerend in Brussel).* Free use also includes the cases in which the participle is the verbal complement of an auxillary verb (*ze hebben ons niets gezegd, we worden morgen ontslagen*). With free use the imperfect participle is sometimes introduced by *al*, as in *al doende leert men*. Infinitives are mainly used as complements of other verbs, mainly auxillary verbs. Besides this free use there is the nominal use (*het schaatsen)* and the prenominal (*de nog*

te lezen post). In prenominal use the infinitive is always preceded by *te*; in nominal use the infinitive is often, but not always, introduced by *het*. In free use we find infinitives both with and without *te*. An advantage of this treatment is that they do not introduce any systematic ambiguity: it is not necessary to allocate different word types to the infinitives in *ze kan lezen, dat was te verwachten, lang wachten is vervelend, het lossen van de duiven* and *de nog te lezen stukken*.

09. BUIGING In prenominal positions the participles— just like the adjectives— show forms with and without *-e (een slapend kind, een slapende man; een getemd paard, een getemde feeks*). In nominal positions participles always take an inflectional *-e (het geschrevene, het vervelende)*. In free (predicative and adverbial) positions the past participle—similarly to the adjective— does not take an inflectional *-e*; the imperfect participle occurs in free positions both with and without schwa (*al doende leert men, hij liep (al) zingend de trap af*), but we regard this schwa as an optional part of the participle affix, and not as an inflected ending, seeing that its occurance is not on the whole determined by the rules which form the basis of the use of the inflectional *-e*. With the infinitives there is less variation, because these almost always end in *-en*, but there are some exceptions (*niet mis te verstane bewoordingen, een niet te weerstane verleiding*).

Unlike adjectives the non-finite verbs do not have *-s* forms: participles with an *-s* suffix are after all regarded as adjectives (see A.) and infinitives are not compatible with this suffix (*iets te eten(*s), niets te melden(*s*)). The suffix, in combinations such as *nog vier uur gaans, wetens en willens, tot ziens* and *tot bloedens toe*, is not an inflected ending, but a derivational suffix which derives adverbs from infinitives or nouns (cf. *deel-s, daag-s*), see ANS-97, p. 738.

10. GETAL-N. In nominal use, participles can take a plural suffix (*de gedupeerden, de wachtenden)*; as with adjectives this is only possible when people are being referred to. The nominally used infinitive is always singular (*het wachten valt me zwaar*).

2.3.4 Implications

[I06] <BUIGING = met-e> ⇒ <POSITIE = (pre)nominaal>
[I07] <WVORM = infinitief, POSITIE = nominaal> ⇒
<BUIGING = zonder, GETAL-N = zonder-n>
[I08] <PVTIJD = conjunctief> ⇒ <PVAGR = enkelvoud>

The first implication concerns the non-finite forms and also applies to adjectives and determiners. The second states that nominally used infinitives cannot take an inflectional s- or plural affix. The third states that conjunctive forms are always singular; the plural forms are after all systematically the same as the present tense.

2.3.5 The tags

If we restrict ourselves to the maximum number of specific combinations, then there are nine for finite forms:

[T301]	WW(pv,tgw,ev)	kom, speel
[T302]	WW(pv,tgw,mv)	komen, spelen

[T303]	WW(pv,tgw,met-t)	komt, speelt
[T304]	WW(pv,verl,ev)	kwam, speelde
[T305]	WW(pv,verl,mv)	kwamen, speelden
[T306]	WW(pv,verl,met-t)	kwaamt, gingt
[T309]	WW(pv,conj,ev)	kome, leve de koning

There are four combinations for the infinitive and five for each of the participles.

[T301]	WW(pv,tgw,ev)	kom, speel
[T302]	WW(pv,tgw,mv)	komen, spelen
[T303]	WW(pv,tgw,met-t)	komt, speelt
[T304]	WW(pv,verl,ev)	kwam, speelde
[T305]	WW(pv,verl,mv)	kwamen, speelden
[T306]	WW(pv,verl,met-t)	kwaamt, gingt
[T309]	WW(pv,conj,ev)	kome, leve de koning

For the infinitive there are four combinations and five for each of the participles:

[T310]	WW(inf,prenom,zonder)	de nog te lezen post
[T311]	WW(inf,prenom,met-e)	een niet te weerstane verleiding
[T312]	WW(inf,nom,zonder,zonder-n)	(het) spelen, (het) schaatsen
[T314]	WW(inf,vrij,zonder)	zal komen
[T315]	WW(vd,prenom,zonder)	een verwittigd man, een gekregen paard
[T316]	WW(vd,prenom,met-e)	een getemde feeks
[T317]	WW(vd,nom,met-e,zonder-n)	het geschrevene, een gekwetste
[T318]	WW(vd,nom,met-e,mv-n)	gekwetsten, gedupeerden
[T320]	WW(vd,vrij,zonder)	is gekomen, een boom versierd met slingers
[T321]	WW(od,prenom,zonder)	een slapend kind
[T322]	WW(od,prenom,met-e)	een piano spelende aap, slapende kinderen
[T323]	WW(od,nom,met-e,zonder-n)	het resterende, een klagende
[T324]	WW(od,nom,met-e,mv-n)	de wachtenden
[T326]	WW(od,vrij,zonder)	liep lachend weg, al doende leert men

The intermediate value '*buigbaar*' for *WVORM* is used to formulate the Declarations (D10), but is not intended for underspecification in tags. Forms which can be both an infinitive and a past participle (*bekomen, vergaan*), are therefore disambiguated.

2.3.6 Lemmatization

We do not use the stem as the lemma, but the infinitive form.

2.4 NUMERALS

For many reasons, it would be preferable from a linguistic point of view to treat cardinal numbers as nouns, further defined as type names, and the ordinals as adjectives. However, for recognition purposes and the subsequent requirement to conform to ANS practice and the EAGLES recommendations, this option was not chosen. As a consequence the criteria for differentiating between the cardinal

numbers and nouns are somewhat artificial; the same applies to allocating the *POSITIE* differentiation to cardinal numbers.

2.4.1 Demarcation

We include all words which have separate forms for cardinal and ordinal numbers in numerals. This includes of course the names of the numerals, such as *twee(de), dertig(ste), zeshonderd(ste), ...*, but also the words *elfendertig(ste), tig(ste), hoeveel(ste), evenveel(ste)* and *zoveel(ste)* Words such as *beide, veel* and *weinig* do not rank under numerals on the basis of this criterion; instead they are treated as indefinite pronouns, see 2.5. We also include the use of *één* in combinations as an indefinite pronoun, for example as in *het één en ander* and *één en al aandacht*; what characterizes these combinations is that *één* is not in complementary distribution with other numerals. Numerals should be differentiated from nouns. For forms with a plural suffix the difference between numeral and noun is explained on the basis of a minimum pair, as in *met z'n zevenen* versus *hij heeft twee zevens*. In the first example *zeven* is treated as a nominally used numeral *(GETAL-N = meervoud-n)* and in the second, as a noun (*GETAL = meervoud).* Similarly, the plural forms in *met z'n tweetjes* and *met z'n honderden tegelijk* are tagged as nominally used numerals, while the plural forms in *ik heb nog twee tientjes* and *honderden kisten* are treated as nouns. The forms without a plural suffix are classed as nouns, when they form the core of a singular *NP*, as in *ze heeft een zes, deze vijf is mooier dan die* and *zes is/* czijn deelbaar door drie*; if the resulting *NP* on the other hand is plural, it is a numeral, as in *deze vijf zijn mooier dan die* and *er zijn/*is er zes ontsnap*t. Numerals should also be differentiated from adverbs. The forms expressing time in *tegen énen* and *na zessen* for example are not plural numerals with a *GETAL-N* ending, as in *met z'n zessen*, but adverbs. The suffix *-en* can therefore be seen as a derivational affix for deriving adverbs. This also occurs in forms such as *voren, achteren, onderen*, see 2.7. In the adverbs we also include forms such as *halfzeven*; after all, they cannot be numerals, as there is no corresponding ordinal number.

Fractions such as *vijf achtste* are treated as combinations of a cardinal number and an ordinal number respectively. If they are written as a single word, as in *een tweederde meerderheid, een viertiende baan, ik werk nu viertiende,* we include them in the adjectives; this way we can also differentiate the adjective *viertiende* from the ordinal number *veertiende*. Combinations such as *anderhalve, zesenhalve, driekwart* are also included in the adjectives; after all they do not have a corresponding ordinal number.

2.4.2 Declarations

[D11] <POS = telwoord> ⇒ <NUMTYPE, POSITIE>
[D05] <POSITIE = nominaal> ⇒ <GETAL-N>
[D12] <NUMTYPE = hoofdtelwoord, POSITIE = nominaal> ⇒ <GRAAD>
[D13] <POS = telwoord, POSITIE = prenominaal> ⇒ <NAAMVAL>

Unlike adjectives, numerals do not have an inflection feature, seeing as this distinction is systematically neutralized: the cardinal numbers after all never take an inflectional *-e* and the ordinal numbers always end in a schwa. The only form for which differentiation could be relevant is *één (die éne keer*), but because this form is also used as an indefinite pronoun and when functioning as such has an inflection feature anyway, we treat the form *éne* as an indefinite pronoun, and not as a numeral.

2.4.3 Partitions

[P14] NUMTYPE = hoofdtelwoord, rangtelwoord.
[P08] POSITIE = prenominaal, nominaal, vrij.
[P10] GETAL-N = met-n, zonder-n.
[P05] GRAAD = basis, diminutief.
[P07] NAAMVAL = standaard, bijzonder.

14. NUMTYPE. The cardinal numbers include for example één, twee, drie, *zoveel.* The ordinal numbers are the corresponding forms with the suffix *-ste* or *-de*, cf. *eerste, tweede, derde, zoveelste.*

08. POSITIE. Cardinal numbers are treated as prenominal when they precede a noun and determine how many of the noun are intended, such as in *één hond* and *vijf kinderen.* In combinations such as *vijf juli* therefore, the numeral is not prenominal, as it does not determine the quantity of *juli's.* As with the adjectives, the prenominal numerals can be used elliptically, such as *twee* in *hij krijgt vijf rode knikkers en jij twee groene*. The ordinal numbers are treated prenominally when they precede a noun and determine which position is occupied by the noun, as in *het vijfde kind*.

Numerals with a plural affix (see *GETAL-N)* or a diminutive suffix (see *GRAAD*) are always nominal. Forms without these affixes are nomina when they occur in the same positions as the forms with such affixes, as in sentences with an *er*, cf. *er is er één(tje) ontsnapt*. It is apparent that this latter use is not prenominal because the diminutive forms cannot be used prenominally for example. Examples of nominally used ordinal numbers are *Lodewijk de Veertiende, de dertiende van elke maand, zij was (de) vierde en altijd (de) tweede eindigen.* The cardinal numbers not used (pre)nominally are regarded as free. This includes not only predicative use, as in *hij wordt zestig*, and adverbial, as in *hij reed minstens honderd*, but also the use in *NPs* where the numeral does not express the quantity of what the nominal denotes as the head of the *NP*; relevant examples involve the use of *twintig* in *twintig juli, de jaren twintig, pagina twintig* and *twee euro twintig*. In these last three examples it is therefore not a question of postnominal use: in *de jaren twintig* for example is not a case of a period of twenty years but of a period of ten years (1920 to 1929), and in *pagina twintig* it is not a case of twenty pages but of 1 page.

If a numeral is split into different words, each of these parts is treated in the same way. In *tweehonderd veertien pagina's* the numerals are therefore both prenominal and in *pagina tweehonderd veertien* they are both free.

10. GETAL-N. With nominal use the numerals can take the plural suffix *-en* cf. *met z'n vieren*. The forms in *wij tweeën* and *zij vijven* are the same (comparable with *wij fietsers*). The ordinal numbers can also take a plural suffix in nominal positions, cf. *de eersten*. Note the semantic limitation with reference to people. Forms such as *na vieren* in the sense of *na vier uur* are completely different. In this case it is not a question of a plural form which is refering to people, but and adverb. It is apparent that this cannot be the plural form of the combination *tegen énen*, for example. A motivation for the classification as an adverb is that the addition of *-en* often generates an adverb: applying this to the preposition *voor*, for example, produces the adverb *voren*, see 2.7.

05. GRAAD. The existence of diminutive forms for some of the cardinal numbers, cf. *op z'n eentje, met z'n tweetjes* is typically Dutch. Note that the

differentiation in quantity is also relevant here.

07. NAAMVAL. Dative forms have the suffix *-en* or *-er* and occur mainly in fixed expressions (*te elfder ure, te enen male*). Genitive forms are very rare; a possible example is the Biblical *eens geestes zijn.* In view of this we only make the differentiation—as with the adjectives—between forms with a case suffix (‘*bijzonder’*) and forms without a case suffix (‘*standaard’*).

2.4.4 Implications

[I09] <NUMTYPE = rangtelwoord> ⇒ <POSITIE ≠ vrij>

The ordinal numbers are only used (pre)nominally.

2.4.5 The tags

There is a maximum of seven specific combinations for the cardinal numbers.

[T401]	TW(hoofd,prenom,stan)	vier cijfers
[T402]	TW(hoofd,prenom,bijz)	eens geestes zijn, ten enen male
[T403]	TW(hoofd,nom,zonder-n,basis)	
[T404]	TW(hoofd,nom,mv-n,basis)	met z’n vieren
[T405]	TW(hoofd,nom,zonder-n,dim)	er is er eentje ontsnapt, op z’n eentje
[T406]	TW(hoofd,nom,mv-n,dim)	met z’n tweetjes
[T407]	TW(hoofd,vrij)	veertig worden, zoveel sneller, pagina vijf, de jaren zestig, zes juli

As there is no *GRAAD* or *free use* for ordinal numbers, they have only four.

