LRE Project 62-050 MULTEXT

Workpackage 1

Milestone B Deliverable D1.6.1B

Common Specifications and Notation for Lexicon Encoding and Preliminary Proposal for the Tagsets

Work in Progress

Task Leader:

PISA-ILC - Nicoletta Calzolari and Monica Monachini

Editors:

Nuria Bel, Nicoletta Calzolari and Monica Monachini

corpmon2@icnucevm.cnuce.cnr.it

Contributors:

CNRS - Jean Veronis, Liliane Khouri and Christine Meunier

FBG - Nuria Bel and Ana Aguilar

ISSCO - Susan Armstrong and Graham Russell

MS - Petra Steiner and Lothar Lemnitzer

SNI-S - Juan Alonso

UT - Dirk Heylen and Louis des Tombe

March 1995

Contents

1 Introduction

The work carried out in this task aims at formulating harmonized specifications and at proposing a notation for the lexica and the tagsets, to be contributed by each language group involved in the MULTEXT Project.

MULTEXT's general aim is to develop tools for corpus annotation which contribute to the standardization of this kind of work in an academic and an industrial environment. These tools will be provided with resources from six different languages to ensure their validity. Resources used to feed the tools are, among others, lexical lists for the six languages, containing the necessary information to run the tools. Tools that will use lexica are mainly those which perfom morphological analysis and generation, and lexical lookup tools. MULTEXT proposes to deliver a morphological tool together with basic morphological rules and a number of base form entries, duly coded with respect to the rules. The morphological tool is intended to expand these base forms into word-form lists, with corresponding morphosyntactic information. These word-forms will, in turn, be used for the tagger, providing that a correspondence between the morphosyntactic information and the tags to be used by the tagger is defined. The morphological tool must guarantee extensibility of the MULTEXT tools, as it is thought to be used by end-users to enlarge lexical material treated by the tools. It is also expected that a morphological analysis will be able to perform a "guess" on at least the category of unknown words and, where possible, on morphosyntactic features. Within MULTEXT, therefore, "lexical list" refers to a list of forms with related information: both to base-form lexica, coded in such a way as to feed the morphological tool, and to the word-form lexica, containing relevant information for corpus annotation purposes.

At the first workpackage coordinators' meeting held in Paris, and as also reported in D1.6.1. (September 1994), it was agreed that in view of the urgent need for lexical lists for the creation of the tools, lexical lists of word-forms in a particular format could be supplied already in the first phase, meanwhile leaving for the second phase the development of base-form morphological lexica, input for the morphological tool. These word-form lexical lists were generated from the resources already available at the different sites. Further work will be done in order to ensure the complete mappability between the results of the morphological tool and the formalism proposed for lexical lists.

The present report is mainly devoted to the definition of the information associated with the word-form lists, from now on referred to as "lexical descriptions". We provide here the notation to be used in the lists corresponding to each language to describe a given word-form. Major effort has been devoted to ensure compatibility between the three different types of information to be associated with a given word: morphological information, morphosyntactic lexical description and TAG label.

The present report is divided into four sections:

- section 2 describes the background considerations taken into account, for the definition of the information and its notation, in these lexical lists;
- section 3 proposes a set of lexical specifications and a common notation for the encoding of the MULTEXT lexical lists of word-forms;

- section 4 considers the relationship between the lexical descriptions and the tags to be used for corpus annotation (this part has to be considered as very preliminary, and having the purpose of highlighting problems and issues arising when trying to arrive at similar ways of defining tagsets for different languages);
- section 5 contains language specific applications of the lexical specifications proposed and still idiosyncratic proposals for the tagsets.

2 Background Considerations

Classification of lexical items relies on the old tradition of Greek and Latin grammar. What is normally referred to as "Parts-of-Speech" distinction for different words is well-known to be a crucial task, but not accurate or universal. Lyons (1981, p.109), for instance, warns the reader about this:

> "It is important to realize, however, that the traditional list of ten or so parts of speech is very heteregeneous in composition and reflects, in many of the details of the definitions that accompany it, specific features of the grammatical structure of Greek and Latin that are far from being universal. Furthermore, the definitions themselves are often logically defective. Some of them are circular; and most of them combine inflectional, syntactic and semantic criteria which yield conflicting results when they are applied to a wide range of particular instances in several languages. ... Like most of the definitions in traditional grammar, they rely heavily upon the good sense and tolerance of those who apply and interpret them."

These difficulties in classifying word classes have been the concern of many linguists and greatly affect computational applications, as one cannot expect from machines the sort of "good sense and tolerance" asked for in applying current classifications. On the other hand, the tools MULTEXT is going to develop will be used by humans sharing similar linguistic backgrounds. It is, therefore, imperative that MULTEXT makes these tools user-frendly. It should not be forgotten that the output of corpus annotation as its main goal, as well as the internal codification used for this purpose, should be easily understandable by the expected end-users of its products. MULTEXT tools will be associated with data for demonstration and validation purposes, but, being public domain tools, one should expect that being allowed to use them for experimentation, end-users will incorporate their own classes and distinctions. It must be ensured that users can take supplied data as guidelines to show the functionalities and behaviour of the tools (the MULTEXT-EAST project evidences the importance of this consideration).

With this aim, MULTEXT proposes to address classification problems by joining forces with the EAGLES initiative (MULTEXT T.A. 1993, p.10) which proposes to address them by highlighting "the area of common ground and some aspects of discrepancy between the different systems for classifying morphological units, in order to provide, after testing with respect to all EC languages, the possibility of elaborating common consensual guidelines for morphosyntactic encoding in lexica and corpora" ("Synopsis and Comparison of Morphosyntactic Phenomena encoded in Lexicons and in Corpora. A Common Proposal and Applications to European Languages", Monachini and Calzolari, Oct. 1994, p.12).

In EAGLES, a bottom-up procedure, looking at existing practices in a large number of lexical and textual projects world-wide (both in lexical specifications and in corpus tagsets), has been followed, thus allowing to highlight the large core of commonalities between lexical and textual large projects with respect to the morphosyntactic phenomena described. The procedure adopted within EAGLES was, in fact:

- to survey a number of encoding practices for morphosyntactic description in lexica (mainly MULTILEX and GENELEX, which in turn, are based on many different lexica for many European languages), and in corpora (i.e. the NERC consensual nucleus of morphosyntactic information encoded by the most well-known existing tagging practices and the preliminary scheme proposed by the EAGLES Corpus working groups) with the aim of finding a consensus from their comparison;
- to work in close cooperation between the groups on linguistic annotation of text corpora and morphosyntactic description in computational lexica, wuth the aim of working out a compatible sets of distinctions;
- to first test the proposal by applying it to the European languages.

