

Manual

*Tagging protocol**

TABLE OF CONTENTS

0. INTRODUCTION	3
1.1 Flection	3
1.2 Position	4
1.3 Case.....	4
1.4 Person.....	4
1.5 Number	4
1.6 Gender.....	5
1.7 Function	5
2. ADJECTIVE	5
2.1 Inflection.....	5
2.2 Position.....	6
2.3 Degrees of comparison	7
3. VERB	7
3.1 Inflection	7
3.2 Position.....	7
3.2.1 Regular order	7
3.2.2 Inversed order.....	8
3.2.3 Position right periphery.....	8
3.2.4 Other positions.....	9
3.3 Verb-clitic-cluster.....	9
3.4 Features	9
3.4.1 Finite	9
3.4.1.1 Present tense, indicative	9
3.4.1.2 Present tense, conjunctive	9
3.4.1.3 Imperative.....	9
3.4.1.4 Past tense indicative	9
3.4.1.5 Past tense conjunctive	10
3.4.1.6 Past tense imperative.....	10
3.4.2 Non-finite	10
3.4.3 Participles	10
3.5 Type.....	10
3.5.1 Auxiliary.....	10
3.5.2 Matrix.....	10
3.5.2.1 Inherent reflexive.....	11
3.5.2.2 Transitive verb.....	11
3.5.2.3 Intransitive.....	11

* This tagging protocol is a manual for performing Parts of Speech tagging. It was developed by Sjeff Barbiers and Guido Vanden Wyngaerd, for the SAND-project (Syntactic Atlas of Dutch Dialects), but can be useful for other dialect research groups/ projects.

4	PRONOUNS	11
4.1	Inflection	11
4.2	Position	12
4.3	Case.....	12
4.4	Person and number	13
4.5	Gender.....	13
4.6	Function	13
4.8	Type.....	13
4.8.1	Personal	14
4.8.1.1	Subject doubling	14
4.8.1.2	Strong and weak (non doubling).....	14
4.8.1.3	Expletive.....	14
4.8.2	Reflexive and reciprocal.....	15
4.8.3	Possessive	15
4.8.5	Interrogative	15
4.8.6	Relative.....	15
4.8.7	R-pronouns	15
4.8.8	Quantifiers.....	16
4.8.9	Determiners	16
5.	ADPOSITIONS: PREPOSITIONS AND POSTPOSITIONS.....	16
5.1	Inflection	16
5.2	Position	16
6.	COMPLEMENTIZERS	17
6.1	Inflection	17
6.2	Complementizer-clitic-cluster.....	18
6.3	Complementizer type.....	18
7.	ADVERBS	19
7.1	Inflection	19
7.2	Position	19
7.3	Type.....	20
7.3.1	Interrogative	20
7.3.2	Quantifier.....	20
7.4	R-pronouns (adverbial).....	20

0. INTRODUCTION

This document provides an overview of the tags that were used during the labeling of parts of speech in the SAND-project (Syntactic Atlas of Dutch Dialects). Every tag is represented both with a number code and with one or more capitals. In the tagging application, which assigns the tags semi automatically, the number and capital codes are two alternative but equivalent ways to assign a tag to the transcription.

Every tag has the following format: Category, Attribute Value Specification Specification. For example: V FEAT FIN PT 1.PL; Category = V (verb); Attribute = FEAT (features); Value = FIN (finite); Specification = PT (present tense); Specification = 1.PL (first person plural). Category, attribute, value and specification are marked in capitals. If these capitals are between brackets, it means that the specific part is optional.

Every tag corresponds to a five digit code. The structure is as follows. The first digit indicates the category (for example 1 = N, i.e. noun). The second digit indicates the attribute (for example 3 = CASE, i.e. case). The third number marks the value of the attribute (for example 1 = OBLIQUE). The fourth and fifth numbers specify the value (for example 2 = DAT, i.e. dative). A zero marks depending on its position no category/ no attribute/ no value/ no specification.

De number code 13120 thus corresponds with the tag N CASE OBLIQUE DAT.

Every tag is followed by a short description of category/ attribute/ value/ specification and, if necessary, an illustration (examples).

In most cases it will be necessary to assign more than one number code to a word. For example: *blackberries* in *a bucket (of) blackberries* gets the code 11100: N INFL –ies (noun with flection –ies), plus the code 12300: N POS POST-N (noun in post nominal position).

The tagset is inspired by the tagset used in Corpus Gesproken Nederlands (Corpus Spoken Dutch – F. Van Eynde, *Part of Speech Tagging en Lemmatisering*, Centre for Computer linguistics, K.U. Leuven, 2000.), that is based on the EAGLES standard for tagsets. The SAND tagset differs in a number of ways. These differences will be mentioned and illustrated in this document, whenever necessary.

0. UNCERTAINTY TAG

00000	O	No tag Use this code if the category of the word is not clear. This code is not to be used if the category is clear, but the attribute, value or specification is not. In this latter case, the 0 is to be inserted at the position of attribute/ value/ specification.
-------	---	---

1. NOUN

10000	N	Noun
-------	---	------

1.1	Flection	
11000	N(INFL)	Inflection. The existence of this attribute indicates the occurrence of an audible morphological marking for categories like: person, number, gender, case, definiteness, etc. Inflection doesn't refer to zero morphemes or diminutive suffixes. The following rule can apply:

flection is only tagged as such, if the word can also occur without the morpheme. The following tags give information about the inflection found:

11100	N (INFL) –(e)n	
11200	N (INFL) –(e)t	
11300	N (INFL) –e	
11400	N (INFL) –(e)s	
11500	N (INFL) –st	
11600	N (INFL) OT	Other morphological marking
1.2	Position	
12000	N POS	Position
12100	N POS PRE-N	Prenominal For example: <i>a few books</i> . We call <i>few</i> prenominal because <i>books</i> is the head of the phrase, as follows from the agreement on the finite verb (<i>there are a few books on the table</i>).
12200	N POS N	Nominal This is the ‘ordinary’ use of the noun, as the head of an (argumentative) noun phrase
12300	N POS POST-N	Post nominal For example: <i>a bucket (of) blackberries</i> . We call <i>blackberries</i> postnominal, because <i>bucket</i> is the head of the phrase, as follows from the agreement on the finite verb (<i>There is a bucket at the table</i>).
12400	N POS FREE	Tag for nouns that are not part of an argumentative noun phrase.
12410	N POS FREE PRED	Predicative For example: <i>John is (a) doctor/ mayor</i> .
12420	N POS FREE ADV	Adverbial For example: <i>Zondags gaat zij naar de kerk</i> , ‘[lit.:] Sundays goes she to the church’.
1.3	Case	
13000	N (Case)	Case Case is assigned to nouns only if the flection is audible. So, nominal nouns don’t get the attribute case.
13100	N (CASE)(OBL)	Oblique All nouns that have audible case flection, that isn’t genitive, will be assigned the value ‘oblique’. This value can be specified as ‘accusative’ or ‘dative’. If it’s not clear whether a constituent is accusative or dative, oblique will not be specified.
13110	N (CASE)(OBL)(ACC)	Accusative
13120	N (CASE)(OBL)(DAT)	Dative
13200	N (CASE)(GEN)	Genitive Value for nouns with genitive flection.
1.4	Person	
14000	N Person	Person
14300	N Person 3	All nouns are third person. This tag will be the default.
1.5	Number	
15000	N Number	

15100	N Number S	Singular
15200	N Number Pl	Plural
1.6 Gender		
16000	N (GENUS)	Grammatical gender As the gender of a word is not always clear, this attribute is optional (can but need not be assigned).
16100	N (GENUS) Z	non-neuter
16110	N (GENUS) Z (M)	masculine
16120	N (GENUS) Z (F)	feminine
16200	N (GENUS) Neu	neuter

1.7 Function

The function of a DP (determiner phrase, noun phrase) is normally not incorporated in the tagging process. However, the assignment of the values subject (SUBJ), object (D-OBJ), indirect object (I-OBJ) and prepositional object (P-OBJ) is essential in order to be able to search the database, at a stage where (full) syntactic annotation is lacking. This value is assigned to the noun.

17000	N (FUNCTION)	Grammatical function
17100	N (FUNCTION) SUBJ	Subject Is assigned if the DP agrees with the finite verb in person/ number
17200	N (FUNCTION) D-OBJ	Direct object Applies when DP is the direct object.
17300	N (FUNCTION) I-OBJ	Indirect object Is assigned when DP is indirect object (without a preposition)
17400	N (FUNCTION) P-OBJ	Object of preposition Applies when DP is the complement of a preposition.

2. ADJECTIVE

20000 A

2.1 Inflection

21000	A (INFL)	Inflection This attribute indicates the existence of an audible morphological marking for categories like: person, number, gender, case, etc. The <i>s</i> (E.g. <i>iets moois</i> , 'something beautiful') is counted as flection. This is not the case for degrees of comparison (comparative, superlative), zero morphemes, diminutive suffixes. The following rule applies: inflection is only tagged as such, if the word can also occur without the morpheme. This attribute has the following values:
21100	A (INFL) -(e)n	
21200	A (INFL) -(e)t	
21300	A (INFL) -e	
21400	A (INFL) -(e)s	
21500	A (INFL) -st	
21600	A (INFL) OT	Other inflectional morphology

2.2	Position	
22000	A POS	Position
22100	A POS PRE-N	Prenominal For example: <i>a <u>beautiful</u> book</i> . Words like <i>many</i> and <i>few</i> are categorized as adjectives (not as quantifiers), because they can be used in degrees of comparison, and with adverbs of degree (<i>very, extremely, almost</i>).
22110	A POS PRE-N ELL	Prenominal with ellipsis As, for example in the second conjunct of ‘ <i>He has a white plate and I a green (one)</i> . The sentence doesn’t need to contain the antecedent per se: <i>Ik heb gisteren een witte telefoon gekocht. Hij past beter bij het interieur dan de groene</i> , ‘I bought a white telephone yesterday. It fits into the interior better than the green (one)’.
22200	A POS N	Nominal Nominal (or substantive used) adjectives are not treated as substantives, but as adjectives. Arguments in favour of this method are a.o.: the existence of comparative and superlative forms (E.g. <i>de <u>ouderen</u></i> , ‘the elderly’, <i>de <u>rijksten</u></i> , ‘the richest (people)’), the compatibility with adverbs of degree (<i>de <u>zeer rijken</u></i> , ‘the very rich’), and the fact that plural marking is different with nominal adjectives than with substantives.
22300	A POS POST-N	Postnominal E.g. <i>kindeke <u>teer</u></i> , ‘child fragile’, <i>alle rivieren bevaarbaar in de winter</i> , ‘all rivers navigatable in winter’, <i>niets <u>bijzonders</u></i> , ‘nothing special’, <i>iets <u>groters</u></i> , ‘something bigger’.
22400	A POS FREE	Adjectives that are not part of a DP
22410	A POS FREE PRED	Predicative Predicative adjectives are adjectives that function as a subject complement, E.g. <i>het schip is <u>schoon</u></i> , ‘the ship is clean’, or as secondary predicate, in other words, as predicative adjunct. There are three types of predicative adjuncts: <i>Hij veegt het schip <u>schoon</u></i> , ‘he wipes the ship clean’ (resultative), <i>Hij vindt Marie <u>aardig</u></i> , ‘[lit.:]He finds Mary nice (He likes Mary)’ (predicative), <i>Hij gaf de tas <u>leeg terug</u></i> , ‘he returned the bag empty’ (depictive). Separable prefixes (from verbs) are also assigned this tag, if they have the same form as an adjective, whether they are separated from the verb or not (E.g. <i>Hij drinkt de beker <u>leeg</u></i> , lit: ‘He drinks the mug empty’, <i>dat hij de beker <u>leeg</u> drinkt</i> , lit: ‘that he the mug empty drinks’, <i>dat het venster open waait</i> , ‘that the window open blows’).
22420	A POS FREE ADV	Adverbial Adverbially used adjectives are not treated as adverbs, but as adjectives (like in the ANS-97, CELEX, RN, WOTAN-2, CGN and the German STTS-95). To distinguish adverbs from adverbially used adjectives, we use the following criterion: if the word is also used in prenominal position with the same meaning, then it is not an adverb but an adjective. <i>Vrij</i> ‘free’ is an

adjective in *Je kan hier vrij rondlopen* ‘You can walk around here freely’, whereas the same word is an adverb in *een vrij warme dag*, [lit.] ‘a free (somewhat) hot day’.