[T408]	TW(rang,prenom,stan)	de vierde man
[T409]	TW(rang,prenom,bijz)	te elfder ure
[T410]	TW(rang,nom,zonder-n)	het eerste, (de) vierde zijn
[T411]	TW(rang,nom,mv-n)	de eersten, iets aan derden verkopen

2.4.6 Lemmatization

We use the base form of the cardinal number as the lemma. Some examples are *drie, zestig, zoveel, tig.*

2.5 PRONOUNS

The pronouns form a heterogeneous class and are consequently put into two separate word types in some tagsets, such as PAROLE; the pronouns, which are generally the core of the a *NP*, and the determiners, which are used more as a determiner for a noun. For CGN we have also decided to classify both as the same word type and differentiate the pronomen/determiner in terms of a separate feature (*PDTYPE).* This choice is in part justified by the fact that neither ANS-97 nor EAGLES uses different word types for pronouns and determiners, and partly because, besides their differences, pronouns and

determiners also have a number of characteristics in common (*VWTYPE, NAAMVAL).*

2.5.1 Demarcation

For the identification of pronouns we have mainly followed ANS-97. However, at some points we have adopted a different approach. One of the first points from which we have deviated is the so-called pronominal adverb *hier, daar, waar, ergens, nergens* and *overal*. The classification of these cases as adverbs in ANS-97 is perhaps prompted by the fact that in a combination such as *ik woon hier/daar al zestien jaar* they are the core of a locative determiner. However, for the CGN tagset this is not the main consideration, because definition of word type is based entirely on form and not on functional criteria; *maandag* and *jaren* in *ze komen maandag* and *ik heb jaren in Portugal gewoond* may be the core of a temporal determiner, but as far as the word type is concerned, they are both nouns. Similarly, we can say that pronominal adverbs are pronouns which are, for example, used as a locative determiner. An advantage of this analysis is that it is also relevant for their use in *PPs*: After all, in combination with a preposition, they take the place of the pronouns *dit, dat, wat, iets, niets* and *alles*, compare the ungrammatical *op wat* with the grammatical *waarop*. Orthograghically this combination may form a single entity, so that the tagger will also treat it as one word, but that is not the case when the pronoun is separated from the preposition, as in *daar wacht ik niet op*. In such cases *daar* must be allocated its own word type and the best candidate for this is '*voornaamwoord'*, not only because of the complementary distribution with the pronouns, but also in view of the relevance of the feature *VWTYPE. Hier en daar* are demonstrative, whereas *vragend of betrekkelijk,* and *ergens, nergens* and *overal* are indefinite. A second point of divergence is the much discussed word *er*. ANS-97 treats it as an adverb and differentiates between four uses for it. Two of these are the same as the use of pronominal adverbs, in particular locative, as in *ik kom er niet graag*, and the *PP* use, as in *ze wacht er al maanden op*, in which *er* is in complementary distribution with the personal pronoun *het*. Similarly to *daar* and *cs.* we also classify *er* as a pronoun. This also applies to the two other uses. One of these is what ANS-97 calls the *presentatieve*. This refers to the temporary or expletive subject in sentences such as *er staat een man voor de deur;* note that this use is also a case of complementary distribution with *het*, compare *er wordt gezegd dat* with *het wordt betreurd dat*. Finally, there is the quantative use, as in *ik heb er vijf*, which is clearly pronominal, perhaps even more so than the other three. Historically it is the genitive form of a personal pronoun, see ANS-97, 464. In short, there are various arguments for classifying *er* as a pronoun, rather than an adverb. A third point of divergence concerns *veel/meer/meest, weinig/minder/ minst* and *beide*. ANS-97 includes them in the numerals, but because these words, in contrast to real numerals, do not have corresponding ordinal numbers, we have not adopted this custom. Instead, we class them as pronouns, more specifically as indefinite determiners. This is more in line with the classification for translated equivalents in other languages.

In addition, there are a number of words which ANS-97 classes as demonstrative or indefinite pronouns (or at least covers them in these sections), but which in CGN are included in other word types. For example, the adjectives *dergelijk(e), soortgelijk(e), dusdanig(e), zo- danig(e), -zelfde* and the adverbs *zelf, genoeg, zat*. Furthermore, , ANS-97 includes a number of multiword combinations such as *deze of gene, dit of dat, een en ander, een paar, ...* (p. 356) in the indefinite pronouns. In CGN such combinations of words are allocated their own word types, such as conjunction, pronoun, article, noun, etc.

To give definitive viewpoint in cases of doubt, a lexicon of all pronouns, including their tag(s) and lemma has been drawn up. This entire lexicon has been

included in the CGN lexicon.

2.5.2 Declarations

The heterogeneity of this class is reflected in the large number of declarations needed for the relevant feature combinations.

[D14] <POS = voornaamwoord> ⇒ <VWTYPE, PDTYPE, NAAMVAL>
[D15] <PDTYPE = pronomen> ⇒ <STATUS, PERSOON, GETAL>
[D16] <VWTYPE = persoonlijk, NAAMVAL = standaard, PERSOON = 3, GETAL = enkelvoud> ⇒ <GENUS>
[D17] <PDTYPE = determiner> ⇒ <POSITIE, BUIGING>
[D05] <POSITIE = nominaal> ⇒ <GETAL-N>
[D18] <PDTYPE = determiner, POSITIE = prenominaal> ⇒ <NPAGR>
[D19] <PDTYPE = gradeerbaar> ⇒ <GRAAD>
[D20] <VWTYPE = bezittelijk> ⇒ <STATUS, PERSOON, GETAL>

All pronouns have the features *VWTYPE*, *PDTYPE* and *NAAMVAL*. In addition, pronouns also have *STATUS*, *PERSOON* and *GETAL*, where necessary supplemented by *GENUS*. In addition to the common three, the determiners also have *POSITIE* and *BUIGING*, where necessary supplemented by *GETAL-N*, *NPAGR* and/or *GRAAD*. A separate declaration has been added for possessive pronouns; as determiners they of course have all the features associated with the determiners but they also have the additional features *STATUS*, *PERSOON* and *GETAL*.

2.5.3 Partitions

[P15] VWTYPE = pr (personal, reflexive), reciprocal, possessive, vb (interrogative, relative), exclamative, demonstrative, indefinite.
[P16] PDTYPE = pronomen (adv-pronomen), determiner (gradeerbaar).
[P07] NAAMVAL = standard (nominatief, oblique), genitive, dative.
[P17] STATUS = full, reduced, stress.
[P18] PERSOON = person (1, 2 (2v, 2b), 3 (3p (3m, 3v), 3o)).
[P04] GETAL = count (singular, plural).
[P06] GENUS = masculine, feminine, neutral.
[P08] POSITIE = prenominal, nominal, free.
[P09] BUIGING = without, with-e.
[P19] NPAGR = agr (evon, rest (evz, mv)), agr3 (evmo, rest3 (evf, mv)).
[P10] GETAL-N = without-n, plural-n.
[P05] GRAAD = base, comparative, superlative, diminutive.

15 VWTYPE. The split into nine types is in line with ANS-97. As in EAGLES, there are common intermediate values for the personal and reflective pronouns on the one hand and for the interrogative and relative pronouns on the other; this saves us postulating over the systematic ambiguity of pronouns which can adopt both roles (*me, mij, ons, je* en *wat, welke,* respectively).

16. PDTYPE. The classification into *pronomina* and determiners depends on the *VWTYPE* classification, but there are nevertheless connections. In short, (1) personal, reflexive and reciprocal pronouns are pronouns; (2) possessive pronouns are determiners; (3) interrogative, relative and relative pronouns are—

with the exception of *welk(e)* and *hetgeen—pronomina;* (4) demonstrative and indefinite pronouns are partly determiners and partly pronouns. Within pronouns we have separate subtypes for pronominal adverbs and *er* (adverbial pronoun) and within the determiners there is a separate subtype for indefinite determiners with comparatives, such as *veel* and *weinig* (degreeable). The *PDTYPE* value for every pronoun is given in the lexicon. The criteria used in allocating these values are explained briefly below.

The differentiation between pronouns and determiners has similarities with the difference ANS-97 makes between nominal and non-nominal pronouns, but is not the same. There are two reasons for this: (1) determiners—like adjectives—are also used independently in particular in nominal positions; (2) the genitive forms of these pronouns—like those of the nouns— are also used non-nominally, in particular in pre- or postnominal positions. It is not sufficient to differentiate the pronouns from the non-nominal or prenominal determiners (A), they must also be differentiated from the independent or nominally used determiners (B).

A. The first difference between prenominally used determiners and pronouns is that the case of the first must agree with the case of the noun which they determine, while the pronouns which are used as a determiner with a noun, always have the genitive form, also when the modified noun is a standard form. The possessive *mijn* is therefore a determiner, because its case cannot be different to that of the noun *mijn/*mijns hoofd* and *mijns/*mijn inziens*; the same goes for the indefinite *alle* in *alle/*allen leden* and *te allen/*alle prijze.* The pronouns in *wiens huis* and *mijns gelijke* on the other hand are pronouns, in view of the fact that the contrast in case (genitive versus standard) does not result in ungrammaticality; the same applies to pronouns which follow the word they determine: in *wie uwer* (*= wie van u*) and *één hunner (= één van hun)* the post-determiners are pronouns not determiners, because the difference in case with the previous word does not cause any incompatibility. A second difference is that the prenominally used determiners must have the same gender and count as the noun they determine, while the pronouns used as a determiner for a noun do not require such agreement. The demonstrative pronoun in *deze tafel/boeken* is therefore not a determiner, because it is only compatible with singular gendered or plural nouns, cf. **deze boek*. The same applies for the indefinite *elke*, which can only be combined with singular gendered nouns *elke gans/*boek/*ganzen* The pronouns in *wiens paarden* and *diens hemden* on the other hand are pronouns, because, although they are both singular and gendered, they are still compatible with plural and neutral nouns.

A third difference can be seen in the following paraphrase test: if by replacing the genitive by a *PP*, both the pronoun and the noun appear in the *PP*, we speak of a determiner; if on the other hand only the pronoun appears in the PP, we speak of a pronoun. This test clearly shows the difference between the possessive *mijns*, which is a determiner, and the personal *mijns*, which is a pronoun .

mijns inziens *===* *naar mijn inzien* *DET*
mijns gelijke *===* *de gelijke van mij* *PRO*

B. To differentiate between pronouns and nominally used determiners other tests are required, seeing that, as far as the above-mentioned criteria are concerned, the nominally used determiners behave in the same way as the pronouns. In *met aller instemming*, for example, only the pronoun has the genitive form, there is no agreement with noun gender and count (*aller* is plural and *instemming* singular) and the PP paraphrase only contains the pronoun *('met de instemming van allen'*); yet *aller* is not a pronoun but a nominally used determiner. In order to make this differentiation the morphological structure of the pronoun is the deciding factor: (1) nominally used determiners actually

always have an inflectional *-e*, and (2) if they have a plural form, this is marked by the suffix *-n*, with the restriction that the resulting form can only refer to people: *aller* meets both criteria. Pronouns on the other hand do not have an inflectional *-e*, and if they have a separate plural form, it is not formed by adding an *–n* but by other forms of suffixation or suppletion. The pronoun in *met wier instemming* is a pronoun, as is does not include an inflectional *-e* and the plural form is not made by adding an *-n*. In short, this is not a case of a two-way difference between nominal and non-nominal, as in ANS-97, but of a three-way difference between prenominal (adjectival) determiners, nominal determiners and pronouns. To illustrate and clarify this further we apply the criteria to a concrete example, *schrijver dezes*. From the paraphrase (*'de schrijver van dit'*) it is apparent that only *dezes* is a genitive; it is therefore definitely not an adjectival determiner, but either a pronoun or a nominal determiner. The choice is determined by the morphological structure: seeing that it includes the inflectional *-e* and has a plural ending in *-n*, which only refers to people (*dezen*), it cannot be a pronoun, but a nominal determiner. The same applies to the genitive in *de 20ste dezer*. In some cases this can produce somewhat unexpected results. The demonstrative *deze* is a determiner, in both nominal as prenominal use, while *dit, dat* and *die* are determiners in prenominal use, but pronouns in nominal use: after all they do not contain the inflectional *-e* and do not have a plural ending in *-n*. At first view this may seem random, but on closer inspection it would seem that nominally used *dit, dat,* and *die* (as opposed to deze) still show a number of other characteristics of pronouns. Nominally used *dit* and *dat* have separate relative forms *(ditte, datte*), just like a number of other pronouns (*ikke, watte*), while such forms do not exist voor the nominal determiners. Here it is also the case that *het* (as opposed to *deze*) is used as a relative pronoun and as such is indisputably a pronoun.

07. NAAMVAL. The standard forms do not have a case suffix and are divided further into nominative and oblique. Nominative forms are those which are used as the subject of finite forms (*ik, jij, men*); oblique are forms can for example be used as the object of verbs and prepositions (*mij, jou, hem).* Pronouns which can be both oblique and nominative (*je, wie, iemand*) are given the value '*standaard'*. With determiners the difference between nominative and oblique is systematically neutralized. The genitive is marked by *-s* (*mijns inziens, zijns gelijke, wiens hoed, elkaars fiets, iemands auto)* or *-(e)r* (*dezer dagen, aller landen, wier hoed, wie uwer zonder zonde is, tot veler verbazing*). The dative is marked by *-(e)n* (*met dien verstande, te allen prijze) of -(e)r (te eniger tijd, te mijner ere*).

Note that the case distinction is also relevant for adverbial pronouns. *Hier, daar, waar, ergens, nergens* and *overal* have the value '*oblique'*, and *er* has the same value for locative and PP use, but '*nominatief*' with presentative use and '*genitief*' with quantative use. With POS tagging we differentiate quantative *er* (*NAAMVAL = genitief*) from the three other uses, but within the last two we make no additional differentiation between nominative and oblique (*NAAMVAL = standaard*).

As is evident from this overview, case plays a much more important role in the analysis of pronouns than in the analyisis of nouns, adjectives and numerals. This is also reflected in the declarations: while for the latter this feature is only given in specific cases (see D02, D06, D07 en D13), it is uniformly given to pronouns (14), therefore also to plural and uninflected forms.

17. STATUS. This feature is only given to pronouns and possessive determiners. In reduced forms we include all forms without a vowel *('t, z'n, d'r,*

...), the monosyllabic forms with schwa (*het, ze, er,...*) and the forms *zich, ie*. In the forms with emphasis we include pronouns with incorporated *-zelf* or *-lie(den*) and the forms *ikke, ditte, datte, watte*. All other forms are give the value '*vol'*.

18. PERSOON. As with *STATUS* this feature is given to pronouns and possessive determiners. This might at first view appear too broad, because the well-known 1-2-3 distinction is only relevant for personal, reflexive and possessive pronouns. The reason why the feature is given a broader scope is because in the third person a number of additional differentiations are made which are also relevant for the other pronouns. The differentiation between pronouns which demand a personal referent (*3p*) and those which demand an neutral referent (*3o*) is also important for interrogative, relative and indefinite pronouns: *wie, (n)iemand, iedereen* for example, demand a personal referent and *wat,(n)iets, alles* a neutral one. In addition, with pronouns with a personal referent a further differentiation is made between male *(3m*) and female referents (*3vr*), and that differentiation is also important for the interrogative and relative pronouns: the genitive forms *wiens* and *wier,* for example,—in the singular— demand a male and female referent respectively. Note that this is not about the morfosyntactic gender of a word, but about the natural sex of the referent. The differentiation can be illustrated by the contrast between *hij* and *hijzelf*. The first form is masculine and can have both a personal and a neutral referent; the second form is also masculine, but can only refer to a male person. Seeing as this classification according to the gender of the referent is more semantically pragmatic than morphosyntactic, it could be said that it does not belong in the CGN tagset; the reason why it is included is that in the pronominal system it plays such an important role that both ANS-97 and EAGLES make or recommend similar distinctions. The same applies to further differentiation within the second person, where we make the well-known differentiation between familiar forms (*je, jij, jou, jullie)* and polite forms (*u*); these are given the values '*2v'* and '*2b'* respectively. With the *gij (ge, gij, u*) forms, used mainly in Flanders the differentiation is neutralised and therefore the value '*2'* is given. For the first person no further subtypes are recognized. Finally, the generic value '*persoon*', is given to the reciprocal pronouns (*elkaar, mekaar, ...*), because their antecedent could be anybody.

04. GETAL. This refers to the number of the referent, i.e. to differentiate between reference to a singular entity (*hij, iemand, jouw*) and reference to a group of entities *(wij, hun, jullie*). The differentiation is relevant for pronouns and possessive pronouns, but not for the other determiners. Forming the plural is not characterized by suffixing, as in nouns, but by suppletion. With a number of pronouns the differentiation is neutralized (*zich, wie*); these get the generic value '*getal'*. This last feature explains the difference between *GETAL* and *GETAL-N*.

06. GENUS. In addition to the natural sex, which is treated as a subpartition of the third person, there is the morphosyntactic gender. This feature is only given to the singular standard forms of third person personal pronouns. The differentiation between masculine and feminine, which is ignored for nouns, is made here.