Thus, the EAGLES proposal (in the already mentioned EAGLES reports "Synopsis and Comparison of Morphosyntactic Phenomena encoded in Lexicons and in Corpora" and the "Morphosyntactic Annotation", Leech and Wilson, 1994) – which is also at the basis of, or is mappable to, the lexical and corpus specifications of the LRE projects DELIS, RENOS, CRATER and MECOLB, MLAP project PAROLE, and the French project GRACE – has been the starting point of Task 1.6 within MULTEXT.
The partners have been asked to:

- evaluate if the features and values presented in the tables for each PoS at Level 1, i.e. the recommended features, suit their respective languages and their established practice (an example of the PoS tables used within EAGLES is given below);

NOUN	Type	Gender	Number	Case	Count	Definitness	Inflection
L0	**N O U N**						
L 1	com prop	m f n	sg pl	nom gen dat acc			
L 2					cou mass		
L 2 b		It c Du f(m) Du cont Sp trns Sp notr	It n	Gr voc Gr ind		Da def Da indf Da unmk	Da/Ge weak Da/Ge strg Da/Ge mix

- add features and values needed at the language specific level.

Reports on the evaluation of the EAGLES specifications have been contributed by the partners involved in this MULTEXT task, and comments, suggestions and critical remarks are being taken into account in the EAGLES proposal which is being accordingly revised. MULTEXT Task 1.6. can be seen as the largest contribution, together with DELIS, to the testing, refining and revising of the EAGLES proposal. An example of this interaction was the major revision of the EAGLES proposal which affected the Pronoun/Determiner category proposed, now split into two different categories.

Experience shows that the process of consensus building is a slow process, because of the different interests to be adjusted. Considerations coming from "re-usability" of existing material, as well as from theoretical and application-oriented arguments, have been raised in discussions under this task and should also be taken into account when evaluating its progress and results. Leading ideas to reach final decisions have been described above and will be examined in detail in the following subsections. They can be summarized by the statement of the MULTEXT strong committment to standardization and harmonization of lexical encoding initiatives, now active in Europe with the aim of sharing public domain resources.

2.1 Lexical descriptions and corpus tags

After discussion on several issues, the MULTEXT partners agreed on the necessity of differentiating corpus tags used for the PoS disambiguator, or tagger, from the information which a lexicon can offer. This is because the former is an application-oriented representation of the information described by the latter and depends very much on the tool used. This decision was also in accordance with the orientations given by the EAGLES Lexicon and Corpus working groups (see Monachini and Calzolari, 1994).
Thus the terminology adopted in MULTEXT reflects this separation:

- the term "feature" is preferred when talking about lexical descriptions;
- the terms "tag" and "tagset" are used for the information to be associated with words for automatic corpus annotation.

Hence, it was agreed that two different objects will be produced for each language:

a. a lexicon where morphosyntactic features for each word form are encoded with fine granularity, as close as possible to the recommended EAGLES level-1.

b. a set of tags for the purpose of automatic disambiguation. In practical terms these tags are to reflect broader categories on the basis of the limitations of a statistical tool. This set will be defined and refined upon experimentations with the tagger tool.

In Task 1.6, it was decided – in accordance to the Technical Annex – to begin taking EAGLES recommendations as input for deciding on the basic morphosyntactic information to be associated with the word-forms contained in the lexicon. The MULTEXT application of EAGLES recommendations needed to ensure that the information contained in the so-called Level-1 were significant for most of the languages to be treated. Thus the work under this task may also be considered as a concrete validation of EAGLES work on electronic lexica. The underlying aim of EAGLES is concerned with the re-usability of electronic lexica, and, following this general tendency, MULTEXT lexical descriptions also had to be (as far as possible) independent from the application, aiming at a general description of each language and containing a basic set of shared information. Also, for the sake of "re-usability" of the lexical material supplied, it was judged that the lexical information to be encoded should be as detailed as possible. Thus fine-granularity of the information would allow other users to rearrange categories, when necessary, without much difficulty.

The actual corpus tags we will be using will depend on at least the following:

1. the lexical features, and
2. the capabilities of the MULTEXT tagger to disambiguate between different lexical decriptions, or different types of typical homographies present in different language types.

We can fix (1), but (2) is highly dependent on the tool. That is why we concentrated on (1) in the first phase.

The corpus tags will be developed for each language with a specific application in mind, i.e. that of producing a corpus tagged for part-of-speech (and possibly other morphosyntactic information) by means of automatic disambiguation. The set of corpus tags will, very likely, be revised many times during the course of the project, in order to find an optimal set for each language.

It would be ideal to tag a corpus with the lexical descriptions themselves for each word. However, it is well known that this is well beyond the capabilities of the state-of-the-art tagging techniques.

Corpus tags are, therefore, to be seen as kinds of underspecified lexical tags. There are two reasons why we may want underspecified corpus tags:

1. Experience shows that some distinctions are difficult to get right with a high accuracy.

For example, in some languages, the disambiguation between indicative present and subjunctive present in a corpus is extremely difficult to achieve by automatic means. If some verbs have different forms for the indicative and the subjunctive (e.g. Fr. venir: indic. = viens, subj. = vienne; It. indic. = vieni, subj. = venga), many have the same form (e.g. Fr. manger: indic., subj. and imper. = mange; It. indic. and subj. = ami). In this latter case, disambiguation can only be achieved with very complex parsing of sentences.

Therefore, lexical entries will contain the following detailed and granular information associated with the word-forms

```
mange (manger) Main verb Indicative present, 1st person sing.
mange (manger) Main verb Indicative present, 3rd person sing.
mange (manger) Main verb Subjunctive present, 1st person sing.
mange (manger) Main verb Subjunctive present, 3rd person sing.
mange (manger) Main verb Imperative present, 2nd person sing.

ami   (amare)  Main verb Indicative present,  2nd person sing
ami   (amare)  Main verb Subjunctive present, 1st person sing.
ami   (amare)  Main verb Subjunctive present, 2nd person sing.
ami   (amare)  Main verb Subjunctive present, 3rd person sing.
```

wheras corpus tags will provide broader categories, collapsing several lexical descriptions.

2. In order to train the tagger, we need statistical tables (based on co-occurrences of tags). If we have a large tagset, we need a very large corpus to train the disambiguator, in order to observe rare co-occurrences. For example, in the proposal for French (see below), there are 249 different lexical descriptions, but only 74 collapsed corpus tags. Experience (Church, Penn Treebank, IBM France, etc.) shows that the tagset should be under 100. Actually the Penn Treebank collapsed many tags compared to the original Brown corpus, and got better results.

Two other observations are of relevance as regards the relation between lexical specifications and corpus tags.

(a) Sometimes tagging classes are in reality different from lexical descriptions. For example, classes for punctuation are needed, certain types of semantic or pragmatic or lexical information can be present in the tags (e.g. the days of the week).

(b) Furthermore, the "collapsing" decisions in TAGS are language dependent, therefore it is not possible to have completely identical tagsets across languages. To illustrate, we can give as an example the differences related to person differentiations in verbal morphology.

