Morpho-Syntactic Agreement and Index Agreement in Dutch NPs

Frank Van Eynde

Centre for Computational Linguistics, K.U. Leuven

Abstract

For the treatment of agreement in Dutch NPs I adopt a distinction, familiar from HPSG, between morphosyntactic agreement and index agreement. In order to determine their respective roles, I make another distinction, proposed in Van Eynde (2003), between marked and unmarked nominals. Employing these distinctions, I will demonstrate that the combination of prenominal adjectives and determiners with unmarked nominals is subject to morphosyntactic agreement, whereas the combination of these prenominals with marked nominals is subject to index agreement.

1   Introduction

HPSG makes a distinction between morphosyntactic agreement and index agreement. An example of the former is case concord; an example of the latter is the agreement in person, number and gender between a reflexive pronoun and its antecedent. Pollard and Sag (1994) argues that subject-verb agreement and determiner-noun agreement are also instances of index agreement, at least for English. This is motivated by examples like the following:

(1)   The faculty have voted themselves a new raise.

(2)   Has that hash browns at table nine paid his check?

While the noun *faculty* is morphologically singular, it is taken to stand for an aggregate of individuals in (1), so that its index is plural, and it is this value which matters for the choice of the finite verb *have* and the reflexive *themselves*. Conversely, while *hash browns* is morphologically plural, it is interpreted as standing for the person who ate the hash browns in (2), so that its index is singular, and it is this value which matters for the choice of the determiner *that*, the finite verb *has* and the possessive *his*. To model this Pollard and Sag (1994) does not treat number and gender as properties of the nouns themselves, but rather as constraints on their mode of individuation. In terms of the TFS notation, they do not belong in the CATEGORY|HEAD value, but rather in the CONTENT|INDEX value, along with the PERSON feature.

$$
\begin{bmatrix}
\text{CATEGORY} \mid \text{HEAD} & \begin{bmatrix} \text{CASE} & \textit{case} \end{bmatrix}_{\textit{noun}} \\[4pt]
\text{CONTENT} \mid \text{INDEX} & \begin{bmatrix} \text{PERSON} & \textit{person} \\ \text{NUMBER} & \textit{number} \\ \text{GENDER} & \textit{gender} \end{bmatrix}_{\textit{index}}
\end{bmatrix}
$$

Kathol (1999) adopts the distinction between both types of agreement, but argues that its application to individual languages yields a more complex picture than the one which is sketched for English in Pollard and Sag (1994). He demonstrates that in many languages the agreement in number and gender is not only a matter of index agreement but also of morphosyntactic agreement. To model this, he uses two pairs of features: besides the number and gender features in the index, he also has number and gender features in the HEAD value, along with CASE and DECLENSION.[1]

The use of separate features for morphosyntactic agreement and index agreement is further motivated and exemplified in Wechsler and Zlatić (2000) for Serbo-Croatian and in Martínez (2003) for Spanish and French. Its relevance for Dutch can be illustrated with a noun like *meisje* 'girl'. Since it contains a diminutive affix, its morphosyntactic gender is unambiguously neuter (*het/*de meisje*), but the GENDER value in its index can be either neuter or feminine.

(3)   [Dat   meisje] heeft geen geluk; [het/ze/*hij] is alweer ontslagen.
      [That girl]   has   no   luck; [it/she/*he] is again fired
      'That girl has no luck; she has been fired again.'

(4)   Daar   staat   [het meisje] [dat/*die]   hij heeft gekust.
      There stands [the girl]   [that/*who] he has   kissed
      'There is the girl that he kissed.'

(5)   [Dat   meisje] heeft [haar/*zijn] fiets verloren.
      [That girl]   has   [her/*his]   bike lost
      'That girl has lost her bike.'

Whether it is the morphosyntactic gender or the natural gender which determines the value of the index and, hence, the choice of a coindexed pronoun depends a.o. on the type of the pronoun: for personal pronouns, both are possible, but for relative pronouns it must be neuter and for possessives it must be feminine.[2]
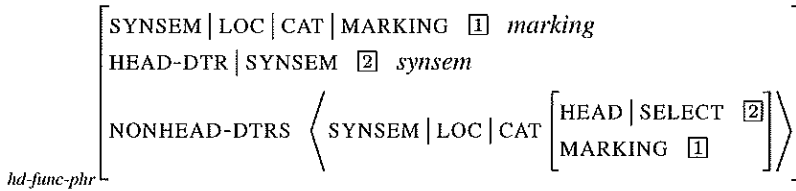
Given the existence of both morphosyntactic agreement and index agreement the obvious question is how they relate to one another. As a first guess, one could assume that the former concerns all aspects of NP-internal agreement, whereas the latter concerns all instances of agreement between a full NP and some other sign, such as another NP or a finite verb. A proposal along these lines is made in Wechsler and Zlatić (2000, 508) who explicitly claim that "index agreement does not apply to NP-internal elements such as determiners and adjectives." In this paper I will argue that this claim is too strong for Dutch. Building on the analysis of Dutch NPs in Van Eynde (2003) (section 2) I will first spell out the details of NP-internal morphosyntactic agreement (section 3) and then discuss two instances of NP-internal index agreement (section 4).

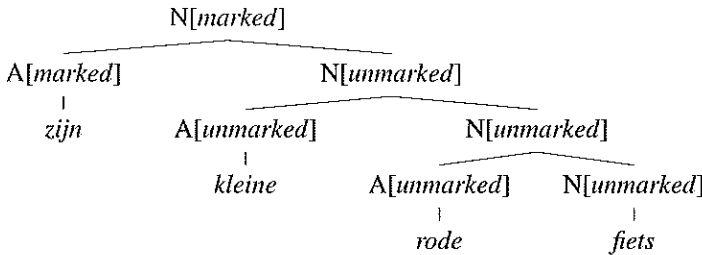[1] Declension is irrelevant for English, but it plays a prominent role in German and Dutch.