08. POSITIE. Determiners are mainly used prenominally (*deze tafel, welke kast*) and nominally (*heb je deze al, welke bedoel je, de zijne*). Postnominal use, as in *kindeke mijn,* is so exceptional that we make no allowances for it. Free use occurs mainly with the indefinite determiners *elk* and *ieder*, as in *ze*

hebben elk/ieder een appel gekregen, and with the determiners which precede the entire *NP*, as in *al de mensen, zulk een ellende, welk een dwaasheid*. Note that this last position is not prenominal, because the determiners show no morphological variation here (and no similarity with the noun). The treatment as free determiner is more suitable, because the *pre-NP* position is typical for adverbally used elements, cf. *zelfs een boswachter, ook de mannen, alleen de kinderen, precies die vlinder.*

09. INFLECTIE . Many determiners have two forms, cf. *welk(e), ieder(e), zulk(e), al(le), ons/onze*.Just like the non-finite verbs, there are no *-s* forms.

10. GETAL-N. With nominal use some determiners take a plural suffix *(dezen, sommigen, enkelen, allen, de zijnen).* As with adjectives the limitation is valid for *personal* referents. Seeing as the feature is only given to nominally used determiners, and therefore not to pronouns, there is only one type of pronoun which has both a *GETAL* and a *GETAL-N* feature especially the nominally used possessive pronouns. We can show that it is useful to have both features with a form such as *de zijnen*: this has '*enkelvoud'* as value voor *GETAL* and '*meervoud-n'* as value for *GETAL-N.*

19. NPAGR. With prenominal use some determiners demand a singular neutral noun (*dit, dat, welk, elk, ieder),* others a singular gendered noun (*elke, iedere*) and others a singular gendered or a plural noun (*deze, die, welke).* So as is the case with *GETAL* and *GETAL-N, NPAGR* has number differentiation. This need not to lead to confusion, as *NPAGR* and *GETAL-N* never occur together; the only pronouns which have both a *GETAL* and a *NPAGR* feature are the prenominally used possessive pronouns. The usefulness of this is shown by a combination such as *ons huis,* in which *ons* has the value '*meervoud'* for *GETAL* and the value '*enkelvoud onzijdig'* for *NPAGR*. A complicating factor is the presence of traces of a three-gender system. Such traces are mainly found in fixed expressions and archaic language use. The dative form in *met dien verstande,* for example, demands a masculine or neutral singular, while the pronoun in *in dier voege* demands a feminine singular or a plural noun. In general the genitive and dative forms follow the old 3 gender system, while the standard forms follow the current 2 gender system.

05. GRAAD. This feature is only given to degreeable determiners. In addition to the base forms *veel, min, weinig,* there are also the comparatives *meer, minder,* the superlatives *meest(e), minst(e)* and the diminutive *minnetjes*.

2.5.4 Implications

[I10] <VWTYPE = persoonlijk> ⇒ <PDTYPE = pronomen>

[I11] <VWTYPE = reflexief> ⇒ <PDTYPE = pronomen, NAAMVAL = oblique>

[I12] <VWTYPE = reciprook> ⇒ <PDTYPE = pronomen, NAAMVAL ≠ nominatief, STATUS = vol, GETAL = meervoud>

[I13] <VWTYPE = bezittelijk> ⇒ <PDTYPE = determiner, POSITIE ≠ vrij>

[I14] <VWTYPE = vragend, PDTYPE = pronomen> ⇒ <STATUS ≠ gereduceerd>

[I15] <VWTYPE = betrekkelijk, PDTYPE = pronomen> ⇒ <STATUS = vol> [I16] <VWTYPE = exclamatief, PDTYPE = pronomen> ⇒ <STATUS = vol> [I06] <BUIGING = met-e> ⇒ <POSITIE = (pre)nominaal>

[I17] <PDTYPE = determiner, POSITIE = prenominaal, NAAMVAL = standaard> ⇒ <NPAGR = agr>

[I18] <PDTYPE = determiner, POSITIE = prenominaal, NAAMVAL = bijzonder> ⇒ <NPAGR = agr3>

The first seven implications refer to specific classes of pronouns. The eighth also applies to adjectives and participles. The last two apply to the prenominal determiners, and in fact also to articles, but because these are classed as a different word type, they have their own implications, see 2.6.

2.5.5 The tags

As the number of combinations is particularly large, we limit ourselves here to an overview in which only the values for VWTYPE, PDTYPE, NAAMVAL and POSITIE are specified. We give the other features a generic value, except when the implications facilitate the allocation of a more specific value. This is how we arrive at a classification of types, which each subsume one or more specific tags. With each type we state how many tags are subsumed by it. A complete overview of the 188 individual tags can be found in the appendix. As this section is about kinds of tags, there is no point in making a differentiation between T- and U - tags. At the end of the section we do state for each feature in which case the use of underspecified values is permitted.

Personal, reflexive and reciprocal pronouns. These pronouns are always pronomina. Their tags therefore, in addition to VW- TYPE, PDTYPE and NAAMVAL contain the feature STATUS, PERSOON and GETAL; for the standard forms of the third person singular personal pronouns the form GENUS is also added. In the third person personal pronouns we also include '*men*' and '*het*'. The quantative '*er*' on the other hand is included in the indefinite pronouns and the other uses of '*er*' in the demonstrative pronouns. We distinguish 8 types which in total subsume 54 tags.

[501-22]	VNW(pers,pron,nomin,status,persoon,getal(,genus))	ik, we, wijzelf, ikke, jij, gij, ge, men, hij, ie
[502-9]	VNW(pers,pron,obl,status,persoon,getal(,genus))	jou, hen, hem, 'm, haar
[503-4]	VNW(pers,pron,stan,status,persoon,getal(,genus))	het, ze, jullie
[504-6]	VNW(pers,pron,gen,vol,persoon,getal)	wie uwer, mijns gelijke
[505-9]	VNW(pr,pron,obl,status,persoon,getal)	me, mij, ons, mezelf
[506-2]	VNW(refl,pron,obl,status,3,getal)	zich, zichzelf
[507-1]	VNW(recip,pron,obl,vol,persoon,mv)	elkaar, mekaar, elkander

The possessive pronouns.

Possessive pronouns are determiners, and therefore, in addition to *VWTYPE*, *PDTYPE* and *NAAMVAL*, have the features *POSITIE* and *BUIGING*. Furthermore, prenominal determiners also have *NPAGR* and nominal determiners *GETAL-N*. Typical of the possessive determiners is that they also have the features which are characteristic of the pronouns, i.e.. *STATUS*, *PERSOON* and *GETAL*. We distinguish 5 types (three prenominal and two nominal) which in total subsume 63 tags; postnominal use, as in *kindeke mijn*, and free use, as in *dat is mijn*, are so exceptional that no tags are provided.

[509-17]	VNW(bez,det,stan,status,persoon,getal,prenom,buiging,agr)	mijn paard, mijne heren, m'n kapsel, je hoed
[510-13]	VNW(bez,det,gen,vol,persoon,getal,prenom,buiging,agr)	mijns inziens, een mijner vrienden
[511-13]	VNW(bez,det,dat,vol,persoon,getal,prenom,met-e,agr)	te mijnen huize, te mijner ere
[512-14]	VNW(bez,det,stan,vol,persoon,getal,nom,met-e,getal-n)	de mijne, de zijnen
[513-6]	VNW(bez,det,dat,vol,persoon,getal,nom,met-e,getal-n)	ten onzent

The interrogative, relative and exclamative pronouns.

Most of these pronouns are pronomina. The *VWTYPE* value is in many cases the intermediate '*vb*': after all there is only one interrogative pronoun which is not also used as a relative pronoun (*watte*). The value for *NAAMVAL* is the intermediate '*standaard'*, except for the inherently oblique adverbial pronomen (*waar*). As these pronouns are topicalized, and reduced forms,cannot be topicalized, the *STATUS* value is '*vol'* or '*nadruk'*. The value for *PERSOON* is '*3*' or a subtype thereof, except for the relative *die*, cf. *ik die hier zo lang gewerkt heb, ...* Common to these pronouns is that they are topicalized. The interrogative pronouns occur in both principal and subordinate clauses, the relative pronouns only occur in subordinate clauses and the exclamatory pronouns only in principal clauses. The *STATUS* value of the interrogative pronouns is '*vol'* or '*nadruk'*, but not '*gereduceerd'*; that of the relative and exclamatory pronouns is always '*vol'*.

[514-1]	VNW(vrag,pron,stan,nadr,3o,ev)	watte
[515-2]	VNW(betr,pron,stan,vol,persoon,getal)	de man die daar staat, het kind dat je daar ziet
[516-2]	VNW(betr,pron,gen,vol,3o,getal)	het warenhuis welks directeur hem een baan had aangeboden
[517-2]	VNW(vb,pron,stan,vol,3,getal)	wie gaat er mee, wat ik niet begrijp is
[518-3]	VNW(vb,pron,gen,vol,3p,getal)	wiens hoed is dit, de vrouw wier hoed daar hangt
[519-1]	VNW(vb,adv-pron,obl,vol,3o,getal)	waar ga je naartoe, de trein waar we op staan te wachten
[520-1]	VNW(excl,pron,stan,vol,3,getal)	wat een dwaasheid, wat kan jij liegen zeg

The only determiner in this class is *welk(e)*. It is mostly used as an interrogative pronoun, but sporadically also as a relative or exclamatory pronoun.

[521-2]	VNW(vb,det,stan,prenom,buiging,agr)	welke stoel, welk kind
[522-1]	VNW(vb,det,stan,nom,met-e,getal-n)	welke vind jij de mooiste, de procedures welke bij zo'n gelegenheid gevolgd worden
[515-2]	VNW(betr,det,stan,nom,buiging,getal-n)	hetgeen ik wil zeggen
[523-1]	VNW(excl,det,stan,vrij,zonder)	welk een dwaasheid

We therefore distinguish 11 types which together subsume 18 tags.

The demonstrative pronouns.

This class covers both pronouns and determiners. We distinguish 11 types which subsume a total of 19 tags. The determiners are usually used prenominally or nominally, but there are also cases of free use. The pronouns, in which we also include the adverbial pronouns *hier, daar, d'r* and the non-quantative *er*, are always in the third person. For more information on the differentiation between pronouns and nominal determiners, see section B, under *PDTYPE*.

[524-3]	VNW(aanw,pron,stan,status,3,getal)	dat(te), dit(te), die
[525-2]	VNW(aanw,pron,gen,vol,3,ev)	diens voorkeur, en dies meer
[526-1]	VNW(aanw,adv-pron,obl,status,3o,getal)	hier, daar, d'r
[527-1]	VNW(aanw,adv-pron,stan,red,3,getal)	het niet-quantative 'er'
[528-4]	VNW(aanw,det,stan,prenom,inflection ,npagr)	dat boek
[529-1]	VNW(aanw,det,gen,prenom,met-e,rest3)	een dezer dagen, de notulen dier vergadering
[530-2]	VNW(aanw,det,dat,prenom,met-e,agr3)	te dien tijde, in dier voege
[531-2]	VNW(aanw,det,stan,nom,met-e,getal-n)	deze(n), gene(n), degene
[532-1]	VNW(aanw,det,gen,nom,met-e,getal-n)	schrijver dezes, de twintigste dezer
[533-1]	VNW(aanw,det,dat,nom,met-e,getal-n)	dat is dan bij dezen beslist
[534-1]	VNW(aanw,det,stan,free,zonder)	zulk een vreemde gedachte

As in ANS-97 we do not include *soortgelijk(e), dergelijk(e), zodanig(e)* and *dus- danig(e)* in the demonstrative pronouns, but in the adjectives. The same applies to compounds with *-zelfde(n).*

The indefinite pronouns.

This class also includes both pronouns and determiners. There are 14 types which together subsume 34 tags. The pronouns, which also cover the adverbial pronouns *(n)ergens, overal* and quantative *er*, are all in the third person. Indefinite pronouns have the same variation as Demonstrative pronouns. On the one hand, there are prenominal, nominal and free determiners, and on the other pronouns, including the adverbial pronouns *overal* and (n)*ergens*. In the latter, we also include quantative *er*. We do not include the different forms of *veel, weinig, beide* in the numerals, but in the degreeable determiners. Note, however, that *zoveel, evenveel, hoeveel* are numerals as they do have corresponding ordinal numbers.

[535-2]	VNW(onbep,pron,stan,vol,3,ev)	(n)iets, (n)iemand, iedereen, alles, wat snoep
[536-1]	VNW(onbep,pron,gen,vol,3p,ev)	andermans, (n)iemands, ieders
[537-1]	VNW(onbep,adv-pron,obl,vol,3o,getal)	(n)ergens, overal [538-1] VNW(onbep,adv-pron,gen,red,3,getal) het quantative 'er'

Indefinite determiners also include the degreeable (*PDTYPE = graad),* which, in addition to the typical determiner features, also have a *GRAAD* feature.

[539-6]	VNW(onbep,det,stan,prenom,inflection ,agr)	beide mannen, elk kind, iedere keer
[540-1]	VNW(onbep,det,gen,prenom,met-e,mv)	proletari¨ers aller landen
[541-2]	VNW(onbep,det,dat,prenom,met-e,agr3)	te allen prijze, te eniger tijd
[542-6]	VNW(onbep,grad,stan,prenom,inflection ,agr,graad)	veel plezier, vele uren, minder werk, de meeste mensen

[543-3]	VNW(onbep,det,stan,nom,met-e,getal-n)	allen zijn tevreden, sommigen zijn gevlucht
[544-1]	VNW(onbep,det,gen,nom,met-e,getal-n)	met aller instemming, tot beider verbazing
[545-5]	VNW(onbep,grad,stan,nom,met-e,getal-n,graad)	velen zijn geroepen, weinigen zijn uitverkoren
[546-1]	VNW(onbep,grad,gen,nom,met-e,getal-n,graad)	tot veler verbazing
[547-1]	VNW(onbep,det,stan,free,zonder)	ze kregen elk/ieder/allebei een bal, al die mensen
[548-3]	VNW(onbep,grad,stan,free,zonder,graad)	minder werken, meer slapen, dat is te weinig, veel groter

Genoeg, *zat*, *allerhande* are not classed as pronouns, but adverbs (after all, pronouns are never used postnominally) and *ietsje*, *(een) weinigje* are classed as nouns (after all, the diminutive forms of pronouns end in *-jes*, not *-je*).

Underspecified combinations.

As with nouns we also allow a limited degree of underspecification. The features this can apply to are as follows:

[P15] VWTYPE = pr (persoonlijk, reflexief) ,reciprook, bezittelijk, vb (vragend, betrekkelijk), exclamatief, aanwijzend, onbepaald.
[P07] NAAMVAL = standaard (nominatief, oblique), genitief, datief.
[P18] PERSOON = persoon (1, 2 (2v, 2b), 3 (3p (3m, 3v), 3o)).
[P04] GETAL = getal (enkelvoud, meervoud).
[P19] NPAGR = agr (evon, rest (evz, mv)), agr3 (evmo, rest3 (evf, mv)).

15. VWTYPE. The intermediate values '*pr*' and '*vb*' are given when the form of the pronoun does not show whether it is personal or reflexive or alternatively interrogative or relative. The forms to which this could apply are those which have been given an intermediate value in the function word lexicon.

07. NAAMVAL. The intermediate value '*standaard*' is used when the form of the pronoun does not show whether it is nominative or oblique. This applies to all determiners and those pronouns for which the nominative/oblique differentiation has been neutralized. See the lexicon for a list

18. PERSOON. The intermediate values '*2*', '*3*' and '*3p*' are given to pronouns which do not have more specific values in the lexicon.

04. GETAL. The generic value '*getal*' is given to the pronouns which do not have a specific value in the lexicon. This includes for example the adverbial pronouns, the reflexive *zich(zelf)* and the *u* forms.