In Spanish, first and third person of different tenses have the same spelling:

```
Yo/El cantaba (Imperfect)
Yo/El cantari'a (Conditional)
Yo/El cante (present of subjunctive)
```

Taking into account that the subject in Spanish is not obligatory, and that the tagger cannot know if the preceeding NP is in fact the subject of the verb, there is no way to discriminate between the two forms. Hence a conflating tag is recommended, marked for instance as "non-second-singular" form or as "first- third singular". Also French has homographs for different verbal persons, but these are the first and the second person of some tenses:

```
Je/Tu viens
Je/Tu e'tais
```

The French tag cannot be the same as the Spanish one, but it could be "non-third-singular" or "first-second-singular". Moreover, having two different tags in French for the homograph could be justified, due to the obligatory presence of a lexical subject, as the tagger will be able to disambiguate among them due to the presence of a pronoun in a near context of most of their occurences.

For some languages (e.g. French, English and Italian) a lot of past experience and empirical evidence exists, which can be used to choose a reasonable initial tagset, that can be seen as preliminary and which can be refined later on in the project. For example, for English, the Penn tagset or the BNC are very good candidates. For French, the IBM tagset is a very good start (the French proposal presented in the following is very close to it). For Italian the tagset based on the DMI (Calzolari et al. 1983) is also a good starting point. These tagsets are the result of years of trial-and-error adjustments, and it seems reasonable not to ignore them. All of these tagsets are, moreover, compatible with the EAGLES proposal, i.e. mappable to it.

2.2 Lexical lists: lemmas and inflected word-forms

As stated in the Introduction, MULTEXT will supply a morphological tool which will need some information on lemmas in order to produce the entire set of associated word-forms. A list of word-forms will constitute another lexicon which, as referred in the Technical Annex, constitutes a value in itself. Hence, the lexica supplied by MULTEXT are of two types:

(a) word-forms, containing

Word-form, morphosyntactic information, lemma, TAG

(b) lemmas, containing

Lemma, morphosyntactic information, inflection information

The information and notation of the lemma dictionary is closely related to the morphological tool used and also on the rules implemented within the tool. Due to the fact, mentioned already in the Introduction, that the availability of word-form lists was considered of priority for corpus annotation tool development, we first concentrated on the definition of the word-form lists following EAGLES recommendations for the morphosyntactic annotation to be encoded, as explained in the preceding section. It was possible to define a representation of morphosyntactic information for these word-form lists independent from a morphological tool, in such a way as to ensure that lemma dictionaries and the output of morphological modules (the ones produced for MULTEXT or others) be compatible and easily mappable to such lists. Following current practices for NLP, the notation used should represent information in attribute/value formalisms (as was done also in EAGLES) and should also be self-informative for human inspection and understanding. Considerations concerning the desirability that these descriptions are able to provide information about language-specific characteristics, where also taken into account. Following these ideas, a notation format was suggested whose main characteristics are:

- attributes are marked by positions;
- values are represented by a single character;
- a special marker reflects the non applicability of a given attribute.

These characteristics make the proposed lexical description notation (see section 3.1 for more details) synonymous with attribute/value pairs used in current unification formalisms. The next sections introduce such formalism and the information to be encoded.

3 MULTEXT lexical specifications and notation

3.1 Notation

The notation format proposed to represent lexical descriptions consists of linear strings of characters representing the morphosyntactic information to be associated with word-forms. The string is constructed following the philosophy of the Intermediate Format proposed in the EAGLES Corpus proposal (Leech and Wilson, 1994), i.e. of having agreed symbols in predefined and fixed positions: the positions of a string of characters are numbered 0, 1,2, etc. in the following way:

a. the agreed character at position 0 encodes part-of-speech;
b. each character at position 1, 2, n, encodes the value of one attribute (person, gender, number, etc.);
c. if an attribute does not apply, the corresponding position in the string contains a special marker, in our case '–' (hyphen).

`Example: Ncms- (noun,common,masculine,singular,nocase)`

This notation adopts the EAGLES Intermediate Format with a small revision: the Intermediate Format encodes information by means of digits, while in MULTEXT characters of a mnemonic nature are preferred.

It is worth noting here that this representation is proposed for word-form lists which will be used for a specific application, i.e. corpus annotation. We have foreseen these lexical descriptions as containing a full description of lexical items. As noted above, the sets of tags, to be used properly for automatic corpus annotation tools, are expected to contain less information.

These lexical descriptions can be seen as notational variants of the feature-based notation in the form of attribute-value pairs. In fact, the string notation proposed, e.g.

`Ex.: Ncms- (noun,common,masculine,singular,nocase)`

is completely synonymous to a feature-structure representation:

`Ex.: {cat=noun, type=common, gender=masculine, number=singular, case=none}`

or

`{cat=noun, type=common, gender=masculine, number=singular}`

The above feature structures are often also represented as follows:

```
+-                     -+
| Cat:    Noun          |
| Type:   common        |
| Gender: masculine     |
| Number: singular      |
+-                     -+
```

Formal characteristics relevant for our applications have been kept. Use of position in the string to encode attributes makes no restrictions on the set of characters to be used as values. It could then be inferred that, if we wanted to keep the formal characteristic of order independent notation, we would have to make sure that the characters meant to represent attribute-values are not ambiguous. As attributes and values are linked by positional criteria, the need of a special marker for void attribute-value pairs is evident if we want to keep descriptions coherent. Thus, the "Ncms-" style can be viewed as a short-hand notation convenient for some users and straightforwardly mappable to the information used in unification-based attribute-value pairs formalisms.

When comparing MULTEXT lexical description representation format with other notations one must keep in mind that they are intended to describe word-forms, and are used in very large lexical lists which contain word-forms. It seems to us relevant to comment on this point because, although it can be justified (and we will do so below) that the same formal operations can be declared in both styles, there is little evidence for justifying the need of operations such as negation and disjunction of features and values when applying them to tagged word-forms as a result of corpus annotation.

3.1.1 The use of '–' ('not-applicable')

We call this marker 'not-applicable'[1], and, as stated above, its function is just to keep the relationship established between attributes and values. It might be used for the following cases (it has been proposed to use the 'not-applicable' marker in order to encode the case of a not-applicable feature for a particular language. However this decision is still under discussion due to the facts reported in section "Comparison of attributes/values used by languages"):

a. not applicable given a particular combination of attributes/values, i.e. although the attribute applies to the category in a given language, it does not apply to a particular subclass of the category.

b. not applicable to a particular lexical item, although the attribute applies to the rest of its paradigm.