[2] The exclusion of the neuter possessive may be due to its homonymy with the masculine *zijn* 'its/his'; using this form for reference to a female individual would be misleading.
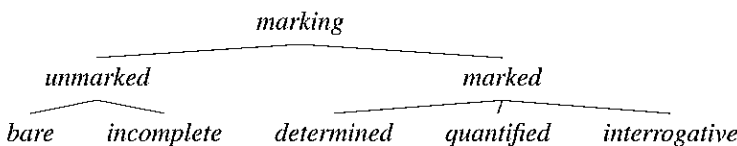
2   The Functor Treatment of Dutch NPs

To model the combination of a noun with its prenominal dependents, including both the adjectives and the determiners, I use the type *head-functor-phrase*, as defined in Van Eynde (1998).[3]

$$\begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{MARKING} \quad \boxed{1} \quad marking \\ \text{HEAD-DTR} \mid \text{SYNSEM} \quad \boxed{2} \quad synsem \\ \text{NONHEAD-DTRS} \quad \left\langle \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \begin{bmatrix} \text{HEAD} \mid \text{SELECT} & \boxed{2} \\ \text{MARKING} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$
hd-func-phr

Typical of this type is that the non-head daughter selects the head daughter ($\boxed{2}$) and that its MARKING value is identified with the one of the mother ($\boxed{1}$).[4] As applied to the NPs, the prenominals are functors which select a nominal as their head and which leave their mark on the nominal's AVM. As a consequence, if we treat the adjectives and the nouns as unmarked, and the determiners as marked, we get the following representation for *zijn kleine rode fiets* 'his small red bike'.

```
                    N[marked]
          ┌────────────┴──────────────┐
     A[marked]                 N[unmarked]
         │               ┌──────────┴──────────┐
       zijn        A[unmarked]            N[unmarked]
                        │             ┌───────┴────────┐
                      kleine    A[unmarked]      N[unmarked]
                                     │                │
                                   rode             fiets
```

In combination with the constraint that the prenominals select an unmarked nominal, this implies that the adjectives can be stacked whereas the determiners cannot, and that the determiners must precede the adjectives. The distinction between determiners and adjectives is, hence, not captured in terms of their part of speech or their syntactic function (specifier vs. adjunct), but in terms of their MARKING value. Anticipating the discussion in the rest of the paper, I will distinguish between two subtypes of unmarked nominals and three subtypes of marked nominals.
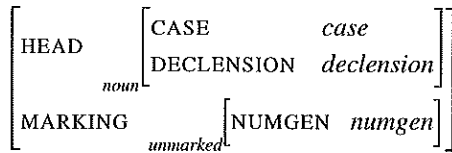
```
                        marking
          ┌────────────────┴────────────────┐
      unmarked                           marked
      ┌───┴────┐              ┌──────────────┼──────────────┐
    bare   incomplete    determined     quantified     interrogative
```

---

[3] A similar version is employed in Allegranza (1998) for the analysis of Italian NPs.

[4] In terms of the hierarchy of phrase types in Pollard and Sag (1994), it subsumes the types *head-specifier* and *head-marker*, as well as the combinations of type *head-adjunct* in which the adjunct precedes the head.
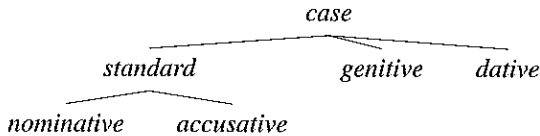
Bare nominals may but need not take a determiner, incomplete nominals must take a determiner and marked nominals do not take a determiner.

3   NP-Internal Morphosyntactic Agreement

In Dutch, the morphosyntactic agreement between a noun and its prenominal dependents concerns case, number, gender and declension. To model it I will assume that the CATEGORY value of each unmarked nominal contains features which capture these distinctions.

$$
\begin{bmatrix}
\text{HEAD} & \text{noun}\begin{bmatrix} \text{CASE} & case \\ \text{DECLENSION} & declension \end{bmatrix} \\
\text{MARKING} & \text{unmarked}\begin{bmatrix} \text{NUMGEN} & numgen \end{bmatrix}
\end{bmatrix}
$$

The HEAD features include CASE and DECLENSION. Since they are propagated throughout the nominal projection, as required by the HEAD FEATURE PRINCIPLE, their values are not only available for modeling NP-internal agreement, but also for modeling constraints on the distribution of the NP as a whole. The inventory of CASE values is the same as for German:

$$
\begin{array}{c}
\textit{case} \\
\diagdown \\
\textit{standard} \qquad \textit{genitive} \qquad \textit{dative} \\
\diagdown \\
\textit{nominative} \qquad \textit{accusative}
\end{array}
$$

The underspecified value *standard* applies to nominals without case affix. Its more specific subtypes can be assigned when the nominal's case value is unified with the case requirements of an external selector: finite verbs, for instance, select a nominative subject and most prepositions require an accusative object. The values *genitive* and *dative* are assigned to (pre)nominals with a case affix, such as the genitives in *'s middags* 'the-GEN noon-GEN' and the datives in *ter plaatse* 'at-the-DAT place-DAT'. Such forms are far less common in Dutch than in German, but they are not limited to fixed expressions either. The Flemish part of the fifth release of the Spoken Dutch Corpus, for instance, contains 3,333 word tokens with a case affix on a total of 1,658,133 word tokens; in other words, for approx. every 500 words there is one with a case affix.[5]

The DECLENSION value signals whether the nominal contains any forms with a declension affix. The nominal *witte paarden* 'white-DECL horses', for instance, is declined, whereas *wit paard* 'white horse' is nondeclined. Nominals which do not contain any form which shows variation for declension, such as *paard* 'horse', have the underspecified value *declension*.[6]

---

[5] I would like to thank Machteld Schouppe for collecting the relevant data. For more information about the Spoken Dutch Corpus, see e.g. Oostdijk et al. (2002).