19. NPAGR. The intermediate values '*agr*', '*agr3*', '*rest*' and '*rest3*' are given to prenominal determiners which do not have a more specific value in the lexicon

2.5.6 Lemmatization

The lemma is identified by the uninflected standard form. Inflectional affixes are removed, except where the base form cannot be used as a stand-alone word, as in

the genitive form *andermans*. With suppletion, such as the nominative *ik* versus the oblique *mij*, both forms are given a different lemma. The reduced forms without the vowel, such as *'t* and *z'n*, are given the corresponding full form as lemma; the reduced forms with the vowel on the other hand are given their own lemma: the lemma value for *me* is not therefore '*mij'*, but '*me'*. If in doubt, check the lexicon.

2.6 ARTICLES

The articles are actually determiners, and can be described in terms of the same features. However, because in Dutch grammars it is more usual to treat the articles as a separate word type (see for example ANS-97), CGN sticks to this tradition.

2.6.1 Demarcation

The definite article *het* has to be differentiated from the homonymic personal pronoun, and the indefinite *een* has to be differentiated from the numeral and the indefinite determiner; the latter is simple because there is clearly a difference in pronunciation (schwa versus a clear *ee*), where we can assume that this is shown in the transcription by the addition or ommission of accents. The genitive form *des* must be differentiated from the homonymic adverb in constructions such as *des te beter*.

2.6.2 Declarations

[D21] <POS = lidwoord> ⇒ <LWTYPE, NAAMVAL, NPAGR>

There is no feature for *BUIGING*, because — in standard language — no differentiation is made between inflected and uninflected articles. Please refer to 2.11.on how to treat the dialectic form in *ne vent*

2.6.3 Partitions

[P20] LWTYPE = bepaald, onbepaald.
[P07] NAAMVAL = standaard, genitief, datief.
[P19] NPAGR = agr (evon, rest), agr3 (evmo, rest3 (evf, mv)).

20. LWTYPE. The exclamative *een* is regarded as '*indefinite'*, as it is in ANS-97.

07. NAAMVAL. In addition to the standard form, the definite article *de* also has the genitive forms *des, der* and *'s (des duivels, der Nederlandse taal,'s avonds*), and the dative forms *der* and *den*; the latter two only occur in fixed expressions, such as in *der minne, op den duur, uit den boze.*

19. NPAGR. *Het* demands a singular neutral noun ('*evon'*) and *de* a singular non-neutral or plural noun ('*rest'*). The indefinite article gives no restrictions cf. *een paard, een man, een mensen dat er waren!*

A complicating factor is the presence of traces of a three-gender system ('*agr3'*). Such traces are mainly found in fixed expressions and archaic language. The genitive form, for example, demands a masculine or neutral singular ('*evmo'*), while *der* demands a femine singular or a plural noun ('*rest3'*).

2.6.4 Implications

[I19] <POS = lidwoord, NAAMVAL = standaard> ⇒ <NPAGR = agr>
[I20] <POS = lidwoord, NAAMVAL = bijzonder> ⇒ <NPAGR = agr3>

The genitive and the dative follow the old three-gender system, while the standard forms follow the current two-gender system.

2.6.5 The tags

As with the pronouns, a comprehensive list of the combinations for the articles varies little from what is offered in the lexicon. However, because the number of articles is so small, we can give the full list here. There are nine combinations in total, including two underspecified.

[T601]LID(bep,stan,evon)	het kind, in ’t geniep
[T602]LID(bep,stan,rest)	de hond(en), de kinderen
[T603]LID(bep,gen,evmo)	des duivels, de plaats des onheils, ’s avonds, ’s maandags
[U604]LID(bep,gen,rest3)	der Nederlandse taal, der Belgen
[T605]LID(bep,dat,evmo)	op den duur, om den brode, uit den boze
[606] LID(bep,dat,evf)	in der minne
[T607]LID(bep,dat,mv)	die in den hemelen zijt
[U608]LID(onbep,stan,agr)	een kind, een mensen dat er waren
[T609]LID(onbep,gen,evf)	de kracht ener vrouw

We know of no examples of masculine or neutral genitive forms for the indefinite article. ANS-97 does mention *eens geestes zijn*, but this is a genitive of the numeral *één*.

2.6.6 Lemmatization

The lemmas for the articles are ‘*de’*, ‘*het’* and ‘*een’*. The indefinite ‘*’n’* is ommitted, because the protocol for orthographic transcription dictates the spelling ‘*een’*; the ‘*’n’* form is only allowed in the combination *zo’n*, but this is generally treated as a whole and as such as a demonstrative pronoun.

2.7 PREPOSITIONS

2.7.1 Demarcation

Prepositions usually take a complement. This can not only be an *NP*, but also a *PP*, an adverb, an adjective, a numeral or a verbal projection (*V, VP, S*). In the latter case they are often classed as conjuctions, which results in systematic ambiguity for words such as *tot, sedert, sinds, voor, na, naar, zonder, met, door, om*. CGN does not adhere to this use and always treats such words as prepositions. We also include *te* which introduces an infinitive, as a preposition, as we do with *aan* in *aan het vissen zijn*, *op* in *op springen staan* and *uit* in *uit vissen gaan*.

Where the complement is an *NP* or a *PP*, it can precede the preposition, as in *loopt overal tegen, rijdt de berg op, de hele dag door, onder de brug door*. In such cases the complement can also be extrapolated, so that only the preposition is left behind, as in *waar denk je aan, rijdt die berg alleen op* and *door die muur kunnen we niet heen*. We only include those words with the postpositional prepositions which can be preceded by an adverbial pronoun; *heen* and *af* are therefore classed as prepositions, but *terug, weg* and *geleden* are not *(*erterug,*

**hierweg, *waargeleden)*; the latter are classed as adverbs. Prepositions can also be used without a complement, in the same way as many transitive verbs can also be used as intransitive. This is the case for example for predicatively used prepositions (*het bier is op, het licht is aan, hij is vroeg op*) and for adverbally used prepositions, such as *boven* in *naar boven gaan*. In intransitively used prepositions we also include separable parts of verbs which have the same form as a preposition (*belt ... op, geeft ... uit*). This avoids creating systematic ambiguity between the preposition *op* and the homonymic particle or adverb. Note: the non-verbal part of a separable compound verb can also be a noun (*haal diep adem*), an adjective (*doe de glazen nog eens vol*) or an adverb (*we komen morgen samen*). With prepositions derived from foreign languages we differentiate between those which are also combined with a Dutch complement, such as *per trein* and *drie à vier glazen*, and those which can only be combined with a complement from the same foreign language, such as *ad hoc* and *en profil*. The first are simply classed as prepositions; the latter are given the tag *SPEC(vreemd).* This mainly concerns expressions derived from Latin (*ab ovo, ad fundum, cum laude, ex machina, ex voto, intra muros, inter alia, post mortem, pro deo, pro domo, salva veritate*) and Italian (*con amore, con brio, sotto voce).*

2.7.2 Declarations

[D22] <POS = voorzetsel> ⇒ <VZTYPE>

2.7.3 Partitions

[P21] VZTYPE = initieel (versmolten), finaal.

21. VZTYPE. Prepositions which precede their complement are given the value '*initieel'*. That complement can be an *NP* and a *PP*, an adverb, an adjective, a numeral or a verbal projection (*V, VP of S*). Prepositions which follow their complement are given the value '*finaal'*; the preceding complement is often an adverbial pronoun *(hij zit er net naast*), but it can also be an *NP* (*de trap af)* or a *PP* (*van het dak af, door de eeuwen heen*). An important difference between *initieel* and *finaal* use is that in the first case the complement must follow the preposition immediately (*modulo parenthese*), while in the second case, the complement can be separated from the preposition by other parts of the sentence: this occurs mainly with adverbial pronouns *([daar] denk ik niet eens [aan]),* but also with *PPs ([door die muur] kom je met dat boortje niet [heen]); f*or more examples of the latter see ANS-97, 509-510. Prepositions without a complement or intransitive prepositions are amongst other things used as the non-verbal part of a separable compound verb *(belt zijn dochter op),* as the complement of another preposition *(naar boven/binnen gaan),* the core of a predicate (*op zijn, binnen zijn, in zijn)* or as the core of a *VP* determiner (*binnen spelen, niet zonder kunnen*). As the group of prepositions which can be used intransitively is a subset of the prepositions which can be used *finaal*, in tagging they are given the value '*finaal'*. In terms of the *VZTYPE* differentiation we can distinguish three classes of prepositions: those which can only be used *initieel*, such as *per, sinds, sedert, te;* those which can only be used *finaal*, such as *af, heen, vandaan*; and those which can occur in both types of positions: *aan* for example is *initieel* in *aan de wand, aan het vissen zijn, aan het zeuren gaan* and *finaal* in *achter de stoet aan, tegen de veertig aan, kondigt een vertraging aan, heeft een zwarte rok aan*. A few prepositions have different forms for initial and final use, such as *met/me(d)e* and *tot/toe.*

Within the inherently initial prepositions we differentiate between a separate class of amalgamated prepositions, i.e. prepositions which form a single entity

with a definite article. In German, French and Italian there are various of such prepositions; Dutch only has the forms *ter* and *ten*. With the exception of the two amalgamated prepositions, the Dutch prepositions are morphologically invariable. Forms such as *voorste, achterste, onderste, bovenste, benedenste, binnenste en buitenste* are not superlative forms of prepositions but adjectives. Something similar applies to the forms *onderen, voren, achteren*; these are not inflected prepositions but adverbs.

2.7.4 Implications

There are no implications.

2.7.5 The tags

There are three possible combinations.

[T701]	VZ(init)	met een lepeltje, met Jan in het hospitaal, met zo te roepen
[T702]	VZ(fin)	liep de trap af, bij de beesten af, speelt het bandje af, kletsen flink wat af
[T703]	VZ(versm)	ten strijde, ten hoogste, ter plaatse

There is no room for underspecification. When tagging a specific value must always be given for *VZTYPE*.

2.7.6 Lemmatization

For prepositions the lemma is identical to word form, except in the case of the two amalgamated prepositions: *ter, ten* are given '*te*' as their lemma.

2.8 CONJUCTIONS

2.8.1 Demarcation

Conjuctions include two classes of words with very different characteristics: coordinate and subordinate conjunctions. Coordinate conjuctions can introduce both sentences and small groups of words and even parts of words. When they introduce a sentence, this can take any order: *V-1, V-2* or *V-finaal*. Coordinate conjuctions form a small closed class: this includes amongst others *en, of, ofwel, noch, maar, want, hetzij*. See the lexicon for a complete list. Subordinate conjuctions can be split into three groups: (1) the complementizers *dat, of, als* and *dan*; (2) combinations of a preposition and a complementizer *(alsof, doordat, nadat, omdat, opdat, totdat, voordat*) or of an adverb and a complementizer *(eerdat, zodat*); (3) the rest ((*ter)wijl, (al)hoewel, (voor)aleer, alvorens, tenzij, zodra, ...*); see the lexicon for a longer list.

A typical characteristic of complementizers is, that they can be preceded by an initial clause as in *leuk dat het was, wie of er gebeld heeft* and *rijk als ze was.* Note that all four complementizers are ambiguous: *dat* is also a demonstrative or relative pronoun, *of* is also a subordinate conjunction, as a preposition and then an adverb. What distinguishes the conjunction *dat* from the relative pronoun is that it does not have an argument function in the subordinate clause: it is neither a subject nor an object, but simply a prelude to the subordinate clause. Compare for example the pronoun in *het feit dat vaak over het hoofd wordt gezien* with the conjunction in *het feit dat dit vaak over het hoofd gezien wordt*. Seeing that the pronoun has to agree with its antecedent in both gender and count, it only occurs

in combination with singular neutral *NPs*, while the conjunction is not subject to this restriction, cf. *de geruchten dat de volgende Paus een Italiaan moet zijn* and *de verwachting dat ze wel zal komen*.

Subordinate conjuctions mainly introduce a subordinate clause. In this case the sentence must follow the typical subordinate clauses word order *(V-final*). This word order criterion can be used to differentiate the conjuctions from the adverbs. Compare for example the conjunction in *wanneer we naar Milaan gaan* with the adverb in *wanneer gaan we naar Milaan?*. We can differentiate the conjunctions *toen, nu* and *dan* from the homonymic adverbs in the same way; *dan* for example is an adverb in *dan gaan we naar Milaan* and in *als dit een droom is, dan word ik liefst niet wakker*,seeing as the clauses which it precedes do not display the typical subordinate clause word order. For the same reason the concessive *al in al is de leugen nog zo snel* is not a conjunction, rather an adverb. The word order criterion can also be used to differentiate the causal conjunction *daar* from the homonymic adverbial pronoun.

Not all words which introduce a verbal projection with the word order of subordinate clauses are classed as conjuctions. If the word is identical to a preposition as far as its form is concerned, such as *voor, na, naar, om, tot, sedert, sinds, zonder,* we do not regard it as ambiguous (either preposition or conjunction) but always as a preposition, see 2.7. A rare case of ambiguity between conjunction and preposition is *als*: this is a conjunction when it introduces a conditional subordinate clause or the second part of a comparision, but not when it introduces a counterfactual conditional statement, as in *als was het een droom*. Subordinate conjuctions can—as their coordinate counterparts—also introduce smaller word groups; this applies amongst others to the use of *als* and *dan* as preludes to the second part of a comparison, cf. *rijker dan Bill* and *zo groot als jij* ; other examples are *behalve* and *hoewel* in a combination such as *hoewel vlijtig en verstandig vond ie geen baan.*

2.8.2 Declarations

[D23] <POS = voegwoord> ⇒ <CONJTYPE>

2.8.3 Partitions

[P22] CONJTYPE = nevenschikkend, onderschikkend.

22. CONJTYPE. Differentiation between coordinate and subordinate conjuctions is not a semantic but a purely syntactic differentiation. The causal *omdat* for example belongs to the subordinate conjuctions, because it demands the word order of a subordinate clause, while the quasi-synonym *want* belongs to the coordinate conjunctions, as it demands the typical word order of a principal clause *(V-2*). The only conjunction which can be both coordinate and subordinate is *of.*

2.8.4 Implications

There are no implications.

2.8.5 The tags

The number of combinations is limited to two.

[T801] VG(neven) Jan en Peter; en toen gebeurde het

[T802] VG(onder) ze komt niet, omdat ze zich niet goed voelt

2.8.6 Lemmatization

The lemma is identical to the word form.

