

Exhibit A2

Dutch PAROLE Lexicon Documentation

Background Information

Institute for Dutch Lexicology (INL)

Language Database Department

P.O. Box 9515

2300 RA Leiden

The Netherlands

www.inl.nl

parole@inl.nl

1 Introduction

The documentation for the Dutch PAROLE lexicon consists of two separate documents. The present document, **Background Information**, contains linguistic information and motivations for design and contents. The document **Contents Specification** contains detailed overviews of the exact contents of the lexicon (number of entries per part of speech, features, functions, etc.) as well as an explanation of the linguistic terminology used.

This documentation serves as an addition to the general documentation accompanying the set of PAROLE lexicons (PAROLE Reports; GENELEX (1993, 1994)). The other languages for which lexicons have been developed are: Catalan, Danish, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish and Swedish. The PAROLE lexicon is delivered as a file in SGML-format compliant with the PAROLE DTD for the lexicon.

Section 2 describes the general design of the lexicon, and the composition process of the entry list. Sections 3 and 4 document the linguistic decisions underlying the morphosyntactic and syntactic descriptions of the entries. Appendix A presents a conceptual graph of the relations between the (SGML) objects in the lexicon.

2 General Design Information

The entry list of the lexicon consists of about 20,200 entries distributed over 13 parts of speech (POS). The entries have been described along the dimensions of morphosyntax and syntax. Morphosyntactic information consists of various lexical properties, like gender, number, case, person, inflection, etc. Syntactic descriptions consist of typical complementation patterns associated with the various lemmata.

The composition of the entry list of the lexicon is based on 3 INL corpora and 2 lexicons. The corpora contain a total of about 54 million words and have been automatically annotated for part of speech and lemma. The lexicons contain morphosyntactic information of various kinds. For verbs, nouns, adjectives and adverbs, lemmata that were covered by at least 2 corpora and the 2 lexicons were selected on the basis of cumulative frequency, coverage (distribution over sources) and inflected forms. For the smaller parts of speech, these selection requirements appeared to be too strict. Entry selection for these parts of speech was based on ranked frequency.

The entry list does not contain

1. Multiword units (e.g. ‘ten laste van’/*to the account of*) and frozen lexical combinations like ‘manisch depressief’/*manic depressive*, ‘bar en boos’/*lit. severe and bad, meaning very bad*).
2. Word forms with a hyphen in their entry form (e.g. ‘minister-president’/*prime minister*, ‘secretaris-generaal’/*secretary-general*) unless the hyphen is a spelling revision for diaeresis (like ‘zo-even’/*just now*, ‘toe-eigenen’/*to appropriate*).

The entries, uniquely defined by the combination of part of speech (e.g. noun) and subtype (e.g. common vs. proper noun), are provided with morphosyntactic information according to the Dutch set of PAROLE categories and features (see **Contents Specification**), and, where available, with syntactic information. Morphosyntactic information is automatically extracted from the INL lexicons. Syntactic data have been collected manually, by inspection of corpus data and - where necessary - consultation of reference works. The corpus consulted consists of the newspaper component and the varied component of the 38 Million Words Corpus 1996 (see: Kruyt & Dutilh, 1997) .

Throughout the lexicon, the most recent spelling rules as implemented in *Woordenlijst Nederlandse taal* (1995), informally called *Groene boekje*, have been taken into account when dealing with texts in an earlier spelling for the composition of the entry list.

-

3 General linguistic decisions

The linguistic decisions were guided by the following reference works: *Algemene Nederlandse Spraakkunst* (ANS, 1987) en Bennis & Hoekstra (1989)

3.1 Morphosyntax

For general information about the composition of the entry list and the origin of the morphosyntactic information in the lexicon, see section 2 above. For POS-specific aspects of the entry list, see the **Entry list** subsections for the various parts of speech in section 4.

3.1.1 Criteria for splitting homographs

For verbs, nouns, pronouns and determiners, the following criteria were applied for distinguishing two (or more) entries for one homograph.