[6] In German grammar, the DECLENSION values are known as *strong* and *weak*. They roughly correspond to respectively *nondeclined* and *declined*, but I prefer the latter terms for Dutch, mainly because

declension

declined   nondeclined
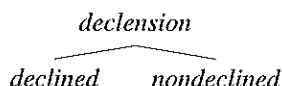
The DECLENSION value is relevant for NP-internal agreement, since declined nominals do not combine in the same way with a determiner as the other nominals. The mass noun *goud* 'gold', for instance, and the nondeclined *wit goud* 'white gold' can both be used without determiner, but the declined *witte goud* 'white-DECL gold' must be preceded by a determiner of type *determined*, i.e. a possessive or a demonstrative.[7]

The MARKING value is not shared between the nominal and its head, but rather between the nominal and its functor. For reasons which will become clear in section 4, I assume that this is the proper place for the information about morphosyntactic number and gender. More specifically, I assume that the nominals of type *unmarked*—but not those of type *marked*—have the feature NUMGEN. The reason for conflating the information about number and gender is that the gender distinction is systematically neutralized in the plural. The inventory of values looks as follows:

numgen

sg   pl

sg-masc   sg-fem   sg-neu
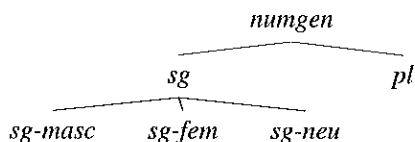
Employing this format I will now model the agreement relations which hold between a noun and its prenominals in standard Dutch.[8] For this purpose, I will distinguish between three types of prenominals: the possessive and demonstrative determiners (3.1), the interrogative and quantifying determiners (3.2) and the prenominal adjectives (3.3).
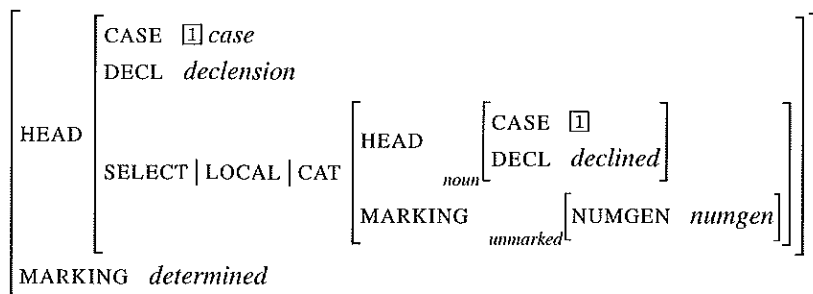
3.1   The Possessive and Demonstrative Determiners

The possessive and demonstrative determiners select an unmarked nominal and turn it into a marked NP of type *determined*.

the role of the *strong/weak* dichotomy in German grammar is not the same as the role of declension in Dutch grammar.

[7] This type also includes the definite article; as a matter of fact, *de* and *het* are the reduced counterparts of the demonstrative *die* and *dat*.

[8] The restriction to standard Dutch is important, since the details of NP-internal agreement are subject to dialectal variation.

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix}
\text{CASE} & \boxed{1}\ case \\
\text{DECL} & declension \\[1em]
\text{SELECT | LOCAL | CAT} & \begin{bmatrix}
\text{HEAD}_{noun} & \begin{bmatrix} \text{CASE} & \boxed{1} \\ \text{DECL} & declined \end{bmatrix} \\[1em]
\text{MARKING}_{unmarked} & \begin{bmatrix} \text{NUMGEN} & numgen \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\[1em]
\text{MARKING} & determined
\end{bmatrix}
$$

Typical of these determiners is that they require the nominal to be declined, also if they are not declined themselves. Given the workings of unification, this implies that they are compatible with declined nominals, as in *ons zwarte paard* 'our black-DECL horse', as well as with nominals with the underspecified value *declension*, as in *ons paard* 'our horse', but not with nominals with the value *nondeclined*, as in **ons zwart paard* 'our black horse'.[9]

The CASE and DECLENSION values of the determiners covary with their morphological form. This consists of a stem optionally followed by the declension affix *-e* and/or a case affix, i.e. *-s*, *-er* or *-n*.

stem	[+D]	[+C]	[+C]	[+D,+C]	
mijn		*mijns*	*mijner*	*mijnen*	my
ons	*onze*	*onzes*	*onzer*	*onzen*	our
zijn		*zijns*	*zijner*	*zijnen*	his
haar		*haars*	*harer*	*haren*	her
hun		*huns*	*hunner*	*hunnen*	their
	deze	*dezer*			this
	gene				yonder

The addition of the affixes triggers the usual phonotactic adjustments. For instance, the addition of the case affix *-s* to a stem which ends in a sibilant triggers the insertion of a sjwa, as in *onz-e-s* 'our-GEN'.

The forms without case affix have the CASE value *standard* and share this value with the selected nominal. The forms which also lack the declension affix have the value *nondeclined*, if and only if they have a [+D] counterpart; if there is no such counterpart, as in the case of *mijn* 'my', the relevant value is the underspecified *declension*. By contrast, the forms with a declension affix are invariably *declined*, also if they lack a counterpart without affix, as in the case of *deze* and *gene*.[10] The declension distinction correlates with the NUMGEN value of the selected nominal:

---

[9] In dialects, combinations like *ons zwart paard* are not ungrammatical; as a matter of fact, many Dutch dialects extend the agreement system of the interrogative and quantifying determiners, described in section 3.2, to all of the determiners.

[10] The [–D] counterpart of *deze* is *dees*, but this form is only used in dialects. In standard Dutch, it is replaced by *dit*, which is not a determiner but rather a pronoun which is used in both nominal and prenominal positions. Other examples of pronouns in prenominal position include the genitive in *wiens paard* 'whose horse' and the *wh*-pronoun in *wat melk* 'some milk'.

while the forms with the value *nondeclined* require a singular neuter nominal, the ones with the value *declined* require a nominal which is not singular neuter. This accounts for the contrast between *ons paard/*ezel(s)* 'our horse/*donkey(s)' and *onze ezel(s)/*paard* 'our-DECL donkey(s)/*horse'. Forms with the underspecified value *declension* do not impose any requirements on their head's NUMGEN value and are, hence, compatible with any nominal in standard case.
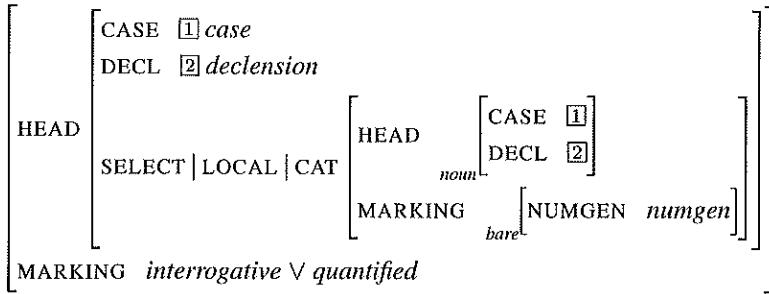
Characteristic of this analysis is the different treatment of case and declension, on the one hand, and number and gender, on the other hand. While the former are present in the HEAD values of both the determiner and the selected nominal, the latter only figure in the MARKING value of the selected nominal. This reflects the fact that the prenominal determiners carry information about case and declension, but not about number or gender. More specifically, the affixes which they take are the ones of declension and case, but not of number or gender.