2.3	Degrees of comparison	
23000	A (DEGREE)	The comparative and superlative, A + diminutive. The positive doesn't get assigned 'DEGREE'. Indirect comparatives (of the type <i>more A</i> , <i>most A</i> , <i>less A</i> , <i>least A</i> ; e.g. <i>most beautiful</i>) will get DEGREE 'comparative' or 'superlative'. In that case, the tag is assigned to the modifier of the adjective (<i>most</i>); the adjective that is the head (<i>beautiful</i>) won't be assigned the tag DEGREE.
23100	A (DEGREE) COMP	Comparative. E.g. <i>bigger</i>
23200	A (DEGREE) SUP	Superlative. E.g. <i>biggest</i>
23300	A (DEGREE) DIM	Diminutive. E.g. <i>dunnetjes</i> '[lit:] thinly-DIM'

3. VERB

The problem of proclisis and enclisis arises when dealing with the finite verb: weak personal pronouns and/ or negative particle, that form one word/ a unity with the finite verb. In such cases, it's not always easy to distinguish between verb and pronoun (E.g. *ganeme*, normal: *gaan we*, 'go-we', *issem*, normal: *is hij*, 'is-he', *zakzekik*, normal *zal ik ze*, 'will-I-them-I-I', etc.). Therefore, we have two different tiers in the transcript: the first one (informant tier or assistant interviewer tier) contains the unanalyzed cluster, while in the second tier (the cluster tier) the cluster is divided into separate parts (E.g. *is em, za k ze kik*). Part of speech tagging is done with both tiers/ at both levels. In the informant or assistant interviewer tier, the word gets tagged as V-clitic-cluster (33000). In the cluster tier the finite verb gets the relevant tags, and of course also the other elements of the cluster get tagged. The following tags can either be assigned to verbs in the informant tier (if the finite verb is not part of a cluster) or in the cluster tier (when it is part of a cluster).

30000	V	Verb
3.1	Inflection	
31000	V (INFL)	Inflection. The occurrence of this attribute indicates the present of an audible affix. The following rule applies: inflection is only marked if the word can also occur without the inflectional morpheme. The attribute has the following values:
31100	V (INFL) -(e)n	
31200	V (INFL) -(e)t	
31300	V (INFL) -e	
31400	V (INFL) -(e)s	
31500	V (INFL) -st	
31600	V (INFL) OT	Other inflection (none of the above)
3.2	Position	
32000	V POS	Position
3.2.1	Regular order	
32100	V POS REG	Regular order

Position of the finite verb (finite form), directly following the subject (followed by the rest of the sentence). Also in case of subject doubling, where the subject precedes and follows the finite verb, the finite verb gets this value.

3.2.2 Inversed order
32200 V POS INV

Inverted order

Position of the finite verb (finite form), directly followed by the subject. All verb initial sentences (e.g. yes/ no questions, imperatives) and non-subject-initial matrix sentences will get this value.

3.2.3 Position right periphery
32300 V POS END

End position

The verb is not in first or second position, but is situated at the end of the sentence. Participles and infinitives are always V-final (unless they are free or used adnominally, see below). If more than one verb is in the final position of the sentence, we call it a verb cluster/ A verb cluster means that there is more than one verb in the final position, e.g. *dat Jan ze wel op zou willen eten*. '[lit.] that John them AFF PRT would want eat'. The hierarchical (not the linear) order/ position of each verb is given, through the specifications mentioned below. A verb cluster can be 'interrupted' by DP's, prepositions, adverbs, elements like *te* (=infinitive indicator: *hij zit te studeren*, '[lit:] he sits to study', meaning: he is studying). etc. but not by complementizers. The underlined parts form a verb cluster in the following sentence: *dat Jan probeert de krant te lezen*, '[lit:] that John tries the paper to read', but the underlined parts do not form a verb cluster in this sentence: *dat Jan probeert om de krant te lezen*, '[lit:] that John tries in order the paper to read'. This last sentence contains a complementizer, marking a new subordinate clause domain, and thus (potentially) a new verb cluster domain.

32310 V POS END(1)

Hierarchically highest verb

The finite verb always gets this specification. E.g. *Ik denk dat Jan de wagen gemaakt zou kunnen hebben*, '[lit:] I think that John the car fixed could can have'.

32320 V POS END(2)

Hierarchically second highest verb

This is always a non-finite verb, even if there is a finite verb in first or second position (in the main clause), counting in the cluster starts (at the end) with this specification.

E.g.: *Ik denk dat Jan de wagen gemaakt zou kunnen hebben*, '[lit:] I think that John the car fixed could can have'.

32330 V POS END(3)

Hierarchically third highest verb

E.g.: *Ik denk dat Jan de wagen gemaakt zou kunnen hebben*, '[lit.:] I think that John the car fixed could can have'.

32340 V POS END(4)

Hierarchically lowest verb

E.g.: *Ik denk dat Jan de wagen gemaakt zou kunnen hebben*, ‘[[lit.:] I think that John the car fixed could have can.’