Example: in the description of pronouns, for personal pronouns the grammatical person is to be encoded, but for demonstrative pronouns it is avoided; in this case '–' is applied following (a). On the other hand, gender cannot be informative for some personal pronouns, but it is still relevant for other personal pronouns; the application of '–' follows (b):

```
Pd-ms     "Este"     Pronoun, demonstrative, masculine, singular.
Pp1-s     "Yo"       Pronoun, personal, first, singular.
Pp1mp     "Nosotros" Pronoun, personal, first, masculine, plural.
Pp1fp     "Nosotras" Pronoun, personal, first, feminine, plural.
```

Their uses are clearly not equivalent, but there would only be meaningful differences would occur in highly typed theories of lexical description. For illustrating this point let'us have the following type system for pronouns:

[1]This section and the following are still under discussion and present proposals coming from different partners which still need to be revised and/or refined.

```
TYPES          SUBTYPES              ATTRIBUTES     VALUES

Pronoun                               gender        masculine
                                                    feminine

                                      number        singular
                                                    plural
               Demonstrative

               Personal                person          1
                                                       2
                                                       3
```

For this system, gender and number attributes belong to the set of features which describe all pronouns. Person will only belong to the set of features which describe personal pronouns – in addition to gender and to number. Applied to this type system, case (a) would mean that the attribute-value pair does not belong to the set of features which describe a subtype, while (b) would mean indeterminacy of a given word-form (which could be expressed as a disjunction of all the values for the particular attribute or leaving a void for the value, being open to unification; this choice mainly depends on the purpose of the description, e.g. syntactic parsing).
Different representations will result: 'este' description corresponds to case (a), and 'yo' description corresponds to case (b) (subtypes are represented 'dem' and 'pers'):

```
|phon                este|
|cat      |gender    masc||
|    'dem'|number    sing||

|phon                yo   |
|cat       |gender   []   ||
|    'pers'|number   sing||
|          |person   1    ||
```

In simpler flat type systems where distinctions are made only for the generic type "pronoun", both cases a. and b. will be treated by unification mechanisms in the same way.

From the conversion point of view, we have to be concerned with the output of the MULTEXT morphological tool, as it will be the source of word-form lexical lists. The Mmorph tool does not incorporate a highly hierarchical typing system and thus no problems are expected in converting Mmorph output into lexical descriptions of the proposed format, if desired. The results from applying the Mmorph tool will probably (it strongly depends on implementation strategies) be the following:

1. a non present attribute in the description attached to the word-form;
2. a disjunction expression, i.e. {gender=masc|fem};
3. encoded as a third possible value, i.e. {gender=none}.

The simplest case for converting would be the third one, as then automatic non-intelligent conversion is possible. In the first two cases the conversion routine will have to make some inferences on type declarations. It is also expected, that when converting from other lexical sources, special conversion routines will have to be used. As seen above, the conversion from "Ncms" lexical description notation into other unification based format will only be difficult if the target formalism is a highly typed system. If this is not the case, the presence of the "not-applicable" marker will have to be converted into a special value or into nothing, leaving it open. For conversion into highly typed system it might be useful to have cases (a) and (b) marked by different characters, in order to guide an intelligent conversion routine to the desired results.

3.1.2 Mapping of lexical descriptions onto corpus tags

The tags (see the examples below) used to exemplify issues and problems to be dealt with in the mapping between lexical descriptions and corpus tags, come from the tagsets proposed in the language-specific applications of four of the MULTEXT partners. These tagsets (containing dfferences among them, because constructed on the basis of tagging practices already used by the partners) should be considered as a preliminary proposal to be discussed for harmozation and refined after experimentations on the MULTEXT tagger.

Mapping of these lexical descriptions into corpus tags has also been taken into account. It is also considered desirable to see whether under-informative corpus tags can be directly mappable to the lexical descriptions each one subsumes.

Decisions about corpus tags are language dependent. The information to be encoded depends on the ability of a given tool to disambiguate between different potential lexical descriptions for a given word-form. We have already mentioned the key concepts to be applied for defining sets of corpus tags in the preceding sections. Therefore one can first assume that the mapping from lexical descriptions onto corpus tags can be done with conversion tables which relate two different items: corpus tags and lexical descriptions. These tables are likely to be modified many times in the course of the project, based on experimentation with the disambiguation tool.

An example of such mappings is:

```
Lex.spec. TAG   Definition

Pp1msa-   P1S   Personal pronoun, first person, masc. sing. accusative
Px1msa-   P1S   Reflexive pronoun, first person, masc. sing. accusative
Pp1fsa-   P1S   Personal pronoun, first person, fem. sing. accusative
Px1fsa-   P1S   Reflexive pronoun, first person, fem. sing. accusative
Pp1msd-   P1S   Personal pronoun, first person, masc. sing. dative
Px1msd-   P1S   Reflexive pronoun, first person, masc. sing, dative
Pp1fsd-   P1S   Personal pronoun, first person, fem. sing., dative
Px1fsd-   P1S   Reflexive pronoun, first person, fem. sing., dative
```

All these lexical descriptions correspond to the Spanish form "me". For this word-form the tags P1S – which conflates all the possible lexical descriptions – has been decided on the basis of the assumption that an automatic tool would have disambiguation problems in assigning the correct analysis among all the lexical descriptions. The correct analyis of this word-form would require syntactic analysis.

The mapping from the lexical descriptions to the corpus tags should be applicative, that is, "each lexical description should map to one and only one corpus tag, while it is not possible to do the reverse" due to the limitations of current tagging techniques. The situation where corpus tags are more precise than a lexical description (i.e. one lexical tag corresponds to more than one corpus tag) should be, in principle, avoided.

In order to avoid redundancy in the conversion tables and to make tag optimization work easier, it has been proposed to study the possibility of having intermediate representations which prepare the conflation of information and which facilitate automatic mapping from lexical descriptions onto tags. This intermediate internal notation makes use of "regular expressions" which incorporate operators in order to sum up the information referred by different lexical descriptions and conflated in a given tag. For the example given above, the resulting regular expression may incorporate two operators: "match any" (.), "list" ([]) – other possible operators proposed are "disjunction" | and negation ^.

`P[px]1.s[ad]-      P1S`

However, the application of such regular expressions is still being studied as its use conveys some requirements on the conflation of lexical descriptions and on the construction of corpus tags. An example will illustrate the issues to be taken into account. For Spanish, first and third person of some tenses are homographs. This can be taken into account when conflating information:

Verbal paradigm	regular exp.	TAG
cantaba, comi'a, veni'a	Vmii[13]s-	VMIIS
cantari'a, comeri'a, vendri'a	Vmcs[13]s-	VMCSS
cante, coma, venga	Vmsp[13]s-	VMSPS
cantara, comiera, viniera	Vmsi[13]s-	VMSIS

For Italian, the conflation of information on homographs also in the verbal paradigm may cause problems to the applicative principle mentioned above:

Verbal paradigm	lex.descr.	regular exp.	TAG
premiate	Vmip2p- Vmmp2p- Vmsp2p- Vmps-pf	Vm([ims]p2p-)|(ps-pf)	VMP2IMCPP
leggete	Vmip2p- Vmmp2p-	Vm[im]p2p-	VMP2IMP
leggiate	Vmsp2p-	Vmsp2p-	VP2CP
lette	Vmps-pf	Vmps-ps	VFPPR

As can be seen, if we use tags such as the ones above which are based on the principle "one graphical form – one tag", there is a violation of the applicative principle, i.e. the same lexical description will correspond to two different tags, because of different conflation clusters.