The forms with a case affix are genitive or dative. More specifically, the *-s* forms are genitives and have the value *nondeclined*. The nominals which they select must have the NUMGEN value *sg-masc* or *sg-neu*, as in *onzes inziens* 'our-GEN opinion-GEN'. The *-er* forms are also genitive and nondeclined, but require a nominal which is either singular feminine or plural, as in *één dezer dagen* 'one this-GEN days'. They can also be dative, and in that use they require a singular feminine nominal, as in *te zijner ere* 'to his-DAT honour-DAT'. The *-en* forms are datives. They are *declined* and require the nominal to have a NUMGEN value which is different from *sg-fem*, as in *te mijnen behoeve* 'to my-DECL-DAT need-DAT'.

Because of the unification of the SELECT|LOCAL|CAT value of the determiner with the CAT value of the selected nominal, the addition of a determiner may have the effect of resolving underspecification. The noun *dagen* 'days', for instance, has the underspecified value *case*, but *deze dagen* 'this-DECL days' can only be *standard* and *dezer dagen* 'this-GEN days' is unambiguously *genitive*. The same applies to gender. As pointed out in Haeseryn et al. (1997, 155–159), Dutch has many nouns, which can be both neuter and nonneuter, sometimes—but not always—with different meanings, as in the neuter *het pad* 'the path' vs. the nonneuter *de pad* 'the toad'. If such a noun is combined with a declined determiner, as in *deze pad*, it is unambiguously identified as nonneuter.

3.2   The Interrogative and Quantifying Determiners

The interrogative and quantifying determiners differ from the demonstrative and possessive determiners in at least three respects. First, the morphosyntactic agreement also concerns declension: they only require the selected nominal to be declined if they are declined themselves. Second, they only combine with bare nominals: an inherently incomplete nominal, such as *zwarte paard* 'black-DECL horse', is compatible with a possessive or a demonstrative, as in *mijn zwarte paard*, but not with an interrogative or quantifying determiner, as in * *welk/geen zwarte paard* 'which/no black-DECL horse'. Third, their MARKING value is different.

$$
\left[
\begin{array}{l}
\text{HEAD}
\left[
\begin{array}{l}
\text{CASE}\quad \boxed{1}\,case \\
\text{DECL}\quad \boxed{2}\,declension \\[2ex]
\text{SELECT} \mid \text{LOCAL} \mid \text{CAT}
\left[
\begin{array}{ll}
\text{HEAD}
\left[
\begin{array}{l}
\text{CASE}\quad \boxed{1} \\
\text{DECL}\quad \boxed{2}
\end{array}
\right]_{noun} \\[3ex]
\text{MARKING}\ \Big[\text{NUMGEN}\quad numgen\Big]_{bare}
\end{array}
\right]
\end{array}
\right] \\[6ex]
\text{MARKING}\quad interrogative \vee quantified
\end{array}
\right]
$$

The values for CASE, DECLENSION and NUMGEN covary with the morphological form in the same way as in the case of the other determiners.

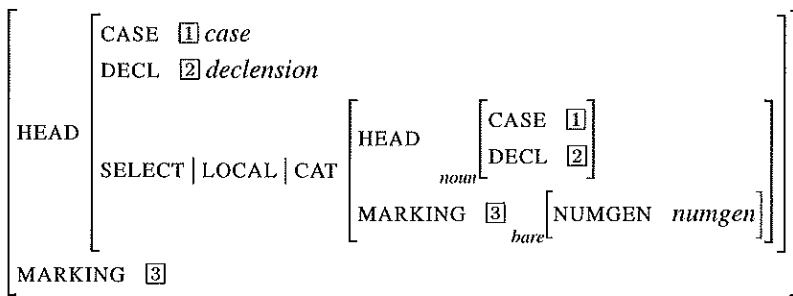| stem	[+D]	[+C]	[+C]	[+D,+C]	
welk	*welke*				which
elk	*elke*				each
ieder	*iedere*				every
geen		*geens-*	*gener*	*genen*	no
enig	*enige*	*enigs-*	*eniger*		any

The only major difference is that the *-s* forms are invariably incorporated, as in *enigszins* 'any-GEN-way-GEN' and *geenszins* 'no-GEN-way-GEN'.

3.3   The Prenominal Adjectives

In terms of morphosyntactic agreement, the prenominal adjectives have much in common with the interrogative and quantifying determiners: there is, for instance, the sharing of both the CASE value and the DECLENSION value, and the paradigm of forms is identical.

stem	[+D]	[+C]	[+C]	[+D,+C]	
goed	*goede*	*goeds-*	*goeder*	*goeden*	good
koel	*koele*			*koelen*	cool
open					open

The main difference concerns the MARKING value. For most of the prenominal adjectives this is not (a subtype of) *marked*, but rather the same value as the one of the selected nominal. From this it follows that they can be stacked.