2.9 ADVERBS

The adverbs form a heterogeneous class, even more so than the pronouns. In order to determine which words belong to the class, we use form based rather than functional criteria. A word which forms the core of an adverbial determiner, is not necessarily an adverb. The temporal clause in *we gaan zondag naar Milaan* for example is not an adverb but a noun, and the mood determiner in *hij praat snel* is not treated as an adverb either, but as an adverbially used adjective (<*POS = adjectief, POSITIE = vrij*> see 2.2). In order to differentiate adverbially used adjectives from adverbs, we use the following criterion: when the adverbial form can also be used prenominally with the same or similar meaning, it is an adjective; otherwise it is an adverb. The form *vrij* in *je kan hier vrij rondlopen* is therefore an adverbially used adjective, while the same form in *een vrij warme dag* is an adverb. As the CGN tagset does not only allow for the fact that adjectives can have an adverbial use, but also participles, numerals and determiners, the heterogeneity of the adverb class is somewhat contained, but even then, their number and diversity are still large. To show the diversity we have taken an inductive approach: using a list of words which are classed in CELEX as adverbs (approx. 850 words) we arrived at a morphologically based classification. Besides the rather large group of (1) free adverbs *(nu, niet, nog, al, hoe, ...)* we distinguish various types of adverbs which are joined – according to form: (2) those with an adverbial suffix (*stomweg, beroepshalve, derwaarts*), (3) those with an adverbial core (*welnu, hierzo*), (4) those with a prepositional core (*tussenin, bovenaan*), (5) those with an incorporated pronominal complement (*daarin, erop, waar- over, desondanks, bovendien*), (6) those with a nominal core (*uitermate, binnens- huis, bergaf*), (7) those with an adjectival core (*stilaan, voluit*), and (8) those with a verbal core (*ongetwijfeld, welteverstaan*). Finally, there is the separate group of (9) foreign words (*incognito, sowieso, normaliter*). For a more complete list and the other subclassifications, see the function word lexicon. Of the adverbs included in the lexicon only a very small proportion shows morphological variation: there are the diminutive forms *strakjes, saampjes, eventjes, ...* and there are the comparative and superlative forms of amongst others *graag*. From this we could decide to enter a *GRAAD* feature for the adverbs, too, but it is doubtful whether this is worth it, because in addition to the fact that the differentiation is only relevant for a very small number of the adverbs, there is also the fact that the non-base forms are lexicalized. *Liever* is a compositional interpretable comparative form of the adjective *lief*, but as a comparative of the adverb *graag* it is actually a unique form. The same applies to the superlative forms in *ze komt liefst alleen, hoogst verleidelijk, eerst doe je dit en dan dat* and *laatst zag ik een merkwaardig tafereel*. This is why we choose to treat the (rare) diminutive, comparative and superlative forms of the adverbs as separate lemmas. Another potential exception to the morphological invariability of the adverbs are the inflectional endings of the intensifiers *heel* and *erg* in combinations such as *een hele lange tafel* and *een erge leuke vakantie*. We do not believe it would be appropriate to introduce a *BUIGING* feature for adverbs in order to treat these two isolated and in addition, not entirely grammatical forms. Instead we treat both forms as prenominally used adjectives. Note, incidentally, that they – in somewhat different meanings— are also used as such, cf. *een hele dag* and *erge pijnen.*

As a result of these two interventions, we can indeed include morphological

invariability as a typical characteristic of adverbs. Seeing as adverbs only get a *POS* feature, there are no implications and there is only one combination.

[T901] BIJW() gisteren, nu, niet, nog, al, hoe

The lemma is identical to the word form, except in the case of truncated forms such as *'ns* and *d'rover;* like all truncated forms, these are reduced to a form without an apostrophe, in this case *eens* and *erover*.

2.10 INTERJECTIONS

In interjections we include those words which can generally be used as a stand-alone linguistic utterance. In line with ANS-97 we distinguish three types: (1) onomatopoeia, such as *kukeleku*; (2) expressions of the speaker's emotions, such as expressions of pain, amazement, frustration, etc.; here we also include swearing and cursing; (3) terms of social exchange, such as greeting, thanking, apologizing, etc. Not all words which can be used as a stand-alone linguistic utterrance are an interjection. Imperatives such as *bijt* and adverbs such as *weg* for example can be used as a stand-alone linguistic utterance, but are classed as verbs and adverbs respectively in the CGN corpus. This follows the general principle that the allocation of POS tags is based on form rather than functionally based criteria. For the same reason we class words such as *munt, kruis* and *Christus* as nouns, even when they are used as an exclamation. For a list of the words which we class as interjections see the CGN lexicon. Like the adverbs, the interjections are only given a *POS* feature. There are therefore no implications and there is only one combination.

[T001] TSW() oei, amai, uh, hoera, AUB

The lemma is identical to the word form.

2.11 DIALECT WORDS

In dialect words we include all words which have been given the word type classification *'*d'*. As the differentiations which apply to the standard language are not always relevant to dialect words, we only allocate *POS* and *XTYPE* features to these words. To differentiate them from the other words, we also include in their tags the value *'dial(ectisch)*'. There are a total of 28 combinations. We give them an *R* number (for regional), because the D numbers are already used for the Declarations.

[R101]	N(soort,dial)	bompa*d, ne*d lange *frak*d*
[R102]	N(eigen,dial)	
[R201]	ADJ(dial)	ne*d *langen*d* toot*d
[R301]	WW(dial)	'k *zen*d* nie*d thuis, 'k *hem*d* gee*d geld
[R401]	TW(hoofd,dial)	
[R402]	TW(rang,dial)	den*d *elfste*d*
[R501]	VNW(pers,pron,dial)	kom *de*d* gij mee, 'k heb *ulie*d* gezien
[R502]	VNW(refl,pron,dial)	
[R503]	VNW(recip,pron,dial)	we zien *malkanderen*d* niet veel
[R504]	VNW(bez,det,dial)	hij heeft *z'ne*d* frak*d vergeten
[R505]	VNW(vrag,pron,dial)	
[R506]	VNW(vrag,det,dial)	
[R507]	VNW(betr,pron,dial)	

[R508]	VNW(betr,det,dial)	
[R509]	VNW(excl,pron,dial)	
[R510]	VNW(excl,det,dial)	
[R511]	VNW(aanw,pron,dial)	
[R512]	VNW(aanw,det,dial)	*diejen*d* boek, *dees*d* week
[R513]	VNW(onbep,pron,dial)	z' have *iet*d* gezien
[R514]	VNW(onbep,det,dial)	ze kan *elken*d* dag vertrekken
[R601]	LID(bep,dial)	het gevecht met *den*d* beer
[R602]	LID(onbep,dial)	*nen*d* toffe gast, *ne*d* vieze vent
[R701]	VZ(init,dial)	*me*d* veel geduld
[R702]	VZ(fin,dial)	
[R801]	VG(neven,dial)	
[R802]	VG(onder,dial)	't schijnt *da*d* we mogen komen
[R901]	BW(dial)	*efkes*d, nie*d*
[R001]	TSW(dial)	*neeje*d, wablieft*d*

The lemma is identical to the word form.

2.12 SPECIAL SIGNS

This group includes signs which cannot be included in the normal word types. They do not have a *POS* feature, but they do have a *SPECTYPE* feature.

[D24] <TOKENTYPE = speciaal> ⇒ <SPECTYPE>

[P23] SPECTYPE = afgebroken, onverstaanbaar, vreemd, deeleigen, meta, commentaar, achtergrond.

23. SPECTYPE. In the broken off tokens we include those which have been given the word type classification *a, as well as the signs which are followed or preceded by a hyphen, as with the first sign in *binnen- en buitenland* and the last in *regeringsvoorstellen en -beslissingen*. In incomprehensible tokens we include those which have been given the word type classifcation *ggg* (non-verbal utterance), *xxx* (incomprehensible) or Xxx (incomprehensible name).

In foreign tokens we include those which have been given the word type classification **v*. This value is also allocated to foreign words which are not morphosyntacally integrated into the Dutch system. With this we mean that they cannot be described in terms of the tags for the ten basic word types, for example because they make differentiations which do not belong to the Dutch system, such as the Latin ablative *anno*. The criterion for allocating *SPEC(vreemd*) is therefore not based on frequency of use, familiarity or conformity with the Dutch pronunciation but only on morphosyntactic integration. *SPEC(vreemd)* is therfore also given to parts of multiword expressions which are established as a whole, but whose individual words cannot be described in terms of the Dutch CGN tags, such as *ad hoc, wishful thinking, al dente, en profil, SVP, CQ*, etc.

The value '*deeleigen'* is given to parts of compound propoer nouns, both Dutch (*Den Haag, Piet De Zager, Hans Van Halteren*) and foreign (*Rio De Janeiro, Yom Kippur, Labour Day, Herald Tribune)*. The value '*comment'* is given to fragments which have been marked as such under word sort; these are bits of comments. The '*achtergrond'* is given to fragments which are marked as such under word sort; these are background sounds. The relevant tags are as follows.

[T002]	SPEC(afgebr)	uitge*a, binnen-
[T003]	SPEC(onverst)	ggg, xxx, Xxx

[T004]	SPEC(vreemd)	whatever*v, ad, hoc, wishful
[T005]	SPEC(deeleigen)	Den, Haag, New, York
[T006]	SPEC(meta)	(het woord) homosexueel
[T008]	SPEC(achter)	voor achtergrondgeluid
[T009]	SPEC(comment)	voor commentaren

To avoid problems with the screen alignment special signs are also given a lemma. The value for this is always an underscore ('_').

2.13 PUNCTUATION MARKS

The three punctuation marks used with the word types are the full stop, ellipsis, and the question mark. They are given the tag *LET().*

[T007] LET() ., ..., ?

In order to avoid problems with the screen alignment a lemma is also allocated to punctuation marks. The values are: for the full stop *('.')*, the question mark *('?')* and for ellipsis *('...').*

3 COMPARISON WITH EAGLES

In the EAGLES standard for tagsets a three way differentiation is made between (1) compulsory attributes and values, (2) recommended attributes and values, and (3) special extensions, which on the one hand comprise optional additions and on the other language specific additions. A separate section is devoted to each; in a fourth and final section we cover the CGN features which do not appear in the EAGLES recommendations.

3.1 COMPULSORY

This category covers only one attribute, namely *POS.* It has thirteen values, ten of which are identical to the ten word types which are distinguished in the CGN tagset. Of the other three, one corresponds to the special signs ('*Residual'*) and one to the punctuation marks ('*Punctuation'*). The thirteenth is called '*Unique'* and is intended for '*categories with a unique or very small membership*', such as the English '*infinitival to'*. CGN does not have an equivalent to '*unique'*: the expletive *er* and the infinitival *te*, which according to EAGLES could fall into this category are included in the ten existing word types, under pronouns and prepositions respectively.

3.2 RECOMMENDED

The recommended features are given for each word type.

NOUNS

The four recommended features (*Type, Gender, Number, Case*) are also included in the CGN tagset. They correspond to *NTYPE, GENUS, GETAL* and *NAAMVAL,* respectively.

ADJECTIVES

Three of the four recommended features (*Degree, Gender, Number, Case*) are also found in the CGN tagset. *Genus* is missing because Dutch adjectives, as opposed to the French, for example, are not marked for gender. Moreover, CGN only gives nominally used adjectives a number value.

VERBS

EAGLES mentions no fewer than eight recommended features. '*Finiteness'* corresponds to the differentiation between finite forms and non-finite forms, *(WVORM),* and '*Tense'* corresponds to the differentiation between present and past tense, (*PVTIJD*). There is no separate feature for '*Verb form/Mood'* in CGN, but the relevant distinctions are spread over *WVORM (infinitief, deelwoord)* and *PVTIJD (conjunctief*). '*Number'* is part of *PVAGR* for the finite forms and for *GETAL-N* for the (nominally used) non-finite forms. There is no equivalent for '*Person'* in the CGN tagset, but it is included in *PVAGR*. Of the three remaining features, two are not relevant for Dutch: '*Voice'* is not relevant because the Dutch passive is not morphologically marked, (as opposed to Greek and Danish), and '*Gender'* is not relevant because Dutch participles do not have variation in gender (as opposed to participles in the Romance languages).

Finally, '*Status'* covers the differentiation between principal and auxillary verbs. A CGN equivalent to this would be the differentiation between principal, auxillary and copulative verbs, but because to do this would require a complete syntactic analysis of the sentence, it has not been included in the tagset: after all morphological characteristics are not enough, because the auxillary and copulative verbs have the same morphological variation as the principal verbs, and lexical characteristics are equally insufficient because all auxillary and copulative verbs are homonyms which can be classed as principal verbs.

NUMERALS

EAGLES gives five recommended features. '*Type'* corresponds to *NUM- TYPE*, '*Case'* with *NAAMVAL* and '*Number'* with *GETAL-N*. '*Gender'* is also not relevant here. The feature '*Function'* with the values '*pronoun', 'determiner'* and '*adjective'* is intended '*to indicate the part-of-speech function of a word within the numeral category*'. However, what exactly is meant by this is not explained.

PRONOUNS

EAGLES gives eight recommended features. Two of these cover the differentiations we have grouped under *VWTYPE*, namely. *'Pronoun-Type' (demonstrative, indefinite, possessive, int/rel, pers/refl)* and '*Determiner-Type'(demonstrative, indefinite, possessive, int/rel, partitive).* A third feature, 'Category' *(pronoun, determiner both*) corresponds with *PDTYPE;* what EAGLES understands by '*both'* is not explained. Note that by separating *PDTYPE* and *VWTYPE* CGN the redundancy of the EAGLES suggestion is avoided. '*Case'* corresponds to *NAAMVAL* and '*Person'* to *PERSOON*. '*Number'* and '*Gender'* correspond to *GETAL* and *GENUS* for the pronoun and *NPAGR* for the prenominal determiners; in addition, for the nominal determiners *'Number'* corresponds to *GETAL-N*.

Finally, *'possessive'*, with the values 'singular' and 'plural', is relevant for the possessive pronouns: in *la nostra casa* for example *nostra* is both (1st person) plural and singular. In CGN this differentiation is made by giving possessive

pronouns both *GETAL* and *NPAGR* (*prenominaal)* or *GETAL-N (nominaal*).

ARTICLES

The four features recommended in EAGLES are also found in the CGN tagset: '*Article-Type'* corresponds to *LWTYPE*, '*Case'* to *NAAMVAL*, and '*Number'* and '*Gender'* to *NPAGR*.

PREPOSITIONS

EAGLES calls prepositions '*adpositions'* and gives only one feature ('*Type'*), with the value '*Preposition'*. This corresponds to the value '*initieel'* of the *VZTYPE* feature.

CONJUCTIONS

The feature '*Type'* with the values '*coordinating'* and '*subordinating'* corresponds to *CONJTYPE* and the values '*nevenschikkend'* and '*onderschikkend'*.

ADVERBS

EAGLES gives '*Degree' (positive, comparative, superlative*). Seeing as CGN treats adverbially used adjectives as adjectives and not as adverbs, there are very few adverbs for which the *GRAAD* variation is relevant. Moreover, the comparatives and the superlatives of such adverbs are often lexicalized. This is why we have we not included this feature in the CGN tagset, see also 2.9.

SPECIAL SIGNS

EAGLES recommends three features. The first (*Type*) has six possible values *(foreign word, formula, symbol, acronym, abbreviation, unclassified);* the latter value is used for incomplete words and '*pause fillers'*. The corresponding CGN feature is *SPECTYPE* with the values '*vreemd', 'afgebroken'* and '*onverstaanbaar'*; There is no CGN equivlaent for '*symbol'* and '*formula'*, because such signs do not occur in the transcribed spoken language: people do not say '*$*', but '*dollar'*; CGN does also not have an equivalent for '*acronym'* because it includes this with the nouns. This also does away with the need for the other two EAGLES features (*Gender* and *Number*).

3.3 OPTIONAL

As mentioned in the introduction, in this group an additional differentiation is made between '*application- or task specific extensions'* and '*language specific extensions'*. In this paragraph we call these *A-* and *L-extensions* respectively. They concern either the addition of new attributes or the addition of extra values to the recommended attributes.

The *L-extensions* are only mentioned in this paragraph where they are relevant to Dutch; therefore attributes or values where it is explicitly said that they are especially intended for Danish, English or another specific European language are not discussed.

NOUNS

The feature '*Countability'*, with the values '*mass'* and '*countable'* is the only *A-extension* mentioned in EAGLES. The CGN tagset does not contain an equivalent to this feature, because it concerns semantic differentiation.