3.2.4	Other positions	
32400	V POS OT	Other verbal positions
32410	V POS OT PRE-N	Prenominal E.g. <i>een gedurfd voorstel</i> , ‘a daring proposition’
32420	V POS OT N	Nominal E.g. <i>de vrijgestelden</i> , ‘the freed/ released/ liberated (people)’, <i>het <u>zingen</u> van de nachtegaal</i> , ‘the singing of the nightingale’.
32430	V POS OT POST-N	Postnominal E.g. <i>een tas gemaakt van leer</i> , ‘a bag made of leather’.
3.3	Verb-clitic-cluster	
33000	V CL-CL	This attribute is assigned at the informant or assistant interviewer tier to verbs that form a cluster with weak pronouns (see introduction of this paragraph).
3.4	Features	
34000	V FEAT	
3.4.1	Finite	
34100	V FEAT FIN	Finite verb (form)
3.4.1.1	Present tense, indicative	
34110	V FEAT FIN PT	Present tense
34111	V FEAT FIN PT 1.s	Present tense, first person singular
34112	V FEAT FIN PT 2.s	Present tense, second person singular
34113	V FEAT FIN PT 2.s-p	Present tense, second person singular, polite form
34114	V FEAT FIN PT 3.s	Present tense, third person singular
34115	V FEAT FIN PT 1.pl	Present tense, first person plural
34116	V FEAT FIN PT 2.pl	Present tense, second person plural
34117	V FEAT FIN PT 2.pl-p	Present tense, second person plural, polite form
34118	V FEAT FIN PT 3.pl	Present tense, third person plural
3.4.1.2	Present tense, conjunctive	
34120	V FEAT FIN PT.CONJ	Present tense conjunctive
34121	V FEAT FIN PT.CONJ 1.s	Present tense conjunctive, first person singular
34122	V FEAT FIN PT.CONJ 2.s	etc.
34123	V FEAT FIN PT.CONJ 2.s-p	
34124	V FEAT FIN PT.CONJ 3.s	
34125	V FEAT FIN PT.CONJ 1.pl	
34126	V FEAT FIN PT.CONJ 2.pl	
34127	V FEAT FIN PT.CONJ 2.pl-p	
34128	V FEAT FIN PT.CONJ 3.pl	
3.4.1.3	Imperative	
34130	V FEAT FIN TT.IMP	Finite imperative
34131	V FEAT FIN TT.IMP S	Finite imperative singular
34132	V FEAT FIN TT.IMP PL	Finite imperative plural
3.4.1.4	Past tense indicative	
34140	V FEAT FIN PastT	Past tense

34141	V FEAT FIN PastT 1.s	Past tense, first person singular, etc
34142	V FEAT FIN PastT 2.s	
34143	V FEAT FIN PastT 2.s-p	
34144	V FEAT FIN PastT 3.s	
34145	V FEAT FIN PastT 1.pl	
34146	V FEAT FIN PastT 2.pl	
34147	V FEAT FIN PastT 2.pl-p	
34148	V FEAT FIN PastT 3.pl	
3.4.1.5	Past tense conjunctive	
34150	V FEAT FIN PAST.CONJ	Conjunctive, past tense
34151	V FEAT FIN PAST.CONJ 1.s	Conjunctive, past tense, first person singular, etc.
34152	V FEAT FIN PAST.CONJ 2.s	
34153	V FEAT FIN PAST.CONJ 2.s-p	
34154	V FEAT FIN PAST.CONJ 3.s	
34155	V FEAT FIN PAST.CONJ 1.pl	
34156	V FEAT FIN PAST.CONJ 2.pl	
34157	V FEAT FIN PAST.CONJ 2.pl-p	
34158	V FEAT FIN PAST.CONJ 3.pl	
3.4.1.6	Past tense imperative	
34160	V FEAT FIN PAST.IMP	Imperative, past tense
34161	V FEAT FIN PAST.IMP s	Imperative, past tense singular
34162	V FEAT FIN PAST.IMP pl	Imperative, past tense plural
3.4.2	Non-finite	
34200	V FEAT INF	Infinitive
34210	V FEAT INF N	Infinitive, used nominally
34220	V FEAT INF FREE	
3.4.3	Participles	
34300	V FEAT PART	Participle
34310	V FEAT PART PAST	Past participle
34311	V FEAT PART PAST +PREFIX	Past participle with prefix <i>ge-</i> or <i>e-</i>
34312	V FEAT PART PAST -PREFIX	Past participle without a prefix
34320	V FEAT PART PRES	Present participle
3.5	Type	
35000	V TYPE	
3.5.1	Auxiliary	
35100	V TYPE AUX	Auxiliary verb
35110	V TYPE AUX PERF	Perfective auxiliary
35120	V TYPE AUX MOD	Modal auxiliary
		E.g.: <i>can, must, want, need, would, could</i>
35130	V TYPE AUX ASP	Aspectual auxiliary
		E.g.: <i>go, come, stand, lay, sit, stay, begin</i>
35140	V TYPE AUX PASS	Passive auxiliary
		E.g.: <i>worden</i> in <i>Er wordt gedanst</i> ‘[lit.:] there is/get dance.part’. If the sentence has a perfective meaning at the same time, then assign tag 35110 (E.g.: <i>Er is gedanst</i> , ‘there has been dancing going on/ [lit.:] there is danced’).
3.5.2	Matrix	

35200	V TYPE HEAD	Matrix verb
3.5.2.1	Inherent reflexive	
35210	V TYPE HEAD REFL	Inherent reflexive matrix verb
3.5.2.2	Transitive verb	
35220	V TYPE HEAD TRANS	Transitive matrix verb
3.5.2.3	Intransitive	
35230	V TYPE HEAD INTR	Intransitive matrix verb
35231	V TYPE HEAD INTR UNACC	Unaccusative verb These verbs use <i>zijn</i> 'to be' in perfective in Dutch, can not be used in an impersonal passive construction. The participle of such verbs can modify a noun that corresponds to the subject, if they are used as prenominal participles (e.g. <i>de gestorven man</i> , '[lit:] the died.part man'). E.g.: <i>sterven</i> , 'to die', <i>vallen</i> 'to fall'.
35232	V TYPE HEAD INTR UNERG	Unergative verb These verbs use <i>hebben</i> 'to have' in perfective in Dutch, they can be used in an impersonal passive construction, but can not modify a noun, that corresponds to the subject as a prenominal participle. E.g.: <i>werken</i> , 'to work', <i>slapen</i> , 'to sleep'.

4 PRONOUNS

Dutch has a number of different elements that are classified as pronouns (the following list is taken from ANS):

- personal pronouns
- reflexive pronouns
- reciprocal pronouns
- possessive pronouns
- demonstrative pronouns
- interrogative pronouns
- relative pronouns
- indefinite pronouns
- definite pronouns

This is a large and rather heterogeneous group that is, in our opinion, insufficient in one respect: determiners and numerals would fit into this group. We decided to include these elements in this category.