(1) Different inflected forms, e.g.

doorbreken~1: inflected form ‘doorbreekt’/*breaks through* vs.

doorbreken~2 : inflected form ‘breekt door’/*breaks apart*

gat~1: plural ‘gaten’/*holes* vs.

gat~2: plural ‘gatten’/*backsides*

(2) Different values for morphological features, e.g.

bal~1: gender *masculine*. vs.

bal~2: gender *neuter*.

(3) Different subtype, e.g.

zijn~1: VERB main *to be* vs.

zijn~2: VERB aux. vs.

zijn~3: VERB copula;

welk~1: DETERMINER:interrogative *what/which* vs.

welk~2: DETERMINER: relative.

For residuals, the criteria were the following.

1. Different article selection, e.g.

ABN~1: RESIDUAL acronym (meaning: ‘Standard Dutch’ language variant) with article ‘het’ vs. **ABN~2** : RESIDUAL acronym (name of Dutch banking company) with article ‘de’.

(2) Different reading, e.g.

t.a.v.~1: RESIDUAL abbreviation ‘ter attentie van’/*for the attention of* vs.

t.a.v.~2: RESIDUAL abbreviation ‘ten aanzien van’/*with regard to*

2. Variant forms of entries

43 entries (nouns and adjectives) in the lexicon have variant forms which are considered equivalent according to Dutch spelling rules (e.g. *literair* vs. *litterair*, *plafond* vs. *plafon*). All variant forms are included in the lexicon as entries and have their own syntactic descriptions. For each entry, the other variants are cross-referenced in the GMU field. E.g. the entry *plafond* has in its GMU field the variant *plafon* and vice versa.

3.1.3 Morphological paradigms

Word forms in the Dutch PAROLE lexicon are not inflected according to general paradigms, but are related to their lemma by a set of string procedures. These procedures are not unique. They can be shared by many other word forms. An example is suffixation with *-e* for adjectives, which produces ‘*goede*’/*good* from ‘*goed*’. Inflected forms can be derived directly by applying the string procedures to the lemma they are connected with.

3.2 Syntax

3.2.1 Lexicon vs. grammar

A well-known problem is the borderline between lexicon and grammar. The Dutch PAROLE lexicon does not describe all grammatical devices for word order, tense, reference

mechanisms, etc. Focus is on the description of entry-specific complementation patterns rather than on exhaustive syntactic pattern listing. This section reports on some general lexicographical decisions to be taken into account when using this lexicon. POS-specific decisions are accounted for in section 4.

Grammatical patterns not accounted for in the lexicon

For nouns, verbs, and adjectives, covering the majority of the lexicon contents, the sections 4.1-4.3 describe which syntactic patterns are considered to belong to the grammar and are, by consequence, not accounted for in the lexicon. Concerning the other POS-categories, the majority of which are closed classes, grammatical behaviour is accounted for if attested in the corpus.

Standard complement order in complementation pattern of an entry

For verbs, adjectives and nouns, a specific standard order of the entry and its complements was decided on. For example, adjectives with a PP complement, are described as PP+adjective, e.g.

- ‘de aan de situatie aangepaste plannen’/lit. *the to the situation adapted plans*

The lexicographers have rephrased corpus examples according to the standard complement order, unless this order cannot reasonably be considered grammatical in Dutch. That is, a deviant description implies that the standard order is not grammatical in Dutch. The standard complement order implies that other possibilities for syntactic patterning are considered to belong to the grammar and are not accounted for in the lexicon.

One thing to be kept in mind is that the syntactic descriptions only explicitly stipulate the order of complements with respect to the entry, not the order of complements with respect to each other. The relative order of complements is only implicitly specified by the order of complement presentation in the descriptions (on the CONSTRUCTION level, that is).

Interpretation of complement names

Certain complement names represent sets of implied complements:

- NP also covers N, NG, PRO, and article-selecting RESIDUALS (e.g. abbreviations like ‘B.V.’, and acronyms like ‘DDR’), e.g. ‘Hij slaapt/de man slaapt’/*He is sleeping/the man is sleeping* (NP+VERB; NP also covers pronoun).
- ART also covers DET, and for quantifiable nouns (like ‘tanker’) also NUM (see section 4.1.2); e.g. ‘70 koeien/de koeien /zijn koeien’/*70 cows/the cows/his cows* (ART+NOUN; ART covers NUM and DET).