In general, it is observed that the use of operators in regular expressions results in a form of marking the information which is not going to be expressed in the corpus tag. Thus, tags would have to contain less information than the regular expression and hence than the lexical description.

Another issue to be considered is the following. Having tags with little lexical information, as in the following French example, may lead to another problematic issue in cases where such regular expressions are also used in helping to recover all possible lexical information from a given "under-specified" corpus tag. The mapping from the regular expression onto lexical descriptions will also have to take into account the word-form in order to reject possible descriptions which do not correspond to the tagged word-forms. Below are some examples from the proposed verbal tags and regular expressions:

```
TAG       Regular expression  Lexical descriptions  Possible word-forms

VM1P      Vm[iscm][pifs]1p--  Vmip1p--             venons
                              Vmii1p--             venions
                              Vmif1p--             viendrons
                               ...                   ....
```

Let us consider that the word "venons" is tagged as "VM1P". If we want to know which are the lexical description to which the tag can be referring to, the explosion of the information contained in the regular expression will also give lexical descriptions which do not correspond to the word "venons", but to other words. Regular expressions can only map a given tag for a word-form into all possible lexical descriptions for such a word-form if the information conflated only reflects ambiguities due to homography. Only with this criterion for defining tags, all the possible lexical descriptions subsumed by the corpus tag and expressed in the regular expressions will be true of a given tagged word.

If the criterion for conflating information is limited to homograph ambiguities, we see – as in the following example – that all possible lexical descriptions expanded from the regular expression are true of a given word-form.

```
TAG       Regular expression  Lexical descriptions  Possible word-forms

VSXICP    Vm(sp.s)|(ip2s)-          Vmip2s-              ami
                                    Vmsp1s-              ami
                                    Vmsp2s-              ami
                                    Vmsp3s-              ami
```

As mentioned in the section "Comparison of Attributes/values used by languages", the application of the proposed operators in regular expressions for avoiding redundancy, in some cases, is not needed if lexical expressions already encode the possibility of having, for a given word-form, more than one possible lexical description. This is the case with the proposed values “common” for gender, “invariant” for number (in Italian), or “object” for case (in French pronouns).

Almost all the languages treated in MULTEXT have nouns, adjectives, determiners (among others) which have the same word-form both for feminine and masculine agreement. The Italian group has proposed a value for gender named “common” which avoids having to write two different entries with the same word-form, but with different lexical descriptions. In fact, this use of a special value advances the possible use of proposed operators in the regular expression.

```
word-form       lexical description regular expression  TAG
insegnante      Nccs-               Nccs-               NNS
```

could also be expressed as:

```
word-form       lexical description regular expression  TAG
insegnante      Ncms-               Nc[mf]s- or Nc.s-   NNS
                Ncfs-               or Nc(m|f)s
```

The need, as well as the consequences, for the mapping between lexical descriptions and corpus tags, of the regular expressions must still be regulated. It should be noted that regular expressions can be regarded as a convenient way to map the lexical descriptions to the corpus tags since, in many cases, the information in the lexicon is more precise than the information we can/want to have in the corpus tag set. Such a mapping still seems very interesting because there are many corpus tag systems, even for the same language, which makes it extremely difficult to relate the one to the other. Regular expressions could act as a common reference for the different systems to make comparison easy. Besides, regular expressions could make translations between the lexical description and corpus tags easier and enable the automatic generation of conversion tables.

3.1.3 Attribute/value tables

The categories listed below with the relevant attributes and values are based on EAGLES documents and are the results of a first testing based on a proposal made by Veronis et al. 1994 for lexical specifications in MULTEXT.

As it has already been mentioned in the section “Background considerations” that propose features for describing lexical items of different languages aiming at defining a set which can be said “common” for all of them is a complex task. The underlying philosophy for this task has then be to lead different groups into a pragmatic solution where the concept of an ”harmonized” set of features could be reached.

The groups have first worked out their lexical descriptions taking as input EAGLES and Veronis et al. (1994) documents. The very general criterion was to encode those proposed features which were considered relevant for the language in question. Therefore MULTEXT also followed EAGLES bottom-up methodology in trying to define extensively the features “used” in the lexical descriptions for each group language, as this procedure will make evident the features commonly used. After this phase, whose result can now be seen in the section

"Comparison of attribute/values used by the groups", a new phase is envisaged as to accomodate language-specific considerations into a general model to be used by MULTEXT. This accomodation must take into account extensibility to other languages and also application motivated arguments, as well as internal coherence. For this new phase more specific criteria would be desirable with respect the addition of new features to the EAGLES Level-1 set. The aimed result is a "harmonized" set of features which properly describe lexical items of the different languages.

Following the general aim of the project, these harmonized specifications – and the related resources – will contribute to the standarization of the corpus annotation work. They are supposed to serve as a user oriented additional characteristic of our tool package in the sense that end-users will have a common ground for inspecting and understanding the resources and tool results independently to a large extent of the language. This common set of features will also be a common ground to perform comparisons of different annotation tool results, because, as mentioned in the previous section, the existence of many lexical description systems is causing nowadays a problem for comparing results.

Therefore the categories and features listed below are the common reference for the work done by the different groups. Further discussion on this first proposal is to be found in the section "Comparison of the attributes/values used by the groups" which is in turn to define criteria for changing this first proposal.

Tables of categories

```
=============== ====
Part-of-Speech  Code
=============== ====
Noun            N
Verb            V
Adjective       A
Pronoun         P
Determiner      D    (for those who do not have a separate category
Article         T     for Articles, these are included in Determiner)
Adverb          R
Adposition      S
Conjunction     C
Numeral         M
Interjection    I
Unique          U
Residual        X
Abbreviation    Y
=============== ====

Each character at  positions  1,  2,  etc.  encodes the  value  of one
attribute  (person,  gender,  number,  etc.),  according to the tables
given below.
```

2.2.2 Attribute/value tables

Abbreviations used:

- P Position (starts with 0 for encoding PoS values)
- ATT Attribute name
- VAL Value
- C Code

1. Nouns (N)

P	ATT	VAL	C
1	Type	common	c
		proper	p
2	Gender	masculine	m
		feminine	f
		neuter	n
3	Number	singular	s
		plural	p
4	Case	nominative	n
		genitive	g
		dative	d
		accusative	a

2. Verbs (V)

P	ATT	VAL	C
1	Type	main	m
		auxiliary	a
		modal	o
2	Mood/VForm	indicative	i
		subjunctive	s
		imperative	m
		conditional	c

		infinitive	n
		participle	p
		gerund	g
		supine	s
		base	b
3	Tense	present	p
		imperfect	i
		future	f
		past	s
4	Person	first	1
		second	2
		third	3
5	Number	singular	s
		plural	p
6	Gender	masculine	m
		feminine	f
		neuter	n