ADJECTIVES

Under the *A-extensions* EAGLES gives the features '*Use'* with the values '*attributive'* and *'predicative'*, and *'NP Function'* with the values *'premodifying', 'postmodifying'* and *'head-function'*. These correspond—approximately—to the *POSITIE* feature: the '*NP Function'* values correspond to '*prenominaal', 'postnominaal'* and '*nominaal'*, respectively. The fourth *POSITIE* feature ('*vrij*) corresponds to '*predicative'*, but has a broader interpretation because it also covers adverbial use. This broadening is useful because in Dutch adverbially used adjectives have exactly the same form have as predicatively used adjectives; in that respect Dutch is different from most other European langauges, in which adverbial use is marked by a suffix, cf. the English -*ly*, the French -*ment* and the Italian -*mente*.

A third *A-extension* concerns the feature '*Inflection-type'* with the values '*weak-flection','strong-flection'*and *'mixed'*. It bears some resemblance to the *BUIGING* feature, but the values are so different that it is in fact a different feature.

VERBS

Under the *A-extensions* EAGLES gives four features. '*Aspect'* is relevant for Greek and other Romance languages, but not for Dutch. '*Separability'* is not included because allocating the values needs a full syntactic analysis; *wacht* for example is 'separable' in *ik wacht voorlopig de resultaten van het experiment af* but not in *ik wacht al weken op de resultaten van het experiment*. '*Reflexivity'* is not included because it concerns the valency of the verbs. Neither is the feature '*Auxiliary'* included which states which auxillary verb is selected for the past tense *(hebben, zijn).*

PRONOUNS

EAGLES gives three *A-extensions*. '*Special Pronoun Type'* has three values *(personal, reflexive, reciprocal*) as does '*Wh-Type'* (*interrogative, relative, relative);* they refer to specific parts of the *VWTYPE* classification in CGN. '*Politeness'* has two values (*polite, familiar*) and is part of the *PERSOONspartitie* inCGN. The *L-extension 'Strength'* with the values '*weak'* and '*strong'* corresponds to *STATUS*.

PREPOSITIONS

The only *A-extension* concerns not a new feature but the addition of the value *'Fused prep-art'* to '*Type'*. This corresponds to the value '*versmolten'* in CGN. Under the *L-extensions* a further two *extra-values* are given, namely *'Postposition'* and '*Circumposition'*. The first corresponds to '*finaal'*, the second does not have an equivalent in CGN, because circumpositions are treated as combinations of an initial preposition, a complement and a final preposition.

CONJUCTIONS

Under *A-extensions* EAGLES gives the feature '*Coord-Type'* with the values '*simple'*, '*correlative'*, *'initial'* and '*non-initial'*. The first version of the CGN tagset included a corresponding feature, but this was removed due to difficulties in interpretation when allocating it. Moreover, recognizing discontinuous conjuctions as parts of a whole (*hetzij ... hetzij ..., noch ... noch ...)* is not a job for the tagger, but for syntactic analysis and/or lexicological linking.

ADVERBS

Under *A-extensions* EAGLES gives *'Adverb-Type'*, *'Polarity'* and '*Wh-type'*. *'Adverb- Type'* has the values '*general'*, *'degree'* and—with *L-extension*—'*particle'* and '*pronominal'*. This classification boils down to the isolation of three small subclasses of adverbs, while all the others are sent to a large *remaining group* ('*general'*). We did not think it appropriate to set up a CGN equivalent, all the more because we treat the so-called pronomimal adverbs as pronouns and because we class most particles as intransitive prepositions, see 2.5 and 2.7.

'*Polarity'*, with the values '*wh-type'* and '*non-wh-type'* the '*wh-adverbs*' are distinguished from the others and *'Wh-type'* further divides the former into *interrogative*, *relative* and *exclamative*. In Dutch the *wh-adverbs* are restricted to *hoe, wanneer* and the combinations of *waar* with a preposition, such as *waarop, waarvan, ...*; the other *wh-woorden* (included *waar*) are pronouns. It did not seem worthwhile to introduce two extra features for such a small section of the adverbs.

3.4 NOT MENTIONED BY EAGLES

Not mentioned in EAGLES, but inlcuded in the CGN tagset, is the distinction between base forms and diminutive forms. The reason this is not covered in EAGLES is perhaps due to the fact that in most European languages forming the diminutive is not a productive process, but in Dutch (and Italian) it is; incidentally, not only for nouns but also with adjectives, numerals and degreeable determiners. Inclusion in the tagset is therefore desirable, because we would otherwise have to postulate separate lemmas for the base forms and diminutive forms, which would be contrary to lexicographical practice. Another point where CGN goes further than the EAGLES recommendations concerns the use of the features *POSITIE* and *BUIGING*. These are covered in EAGLES under the *A-extensions* for adjectives, but in CGN they are also used for the non-finite verbs and the determiners. The wider use of *BUIGING* has to do with the fact that in other Euorpean lanaguges the variation (*without, with-e, with-s*) corresponds to the gender and number differentiation.

4 SUMMARY

Formally speaking, the CGN tagset is a sextuple <*A,W,P,D,I,T*>, in which *A* is a collection of attributes, *W* of values, *P* of partitions, *D* of declarations, *I* of implications and *T* of tags. Features are those combinations of attributes and values which meet the restrictions given by the partitions. Tags are lists of features which meet the restrictions laid down by the declarations and the implications.

4.1 THE PARTITIONS

[P01] TOKENTYPE	= woord, speciaal, leesteken.
[P02] POS	= substantief, adjectief, werkwoord, telwoord, voornaamwoord,lidwoord, voorzetsel, voegwoord, bijwoord, tussenwerpsel.
[P03] NTYPE	= soortnaam, eigennaam.
[P04] GETAL	= getal (enkelvoud, meervoud).
[P05] GRAAD	= basis, comparatief, superlatief, diminutief.
[P06] GENUS	= genus (zijdig (masculien, feminien), onzijdig).
[P07] NAAMVAL	= standaard (nominatief, oblique), bijzonder (genitief, datief).
[P08] POSITIE	= prenominaal, nominaal, postnominaal, vrij.
[P09] BUIGING	= zonder, met-e, met-s.
[P10] GETAL-N	= zonder-n, meervoud-n.
[P11] WVORM	= persoonsvorm, buigbaar (infinitief, onvdw, voltdw).
[P12] PVTIJD	= tegenwoordig, verleden, conjunctief.
[P13] PVAGR	= enkelvoud, meervoud, met-t.
[P14] NUMTYPE	= hoofdtelwoord, rangtelwoord.
[P15] VWTYPE	= pr (persoonlijk, reflexief), reciprook, bezittelijk, vb (vragend, betrekkelijk), exclamatief, aanwijzend, onbepaald.
[P16] PDTYPE	= pronomen (adv-pronomen), determiner (gradeerbaar).
[P17] PERSOON	= persoon (1, 2 (2v, 2b), 3 (3p (3m, 3v), 3o)).
[P18] STATUS	= vol, gereduceerd, nadruk.
[P19] NPAGR	= agr (evon, rest (evz, mv)), agr3 (evmo, rest3 (evf, mv)).
[P20] LWTYPE	= bepaald, onbepaald.
[P21] VZTYPE	= initieel (versmolten), finaal.
[P22] CONJTYPE	= nevenschikkend, onderschikkend.
[P23] SPECTYPE	= afgebroken, onverstaanbaar, vreemd, deeleigen, meta, commentaar, achtergrond.

4.2 THE DECLARATIONS

[D00] <TOKENTYPE = woord> ⇒ <POS>
[D01] <POS = substantief> ⇒ <NTYPE, GETAL, GRAAD>
[D02] <POS = substantief, GETAL = enkelvoud> ⇒ <NAAMVAL>
[D03] <POS = substantief, GETAL = enkelvoud, NAAMVAL = standaard> ⇒ <GENUS>
[D04] <POS = adjectief> ⇒ <POSITIE, GRAAD, BUIGING>
[D05] <POSITIE = nominaal> ⇒ <GETAL-N>
[D06] <POS = adjectief, POSITIE = nominaal, BUIGING = met-e, GETAL-N = zonder-n> ⇒ <NAAMVAL>
[D07] <POS = adjectief, POSITIE = prenominaal, BUIGING = met-e> ⇒ <NAAMVAL>
[D08] <POS = werkwoord> ⇒ <WVORM>
[D09] <WVORM = persoonsvorm> ⇒ <PVTIJD, PVAGR>
[D10] <WVORM = buigbaar> ⇒ <POSITIE, BUIGING>
[D11] <POS = telwoord> ⇒ <NUMTYPE, POSITIE>
[D12] <NUMTYPE = hoofdtelwoord, POSITIE = nominaal> ⇒ <GRAAD>
[D13] <POS = telwoord, POSITIE = prenominaal> ⇒ <NAAMVAL>
[D14] <POS = voornaamwoord> ⇒ <VWTYPE, PDTYPE, NAAMVAL>
[D15] <PDTYPE = pronomen> ⇒ <STATUS, PERSOON, GETAL>
[D16] <VWTYPE = persoonlijk, NAAMVAL = standaard, PERSOON = 3, GETAL = enkelvoud> ⇒ <GENUS>
[D17] <PDTYPE = determiner> ⇒ <POSITIE, BUIGING>
[D18] <PDTYPE = determiner, POSITIE = prenominaal> ⇒ <NPAGR>
[D19] <PDTYPE = gradeerbaar> ⇒ <GRAAD>
[D20] <VWTYPE = bezittelijk> ⇒ <STATUS, PERSOON, GETAL>
[D21] <POS = lidwoord> ⇒ <LWTYPE, NAAMVAL, NPAGR>

[D22] <POS = voorzetsel> ⇒ <VZTYPE>
[D23] <POS = voegwoord> ⇒ <CONJTYPE>
[D24] <TOKENTYPE = speciaal> ⇒ <SPECTYPE>

4.3 THE IMPLICATIONS

[I01] <POS = substantief, GRAAD = diminutief, GETAL = enkelvoud> ⇒ <NAAMVAL ≠ datief>
[I02] <POS = substantief, GRAAD = diminutief, GETAL = enkelvoud, NAAMVAL = standaard> ⇒ <GENUS = onzijdig>
[I03] <POS = adjectief, GRAAD = superlatief> ⇒ <POSITIE ≠ postnominaal>
[I04] <POS = adjectief, GRAAD = diminutief> ⇒ <POSITIE = vrij>
[I05] <BUIGING = met-s> ⇒ <POSITIE = postnominaal>
[I06] <BUIGING = met-e> ⇒ <POSITIE = (pre)nominaal>
[I07] <WVORM = infinitief, POSITIE = nominaal> ⇒ <BUIGING = zonder, GETAL-N = zonder-n>
[I08] <PVTIJD = conjunctief> ⇒ <PVAGR = enkelvoud>
[I09] <NUMTYPE = rangtelwoord> ⇒ <POSITIE ≠ vrij>
[I10] <VWTYPE = persoonlijk> ⇒ <PDTYPE = pronomen>
[I11] <VWTYPE = reflexief> ⇒ <PDTYPE = pronomen, NAAMVAL = oblique>
[I12] <VWTYPE = reciprook> ⇒ <PDTYPE = pronomen, NAAMVAL ≠ nominatief, STATUS = vol, GETAL = meervoud>
[I13] <VWTYPE = bezittelijk> ⇒ <PDTYPE = determiner, POSITIE ≠ vrij>
[I14] <VWTYPE = vragend, PDTYPE = pronomen> ⇒ <STATUS ≠ gereduceerd>
[I15] <VWTYPE = betrekkelijk, PDTYPE = pronomen> ⇒ <STATUS = vol>
[I16] <VWTYPE = exclamatief, PDTYPE = pronomen> ⇒ <STATUS = vol>
[I17] <PDTYPE = determiner, POSITIE = prenominaal, NAAMVAL = standaard> ⇒ <NPAGR = agr>
[I18] <PDTYPE = determiner, POSITIE = prenominaal, NAAMVAL = bijzonder> ⇒ <NPAGR = agr3>
[I19] <POS = lidwoord, NAAMVAL = standaard> ⇒ <NPAGR = agr>
[I20] <POS = lidwoord, NAAMVAL = bijzonder> ⇒ <NPAGR = agr3>

4.4 THE TAGS

POS	T	U	TOTAL
nouns	16	2	18
adjectives	30		30
verbs	21		11
numerals	11		11
pronouns	43	145	188
articles	7	2	9
prepositions	3		3
conjunctions	2		2
adverbs	1		1
interjections	1		1
dialect words		28	28
special signs	7		7
punctuation marks	1		1
TOTAL	143	177	320

[T101]	N(soort,ev,basis,zijd,stan)	die stoel, deze muziek, de filter
[T102]	N(soort,ev,basis,onz,stan)	het kind, ons huis, het filter
[T103]	N(soort,ev,dim,onz,stan)	dit stoeltje, op ’t nippertje
[T104]	N(soort,ev,basis,gen)	’s avonds, de heer des huizes
[T105]	N(soort,ev,dim,gen)	vadertjes pijp
[T106]	N(soort,ev,basis,dat)	ter plaatse, heden ten dage
[T107]	N(soort,mv,basis)	stoelen, kinderen, hersenen
[T108]	N(soort,mv,dim)	stoeltjes, huisjes, hersentjes
[T109]	N(eigen,ev,basis,zijd,stan)	de Noordzee, de Kemmelberg, Karel
[T110]	N(eigen,ev,basis,onz,stan)	het Hageland, het Nederlands
[T111]	N(eigen,ev,dim,onz,stan)	het slimme Kareltje
[T112]	N(eigen,ev,basis,gen)	des Heren, Hagelands trots
[T113]	N(eigen,ev,dim,gen)	Kareltjes fiets
[T114]	N(eigen,ev,basis,dat)	wat den Here toekomt
[T115]	N(eigen,mv,basis)	de Ardennen, de Middeleeuwen
[T116]	N(eigen,mv,dim)	de Maatjes
[U117]	N(soort,ev,basis,genus,stan)	een riool, geen filter
[U118]	N(eigen,ev,basis,genus,stan)	Linux, Esselte
[T201]	ADJ(prenom,basis,zonder)	een mooi huis, een houten pot
[T202]	ADJ(prenom,basis,met-e,stan)	mooie huizen, een grote pot
[T203]	ADJ(prenom,basis,met-e,bijz)	zaliger gedachtenis, van goeden huize
[T204]	ADJ(prenom,comp,zonder)	een mooier huis
[T205]	ADJ(prenom,comp,met-e,stan)	mooiere huizen, een grotere pot
[T206]	ADJ(prenom,comp,met-e,bijz)	van beteren huize
[T207]	ADJ(prenom,sup,zonder)	een alleraardigst mens
[T208]	ADJ(prenom,sup,met-e,stan)	de mooiste keuken, het grootste paard