$$
\left[
\begin{array}{l}
\text{HEAD}
\left[
\begin{array}{l}
\text{CASE}\quad \boxed{1}\,case \\
\text{DECL}\quad \boxed{2}\,declension \\[2ex]
\text{SELECT} \mid \text{LOCAL} \mid \text{CAT}
\left[
\begin{array}{ll}
\text{HEAD}
\left[
\begin{array}{l}
\text{CASE}\quad \boxed{1} \\
\text{DECL}\quad \boxed{2}
\end{array}
\right]_{noun} \\[3ex]
\text{MARKING}\ \boxed{3}\,\Big[\text{NUMGEN}\quad numgen\Big]_{bare}
\end{array}
\right]
\end{array}
\right] \\[6ex]
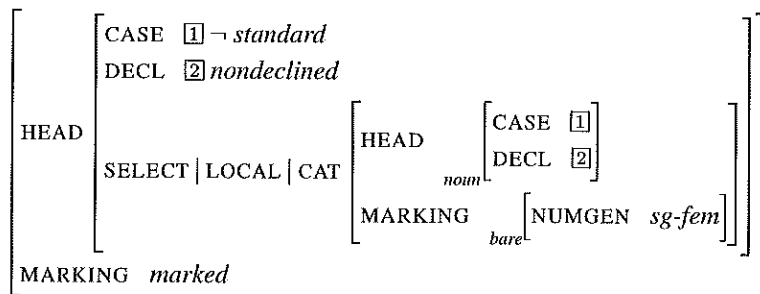\text{MARKING}\quad \boxed{3}
\end{array}
\right]
$$

As usual, the forms without case affix have the CASE value *standard* and share this value with the selected nominal. Moreover, the [+D] forms have the value *declined* and hence require a singular nonneuter or plural nominal, as in *grote zwarte ezel(s)* 'large-DECL black-DECL donkey(s)', whereas the forms without declension affix have the value *nondeclined* if they have a [+D] counterpart, and the underspecified value *declension* otherwise. In the former case they require a singular neuter nominal; in the latter they combine with any nominal in standard case.

A special case are the combinations in which an adjective without declension affix occurs in singular nonneuter and plural nominals, even though it has a [+D] counterpart. A relevant example is the combination *industrieel ingenieur* 'industrial engineer'. In spite of the existence of the [+D] form *industriële* and in spite of the fact that *ingenieur* is masculine, the adjective is not declined. This anomaly correlates with a semantic idiosyncracy: in this combination the adjective and the noun do not denote distinct properties of the same entity; instead, they form a single multi-word expression which denotes one property. A criterion for identifying such multi-word expressions is the relative clause test: while a regular [Adj N] combination, such as *zuiver water* 'clean water', can be paraphrased in terms of a relative clause, as in *water dat zuiver is* 'water that is clean', a multi-word expression, such as *een industrieel ingenieur* 'an industrial engineer', cannot be paraphrased as *een ingenieur die industrieel is* 'an engineer that is industrial'. This demonstrates that the declension distinction loses its relevance in multi-word expressions; in terms of the TFS notation, the adjectives in such expressions have the underspecified value *declension*, just like the ones which lack a [+D] counterpart. For more examples, see Haeseryn et al. (1997, 408–412).
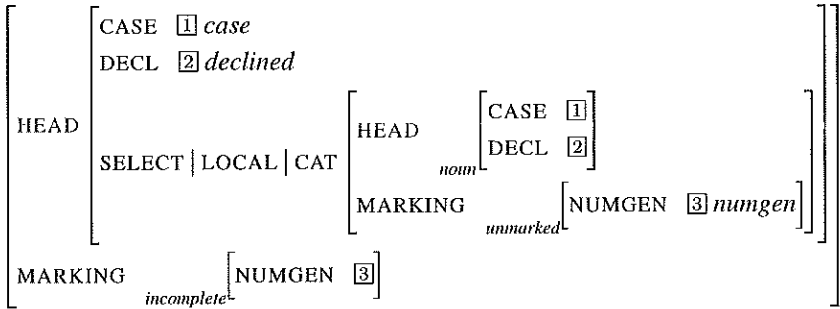
As for the forms with a case affix, the ones with the *-en* affix are declined and occur in singular nonfeminine datives, as in *van goeden huize* 'of good-DECL-DAT house-DAT' and *ten eeuwigen dage* 'to-the-DAT eternal-DECL-DAT day-DAT'. The *-er* forms are special, in that they cannot be preceded by any other prenominal, as illustrated by the genitive in *(*de) zaliger gedachtenis* '(*the) holy-GEN remembrance' and the dative in *van (*mijn) ganser harte* 'of (*my) whole-DAT heart-DAT'. Their MARKING value is, hence, of type *marked*.[11]

$$
\begin{bmatrix}
\text{HEAD} & \begin{bmatrix}
\text{CASE} & \boxed{1} \neg \textit{standard} \\
\text{DECL} & \boxed{2}\,\textit{nondeclined} \\
\text{SELECT} \mid \text{LOCAL} \mid \text{CAT} & \begin{bmatrix}
\text{HEAD}_{\textit{noun}} & \begin{bmatrix} \text{CASE} & \boxed{1} \\ \text{DECL} & \boxed{2} \end{bmatrix} \\
\text{MARKING}_{\textit{bare}} & \begin{bmatrix} \text{NUMGEN} & \textit{sg-fem} \end{bmatrix}
\end{bmatrix} \\
\text{MARKING} & \textit{marked}
\end{bmatrix}
\end{bmatrix}
$$

---

[11] The NUMGEN value is constrained to *sg-fem*, since the *-er* forms of the adjectives are not used in plural genitives.

The prenominal *-s* forms are invariably incorporated, as in *goedsmoeds* 'good-GEN-courage-GEN' and *blootsvoets* 'bare-GEN-foot-GEN'.

While the above is a straightforward extension of the determiner system to the prenominal adjectives, it covers only part of the latter's uses. What it does not cover are some secondary uses of the [+D] forms, i.e. the ones with the affixes *-e* and *-en*. As a first example, let us take the use of the *-e* form in *zwarte paard* 'black-DECL horse'. This use of *zwarte* is not covered by the standard AVM of the *-e* forms, since *paard* is neither plural nor nonneuter. To accomodate it, one could drop the NUMGEN constraint on the [+D] forms, but this would be misleading, for while *zwarte ezels* 'black-DECL donkeys' and *zwarte olie* 'black-DECL oil' are impeccable combinations as they stand, *zwarte paard* 'black-DECL horse' is not. Instead, it is only well-formed, if it is preceded by a determiner of type *determined*, as in *ons zwarte paard* 'our black-DECL horse'. To make this explicit in the notation, I give it the MARKING value *incomplete*.