Whenever implied complements actually do occur in syntactic descriptions, they are very characteristic for the syntactic patterning of the specific entry and were attested as such in the corpus. This particularly applies to POS-categories which normally are constituting components of phrasal complements (like determiners and numerals in NP’s).

E.g.

- 30 stuks /*30 heads (e.g. of cattle)* /*de stuks/*zijn stuks: (NUM+NOUN; ART and DET are not possible).

The category CONJ may refer both to the conjunction itself and to its complement, e.g. the conjunction complement for the entry ‘lopen’/*to walk* in

- ‘lopen als een olifant’/*walk like an elephant*

refers to ‘als een olifant’. This is expressed by the conjunction ‘als’, which, in turn, implicitly takes the NP ‘een olifant’ as its complement. This type of description contrasts with the following:

- ‘als zodanig’/*as such*

where the pronominal entry ‘zodanig’ takes the bare conjunction ‘als’ as its complement (which has no further complements).

3.2.2 Optionality

In many cases, the distinction between obligatory and optional complements has proved difficult to make. As a rule of thumb, the following criterion was applied: a complement is considered to be obligatory if omission from the sentence results in an ungrammatical or very marked sentence, or in a change of the lexical meaning of the entry. Still, very often the decision on optionality is open to subjectivity and inconsistency. It should therefore be regarded as an indication of a tendency rather than a rule.

3.2.3 Syntactic functions

For noun and verb entries, the complements can have various syntactic functions. This is not the case for the other POS-categories.

If an entry subcategorizes for complements having different syntactic functions, each alternative has been described. For example, the noun lemma ‘aanduiding’/*indication* has been assigned the following syntactic descriptions (see Exhibit A1, Appendix B for an explanation of the terminology):

- De **aanduiding** voor kampeerterein/ *the indication for a camp site*

Complementation pattern: ARTICLE <aanduiding> PP with function NPREPCOMP.

- De **aanduiding** van de plaats/*the indication of the place*

Complementation pattern: ARTICLE <aanduiding> PP with function NOFCOMP.

3.2.4 Feature SSRELATIVE

The lexicon does not contain clausal complements with the feature SSRELATIVE. Relative sentences are considered to belong to the grammar and are hence not accounted for in the lexicon. For the POS-categories adposition and pronoun, some entries do have clausal complements that strictly speaking are relative clauses; they have the feature SUBORDONNEE (see 4.5 and 4.8) .

3.2.5 Introducers

An introducer is the lexical element which introduces complements, e.g. the preposition ‘op’ introducing a PP as in:

- ‘Hij rekt op zijn medewerking’ /*He is relying on his cooperation*

Introducers are only specified in cases with a limited number of possibilities from a whole set. In the example just given, ‘op’ is specified as introducer of the PP ‘op zijn medewerking’ because ‘op’ is the only possible preposition for this verb (in this meaning). No introducer would have been specified if many other prepositions might have been filled in as well.

"Om te"

Infinitival clauses are sometimes introduced by ‘om (te)’. This ‘om’ can be obligatory, optional or not possible at all.

1. If 'om' is obligatory, then it is specified as the introducer for the infinitival clause.
2. If 'om' is possible, but not obligatory, then no introducer for the infinitival clause is specified.
3. If 'om' is not possible, then 'te' is specified as the introducer for the infinitival clause.

The choice between these three possibilities in the syntactic descriptions has been made on the basis of corpus evidence, not by what is considered to be 'correct' Dutch.

3.2.6 Comment fields

For some POS categories the comment field is used for clarification. See the relevant POS-specific sections. Entries displaying no characteristic syntactic behavior other than implied by the grammar of Dutch have been assigned an empty description with the comment 'Geen bijzonder patroon gevonden.'/*No characteristic pattern attested.*

4 Word class-specific linguistic decisions

4.1 Nouns

4.1.1 Grammar

Almost every noun can be preceded by an article and by one or more adjectives. Complementation patterns that only consist of an article and (optionally) an adjective are considered to belong to the grammar, and have not been accounted for in the lexicon.