[T209]	ADJ(prenom,sup,met-e,bijz)	bester kwaliteit
[T210]	ADJ(nom,basis,zonder,zonder-n)	in het groot, het groen
[T211]	ADJ(nom,basis,zonder,mv-n)	de timiden, dezelfden
[T212]	ADJ(nom,basis,met-e,zonder-n,stan)	het leuke is dat, een grote met tartaar
[T213]	ADJ(nom,basis,met-e,zonder-n,bijz)	hosanna in den hogen
[T214]	ADJ(nom,basis,met-e,mv-n)	de rijken
[T215]	ADJ(nom,comp,zonder,zonder-n)	
[T216]	ADJ(nom,comp,met-e,zonder-n,stan)	een betere
[T217]	ADJ(nom,comp,met-e,zonder-n,bijz)	
[T218]	ADJ(nom,comp,met-e,mv-n)	de ouderen
[T219]	ADJ(nom,sup,zonder,zonder-n)	op z'n best, om ter snelst
[T220]	ADJ(nom,sup,met-e,zonder-n,stan)	het leukste is dat, het langste blijven
[T221]	ADJ(nom,sup,met-e,zonder-n,bijz)	des Allerhoogsten
[T222]	ADJ(nom,sup,met-e,mv-n)	de slimsten
[T223]	ADJ(postnom,basis,zonder)	rivieren bevaarbaar in de winter
[T224]	ADJ(postnom,basis,met-s)	iets moois
[T225]	ADJ(postnom,comp,zonder)	een getal groter dan 3
[T226]	ADJ(postnom,comp,met-s)	iets gekkers kon ik niet bedenken
[T227]	ADJ(vrij,basis,zonder)	die stok is lang, lang slapen
[T228]	ADJ(vrij,comp,zonder)	deze stok is langer, langer slapen
[T229]	ADJ(vrij,sup,zonder)	die stok is het langst, het langst slapen
[T230]	ADJ(vrij,dim,zonder)	het is hier stilletjes, stilletjes weggaan
[T301]	WW(pv,tgw,ev)	ik kom, speel je, hij is, zwijg
[T302]	WW(pv,tgw,mv)	komen, spelen
[T303]	WW(pv,tgw,met-t)	jij komt, hij speelt, zwijgt
[T304]	WW(pv,verl,ev)	kwam, speelde
[T305]	WW(pv,verl,mv)	kwamen, speelden
[T306]	WW(pv,verl,met-t)	kwaamt, gingt
[T309]	WW(pv,conj,ev)	kome, leve de koning
[T310]	WW(inf,prenom,zonder)	de nog te lezen post
[T311]	WW(inf,prenom,met-e)	een niet te weerstane verleiding
[T312]	WW(inf,nom,zonder,zonder-n)	(het) spelen, (het) schaatsen
[T314]	WW(inf,vrij,zonder)	zal komen
[T315]	WW(vd,prenom,zonder)	een verwittigd man, een gekregen paard
[T316]	WW(vd,prenom,met-e)	een getemde feeks
[T317]	WW(vd,nom,met-e,zonder-n)	het geschrevene, een gekwetste
[T318]	WW(vd,nom,met-e,mv-n)	gekwetsten, gedupeerden
[T320]	WW(vd,vrij,zonder)	is gekomen
[T321]	WW(od,prenom,zonder)	een slapend kind
[T322]	WW(od,prenom,met-e)	een piano spelende aap, slapende kinderen
[T323]	WW(od,nom,met-e,zonder-n)	het resterende, een klagende
[T324]	WW(od,nom,met-e,mv-n)	de wachtenden
[T326]	WW(od,vrij,zonder)	liep lachend weg, al doende leert men
[T401]	TW(hoofd,prenom,stan)	vier cijfers
[T402]	TW(hoofd,prenom,bijz)	eens geestes zijn, te enen

		male
[T403]	TW(hoofd,nom,zonder-n,basis)	er is er een ontsnapt
[T404]	TW(hoofd,nom,mv-n,basis)	met z’n vieren
[T405]	TW(hoofd,nom,zonder-n,dim)	er is er eentje ontsnapt, op z’n eentje
[T406]	TW(hoofd,nom,mv-n,dim)	met z’n tweetjes
[T407]	TW(hoofd,vrij)	veertig worden, honderd rijden, hoeveel sneller
[T408]	TW(rang,prenom,stan)	de vierde man
[T409]	TW(rang,prenom,bijz)	te elfder ure
[T410]	TW(rang,nom,zonder-n)	het eerste, (de) vierde eindigen, Karel de Vijfde
[T411]	TW(rang,nom,mv-n)	de eersten, iets aan derden verkopen
[T501a]	VNW(pers,pron,nomin,vol,1,ev)	ik
[T501b]	VNW(pers,pron,nomin,nadr,1,ev)	ikzelf, ikke
[T501c]	VNW(pers,pron,nomin,red,1,ev)	’k
[T501d]	VNW(pers,pron,nomin,vol,1,mv)	wij
[T501e]	VNW(pers,pron,nomin,nadr,1,mv)	wijzelf
[T501f]	VNW(pers,pron,nomin,red,1,mv)	we
[T501g]	VNW(pers,pron,nomin,vol,2v,ev)	jij
[T501h]	VNW(pers,pron,nomin,nadr,2v,ev)	jijzelf
[T501i]	VNW(pers,pron,nomin,red,2v,ev)	je
[U501j]	VNW(pers,pron,nomin,vol,2b,getal)	u
[U501k]	VNW(pers,pron,nomin,nadr,2b,getal)	uzelf
[U501l]	VNW(pers,pron,nomin,vol,2,getal)	gij
[U501m]	VNW(pers,pron,nomin,nadr,2,getal)	gijzelf
[U501n]	VNW(pers,pron,nomin,red,2,getal)	ge
[U501o]	VNW(pers,pron,nomin,vol,3,ev,masc)	hij
[T501p]	VNW(pers,pron,nomin,nadr,3m,ev,masc)	hijzelf
[U501q]	VNW(pers,pron,nomin,red,3,ev,masc)	ie
[U501r]	VNW(pers,pron,nomin,red,3p,ev,masc)	men
[T501s]	VNW(pers,pron,nomin,vol,3v,ev,fem)	zij
[T501t]	VNW(pers,pron,nomin,nadr,3v,ev,fem)	zijzelf
[U501u]	VNW(pers,pron,nomin,vol,3p,mv)	zij
[U501v]	VNW(pers,pron,nomin,nadr,3p,mv)	zijzelf
[T502a]	VNW(pers,pron,obl,vol,2v,ev)	jou
[U502b]	VNW(pers,pron,obl,vol,3,ev,masc)	hem
[T502c]	VNW(pers,pron,obl,nadr,3m,ev,masc)	hemzelf
[U502d]	VNW(pers,pron,obl,red,3,ev,masc)	’m
[U502e]	VNW(pers,pron,obl,vol,3,getal,fem)	haar
[U502f]	VNW(pers,pron,obl,nadr,3v,getal,fem)	haarzelf
[U502g]	VNW(pers,pron,obl,red,3v,getal,fem)	’r, d’r
[U502h]	VNW(pers,pron,obl,vol,3p,mv)	hen, hun
[U502i]	VNW(pers,pron,obl,nadr,3p,mv)	henzelf, hunzelf
[U503a]	VNW(pers,pron,stan,nadr,2v,mv)	jullie
[U503b]	VNW(pers,pron,stan,red,3,ev,onz)	het, ’t
[U503c]	VNW(pers,pron,stan,red,3,ev,fem)	ze
[U503d]	VNW(pers,pron,stan,red,3,mv)	ze
[T504a]	VNW(pers,pron,gen,vol,1,ev)	mijns gelijke, gedenk mijner
[T504b]	VNW(pers,pron,gen,vol,1,mv)	ons gelijke, velen onzer
[U504c]	VNW(pers,pron,gen,vol,2,getal)	uws gelijke, wie uwer
[T504d]	VNW(pers,pron,gen,vol,3m,ev)	zijns gelijke, zijner
[U504e]	VNW(pers,pron,gen,vol,3v,getal)	haars gelijke, harer
[U504f]	VNW(pers,pron,gen,vol,3p,mv)	huns gelijke, een hunner
[U505a]	VNW(pr,pron,obl,vol,1,ev)	mij
[U505b]	VNW(pr,pron,obl,nadr,1,ev)	mezelf, mijzelf
[U505c]	VNW(pr,pron,obl,red,1,ev)	me
[U505d]	VNW(pr,pron,obl,vol,1,mv)	ons
[U505e]	VNW(pr,pron,obl,nadr,1,mv)	onszelf

[U505f]	VNW(pr,pron,obl,red,2v,getal)	je
[U505g]	VNW(pr,pron,obl,nadr,2v,getal)	jezelf
[U505h]	VNW(pr,pron,obl,vol,2,getal)	u
[U505i]	VNW(pr,pron,obl,nadr,2,getal)	uzelf
[U506a]	VNW(refl,pron,obl,red,3,getal)	zich
[U506b]	VNW(refl,pron,obl,nadr,3,getal)	zichzelf
[U507a]	VNW(recip,pron,obl,vol,persoon,mv)	elkaar, mekaar, elkander
[U508a]	VNW(recip,pron,gen,vol,persoon,mv)	elkaars, mekaars, elkanders
[U509a]	VNW(bez,det,stan,vol,1,ev,prenom,zonder,agr)	mijn paard(en)
[U509b]	VNW(bez,det,stan,vol,1,ev,prenom,met-e,rest)	mijne heren
[U509c]	VNW(bez,det,stan,red,1,ev,prenom,zonder,agr)	m'n paard(en)
[U509d]	VNW(bez,det,stan,vol,1,mv,prenom,zonder,evon)	ons paard
[U509e]	VNW(bez,det,stan,vol,1,mv,prenom,met-e,rest)	onze paarden
[U509f]	VNW(bez,det,stan,vol,2,getal,prenom,zonder,agr)	uw paard(en) [
U509g]	VNW(bez,det,stan,vol,2,getal,prenom,met-e,rest)	uwe heiligheid
[U509h]	VNW(bez,det,stan,vol,2v,ev,prenom,zonder,agr)	jouw paard(en)
[U509i]	VNW(bez,det,stan,red,2v,ev,prenom,zonder,agr)	je paard(en)
[U509j]	VNW(bez,det,stan,nadr,2v,mv,prenom,zonder,agr)	jullie paard(en)
[U509k]	VNW(bez,det,stan,vol,3,ev,prenom,zonder,agr)	zijn paard(en), haar kind
[U509l]	VNW(bez,det,stan,vol,3m,ev,prenom,met-e,rest)	zijne excellentie
[U509m]	VNW(bez,det,stan,vol,3v,ev,prenom,met-e,rest)	hare majesteit
[U509n]	VNW(bez,det,stan,red,3,ev,prenom,zonder,agr)	z'n paard
[U509o]	VNW(bez,det,stan,vol,3,mv,prenom,zonder,agr)	hun paarden
[U509p]	VNW(bez,det,stan,vol,3p,mv,prenom,met-e,rest)	hunne
[U509q]	VNW(bez,det,stan,red,3,getal,prenom,zonder,agr)	'r paard, d'r paard
[T510a]	VNW(bez,det,gen,vol,1,ev,prenom,zonder,evmo)	mijns inziens
[U510b]	VNW(bez,det,gen,vol,1,ev,prenom,met-e,rest3)	een mijner vrienden
[T510c]	VNW(bez,det,gen,vol,1,mv,prenom,met-e,evmo)	onzes inziens
[U510d]	VNW(bez,det,gen,vol,1,mv,prenom,met-e,rest3)	een onzer vrienden
[U510e]	VNW(bez,det,gen,vol,2,getal,prenom,zonder,evmo)	uws
[U510f]	VNW(bez,det,gen,vol,2,getal,prenom,met-e,rest3)	een uwer vrienden
[U510g]	VNW(bez,det,gen,vol,2v,ev,prenom,met-e,rest3)	een jouwer vrienden
[U510h]	VNW(bez,det,gen,vol,3,ev,prenom,zonder,evmo)	zijns inziens
[U510i]	VNW(bez,det,gen,vol,3,ev,prenom,met-e,rest3)	een zijner vrienden
[T510j]	VNW(bez,det,gen,vol,3v,ev,prenom,zonder,evmo)	haars inziens
[U510k]	VNW(bez,det,gen,vol,3v,ev,prenom,met-e,rest3)	een harer vrienden
[U510l]	VNW(bez,det,gen,vol,3p,mv,prenom,zonder,evmo)	huns inziens
[U510m]	VNW(bez,det,gen,vol,3p,mv,prenom,met-e,rest3)	een hunner vrienden
[T511a]	VNW(bez,det,dat,vol,1,ev,prenom,met-e,evmo)	te mijnen huize
[T511b]	VNW(bez,det,dat,vol,1,ev,prenom,met-e,evf)	te mijner ere
[T511c]	VNW(bez,det,dat,vol,1,mv,prenom,met-e,evmo)	te onzen behoeve
[T511d]	VNW(bez,det,dat,vol,1,mv,prenom,met-e,evf)	te onzer ere
[U511e]	VNW(bez,det,dat,vol,2,getal,prenom,met-e,evmo)	te uwen behoeve
[U511f]	VNW(bez,det,dat,vol,2,getal,prenom,met-e,evf)	te uwer ere
[T511g]	VNW(bez,det,dat,vol,2v,ev,prenom,met-e,evf)	te jouwer nagedachtenis
[U511h]	VNW(bez,det,dat,vol,3,ev,prenom,met-e,evmo)	zijnen
[U511i]	VNW(bez,det,dat,vol,3,ev,prenom,met-e,evf)	te zijner tijd
[T511j]	VNW(bez,det,dat,vol,3v,ev,prenom,met-e,evmo)	haren
[T511k]	VNW(bez,det,dat,vol,3v,ev,prenom,met-e,evf)	te harer ere
[U511l]	VNW(bez,det,dat,vol,3p,mv,prenom,met-e,evmo)	hunnen
[U511m]	VNW(bez,det,dat,vol,3p,mv,prenom,met-e,evf)	te hunner ere
[U512h]	VNW(bez,det,stan,vol,1,ev,nom,met-e,zonder-n)	het mijne
[U512i]	VNW(bez,det,stan,vol,1,mv,nom,met-e,zonder-n)	de onze
[U512j]	VNW(bez,det,stan,vol,2,getal,nom,met-e,zonder-n)	het uwe
[U512k]	VNW(bez,det,stan,vol,2v,ev,nom,met-e,zonder-n)	de jouwe
[U512l]	VNW(bez,det,stan,vol,3m,ev,nom,met-e,zonder-n)	het zijne
[U512m]	VNW(bez,det,stan,vol,3v,ev,nom,met-e,zonder-n)	de hare
[U512n]	VNW(bez,det,stan,vol,3p,mv,nom,met-e,zonder-n)	het hunne
[U512o]	VNW(bez,det,stan,vol,1,ev,nom,met-e,mv-n)	de mijnen
[U512p]	VNW(bez,det,stan,vol,1,mv,nom,met-e,mv-n)	de onzen