$$
\begin{bmatrix}
\text{HEAD} \begin{bmatrix}
\begin{bmatrix}
\text{CASE} & \boxed{1}\ case \\
\text{DECL} & \boxed{2}\ declined
\end{bmatrix} \\
\text{SELECT} \mid \text{LOCAL} \mid \text{CAT} \begin{bmatrix}
\text{HEAD}_{noun} \begin{bmatrix}
\text{CASE} & \boxed{1} \\
\text{DECL} & \boxed{2}
\end{bmatrix} \\
\text{MARKING}_{unmarked} \begin{bmatrix}
\text{NUMGEN} & \boxed{3}\ numgen
\end{bmatrix}
\end{bmatrix} \\
\text{MARKING}_{incomplete} \begin{bmatrix}
\text{NUMGEN} & \boxed{3}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Since *incomplete* is a subsort of *unmarked*, stacking is correctly allowed, as in *ons kleine zwarte paard* 'our small-DECL black-DECL horse'. To turn the incomplete nominal into a complete one the only option is to add a demonstrative or possessive determiner; adding an interrogative or quantifying determiner is impossible since they require a bare nominal, and adding a prenominal adjective is insufficient, since the ones which are compatible with an incomplete nominal are precisely those which yield an incomplete nominal again.

The secondary use of the *-e* forms is not limited to nominals in standard case; it also occurs in nominals with an *-er* determiner, as in the genitive adjuncts of *woordenboek der Nederlandse taal* 'dictionary the-GEN Dutch-DECL language' and *problemen der rijke landen* 'problems the-GEN rich-DECL countries', and in the dative *ter meerdere eer en glorie* 'to-the-DAT more-DECL honour and glory'. Normally, one would expect the *-er* form of the adjective in these positions, but since this form may not be preceded by any other prenominal, it has to be replaced by another form, and this happens to be the *-e* form. Also here, stacking is allowed, as in *der kleine zwarte paarden* 'the-GEN small-DECL black-DECL horses'.

The secondary use of the *-en* forms is also due to a gap in the 'regular' system: since the *-s* forms of the prenominal adjectives are invariably incorporated, one has to employ another form if the adjective is a separate word. As exemplified

by the genitive adjunct in *de geneugten des goeden levens* 'the pleasures the-GEN good-DECL-GEN life-GEN', it is the *-en* form which fulfills this function.[12] Since this form must be preceded by the *-s* form of a determiner of type *determined*, the resulting nominal is inherently incomplete.

In sum, the combination of a prenominal adjective with a nominal can have three possible MARKING values. In the majority of cases, it will be *bare*, which means that it may but need not take a determiner, but the *-er* forms are *marked* and the [+D] forms have some secondary uses in which they have the value *incomplete*. The postulation of a systematic ambiguity for the [+D] forms may be seen as undesirable. However, the alternative of making the AVMs so general that they subsume both uses will inevitably lead to massive overgeneration. Whether this is acceptable depends on one's goals: for a parser it might be acceptable, but for a grammar checker or a generator one needs to distinguish the nominals which are inherently incomplete and hence ill-formed from the nominals which are bare or marked.

3.4 Summing Up

In Dutch NPs prenominals show agreement with the nominals they select in terms of case, declension, number and gender. To model this I have included the relevant features in the CAT values of the (pre)nominals. The resulting treatment is similar to the one which Kathol (1999, 251–263) proposes for German, but differs from it in two respects. One concerns the conflation of the number and gender distinctions in one feature called NUMGEN. The other concerns the status of this feature: instead of including it in the HEAD value of all (pre)nominals, along with CASE and DECLENSION, as in Kathol (1999), I include it in the MARKING value of the unmarked nominals. This difference will be motivated in section 4.

4 NP-Internal Index Agreement

As pointed out in the introduction, the indices contain information about person, number and gender.

$$\begin{bmatrix} \text{PERSON} & person \\ \text{NUMBER} & number \\ \text{GENDER} & gender \end{bmatrix}_{index}$$

The inventories of values are familiar.

person			number		gender		
1	*2*	*3*	*singular*	*plural*	*masculine*	*feminine*	*neuter*

---

[12] Also this form is sometimes incorporated, as in *'s anderendaags* 'the-GEN other-GEN-day-GEN' and *grotendeels* 'large-GEN-part-GEN'.

For most nominals, the values of NUMBER and GENDER correspond to the MARKING|NUMGEN values, but there are exceptions. The NUMGEN value of *meisje* 'girl-DIM', for instance, is unambiguously *sg-neu*, but its GENDER value can be either *neuter* or *feminine*, see (3–5). Another difference is that the NUMGEN feature only occurs in the AVMs of the unmarked nominals, whereas the NUMBER and GENDER values in the index are shared throughout the nominal projection, and hence also by the marked nominals. So far, this was mainly a matter of stipulation; in this section it will be motivated on empirical grounds.

4.1   Marked Nouns

The examples in the previous sections were all NPs with an unmarked common noun as their head. When extending the treatment to proper nouns, it is important to distinguish between two kinds. On the one hand, there are the proper nouns which do not take any prenominals to form a full NP; they include names of people (*Jan, Lea*), towns (*Brussel, Geraardsbergen*) and countries (*Duitsland, Denemarken*). On the other hand, there are the proper nouns which require the presence of a determiner; they include names of geographical regions (*het Hageland*), mountains (*de Pellenberg*), waterways (*de Schelde, het Albertkanaal*), lakes (*het Zilvermeer*) and seas (*de Noordzee*); they also include a large class of plural proper nouns (*de Schoolbergen, de Kempen, de Ardennen*). Since the determiner can be separated from the noun by any number of prenominal adjectives, as in *de stille Kempen* 'the quiet-DECL Kempen' and *het bosrijke en gastvrije Hageland* 'the forest-rich-DECL and hospitable-DECL Hageland', the combination of the determiner and the proper noun cannot be treated as a single word. Instead, these proper nouns have a MARKING value of type *unmarked* and require the presence of an appropriate determiner to become *marked*. In this respect, they differ from the pronouns of the first type, which are inherently *marked* or—more specifically—*determined*.