4.1.2 Standard complement order

The standard complement order for nouns is article+noun+other complements. This implies that the initial adposition ‘bij’ in ‘bij gebrek_{NOUN} aan (voedsel)’/for want of (food) has not been described; the noun ‘gebrek’ only selects an article and the PP headed by ‘aan’.

4.1.3 Complement numeral and determiner

As said in section 3.2.1, articles also cover DET and NUM. In some contexts, determiners and numerals are very specific complements, e.g.

- Zijn_{DET} echtgenote Sonja/his wife Sonja vs. *de echtgenote Sonja/the wife Sonja
- 30_{NUM} stuks rundvee/30 heads of cattle vs. *de stuks rundvee/the heads of cattle

4.1.4 Comment field

Entries having equivalent variants (see 3.1.2) have a comment field

- ‘Zie ook <variant form>’/See also <variant form>

Most nouns which have a plural can be preceded by the complement ‘numeral’ (NUM). This belongs to the grammar and for these nouns the complement ART also refers to NUM (in addition to other complement types; see section 3.2.1). Some nouns, though in plural form, usually are not combined with a numeral (like ‘gebroeders’/the brothers); these nouns have been assigned the following comment line

- ‘Meervoudig lemma; NUM niet gebruikelijk’/Plural lemma; NUM not common.

Similarly, some nouns which do not have plural forms (like ‘moed’/courage) have the following comment

- ‘Enkelvoudig lemma; NUM niet gebruikelijk’/Singular lemma; NUM not common

Combinations of comments are of course possible.

4.2 Adjectives

4.2.1 Grammar

The following syntactic patterns with adjectives are considered to be regular and are generally not accounted for in the lexicon:

- a. adjectives followed by a noun
- b. adjectives preceded by an adverb; only marked cases are accounted for.
- c. adjectives used in the grammatical function of adverb

Adjectives followed by proper nouns are described, however, like in:

- ‘Jean Orsini bijgenaamd Carlo’/*Jean Orsini nicknamed Carlo*

which is described as: noun(proper) + adjective + noun(proper).

4.2.2 Standard complement order

The standard order for complements belonging to adjective entries is complement + adjective, e.g.

- ‘als arts werkzaam’/(lit.) *as doctor working* (conjunction + adjective).

Thus, predicative constructions like

- ‘(Hij is) als arts werkzaam’/(lit.) *(He is) as doctor working*

are rephrased to this pattern.

In some cases, however, this rephrasing is not possible, e.g.

- ‘hij is wars van vrouwen’/*he is averse to women* (adjective + PP) vs. the ungrammatical
- * ‘de van vrouwen warse man’.

These cases are accounted for in the lexicon by the deviant syntactic pattern adjective+complement.

The decision whether or not it is grammatically possible to apply the standard complement order is to some extent subjective and hence not consistent.

4.2.3 Participle as adjective

Present and past participles of verbs are included in the lexicon as adjectives if they were listed in the source lexicons as an adjective, and if they conformed to the criteria listed in section 1. With respect to participles of separable verbs (see 4.3.4) having an orthographic variant in written language use (written as one or two words), the Dutch spelling guide ‘Groene boekje’ has been determinative whether or not to include these participles in the entry list.

Due to transcategorisation, past participles can be used in attributive use (pattern (a) of 4.2.1). This is not described unless the adjectival past participle has a PP-complement, often introduced by ‘door’. E.g.

- ‘de door de bliksem getroffen auto/(lit.) *the by lightning struck car*

is described as a PP with introducer ‘door’ preceding the adjectival past participle. In predicative usage, however, the past participle is very often ambiguous between verb form and adjective, especially when used with auxiliary verb ‘zijn’. This type of usage generally has no syntactic description.

4.2.4 Complement PP vs. pronominal adverb

Generally, when an adjective combines with a PP, the PP can be replaced by a pronominal adverb if the noun in the PP does not refer to a person:

- ‘de op de auto verliefde jongen’/(lit.) *the with the car in love boy* and
- ‘de daarop verliefde jongen’/(lit.) *the therewith in love boy*.

This use is only described in marked cases attested in the corpus.