40000 PRON

4.1 Inflection

41000 PRON (INFL)

Inflection

This attribute indicates the presence of an audible morpheme for categories like: person, number, gender, case, mode, definiteness etc. For example the *-e* suffix attached to possessive pronouns. Differences between *hen* vs. *hun* ('them.ACC' vs. 'them.DAT') or *hem* vs. *hen* ('he.ACC' vs. 'them.ACC') are not marked (tagged) through this attribute. The following rule

applies: inflection is only marked if the word can also occur without the inflectional morpheme. The following values give information regarding the nature of the inflection:

41100	PRON (INFL) –(e)n
41200	PRON (INFL) –(e)t
41300	PRON (INFL) –e
41400	PRON (INFL) –(e)s
41500	PRON (INFL) –st
41600	PRON (INFL) –OT

Other inflection

4.2 Position

42000 PRON POS

42100 PRON POS PRE-N

Prenominal

Determiners are always prenominal. Interrogative pronouns, relative pronouns, demonstrative pronouns, possessive pronouns and quantifiers can be prenominal.

E.g. *which N, that N, my N, all/ three/ some N.*

42110 PRON POS PRE-N (ELL)

Prenominal with ellipsis

We speak of ellipsis if the noun can be added (but is not), in constructions like *ik heb deze boeken gekocht, en hij die*. ‘I bought these books, [lit.:] and he those’. There is no ellipsis in constructions like the following: *De mijne (*boeken) liggen nog op mijn kamer*, ‘[lit.:] The mine lay still at my room’. Another example of ellipsis: *Ik heb die rode gekocht*, ‘[lit.:] I have those red bought’.

42200 PRON POS N

Nominal

A pronoun is used nominally, if it is the head of an NP, and if there are (or can be) other elements inside the same NP. If this last part of the definition is not the case, we call the position free.

E.g.: *dit alles*, ‘this all’, *dat allemaal*, ‘that all’, *de mijne*, ‘the mine’.

42300 PRON POS POST-N

Postnominal

E.g.: *Zij allen/ beiden hebben het geweigerd*, ‘[lit.:] they all/ both have it refused’.

42400 PRON POS FREE

Free

A pronoun is free if it forms an NP on its own.

Personal pronouns, reflexive pronouns, reciprocal pronouns and R-pronouns are always free.

Interrogative pronouns, relative pronouns, possessive pronouns, demonstrative pronouns, and quantifiers can be free (*who, what, this, that, all, nothing, nobody*).

Pronouns that are marked for genitive case (E.g. *wiens*, ‘whose’) are free.

42410 PRON POS FREE (PRED)

Predicative pronouns

Predicative pronouns are found in constructions of the type *Die fiets is mijns*, ‘that bicycle is mine.GEN’, *Jan is daar*, ‘John is there’.

4.3 Case

Just like with nouns, pronouns are labelled with an attribute for case if the case is morphologically visible.

43000	PRON (CASE)	
43100	PRON (CASE)(NOM)	Nominative This is normally the pronoun that is used as the subject and that agrees with the finite verb in person and number.
43200	PRON (CASE)(OBL)	Oblique This means not nominative or genitive. This value can be further specified as accusative or dative. If it is not entirely clear if the pronoun is in accusative or dative case, the value oblique is sufficient, and no further specification is given.
43210	PRON (CASE)(OBL)(ACC)	Accusative
43220	PRON (CASE)(OBL)(DAT)	Dative
43300	PRON (CASE)(GEN)	Genitive

4.4 Person and number

44000	PRON (FEAT)	
44100	PRON (FEAT) 1.S	First person singular
44200	PRON (FEAT) 2.S	Second person singular
44300	PRON (FEAT) 2.S-P	Second person singular, polite form
44400	PRON (FEAT) 3.S	Third person singular
44500	PRON (FEAT) 1.PL	First person plural
44600	PRON (FEAT) 2.PL	Second person plural
44700	PRON (FEAT) 2.PL-P	Second person plural, polite form
44800	PRON (FEAT) 3.PL	Third person plural

4.5 Gender

45000	PRON (GEND)	
45100	PRON (GEND) Z	Non-neuter
45110	PRON (GEND) Z (M)	Masculine. This specification is only added in clear cases.
45120	PRON (GEND) Z (F)	Feminine. This specification is only added in clear cases
45200	PRON (GEND) N	Neuter

4.6 Function

Grammatical function has strictly speaking, no place in parts of speech tagging. However, it is necessary to add this attribute in order to be able to search the database, when syntactic annotation is not been done (yet).

46000	PRON (FUNCT)	
46100	PRON (FUNCT) SUBJ	Subject Constituent that agrees with finite verb.
46200	PRON (FUNCT) D-OBJ	Direct object
46300	PRON (FUNCT) I-OBJ	Indirect object Indirect object, without a preposition.
46400	PRON (FUNCT) P-OBJ	Prepositional object

4.8 Type

All types of pronouns are tagged under this attribute. For most of them we follow ANS (except for determiner, R-pronoun and quantifier, that are not classified as pronouns in ANS).