3. Adjectives (A)

P	ATT	VAL	C
1	Type	qualificative	f
		ordinal	o
		cardinal	c
		indefinite	i
		possessive	s
2	Degree	positive	p
		comparative	c
		superlative	s
3	Gender	masculine	m
		feminine	f
		neuter	n
4	Number	singular	s
		plural	p
5	Case	nominative	n

		genitive	g
		dative	d
		accusative	a

4. Pronouns (P)

P	ATT	VAL	C
1	Type	personal	p
		demonstrative	d
		indefinite	i
		possessive	s
		interrogative	t
		relative	r
		exclamative	e
		reflexive	x
		reciprocal	l
2	Person	first	1
		second	2
		third	3
3	Gender	masculine	m
		feminine	f
		neuter	n
4	Number	singular	s
		plural	p
5	Case	nominative	n
		genitive	g
		dative	d
		accusative	a
		oblique	o
		object	j
6	Possessor	singular	s
		plural	p

5. Determiners (D)

= ============== ============== =

P	ATT	VAL	C
1	Type	demonstrative	d
		indefinite	i
		possessive	s
		interrogative	t
2	Person	first	1
		second	2
		third	3
3	Gender	masculine	m
		feminine	f
		neuter	n
4	Number	singular	s
		plural	p
5	Case	nominative	n
		genitive	g
		dative	d
		accusative	a
		oblique	o
6	Possessor	singular	s
		plural	p

6. Articles (T)

P	ATT	VAL	C
1	Type	definite	d
		indefinite	i
2	Gender	masculine	m
		feminine	f
		neuter	n
3	Number	singular	s
		plural	p
4	Case	nominative	n
		genitive	g
		dative	d

		accusative	a

7. Adverbs (R)

P	ATT	VAL	C
1	Type	general	g
		particle	p
2	Degree	positive	p
		comparative	c
		superlative	s

8. Adpositions (S)

P	ATT	VAL	C
1	Type	preposition	p
		postposition	t
		circumposition	c
2	Formation	simple	s
		compound	c

9. Conjunctions (C)

P	ATT	VAL	C
1	Type	coordinating	c
		subordinating	s

10. Numerals (M)

P	ATT	VAL	C

```
1 Type           cardinal      c
                 ordinal       o
- -------------- ------------- -
2 Gender         masculine     m
                 feminine      f
                 neuter        n
- -------------- ------------- -
3 Number         singular      s
                 plural        p
- -------------- ------------- -
5 Case           nominative    n
                 genitive      g
                 dative        d
                 accusative    a
= ============== ============= =
```

```
11. Interjections (I)
```

```
12. Unique membership class (U)
```

```
13. Residual (X)
```

```
14. Abbreviations (Y)
```

3.1.4 Comparison of Attribute/value features used by the groups

The following tables reflect the attributes and values used for lexical description in MULTEXT. They take into account input supplied by different groups which the reader can find further detailed in specific language application annexes (see section 5). It is worth noting that the tables reflect both features used by five groups using lexical descriptions as proposed in previous versions of this document and also features for Dutch resulting from morphological generation with the "mmorph" tool. The comparison is certainly of help in order to have a clear picture of the level of consensus reached with respect to harmonization when elaborating lexical lists. It has already been mentioned in previous sections that the project is working towards defining criteria for the application of EAGLES guidelines on standardization of lexical resources for easy re-usability.

Looking at the tables below some very general issues arise with respect to the application criteria of the general tables to the particular languages and the interpretation of the general guidelines. Until now groups have been working on the assumption that they must encode recommended Level-1 EAGLES features if they are relevant to their languages. The possibility of adding new values and new attribute/value pairs was also foreseen if recommended features were not enough to describe lexical items with fine-granularity of lexical descriptions. It was also found useful in view of supplying lexical material to be used by other tools than the MULTEXT ones. This openness has led to a number of incoherencies with respect to application criteria which we summarize in the points below. A decision with respect to general criteria for application must be reached in the next phase. Hence, here the issues concerning harmonization which arise from comparing application sections follow.

There is an unbalanced treatment of features considered as "general" for the different categories. The presence of a particular attribute seems to be mainly justified for two reasons:

– representative in most of the studied languages;
– linguistic tradition.

We see in the comparative tables that a particular language is allowed to add a new attribute because of its relevance for the lexical description of a given category (i.e. when the language items belonging to that category are inflected or marked with respect to it). The most evident case is the proposal made in order to encode Possessor-gender (among other features) for Pronouns and Determiners. It is obvious that this feature cannot be used by languages which do not have different forms regarding this particular distinction.

On the other hand, note that "case" as a feature recommended as "general" for describing Nouns, Adjectives, Pronouns, Determiners, Articles and Numerals, is in fact used only for Pronouns by most of the languages, and only German can apply it for the rest of the categories.

What we mean by "unbalanced treatment" has to do with the fact that features being used by just a few languages, or even just one, receive different treatment when considering them "general" or "language specific".

Also arising from the possibility of adding language specific attributes and values where relevant for a given description, the procedure followed in this task has shown that it has not been easy to reach a consensus in order to harmonize a number of specific features and values considered by a given language. One of cases of such proliferation of features is seen, in fact, when considering the comparative table for Determiners. One of the groups suggests having language specific types to refer to "definite article" and "indefinite article", while other groups prefer to have a general type "article" and other attributes, i.e. Quantification or Definiteness to encode this distinction at a lower level. The particular features suggested by the groups which, in our opinion, could be adapted to the EAGLES model will be discussed during the next phase.

Because of this openness with respect to adding attibutes and values, we would like to point out the case for the values "common" and "invariant" added by the Italian descriptions to all nominal inflected categories for the attributes "gender" and "number" respectively (where a disjunction of values could be used instead). It is a fact that most of the languages could easily adopt this value for the forms which are identical for masculine/feminine, singular/plural agreement features, but this issue has certainly to be clarified and further discussed. Probably a decision with respect to the "fine granularity" of lexical descriptions should be devised. In fact there is another example in another category of the same strategy, that is to conflate in a new value for a given attribute a homography which causes explosion of entries. The French group has suggested conflating accusative/dative values for pronominal case into "object" as a generic value. The new division proposed would also apply to other romance languages but it might compromise the "fine granularity" tendency the project aimed at for lexical descriptions.

There can also be observed a certain confrontation of two different traditions when some groups propose to add a new attribute to characterise an element while others propose to add a new value to label a new class under a general, already available attribute such as "type". To add a new attribute would correspond to the unification based grammar practice, and a label for a class would correspond to the so called "taxonomic" theories. We see an example of this confrontation in the proposal of having an attribute/value "wh" for marking relative particles in different categories: pronouns, determiners, adverbs. EAGLES level-1 seems to prefer separating relatives with a different value for the attribute "type" of pronouns and determiners. Marking as an additional feature the relative characteristics of a given pronominal would help for instance to specifically characterise items such as the English "whose" or Spanish "cuyo, cuya, cuyos, cuyas" which are normally described as Possessive relative pronouns. Under the current classification a decision must be made either to put them under the Possessive or the Relative value of the attribute "type". It has also to be mentioned that no special treatment can be made for relative adverbs which are not taken as a separate class under Adverb type in the EAGLES proposal. Thus, from the comparison made, it is worth mentioning that a new attribute "wh" for adverbs or, as suggested by the German group, a new value for interrogative – and also for relatives – adverbs should be devised.