4.2.5 Comment field

Entries having equivalent variants (homographs, see 3.1.1) have a comment field: ‘Zie ook <variant form>’/See also <variant form>.

3. Verbs

4.3.1 Grammar

Optional adverbs and adjuncts (e.g. adjuncts of time, place etc.) are considered to belong to the grammar and are not accounted for in the lexicon.

4.3.2 Standard complement order

The following standard complement order has been applied for the description of complementation patterns of verbs:

Subject + verb + indirect object (if any) + object (if any) + adverb/adjunct (if any)

For verbs with a both a grammatical and a real subject, like 'bedroeven' in

- Het bedroeft mij dat hij zich zo gedraagt./*It saddens me that he is behaving this way*
- Er kwam een tip binnen./*there came in a tip*

the standard complement order is subject+verb+indirect object (if any) real subject.

4.3.3 Entry list

Reflexive verbs, such as "zich voelen", are described under the verb entry, in this case "voelen"/*to feel (oneself)*.

Verbs which can only be used as an infinitive complemented by an auxiliary are not included in the entry list (cf. 4.3.9).

4.3.4 Complement: indirect object

In case a verb has more than one syntactic description, one of which includes an optional indirect object, then, generally, an (optional) indirect object has been added to the other descriptions as well.

4.3.5 Reversible complements

If a verb has one obligatory PP-complement but may have two, and there are two possibilities for their mutual order in the sentence, then often (but not consistently) both options are described, e.g.

- ‘Hij babbelt met de kinderen (over de vakantie)’/*He is chatting with the children (about the holidays)*
- ‘Hij babbelt over de vakantie (met de kinderen)’/*He is chatting about the holidays (with the children)*

4.3.6 Complement adjective vs. adverb

A complement adjective is specified as an adjective if it functions as a predicative subject or object complement, in the syntactic descriptions represented by the syntactic functions SUBJPRED and OBJPRED.

- De vakbeweging acht die maatregelen wenselijk (adjective ‘wenselijk’: OBJPRED)/*The trade union considers those measures desirable*

A complement adjective is specified as an adverb if it exclusively functions as a verb specifier. The latter use is only described in marked cases:

- De bijen zoemen zachtjes (adverb 'zachtjes': ADVERBIAL)/*The bees are buzzing softly*

4.3.7 Complement PP: adverbial vs. prepositional object

PP's can have the function of adverbial or prepositional object. The function of prepositional object is used for PP's which are more or less standardly connected with a particular verb. These PP's often (but not always) have an introducer. To more 'loosely connected' PP's, the function of adverbial is assigned. This distinction however is not clear-cut, and therefore should be considered as an indication of a tendency rather than a rule.

4.3.8 Clausal complements: interrogative vs. subordinate

As some introducers of interrogative and subordinate clauses are identical ('dat', 'of', 'wie', etc.), the discrimination between interrogative and subordinate clauses is often difficult to make. The following principle has guided the decision: when the meaning of the verbal entry has an interrogative aspect, then the clausal complement was considered to bear the interrogative (syntactic) feature SSINTERROGATIVE.

4.3.9 Complement: auxiliaries

Auxiliaries have not been described as verbal complements.

4.3.10 Complement control

Basic control is described at the construction level for verbs only (subject, object, and indirect object control). A deliberately global definition of control was adopted, basically addressing the binding of the empty pronominal subject in embedded infinitival clauses

4.3.11 Verbs of the type ‘polsen’

Some verbs like ‘polsen’/to *sound someone*, may take a complement which seems to be a direct object. E.g.

- ‘Hij polst de Commissaris van de Koningin of hij er wat voor voelt’/whether he is *interested*
- ‘Hij polst de Commissaris van de Koningin hoe de bevolking er tegenover staat’/what *the people think about it*

However, the underlined clauses can be replaced by the pronominal adverb *daarover* and for this reason they are considered to be adverbials rather than direct objects.

4.3.12 Comment field

So-called ‘separable verbs’ are a kind of compound verbs with a first component (noun, adjective, preposition etc.) which can be separated from the verb component in the sentence, depending on certain grammatical conditions. Example: ‘*uitlaten*’ in ‘hij *laat* de hond *uit*’/(lit.) *he is letting the dog out (i.e. walking the dog)*. Because of this special grammatical behaviour, this category of verbs has been assigned a comment field ‘Scheidbaar werkwoord’/Separable verb.