48000	PRON TYPE	
-------	-----------	--

- 4.8.1 Personal
- 48100 PRON TYPE PERS Personal pronouns
- 4.8.1.1 Subject doubling
- 48110 PRON TYPE PERS(DOUBL) Personal pronoun in a subject doubling constructions.
- 48111 PRON TYPE PERS(DOUBL) 1-STRONG Linearly first person pronoun in a doubling construction, if it is strong.
E.g. *Zij heeft ze me niet gebeld*, ‘She.STRONG has she.WEAK me not called’.
- 48112 PRON TYPE PERS(DOUBL) 1-WEAK Linearly first person pronoun in a doubling construction, if it is weak.
E.g. *Ze heeft zij me niet gebeld*, ‘She.WEAK has she.STRONG me not called’.
- 48113 PRON TYPE PERS(DOUBL) 2-STRONG Linearly second person pronoun in a subject doubling construction if it is strong.
E.g. *Ze heeft zij me niet gebeld*, ‘She.WEAK has she.STRONG me not called’.
- 48114 PRON TYPE PERS(DOUBL) 2-WEAK Linearly second person pronoun in a subject doubling construction, if it is weak.
E.g. *Zij heeft ze me niet gebeld*, ‘She.STRONG has she.WEAK me not called’.
- 48115 PRON TYPE PERS(DOUBL) 3-STRONG Linearly third person pronoun in a subject doubling (or: tripling) construction, if it is strong.
E.g. *Marie heeft ze zij me niet gebeld*, ‘Marie has she.WEAK she.STRONG me not called’.
- 48116 PRON TYPE PERS(DOUBL) 3-WEAK Linearly third person pronoun in a subject doubling (or: tripling) construction, if it is weak.
E.g. *Marie heeft zij ze me niet gebeld*, ‘Marie has she.STRONG she.WEAK me not called’.
- 4.8.1.2 Strong and weak (non doubling)
- 48120 PRON TYPE PERSSTATUS This tag distinguishes between strong and weak pronouns in a non doubling construction. It is assigned only if the distinction is clear. When in doubt, don’t assign this tag.
- 48121 PRON TYPE PERSSTATUS STRONG Strong pronoun, no doubling
- 48122 PRON TYPE PERSSTATUS WEAK Weak pronoun, no doubling
- 4.8.1.3 Expletive
- 48130 PRON TYPE PERSEXPL This tag is assigned to pronouns that function as expletive
E.g. *Het is jammer dat Piet gaat*, ‘It is a shame that Pete leaves’. *Er loopt een poes in de tuin*, ‘[lit.:] There walks a cat in the garden’.
- 48131 PRON TYPE PERSEXPL STRONG This specification is only assigned to strong pronouns that function as expletives.
E.g.: *Dat schijnt dat Piet komt*, ‘[lit.:] That (it.strong) seems that Pete comes’. *Daar was eens een prinses die trouwen wilde*, ‘[lit.:] There was once a princess that marry wanted’ (‘Once upon a time there was a princess that wanted to get married.’).

4.8.2	Reflexive and reciprocal	
48200	PRON TYPE REFL	
48210	PRON TYPE REFL	Reflexive pronoun E.g.: <i>Jan kent <u>zichzelf</u></i> , '[lit.:] John knows REFL/himself'.
48211	PRON TYPE REFL SIMPL	Reflexive pronoun consisting of one morpheme. E.g.: <i>Jan wast <u>zich</u></i> , 'John washes REFL'.
48212	PRON TYPE REFL COMPL	Reflexive pronoun consisting of two morphemes. E.g.: <i>Jan wast <u>zichzelf</u></i> 'John washes REFL', <i>Piet schaamt <u>z'n eigen</u></i> '[lit.:] Pete shames HIS OWN'.
48220	PRON TYPE REFL REC	Reciprocal pronoun. E.g. <i>Jan en Marie ontmoetten elkaar in 1933</i> , 'John and Mary met each other in 1933'.
4.8.3	Possessive	
48300	PRON TYPE POSS	Possessive pronoun
48310	PRON TYPE POSSSTRONG	Strong possessive pronoun E.g.: <i>Marie heeft <u>haar</u> huis verkocht</i> 'Marie sold her house'.
48320	PRON TYPE POSSWEAK	Weak possessive pronoun E.g.: <i>Marie heeft <u>d'r</u> huis verkocht</i> 'Marie sold her house'.
4.8.4	Demonstrative	
48400	PRON TYPE DEM	Demonstrative pronoun
48410	PRON TYPE DEM DEF	Definite demonstrative pronoun. E.g.: <i><u>die</u> fiets</i> , 'that bicycle'
48420	PRON TYPE DEM INDEF	Indefinite demonstrative pronoun E.g.: <i><u>zulke</u> boeken</i> 'such books', <i><u>zo 'n</u> brug</i> 'such-a bridge'.
4.8.5	Interrogative	
48500	PRON TYPE WH	Interrogative pronoun E.g. <i>who, what, which</i>
4.8.6	Relative	
48600	PRON TYPE REL	Relative pronoun
48610	PRON TYPE REL W	Relative pronoun starting with a /w/ E.g. <i>Alles <u>wat</u> Jan weet</i> 'Everything what John knows', <i>Het meisje met <u>wie</u> hij uitging</i> 'The girl with who he dated', <i>De bal <u>waarmee</u> zij speelden</i> 'The ball where-with they played'.
48620	PRON TYPE REL D	Relative pronoun starting with a /d/ E.g. <i>Alles <u>dat</u> Jan weet</i> 'Everything that John knows', <i>Hij is iemand <u>die</u> graag praat</i> 'He is someone that gladly talks'.
4.8.7	R-pronouns	
48700	PRON TYPE R-PRON	This tag is assigned to <i>er</i> 'there', <i>daar</i> 'there', <i>overal</i> 'everywhere', <i>ergens</i> 'somewhere', <i>nergens</i> 'nowhere', <i>hier</i> 'here', <i>waar</i> 'where' whenever they are used non-adverbially. E.g.: <i>Hij denkt <u>nergens</u> meer aan</i> , 'He thinks of nothing anymore'.
48710	PRON TYPE R-PRON STRONG	Strong R-pronoun

		E.g. <i>Kijk <u>daar</u> loopt een adelaar</i> , ‘Look, there walks an eagle’.
48720	PRON TYPE R-PRON WEAK	Weak R-pronoun E.g. <i>D'r staat een paard in de gang</i> ‘[lit.:] There.weak stands a horse down the hallway’.
4.8.8	Quantifiers	
48800	PRON TYPE QUANT	This category resembles the ‘indefinite pronouns’ used in ANS, but is enlarged with numerals.
48810	PRON TYPE QUANT NUM	Numerals, both ordinals and cardinals. <i>Veel</i> ‘many’ and <i>weinig</i> ‘less’ do not belong to this class, but are adjectives.
48820	PRON TYPE QUANT UNIV	Universal quantifiers: <i>each, every, all, altogether, everything, everybody, both</i> .
48830	PRON TYPE QUANT EXIS	Existential quantifiers: <i>any, some</i> .
48840	PRON TYPE QUANT NEG	Negative quantifiers: <i>nobody, nothing, no</i> .
4.8.9	Determiners	
48900	PRON TYPE ART	
48910	PRON TYPE ART DEF	Definite articles: <i>de</i> ‘the.MASC/FEM’, <i>het</i> ‘the.NEUT’
48920	PRON TYPE ART IND	Indefinite articles: <i>een</i> ‘a’