As a consequence, we expect that the unmarked proper nouns are subject to the same kind of morphosyntactic agreement as the common nouns. This expectation is indeed borne out. For a start, their NUMGEN value has to be compatible with the constraints which are imposed by the prenominals. Compare, for instance, the singular neuter *het/*de Hageland* with the singular nonneuter *de/*het Pellenberg* and the plural *de/*het Kempen*. The similarity with the unmarked common nouns is especially clear in compound proper nouns, for in such cases, the NUMGEN value of the noun is identical to the one of its common noun head: *Hageland*, for instance, is singular neuter, since the common noun *land* 'land' is singular neuter, and *Schoolbergen* is plural, since *bergen* 'mountains is plural. Second, there is obligatory case concord: prenominals which combine with an unmarked proper noun are subject to case agreement, as in the genitive of *de wijngaarden* [*des/*het Hagelands*] 'the vineyards the-GEN/*the Hageland-GEN'.

Also in terms of index agreement we expect the unmarked proper nouns to side with the unmarked common nouns. This implies that the number and gender values in the index may but need not correspond to the NUMGEN value. As an

example, let us take *Ardennen*. Its NUMGEN value is unambiguously plural, but its index may also be singular, if it is presented as a single entity.

(6)   De   Ardennen liggen/?ligt in België.
      The Ardennes lie/?lies   in Belgium
      'The Ardennes are in Belgium.'

(7)   De   Ardennen is/?zijn een mooie   streek.
      The Ardennes is/?are a   beautiful region
      'The Ardennes is a beautiful region.'

This is comparable to the gender values of *meisje* 'girl'.

The marked proper nouns, by contrast, are not subject to morphosyntactic agreement. If they are preceded by a prenominal, as in *het/*de rustige Geraardsbergen* 'the quiet-DECL Geraardsbergen', then the choice of the determiner is exclusively determined by the index value of the noun, and not by what its morphology suggests. In this case, for instance, the presence of the plural affix in *Geraardsbergen* is not relevant, for if it were, the determiner would be *de* instead of *het*. Instead, what matters is the NUMBER and GENDER values in the index, and these are determined on semantic grounds: all of the marked proper nouns are singular, since they are names of single entities, and their gender value is determined by the natural gender of the referent, i.e. feminine for reference to female individuals, masculine for reference to male individuals, and neuter for reference to inanimate entities, such as towns and countries. From this it follows that the NUM-GEN feature has no relevance for the analysis of the marked nouns. An interesting example to further illustrate this point concerns the noun *Pellenberg*. When it is the name of a hill, it must be preceded by a determiner and has the NUMGEN value *sg-masc*, just like the common noun *berg* 'mountain'; this accounts for *de/*het steile Pellenberg* 'the steep-DECL Pellenberg. The same noun, though, is also used as the name of the village in which that hill is situated, and in that use it is marked; this implies that it lacks the NUMGEN value and that it combines with prenominals which select a nominal with a singular neuter index, as in *het/*de rustige Pellenberg* 'the quiet-DECL Pellenberg'.[13] The absence of morphosyntactic agreement is further confirmed by the lack of case concord. In contrast to the unmarked nouns, the marked proper nouns do not show case agreement with their prenominal dependents. In genitives, for instance, the determiner takes the standard form, as in [*onze/*onzes Karels*] *fiets* 'our/*our-GEN Karel-GEN bike'; this is remarkable, since the possessive has a special form for the genitive.[14]

The contrast with the unmarked nouns also shows in the case of index agreement. Whereas grammatical and semantic considerations may be in competition in

---

[13] The same phenomenon can be observed in Italian, where names of towns are invariably treated as singular feminine, presumably because the noun *città* 'town' is feminine. *Torino*, for instance, has the typical ending of a singular masculine noun, but if it takes any prenominals, they must be feminine, as in *la bella Torino* 'the-FEM-SG beautiful-FEM-SG Torino'.

[14] In spite of the absence of case concord, the CASE value of the proper noun has to be included in its HEAD feature, since it is relevant for the external distribution of the NP.

the determination of the index values of the unmarked nouns, the index values of the marked proper nouns are exclusively determined by semantic considerations. Treating the name *Geraardsbergen* as anything else than singular and neuter is not just unusual, but downright ungrammatical.

(8)   Geraardsbergen ligt/*liggen ten   westen van Brussel.
      Geraardsbergen lies/*lie   to-the west   of   Brussels.

(9)   Waar   ligt Geraardsbergen? Het/*hij/*zij ligt in Vlaanderen.
      Where lies Geraardsbergen? It/*he/*she   lies in Flanders.

In sum, while the unmarked proper nouns show the same properties as the unmarked common nouns in terms of morphosyntactic agreement and index agreement, the marked proper nouns behave radically different. This difference is made explicit by the absence of the NUMGEN feature in their AVMs. Further evidence for this absence is provided by the genitive -*s* forms. In the case of the unmarked nouns, it is only the singular nonfeminine nouns which take this affix, as in the masculine *'s middags* 'the-GEN noon-GEN' and the neuter *de heer des huizes* 'the lord the-GEN house-GEN'; feminine nouns and nouns with a plural affix cannot take the -*s* affix and hence lack separate forms for the genitive. In the case of the marked proper nouns, by contrast, the NUMGEN restriction is apparently irrelevant, since these nouns have also -*s* forms for nouns with a plural affix, as in *Denemarkens inwoners* 'Denmark's inhabitants', and for nouns with a feminine index, as in *Lea's fiets* 'Lea's bike'.
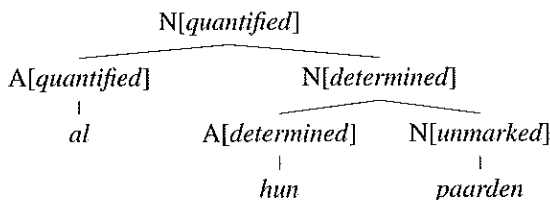
Interestingly, there is a small set of common nouns, which—in some of their uses—behave as marked. They include a number of nouns which are often used as vocatives, such as *vader* 'father' and *moeder* 'mother'. When they are used as unmarked common nouns, they are subject to the familiar patterns of morphosyntactic agreement, as in *elke/*elk vader* 'each-DECL father' and *het huis [mijns/ *mijn vaders]* 'the house my-GEN father-GEN'. In that use, they are also incompatible with the genitive -*s* affix, when they are plural or feminine. By contrast, when they are used as marked nouns, they do not show case agreement, as in *[mijn/*mijns vaders] huis* 'my father-GEN house', and they have a separate form for the genitive, even when plural or feminine. Compare, for instance, *[mijn moeders] huis* 'my mother-GEN house' with * *het huis [mijns/mijner moeders]* 'the house my-GEN mother-GEN'.