4.4 Adverbs

4.4.1 Grammar

The fact that adjectives can be used as adverbs is considered to be accounted for by the grammar.

4.4.2 Entry list

The entry list mainly contains "real" adverbs and pronominal adverbs, with the exception of a few adjectives very frequently used as adverbs (e.g. 'lang'/long).

The prepositions 'boven'/above, 'onder'/under and 'achter'/behind can be followed by another preposition introducing a PP (constituting a succession of two prepositions). According to Dutch grammar (ANS, 1987), they are considered to be adverbs and hence are included in the adverbs entry list.

4.4.3 Clausal complements

Adverbs can introduce a clausal complement: e.g.

- '(de cultuur) *waarbinnen* we leven'/the culture *wherein* we live

4.5 Pronouns

4.5.1 Entry list

Some entries are ambiguous for pronoun or determiner. In nominal use these entries are classified as pronomina (like 'ons' in 'hij beschuldigt ons'/he is *accusing us*), in attributive use (like 'ons' in 'ons huis'/our *house*) they are classified as determiner.

4.5.2 Clausal complements

Personal, indefinite and demonstrative pronouns are often complemented by subordinate clauses introduced by relative pronouns e.g.

- ‘al wie interesse heeft’/*all who are interested*
- ‘al wat nodig is’/*all that is needed*
- ‘allen die komen’/*all who are coming*

These clauses are described in the lexicon as subordinate clauses (feature SUBORDONNEE).

4.6 Determiners

Some entries are ambiguous between determiner or pronoun. In nominal use these entries are classified as pronouns (like 'ons' in *hij beschuldigt ons/he is accusing us*), in attributive use (like ‘ons’ in ‘ons huis’/*our house*) they are classified as determiner.

4.7 Conjunctions

4.7.1 Entry list

Some prepositions, e.g. ‘voor’/*for, before*, ‘met’/*with*, ‘na’/*after*, ‘naar’/*to*, ‘om’/*around, about*, ‘tot’/*to*, ‘zonder’/*without* and the adverb ‘nu’/*now* also function as conjunctions, e.g.

- ‘voor hij naar zijn werk gaat (laat hij de hond uit)’/*before he goes to work, (he walks the dog)*

Consequently, these entries are ambiguous for part of speech and have been assigned two POS.

4.8 Adpositions

4.8.1 Grammar

PP-complements of adjectives are usually described under the respective adjectives. An exception is made for the adpositions *van* and *voor*, which, in restrictive readings like ‘vrolijk *van aard*’/(lit.) *cheerful of disposition* en ‘klein *voor zijn leeftijd*’/small for his age, take adjectives as complement, instead of vice versa.

4.8.2 PP Complements

A few adpositions, *tot*, *van* and *voor*, can directly be followed by a PP:

- ‘*tot aan de grens*’/up to the border
- ‘*tot voor de grens*’/up to before the border

- ‘*van over de grens*’/from across the border

- ‘*voor onder de kerstboom*’/for underneath the Christmas tree

These patterns are described for both adpositions involved. However, the adpositions ‘*achter*, *boven*’ and ‘*onder*’, when followed by a PP, are considered to be adverbs according to Dutch grammar (ANS, 1987; see section 4.4). Patterns like ‘*boven aan de lijst*’/(lit.) *above at the list*, i.e. *at the top of the list* are described as adverbial complementation patterns only.

4.8.3 Clausal complements

Many adpositions can introduce a relative subclause. In these cases, the feature SUBORDONNEE is used (see section 3.2.4, **Contents Specification**).

4.9 Numerals

4.9.1 Entry list

Indefinite numerals are described as POS categories pronoun or determiner.

4.9.2 Grammar

The predicative use of numerals , e.g. in

- ‘er zijn er drie’) /*there are three*

is considered to belong to the grammar and is not accounted for in the lexicon.