[U512q]	VNW(bez,det,stan,vol,2,getal,nom,met-e,mv-n)	de uwen
[U512r]	VNW(bez,det,stan,vol,2v,ev,nom,met-e,mv-n)	de jouwen
[U512s]	VNW(bez,det,stan,vol,3m,ev,nom,met-e,mv-n)	de zijnen
[U512t]	VNW(bez,det,stan,vol,3v,ev,nom,met-e,mv-n)	de haren
[U512u]	VNW(bez,det,stan,vol,3p,mv,nom,met-e,mv-n)	de hunnen
[T513a]	VNW(bez,det,dat,vol,1,ev,nom,met-e,zonder-n)	te mijnent
[T513b]	VNW(bez,det,dat,vol,1,mv,nom,met-e,zonder-n)	ten onzent
[U513c]	VNW(bez,det,dat,vol,2,getal,nom,met-e,zonder-n)	ten uwent
[T513d]	VNW(bez,det,dat,vol,3m,ev,nom,met-e,zonder-n)	te zijnent
[T513e]	VNW(bez,det,dat,vol,3v,ev,nom,met-e,zonder-n)	ten harent
[U513f]	VNW(bez,det,dat,vol,3p,mv,nom,met-e,zonder-n)	ten hunnent
[U514a]	VNW(vrag,pron,stan,nadr,3o,ev)	watte
[U515a]	VNW(betr,pron,stan,vol,persoon,getal)	de man die daar staat
[U515b]	VNW(betr,pron,stan,vol,3,ev)	het kind dat je daar ziet
[U515c]	VNW(betr,det,stan,nom,zonder,zonder-n)	hetgeen je daar ziet, het feest tijdens hetwelk
[U515d]	VNW(betr,det,stan,nom,met-e,zonder-n)	op hetgene de gemeente doet
[T516a]	VNW(betr,pron,gen,vol,3o,ev)	het warenhuis welks directeur hem een baan had aangeboden
[U516b]	VNW(betr,pron,gen,vol,3o,getal)	de kathedraal welker gewelven
[U517a]	VNW(vb,pron,stan,vol,3p,getal)	wie gaat er mee
[U517b]	VNW(vb,pron,stan,vol,3o,ev)	wat ik niet begrijp is
[U518a]	VNW(vb,pron,gen,vol,3m,ev)	wiens hoed is dit
[U518b]	VNW(vb,pron,gen,vol,3v,ev)	de vrouw wier hoed daar hangt
[U518c]	VNW(vb,pron,gen,vol,3p,mv)	de studenten tegen wier houding ...
[U519a]	VNW(vb,adv-pron,obl,vol,3o,getal)	waar ga je naartoe, de trein waar we op staan te wachten
[U520a]	VNW(excl,pron,stan,vol,3,getal)	wat een dwaasheid, wat kan jij liegen zeg
[U521a]	VNW(vb,det,stan,prenom,zonder,evon)	welk kind
[U521b]	VNW(vb,det,stan,prenom,met-e,rest)	welke kinderen
[U522a]	VNW(vb,det,stan,nom,met-e,zonder-n)	welke vind jij de mooiste
[U523a]	VNW(excl,det,stan,vrij,zonder)	welk een dwaasheid
[U524a]	VNW(aanw,pron,stan,vol,3o,ev)	dat, dit, zulks
[U524b]	VNW(aanw,pron,stan,nadr,3o,ev)	datte, ditte
[U524c]	VNW(aanw,pron,stan,vol,3,getal)	die
[T525a]	VNW(aanw,pron,gen,vol,3m,ev)	diens voorkeur
[T525b]	VNW(aanw,pron,gen,vol,3o,ev)	en dies meer
[U526a]	VNW(aanw,adv-pron,obl,vol,3o,getal)	hier, daar
[U527a]	VNW(aanw,adv-pron,stan,red,3,getal)	d’r, het niet-kwantitatieve ‘er’
[U528a]	VNW(aanw,det,stan,prenom,zonder,evon)	dat boek, dit dier, ginds bos, zulk hout
[U528b]	VNW(aanw,det,stan,prenom,zonder,rest)	die stoel(en)
[U528c]	VNW(aanw,det,stan,prenom,zonder,agr)	zo’n boek(en)
[U528d]	VNW(aanw,det,stan,prenom,met-e,rest)	deze man, gene zijde, gindse heuvel, zulke balken
[U529a]	VNW(aanw,det,gen,prenom,met-e,rest3)	een dezer dagen, de notulen dier vergadering
[T530a]	VNW(aanw,det,dat,prenom,met-e,evmo)	te dien tijde
[T530b]	VNW(aanw,det,dat,prenom,met-e,evf)	in dier voege
[U531b]	VNW(aanw,det,stan,nom,met-e,zonder-n)	deze, gene, datgene, degene, diegene
[U531c]	VNW(aanw,det,stan,nom,met-e,mv-n)	dezen, genen, degenen,

		diegenen
[T532a]	VNW(aanw,det,gen,nom,met-e,zonder-n)	schrijver dezes, de twintigste dezer
[T533a]	VNW(aanw,det,dat,nom,met-e,zonder-n)	dat is dan bij dezen beslist
[U534a]	VNW(aanw,det,stan,vrij,zonder)	zulk een vreemde gedachte
[U535a]	VNW(onbep,pron,stan,vol,3p,ev)	alleman, (n)iemand, iedereen, elkeen, menigeen
[U535b]	VNW(onbep,pron,stan,vol,3o,ev)	alles, (n)iets, niks, wat, zoiets
[U536a]	VNW(onbep,pron,gen,vol,3p,ev)	allemans, andermans, (n)iemands, (een)ieders
[U537a]	VNW(onbep,adv-pron,obl,vol,3o,getal)	(n)ergens, overal
[U538a]	VNW(onbep,adv-pron,gen,red,3,getal)	het kwantitatieve ‘er’
[U539a]	VNW(onbep,det,stan,prenom,zonder,evon)	elk huis, ieder kind, enig benul, een enkel woord, sommig bier
[U539b]	VNW(onbep,det,stan,prenom,zonder,agr)	geen kind(eren), menig politicus
[U539c]	VNW(onbep,det,stan,prenom,met-e,evz)	elke hond, iedere keer, ene mijnheer X, menige
[U539d]	VNW(onbep,det,stan,prenom,met-e,mv)	ettelijke
[U539e]	VNW(onbep,det,stan,prenom,met-e,rest)	sommige, enige, enkele
[U539f]	VNW(onbep,det,stan,prenom,met-e,agr)	alle mensen, hoop, vee
[T540a]	VNW(onbep,det,gen,prenom,met-e,mv)	proletari¨ers aller landen
[T541a]	VNW(onbep,det,dat,prenom,met-e,evmo)	te allen prijze
[T541b]	VNW(onbep,det,dat,prenom,met-e,evf)	te eniger tijd
[U542a]	VNW(onbep,grad,stan,prenom,zonder,agr,basis)	veel plezier, weinig geld
[U542b]	VNW(onbep,grad,stan,prenom,met-e,agr,basis)	het vele plezier, de weinige toeschouwers
[U542c]	VNW(onbep,grad,stan,prenom,met-e,mv,basis)	beide mannen
[U542d]	VNW(onbep,grad,stan,prenom,zonder,agr,comp)	meer tijd, minder werk
[U542e]	VNW(onbep,grad,stan,prenom,met-e,agr,sup)	de meeste mensen, het minste tijd
[U542f]	VNW(onbep,grad,stan,prenom,met-e,agr,comp)	in mindere mate
[U543a]	VNW(onbep,det,stan,nom,met-e,mv-n)	allen, sommigen, enkelen, de enen
[U543b]	VNW(onbep,det,stan,nom,met-e,zonder-n)	het ´ene ... het andere
[U543c]	VNW(onbep,det,stan,nom,zonder,zonder-n)	het ´e´en en ander
[T544a]	VNW(onbep,det,gen,nom,met-e,mv-n)	met aller instemming
[U545a]	NW(onbep,grad,stan,nom,met-e,zonder-n,basis)	het weinige
[U545b]	VNW(onbep,grad,stan,nom,met-e,mv-n,basis)	velen, weinigen, beiden
[U545d]	VNW(onbep,grad,stan,nom,met-e,zonder-n,sup)	het minste wat je kan zeggen, de meeste
[U545e]	VNW(onbep,grad,stan,nom,met-e,mv-n,sup)	de minsten, de meesten
[U545f]	VNW(onbep,grad,stan,nom,zonder,mv-n,dim)	met z’n beidjes
[T546a]	VNW(onbep,grad,gen,nom,met-e,mv-n,basis)	tot veler verbazing, met beider instemming
[U547a]	VNW(onbep,det,stan,vrij,zonder)	ze kregen elk/ieder/allebei een bal, al die mensen
[U548a]	VNW(onbep,grad,stan,vrij,zonder,basis)	dat is te weinig, veel groter
[U548b]	VNW(onbep,grad,stan,vrij,zonder,sup)	de minst gevraagde, de meest gezochte
[U548c]	VNW(onbep,grad,stan,vrij,zonder,comp)	minder werken, meer slapen
[T601]	LID(bep,stan,evon)	het kind, in ’t geniep
[T602]	LID(bep,stan,rest)	de hond(en), de kinderen
[T603]	LID(bep,gen,evmo)	des duivels, ’s avonds,
[U604]	LID(bep,gen,rest3)	der Nederlandse taal, der Belgen

[T605]	LID(bep,dat,evmo)	op den duur, om den brode
[T606]	LID(bep,dat,evf)	in der minne
[T607]	LID(bep,dat,mv)	die in den hemelen zijt
[U608]	LID(onbep,stan,agr)	een kind, een mensen dat er waren
[T609]	LID(onbep,gen,evf)	de kracht ener vrouw
[T701]	VZ(init)	met een lepeltje, met Jan in het hospitaal, met zo te roepen
[T702]	VZ(fin)	liep de trap af, bij de beesten af, speelt het bandje af
[T703]	VZ(versm)	ten strijde, ten hoogste, ter plaatse
[T801]	VG(neven)	Jan en Peter; en toen gebeurde het
[T802]	VG(onder)	omdat ze zich niet goed voelt
[T901]	BW()	gisteren, nu, niet, nog, al, hoe
[T001]	TSW()	oei, amai, uh, hoera

DIALECT WORDS

[R101]	N(soort,dial)	bompa*d
[R102]	N(eigen,dial)	
[R201]	ADJ(dial)	ne*d langen*d toot*d
[R301]	WW(dial)	’k zen*d nie*d thuis, ’k hem*d gee*d geld
[R401]	TW(hoofd,dial)	
[R402]	TW(rang,dial)	den*d elfste*d
[R501]	VNW(pers,pron,dial)	kom de*d gij mee, ’k heb ulie*d gezien
[R502]	VNW(refl,pron,dial)	
[R503]	VNW(recip,pron,dial)	we zien malkanderen*d niet veel
[R504]	VNW(bez,det,dial)	hij heeft z’ne*d frak*d vergeten
[R505]	VNW(vrag,pron,dial)	
[R506]	VNW(vrag,det,dial)	
[R507]	NW(betr,pron,dial)	
[R508]	VNW(betr,det,dial)	
[R509]	VNW(excl,pron,dial)	
[R510]	VNW(excl,det,dial)	
[R511]	VNW(aanw,pron,dial)	
[R512]	VNW(aanw,det,dial)	diejen*d boek, dees*d week
[R513]	VNW(onbep,pron,dial)	z’ hebben iet*d gezien
[R514]	VNW(onbep,det,dial)	ze kan elken*d dag vertrekken
[R601]	LID(bep,dial)	het gevecht met den*d beer
[R602]	LID(onbep,dial)	nen*d toffe gast, ne*d vieze vent
[R701]	VZ(init,dial)	me*d veel geduld
[R702]	VZ(fin,dial)	
[R801]	VG(neven,dial)	
[R802]	VG(onder,dial)	’t schijnt da*d ze nie*d kunnen komen

[R901]	BW(dial)	efkes*d, nie*d
[R001]	TSW(dial)	neeje*d, wablieft*d

SPECIAL TOKENS

[T002]	SPEC(afgebr)	uitge*a, binnen-
[T003]	SPEC(onverst)	ggg, xxx, Xxx
[T004]	SPEC(vreemd)	whatever*v, ad, hoc, wishful
[T005]	SPEC(deeleigen)	Den, Haag, New, York
[T006]	SPEC(meta)	(het woord) homosexueel
[T008]	SPEC(achter)	voor achtergrondgeluid
[T009]	SPEC(comment)	voor commentaren

PUNCTUATION MARKS

[T007]	LET()	., ..., ?

5 REFERENCES

[EAGLES] Expert Advisory Group on Language Engineering Standards (1996). Recommendations for the Morphosyntactic Annotation of Corpora. EAGLES Do- cument EAG - TCWG - MAC/R. Version of March, 1996.

[ANS-97] W. Haeseryn, K. Romijn, G. Geerts, J. de Rooij & M.C. van den Toorn (1997), Algemene Nederlandse Spraakkunst. 2de geheel herziene druk. Martinus Nijho , Groningen & Wolters Plantyn, Deurne.

[WOTAN-2] H. van Halteren (1999), The WOTAN2 Tagset Manual (under construction). Nijmegen, april 1999.

[STTS-95] A. Schiller, S. Teufel & C. Thielen (1995), Guidelines f¨ur das Tagging deutscher Textcorpora mit STTS (Stuttgart-T¨ubingen Tagset).

G. Booij & A. van Santen (1998), Morfologie. De woordstructuur van het Neder- lands. 2de geheel herziene druk. Amsterdam University Press, 1998.

I. Schuurman (1998), POS taggers. Leuven, november 1998.

F. Van Eynde (2001), CGN-Functionwordslexicon. Leuven, juni 2001 (integraal opgenomen in het CGN-lexicon).

F. Van Eynde, J. Zavrel & W. Daelemans (2000), Part of Speech Tagging and Lem- matisation for the Spoken Dutch Corpus. In M. Gavrilidou et al. (eds), Proceedings of the Second International Conference on Language Resources and Evaluation. Vo- lume III. p. 1427-1433. Athens, 2000.

J. Zavrel (1999), Annotator-overeenstemming bij het manuele taggingexperiment. Tilburg, juni 1999.

J. Zavrel & W. Daelemans (1999), Evaluatie van Part-of-Speech taggers voor het Corpus Gesproken Nederlands. Tilburg, juli 1999.

TAGLIST translation Dutch - English

Tag	Dutch	English
aanw	aanwijzend	demonstrative
achter	achtergrond	background
ADJ	adjectief	adjective
afgebr	afgebroken	broken off
basis	basis	base
bep	bepaald	definite
betr	betrekkelijk	relative
bez	bezittelijk	possessive
BW	bijwoord	adverb
bijz	bijzonder	special
buigbaar	buigbaar	non-declinable
BUIGING	buiging	inflection
comment	commentaar	comment
comp	comparatief	comparative
conj	conjunctief	conjunctive
dat	datief	dative
deeleigen	deel van een eigennaam	part of proper noun
det	determiner	determiner
dial	dialect	dialect
dim	diminutief	diminutive
eigen	eigennaam	proper noun
ev	enkelvoud	singular
excl	exclamatief	exclamatory
fem	feminien	feminine
fin	finaal	final
gen	genitief	genitive
GENUS	genus	genus
gereduceerd	gereduceerd	reduced
GETAL	getal	number
grad	gradeerbaar	degreeable
hoofd	hoofdtelwoord	cardinal number
inf	infinitief	infinitive
init	initieel	initial
LET	leestekens	punctuation marks
LID	lidwoord	article
masc	masculien	masculine

mv	meervoud	plural
met-	met-	with-
meta	metadata	metadata
NAAMVAL	naamval	case
nadr	nadruk	stress
neven	nevenschikkend	coordinate
nomin	nominaal	nominal
nom	nominatief	nominative
obl	oblique	oblique
onbep	onbepaald	indefinite
onder	onderschikkend	subordinate
onverst	onverstaanbaar	incomprehensible
onvdw/od	onvoltooid deelwoord	present participle
onz	onzijdig	neutral
POS	part-of-speech	part of speech
pers	persoonlijk	personal
pv	persoonsvorm	finite verb
POSITIE	positie	position
postnom	postnominaal	postnominal
prenom	prenominaal	prenominal
pron	pronomen	pronoun
rang	rangtelwoord	ordinal number
recip	reciprook	reciprocal
refl	reflexief	reflexive
soort	soortnaam	type name
SPEC/spec	speciaal	special
stan	standaard	standard
STATUS	status	status
N	substantief	noun
sup	superlatief	superlative
tgw	tegenwoordig	present
TW	telwoord	numeral
TSW	tussenwerpsel	interjection
verl	verleden	past
versm	versmolten	fused
VG	voegwoord	conjunction
vol	vol	full
voltdw/vd	voltooid deelwoord	past participle
VNW	voornaamwoord	pronoun
VZ	voorzetsel	preposition

vrag	vragend	interrogative
vreemd	vreemd	foreign
vrij	vrij	free
WW	werkwoord	verb
zijd	zijdig	gendered
zonder-	zonder-	without-