As we have seen, the EAGLES recommendations lack in some cases the desired fine-grained distinctions which groups working in MULTEXT consider desirable for our applications. Another example of this case is raised by German and English. The groups dealing with these languages – and it could also be applied to the rest of the languages – have suggested a specific value for comparative conjunctions. This addition seems reasonable under the argument that it is an important feature with respect to distributional criteria and can be of great importance for tagging purposes. Again, some guidelines must be defined for considering the addition of features not contained in EAGLES level-1, but it is worth noting that several of the features added for language specific reasons could be considered as applicable to the rest of the languages.

We recommend a new round of discussions on the new features suggested in specific language applications to see whether they can be of use in our concrete application and applicable to the rest of the languages. Once this discussion has led to conclusions, the approved features must be included in the general model. Besides linguistic considerations, having an agreed set of general features is of great concern for the chosen notation style in lexical descriptions. There must be regulations with respect to the encoding of language specific attributes by the other groups or on the ways of differentiating them from those of the general model. This is especially relevant if general conversion routines are to be developed. And because of theroretical coherence, the treatment given to these features must take into account the above mentioned "unbalanced treatment of features". Some other doubts remain in connection with theoretical coherence and the applied nature of the lexica to be supplied. We would only mention one of them to illustrate the kind of issues which must be taken into account in the next phase. It has to do with agreement features of person, number and gender. Are they to be encoded with respect to grammatical agreement or with respect to semantic differentiations. As it is now, following EAGLES recommendations, it seems as if only semantic considerations are taken into account, i.e. Possessive-person of determiners is taken as the "possessor person" for most of the languages which in fact does not trigger agreement.

A decision must be taken with respect to these cases and more specific guidelines must be established for further development of lexical descriptions. It seems from the comparison made that the general criteria "relevant for your language" is not enough. New guidelines must also take the application side into account.

Comparison tables

```
Abbreviations used:
    P   = Position (starts with 0 for encoding PoS values)
    ATT = Attribute name
    VAL = Value
    C   = Code

    x   = value marked by a given group (any character other than x
          means that a given 'language group' codes, in their
          application, the relevant value with that character, not using
          the agreed one).
          The column of characters is left empty in correspondence of
          language specific attributes/values of
          Dutch: they  are attested, in fact, among the set of attributes
          and values for Dutch implementation of Mmorph, where they are
          not represented by means of single codes.
```

1. Nouns (N)

Features used by the groups IT DE ES FR NL EN

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	common	c	x	x	x	x	x	x
		proper	p	x	x	x	x	x	x
2	Gender	masculine	m	x	x	x	x		x
		feminine	f	x	x	x	x		x
		neuter	n		x				x
	l-s.	common	c	x					
	l-s.	De						x	
	l-s.	Het						x	
	l-s.	None						x	

P	ATT	VAL	C	IT	GE	SP	FR	DU	EN
3	Number	singular	s	x	x	x	x	x	x
		plural	p	x	x	x	x	x	x
	l-s.	invariant	n	x					
4	Case	nominative	n		x				
		genitive	g		x				
		dative	d		x				
		accusative	a		x				
5	Sem-gender	M						x	
		F						x	
		N						x	

2. Verbs (V)

Features used by the groups

P	ATT	VAL	C	IT	GE	SP	FR	DU	EN
1	Type	main	m	x	x	x	x	x	v
		auxiliary	a	x	x	x	x	x	x
		modal	o		x	x			m
	l-s.	copula						x	
	l-s.	impersonal						x	
2	Mood/VForm	indicative	i	x	x	x	x		
		subjunctive	s	x	x	x	x		
		imperative	m	x	x	x	x		
		conditional	c	x		x	x		
		infinitive	n	x	x	x	x	x	
		participle	p	x	x	x	x		
		gerund	g	x		x			
		supine	s						
		base	b						x
	l-s.	inf. + particle	u		x				
	l-s.	ImPart						x	
	l-s.	Past participle							
	l-s.	Present participle							
	l-s.	PerfPart						x	
	l-s.	Fin						x	
3	Tense	present	p	x	x	x	x	x	x
		imperfect	i	x	x	x	x		

		future	f	x		x	x		
		past	s	x		x	x	x	x
4	Person	first	1	x	x	x	x	x	x
		second	2	x	x	x	x	x	x
		third	3	x	x	x	x	x	x
5	Number	singular	s	x	x	x	x	x	x
		plural	p	x	x	x	x	x	x
6	Gender	masculine	m	x		x	x		
		feminine	f	x		x	x		
		neuter	n						
	l-s.	common	c	x					
7	Clitic l-s.	no	n	x	x				
		yes	y	x	x				
8	Clitic l-s.	both	t			x			
		accusa	a			x			
		dative	d			x			

3. Adjectives (A)

Features used by the groups

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	qualificative	f		x	x	x		
		ordinal	o		x		x		
		cardinal	c		x		x		
		indefinite	i				x		
		possessive	s		x	x	x		
	l-s.	part1	1		x				
		part2	2		x				
2	Degree	positive	p	x	x	x	x	x	x
		comparative	c	x	x	x	x	x	x
		superlative	s	x	x	x		x	x
3	Gender	masculine	m	x	x	x	x		
		feminine	f	x	x	x	x		
		neuter	n		x				
	l-s.	common	c	x					

4	Number		singular	s	x	x	x	x		
			plural	p	x	x	x	x		
		l-spc.	invariant	n	x					
5	Case		nominative	n		x				
			genitive	g		x				
			dative	d		x				
			accusative	a		x				
6	Position	l-spc.	attributive	a						x
			predicative	p						x

4. Pronouns (P)

Features used by the groups

P	ATT		VAL	C	IT	DE	ES	FR	NL	EN
1	Type		personal	p	x	x	x	x	x	x
			demonstrative	d	x	x	x	x	x	
			indefinite	i	x	x	x	x		
			possessive	s	x		x	x		x
			interrogative	t	x	x	x	x	x	
			relative	r	x	x	x	x	x	
			exclamative	e	x					
			reflexive	x		x			x	x
			reciprocal	l					x	
		l-s.	general	g						x
		l-s.	quantificational	l					x	
2	Person		first	1	x	x	x	x	x	x
			second	2	x	x	x	x	x	x
			third	3	x	x	x	x	x	x
3	Gender		masculine	m	x	x	x	x		
			feminine	f	x	x	x	x		
			neuter	n		x	x	x		
		l-s.	common	c	x					
4	Number		singular	s	x	x	x	x	x	x
			plural	p	x	x	x	x	x	x
		l-s.	invariant	n	x					