From a cross-linguistic point of view, the special status of these nouns is not surprising. Their Italian equivalents, for instance, also show some special properties: while the vast majority of the common nouns take a determiner in combination with the possessive, as in *il mio giardino* 'the my garden, these exceptional nouns do not, cf. *(*il) mio zio* '(*the) my uncle'. This can be seen to follow from the fact that the article requires an unmarked nominal, whereas the noun *zio* 'uncle' is inherently marked in this use.

Summing up, there is a large set of proper nouns and a small set of common nouns which are inherently marked. They are exempt from the usual constraints on morphosyntactic agreement, but they show index agreement. Moreover their index values are exclusively determined by semantic considerations.

4.2   Predeterminers

An interesting test case for the claims in the previous paragraph are the nominals
with a predeterminer, as in *al hun paarden* 'all their horses' and *(ge)heel de stad*
'whole the town'. For the analysis of such NPs I assume a right branching structure:

```
                        N[quantified]
                _____|_____
          A[quantified]            N[determined]
                |              _____|_____
                al        A[determined]    N[unmarked]
                               |               |
                              hun           paarden
```

The reason for preferring it over the left branching [[*al hun*]*paarden*] is that
the first chunk in this structure is an artificial construct which does not occur in
any other environment, whereas the constituents in the right branching structure
are independently motivated. Notice, for instance, that the predeterminer can be
separated from the rest of the NP, whereas the first chunk in the left branching
structure cannot:

(10)   Geheel de stad is omsingeld.
        entire   the town is surrounded

(11)   De stad is geheel   omsingeld.
        the town is entirely surrounded

(12)   * Stad is geheel de omsingeld.
        * town is entirely the surrounded

The predeterminers are, hence, functors which select a nominal of type *determined*
and which yield a nominal of type *quantified*.

Since the nominals of type *determined* are marked, we expect there to be no
morphosyntactic agreement between the predeterminer and the rest of the NP. In-
terestingly, this expectation is indeed borne out. Compare, for instance, the deter-
miner in *proletariërs aller landen* 'proletarians all-GEN countries' with the prede-
terminer in *de proletariërs al/*aller dezer landen* 'the proletarians all these-GEN
countries': while the former shows agreement in number and case, the latter is
morphologically invariant and shows no such agreement. Equally telling is the
contrast between the adjective *heel* and the homonymous predeterminer. The for-
mer is subject to morphosyntactic agreement, as illustrated by the contrast between
*een heel/*hele brood* 'a whole/*whole-DECL bread' and *het hele/*heel brood* 'the
whole-DECL/*whole bread', but the latter is morphologically invariant and shows
no such agreement *heel/*hele het brood*.

At the same time, the predeterminers do show index agreement. In combination
with count nouns, for instance, *al* requires an NP with a plural index, whereas
(ge)heel requires an NP with a singular index. This also accounts for the fact that
the latter is compatible with marked proper nouns, whereas the former is not, cf.
*heel/*al Denemarken* 'whole/all Denmark'.

In sum, the data about the predeterminers confirm the observation that the combination of a prenominal with a marked nominal is not constrained by morphosyntactic agreement, but by index agreement.

5   Conclusion

For the treatment of agreement in Dutch NPs I have adopted the distinction, familiar from HPSG, between morphosyntactic agreement and index agreement (section 1). In order to determine their respective roles, I have made a distinction, proposed in Allegranza (1998) and Van Eynde (2003), between marked and unmarked nominals (section 2). Combining these two distinctions, the conclusions of this paper can be summed up as follows. First, the combination of prenominals with unmarked nominals is subject to morphosyntactic agreement in case, declension, number and gender (section 3). Second, the combination of prenominals with marked nominals is not subject to morphosyntactic agreement, but to index agreement (section 4). The latter provides evidence against the commonly made assumption that index agreement only applies NP-externally.

References

Allegranza, V. (1998), Determiners as functors: NP structure in Italian, *in* S. Balari and L. Dini (eds), *Romance in HPSG*, CSLI Publications, Stanford, pp. 55–107.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. and van den Toorn, M. (1997), *Algemene Nederlandse Spraakkunst*, Nijhoff and Wolters Plantijn, Groningen and Deurne.

Kathol, A. (1999), Agreement and the syntax-morphology interface in HPSG, *in* R. Levine and G. Green (eds), *Studies in Contemporary Phrase Structure Grammar*, Cambridge University Press, Cambridge, pp. 223–274.

Martínez, L. C. (2003), Gender mismatches in Spanish and French N_1/A de N_2, *in* J.-B. Kim and S. Wechsler (eds), *The Proceedings of the 9th International Conference on HPSG*, CSLI Publications, Stanford, pp. 1–17.

Oostdijk, N., Goedertier, W., Van Eynde, F., Boves, L., Martens, J.-P., Moortgat, M. and Baayen, H. (2002), Experiences from the Spoken Dutch Corpus Project, *in* M. G. Rodríguez and C. P. S. Araujo (eds), *Proceedings of the Third International Conference on Language Resources and Evaluation*, ELRA, pp. 340–347.

Pollard, C. and Sag, I. (1994), *Head-driven Phrase Structure Grammar*, CSLI Publications and University of Chicago Press, Stanford and Chicago.

Van Eynde, F. (1998), The immediate dominance schemata of HPSG, *in* P.-A. Coppen, H. van Halteren and L. Teunissen (eds), *Computational Linguistics in the Netherlands 1997*, Rodopi, Amsterdam, pp. 119–133.

Van Eynde, F. (2003), Prenominals in Dutch, *in* J.-B. Kim and S. Wechsler (eds), *Proceedings of the 9th International Conference on Head-driven Phrase Structure Grammar*, CSLI Publications, Stanford, pp. 335–356.

Wechsler, S. and Zlatić, L. (2000), A theory of agreement and its application to Serbo-Croatian, *Language* **76**, 799–832.