4.9.3 Complement categories

Cardinal numerals generally do not have inflected forms, with the exception of some prepositional expressions of time (only relevant for 1-12) and quantification, e.g.

- ‘voor achten’/before eight (o’clock)
- ‘met z’n twintigen’/with twenty people (lit. ‘with his twenties’)

4.10 Residuals

4.10.1 Entry list

The formal difference between acronyms and abbreviations is that the latter have one or more dots (‘.’) in their entry form: ‘UNICEF’ vs. ‘blz.’/p. (*i.e. page*)

4.10.2 Introducer 'de' and 'het'

Both acronyms and abbreviations can take an article complement. The residuals do not carry any morphosyntactic information other than subtype (acronym or abbreviation). From a combinatorial point of view (e.g. parsing), it is desirable to at least have information about article selection in the lexicon. In order to compensate for this, residuals taking articles (like *B.V.*, *ANC*) have been assigned descriptions in which the articles have been specified lexically, i.e. with introducers (like *het* for *ANC*).

4.10.3 Comment field

For abbreviations, the reading is explained in the comment field.

4.11 Interjections

It is in the nature of interjections that they have no syntactic pattern.

12. Articles

No comments

4.13 Unique membership

No comments

Appendix A – Conceptual relations

The lexicon is set up as an SGML file (over 30 MB of plain ASCII). Its contents have been encoded in a distributed manner: all formative entities (like lemmata, syntactic phrases, feature bundles) are SGML entities, related by a pointer mechanism to other entities. The exact specification of the conceptual model underlying the PAROLE lexica can be found in GENELEX (1993). An excerpt of the Dutch instantiation of this model is given in figure 1 below.

Lemma (UM)

1-n 1-n

Graphical form (Gmu) Syntactic unit (Usyn)

1 1

Morphological inflection paradigm (Mfp) Description

0-1

Construction

1-n

Position (PositionC)

1-n

Syntagms: Terminal (SyntagmaT)

Non-terminal (SyntagmaNTC)

Figure 1: Excerpt of the Dutch instantiation of the GENELEX/PAROLE lexicon model (where necessary, SGML object names appear in brackets).

Lemmata have 1 to n graphical manifestations (graphical forms), each of which leads to exactly one morphological inflection pattern. Each orthographic variant of a lemma corresponds to one graphical form.

Every lemma can have 1 to n syntactic patterns. These patterns are represented by a hierarchy of objects, the top node of which is the syntactic unit (Usyn). Every Usyn has exactly one Description associated with it, which consists of an example sentence, an optional comment, and an optional pointer to a syntactic conglomerate object, the Construction. In a number of cases, the Construction is not present; these are cases for which only a comment was added (see 3.2.6 above). A Construction, finally, consists of 1 to n pointers to syntagms, through the intermediate object Position. Positions carry functional complement information (like SUBJECT). Syntagms can be either terminal (basic parts of speech) or non-terminal (syntactic

phrases). They can have morphosyntactic (e.g. number), lexical (heads, or introducers) and syntactic (e.g. control) properties, which are specified by features.

References

G. Geerts, W. Haeseryn, J. de Rooij & M.C. van den Toorn (eds.), *Algemene Nederlandse Spraakkunst* (ANS), Wolters-Noordhoff/Wolters, Groningen/Leuven 1987.

GENELEX (1993), Eureka Project GENELEX, report on the syntactic layer. Version 4.0, GENELEX consortium.

GENELEX (1994), Eureka Project GENELEX, report on the morphological layer. Version 3.3, GENELEX consortium.

Kruyt, J. G. & M.W.F. Dutilh (1997), A 38 Million Words Dutch Text Corpus and its users. In: *Lexikos* (Afrilex Reeks/Series 7), pp. 229-244.

PAROLE Report on the Morphological Layer, Document ID P-WP1.1-MEMO-ERLI-32.

PAROLE Report on the Syntactic Layer, Document ID P-WP1.1-MEMO-ERLI-33.

Woordenlijst Nederlandse taal (1995), SDU Uitgevers/Standaard Uitgeverij, Den Haag/Antwerpen.

Bennis, H. & T. Hoekstra (1989), *Generatieve grammatica*, Foris Publications, Dordrecht.