5. ADPOSITIONS: PREPOSITIONS AND POSTPOSITIONS

Adpositions normally take a complement. This can be a DP, but also a PP, an adverb, an adjective, a numeral or a verbal projection (V, VP, IP, CP). In the latter case, adpositions are treated as complementizers, which results in a systematic ambiguity for words like *tot* ‘until’, *sedert, sinds* ‘since’, *voor* ‘before’, *na* ‘after’, *naar* ‘to’, *zonder* ‘without’, *met* ‘with’, *door* ‘through’, *om* ‘because of/ in order to’. Just like CGN, the SAND doesn’t follow this strategy, and classifies these words as adpositions. This also holds for *te*, that introduces an infinitive, and *aan* ‘[lit.:] on’ in constructions like *aan het vissen* ‘fishing, busy with fishing’, *op* ‘on’ in constructions like *op springen staan* ‘[lit.:] on jump standing, be about to explode’, and *uit* ‘out’ in *uit vissen gaan* ‘go out fishing’.

50000 P Adpositions

5.1 Inflection

We talk about inflection if there is an audible morpheme attached to P, for categories like: person, number, gender, case, mode, definiteness etc. The following rule applies: inflection is only marked if the word can also occur without the inflectional morpheme. There are 6 specifications:

51000	P (INFL)	
51100	P (INFL) –(e)n	
51200	P (INFL) –(e)t	
51300	P (INFL) –e	
51400	P (INFL) –(e)s	
51500	P (INFL) –st	
51600	P (INFL) OT	Other inflectional morpheme

5.2 Position

52000	P POS	
52100	P POS PREP	Prepositional, E.g. <i>op de brug</i> ‘on the bridge’.

52110	P POS PREP FUSION	Prepositional and fused with (a part of) its complement. E.g. <i>ter plaatse</i> ‘on.the spot’, <i>ten geleide</i> ‘[lit.:] at.the guard, preface’
52200	P POS POST	Postpositional, E.g. <i>onder de brug door</i> ‘[lit.:] under the bridge through’. Adpositions accompanied by an R-pronoun are postpositional, even if the R-pronoun is at a distance/ doesn’t directly precede or follow the preposition. E.g. <i>dat ik er gisteren met Jan over gesproken heb</i> , ‘[lit.:] that I there yesterday with John about spoken have (that I’ve talked to John about it yesterday)’.
52300	P POS FREE	Adpositions can also occur without a complement. This is the case for adverbially or predicatively used prepositions. E.g. <i>het bier is op</i> , ‘[lit.] the beer is on/ we ran out of beer’, <i>het licht is aan</i> , ‘[lit.:] the light is on’. Separable prefixes of a verb also count as predicatively used prepositions, whether they are separated from the verb or not. (E.g.: <i>Hij belt haar op</i> , ‘[lit.:] he calls her up; he calls her’, ... <i>dat hij haar opbelt</i> , ‘that he her up calls; that he calls her’.)
52320	P POS FREE ADV	Adverbially used adpositions. E.g. <i>Dat hij liever binnen werkt</i> , ‘that he rather works inside/ indoors’.

6. COMPLEMENTIZERS

On the one hand, there are complementizers in the strict sense: *that*, *or*, *if*, *then*, *to* and on the other hand there are combinations of a preposition and a complementizer: *doordat* ‘through.that/ because of’, *nadat* ‘after.that/ after’, *omdat* ‘to.that/ because’, *opdat* ‘on.that/ so that’, *totdat* ‘until.that/ until’, *voordat* ‘before.that/ before/ until’. Furthermore there are complementizers of the type *terwijl* ‘meanwhile’, *alhoewel* ‘although’, *tenzij* ‘unless’, *alvorens* ‘all/already.forward/ before’. In the CGN tagset, these types are all classified as complementizers. This doesn’t seem to be an optimal situation: *voor* ‘for’ is from this point of view a preposition, but *voordat* ‘for.that’ a complementizer. We therefore decided to assign two different tags to the combination preposition-complementizer; and for this reason we split these combinations in the transcriptions (*doordat* will be transcribed as *door dat*).

60000 C

6.1 Inflection

Complementizer agreement constructions encode, just like inflected verbs, person and number features. We therefore expect that the tagging of complementizers corresponds at a high degree with finite verbs in this respect. The question then arises whether these features should also be indicated on complementizers in dialects that lack complementizer agreement. From experience from the written questionnaire, we know it is sometimes difficult to see whether there is agreement or not. This is an argument in favour of treating all complementizers alike: not assigning person and number features to any complementizer. (We expect..) The complementizer (to be) is part of a cluster with the subject pronoun in many cases. The subject pronoun has the same feature set and so the assignment of these features to the complementizer is redundant. Therefore we decided to select a minimal feature set for complementizers. The subject close to the complementizer provides the necessary person and number features. The subject is recognizable because it is assigned the SUBJ tag.

The attribute INFL indicates the presence of an audible morphological marking for categories like person, number, gender, case, mode etc. The following rule applies: INFL is only assigned if the word can also occur without the inflectional morpheme.

61000 C (INFL)

The following tags provide information regarding the encountered inflection.

61100 C (INFL) –(e)n

61200 C (INFL) –(e)t

61300 C (INFL) –e

61400 C (INFL)–(e) s

61500 C (INFL) –st

61600 C (INFL) OT Other inflectional morphology

6.2 Complementizer-clitic-cluster

This is about enclisis: weak personal pronouns, that are phonologically connected to the complementizer, and, moreover, that form one word with that complementizer. In such cases, it is not always clear where the (word) boundaries are, especially if you are dealing with dialects that have visible agreement (E.g. *attem* ‘[lit.:] as-t-he.ACC(used NOM)’, *dakzekik* (Standard Dutch: *dat ik ze*) ‘[lit.:] that.I-them.ACC-I-I’, *ovveme* ‘if.-e-we.ACC(used NOM)’. We use two tiers in the transcriptions: in one (informant or assistant interviewer tier) we transcribe the non-analyzed cluster (non-divided cluster), and the cluster tier contains the divided/ split cluster. The assignment of tags is done at both levels/ in both tiers: the undivided cluster gets the tag for C-clitic-cluster, and the complementizer and other elements (of the divided cluster) will get separate tags in the cluster tier.

62000 C CL-CL

6.3 Complementizer type

These tags can be assigned at the informant tier or assistant interviewer tier (if the complementizer is not part of a cluster) or at the cluster tier (if that is the case).

63000 C TYPE

63100 C TYPE COORD Coordinating complementizer, E.g. *and*, *or*

63200 C TYPE SUBORD Subordinating complementizer, E.g. *whether*

63210 C TYPE SUBORD FIN Finite subordinating complementizer

63211 C TYPE SUBORD FIN Q Interrogative finite subordinating complementizer. E.g. *I asked whether you knew*. Words like *why*, *when*, *how* are treated as interrogative pronouns and not as complementizers.

63212 C TYPE SUBORD FIN DECL Declarative finite subordinating complementizer. E.g. *if*, *that* and their dialectal counterparts. Also *then*, *while*, *now*, *as soon as*, *since*, *before*, *if*, *unless*. *Then* is a complementizer following a comparative, but not when it is initiating a ‘result’-sentence in an ‘if’-sentence (*if... – then...*); *then* is an adverb, because inversion follows (in Dutch: *als je niet ophoudt, dan word ik boos* ‘[lit.:] if you not stop, then will.be I angry’). Words that can be followed by a DP (*since*, *before*, *except*, *without*) are considered as adpositions. Potentially occurring complementizers in relative clauses (E.g. *de man die dat ik gezien heb* ‘[lit.:] the man that/who that I seen have’) are declarative as well.

63220 C TYPE SUBORD INF Subordinating complementizer, introducing a

infinitival subordinate clause. In Standard Dutch *om*, whereas many dialects employ *van*, and others (also) *voor*.

- 63221 C TYPE SUBORD INF Q Subordinating complementizer that introduces a interrogative infinitival subordinate clause. Possibly *if* occurs in these constructions.
- 63230 C TYPE SUBORD NON-S Complementizers can introduce subordinate clauses, but also smaller phrases, for example in comparatives: *richer than Fons, not as big as John*. In the latter case, the distinction finite vs. infinitival doesn't apply and the tag non-sentential is assigned.

7. ADVERBS

Adjectives and nouns that are used adverbially are not treated as adverbs, but as adjectives and nouns. In order to distinguish between adverbs and adverbially used adjectives/ nouns, we use the following criterion: if the word in question can also be used in pronominal position – having the same meaning – then it is not an adverb, but an adjective.

70000 ADV

7.1 Inflection

This attribute indicates the presence of an audible morphological marking for categories like person, number, gender, case, mode etc. The following rule applies: INFL is only assigned if the word can also occur without the inflectional morpheme.

71000 ADV (INFL)

The following tags give information regarding the form/ type of the encountered inflection.

- 71100 ADV (INFL) –(e)n
 71200 ADV (INFL) –(e)t
 71300 ADV (INFL) –e
 71400 ADV (INFL) –(e)s
 71500 ADV (INFL) –st
 71600 ADV (INFL) OT Other morphological inflection
- 7.2 Position
- 72000 ADV POS
 72100 ADV POS AD-A Ad-adjectival, E.g. *a really rich baker*
 72200 ADV POS AD-P Ad-adpositional, E.g. *recht op zijn hoofd*, ‘right on his head’
 72300 ADV POS AD-ADV Ad-adverbial, E.g. *een heel erg rijke man* ‘[lit.:] a whole really rich man/ a terribly very rich man’.
- 72400 ADV POS FREE
 72410 ADV POS FREE ADV Adverbs that modify a verb, a part of a sentence or a complete clause: *Hij wil nu naar huis*, ‘he wants to go home now’, *hij heeft het wel gedaan* ‘he has it AFF done/ he did do it’, *hij wil het niet doen* ‘he does not want to do it’, *Als het regent, dan zijn er wolken*, ‘if it’s raining, then there are clouds’.
- 72420 ADV POS FREE PRED Adverbs that are nominal predicates: *Dat is jammer*, ‘it’s a shame’, *het feest is nu*, ‘the party is now’. The same holds for adverbial prefixes of separable verbs,

whether they are separated from the verb or not. E.g. *Ik drukte de eierschalen tussen mijn vingers samen* ‘[lit.:] I pressed the eggshells between my fingers together’, *dat ik .. samen drukte* ‘[lit.:] that I ... together pressed’.

7.3 Type

Not all adverbs will fit in with the types mentioned below, therefore this attribute is optional.

73000 ADV (TYPE)

7.3.1 Interrogative

73100 ADV (TYPE) Q

Interrogative adverbs.

E.g. *when, how, why, and where*. Note that these words can also belong to other classes: in *de stad waar ik woon* ‘the city where I live’, *where* is a relative pronoun. In *Waar kijk je naar* ‘[lit: where] what are you looking at’, *where* is an interrogative pronoun, and in *Waar woont hij* ‘where does he live’, *where* is a interrogative adverb.

7.3.2 Quantifier

73200 ADV (TYPE) QUANT

Quantificational adverb/ adverbial quantifier

73210 ADV (TYPE) QUANT UNIV

Universal adverbial quantifiers. E.g. *always, constantly, repeatedly*.

73220 ADV (TYPE) QUANT NON-UNIV

Non-universal adverbial quantifiers. E.g. *once, ever, sometimes, often*.

73230 ADV (TYPE) QUANT NEG

Negative adverbial quantifiers. E.g. *never, rarely, nowhere*.

7.4 R-pronouns (adverbial)

74000 ADV FORM R-PRON

To this class belong words like *here, there, somewhere, nowhere, everywhere* when used adverbially. E.g. *Er zitten hier nergens muizen* ‘[lit.:] EXPL sit here nowhere mice’. This tag is not assigned to R-pronouns that belong to an adposition, as in *Hier zit hij op* ‘[lit.:] here sit he on/ this is what he does (in Standard Dutch: Hij zit *hierop*). Those R-pronouns receive the following tag: PRON TYPE R-PRON.