5	Case	nominative	n		x	x	x		
		genitive	g		x				
		dative	d		x	x			
		accusative	a		x	x			
		oblique	o			x	x		
		object	j				x		
	l-s.	1						x	
	l-s.	4						x	
6	Possessor	singular	s			x	x		x
		plural	p			x	x		x
7	Wh	Not-wh	n						x
		Relative	r						x
		Int	q						x
8	Poss-person	First	1						x
		Second	2						x
		Third	3						x
9	Poss-gender	Masculine	m						x
		Femenine	f						x
		Neuter	n						x
10	Sem-gender	M						x	
		F						x	
		N						x	

5. Determiners (D)

Features used by the groups

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	demonstrative	d	x	x	x	x	x	
		indefinite	i	x	x	x	x		
		possessive	s	x	x	x	x	x	x
		interrogative	t	x	x	x	x		
		exclamative	e	x					
		relative	r	x					
		article	a				x	x	
	l-s.	Def-article	t						x

				IT	DE	ES	FR	NL	EN
	l-s.	Indef-article	a						x
	l-s.	General	g						x
	l-s.	quantificational						x	
2	Person	first	1	x	x	x	x		
		second	2	x	x	x	x		
		third	3	x	x	x	x		
3	Gender	masculine	m	x	x	x	x	x	
		feminine	f	x	x	x	x	x	
		neuter	n		x	x		x	
	l-s	common	c	x					
4	Number	singular	s	x	x	x	x	x	x
		plural	p	x	x	x	x	x	x
	l-s	invariant	n	x					
5	Case	nominative	n		x				
		genitive	g		x				
		dative	d		x				
		accusative	a		x				
		oblique	o						
6	Possessor	singular	s			x	x		
		plural	p			x	x		
7	Quantif./or Defness	definite	d				x	x	
		indefinite	i				x	x	
8	Wh	Not-wh	n						x
		Relative	r						x
		Int/Ecl	q						x
9	Poss-person	First	1						x
		Second	2						x
		Third	3						x
10	Poss-gender	Masculine	m						x
		Feminine	f						x
		Neuter	n						x

6. Articles (T)

Features used by the groups
IT DE ES FR NL EN

P	ATT	VAL	C			
1	Type	definite	d	x	x	x
		indefinite	i	x	x	x
2	Gender	masculine	m	x	x	x
		feminine	f	x	x	x
		neuter	n	x	x	x
	l-s.	common	c	x		
3	Number	singular	s	x	x	x
		plural	p	x	x	x
4	Case	nominative	n		x	
		genitive	g		x	
		dative	d		x	
		accusative	a		x	

7. Adverbs (R)

Features used by the groups

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	general	g			x	x		
		particle	p			x	x		
	l-s.	degree	d		x				
	l-s.	interrogative	i		x				
	l-s.	conjunction	c		x				
	l-s.	modal	m		x				
	l-s.	pronom	p		x				
	l-s.	temporal	t		x				
	l-s.	place	l		x				
2	Degree	positive	p	x	x	x	x		x
		comparative	c		x	x	x		x
		superlative	s	x	x	x			x
	l-s.	negative	n				x		
3	Function	mod							x
		spe							x

4	Wh-ness	interrogative	q						x
		relative	r						x
		no	n						x

8. Adpositions (S)

Features used by the groups

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	preposition	p	x	x	x	x	x	x
		postposition	t		x			x	x
		circumposition	c		x				
	l-s.	part1	a		x				
	l-s.	part2	z		x				
2	Formation	simple	s	x	x	x			
		compound	c	x	x				
3	Gender	masculine	m	x					
		femenine	f	x					
		common	c	x					
4	Number	singular	s	x					
		plural	p	x					

9. Conjunctions (C)

Features used by the groups

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	coordinating	c	x	x	x	x	x	x
		subordinating	s	x	x	x	x	x	x
	l-spc.	compar	v		x				x
	l-spc.	infinitive	i		x				
	l-spc.	part1	a		x				
	l-spc.	part2	z		x				

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
2	ctype	finite	f						x
		that	t						x
		subjunctive	s						x
3	coord-posit.	initial	i						x
		non-initial	n						x

10. Numerals (M)

Features used by the groups

P	ATT	VAL	C	IT	DE	ES	FR	NL	EN
1	Type	cardinal	c	x		x	x		x
		ordinal	o	x		x			x
2	Gender	masculine	m	x		x	x		
		feminine	f	x		x	x		
		neuter	n						
3	Number	singular	s	x		x	x		
		plural	p	x		x	x		
5	Case	nominative	n						
		genitive	g						
		dative	d						
		accusative	a						

Categories used by the groups

	IT	DE	ES	FR	NL	EN
11. Interjections (I)	x	x	x	x	x	
12. Unique membership class (U)						
13. Residual (X)	x	x		x		
14. Particle (Q)		x				
15. Punctuation (F)	x	x				
16. Abreviations (Y)		x				

4 Comments on labels for corpus tags

Current enconding practices use widely different naming conventions for corpus tags. We can find different sets of labels also for the same language – for example SUBSMS, SBMS, NCMS, Nms, etc. can represent "Common noun, masc. sing." in different systems for the very same language.

It has been found, as already mentioned, that corpus tags are strongly committed to the tool and to the language. Therefore, each language will have its own set based on different considerations. However it was considered helpful to suggest some naming conventions for the sake of harmonization. The following is an attempt done by the French partners to give simple general guidelines for achieving a coherent naming convention within the project.

1. Corpus tags should be all upper case, in order to distinguish them from lexical descriptions.
2. Part-of-Speech should be encoded in a single character by the first letter using the same convention as the lexical categories used for lexical descriptions.
3. If possible, each of the characters after the first one should encode one by one the attribute-values using the same letter as in the lexical descriptions but upper case.
4. In case of ambiguity, or merging of values, one should use new characters, not already in the set of possible values for that attribute.

This is not a formal system, and may lead to ambiguities. In order to have a final set of tags a thourough testing must be performed as experimentation is going to show the behaviour of a given set. Also considerations coming from the decision taken with respect to the need and usefulness of special devices for automatic conversion are expected to have some impact in the concrete tags given for a language. Thus, the tags proposed by each group for the time being must be considered temptative until the end of the experimentation phase.

5 Language specific applications

5.1 Application to Italian

In this section the MULTEXT set of lexicon specifications is applied to Italian (Calzolari and Monachini 1994). The language-specific values added for Italian are highligthed with the code 'l-spec'.

Furthermore, a preliminary tagset for Italian is proposed. This is based on the tagset used by our tagger, but also takes into account the criteria expressed above for the construction of the tagset, and the results of a first cycle of experimentations on the MULTEXT tagger. A table containing the the translation of the tag into the regular expression and its definition is presented, i.e.

TAG	Reg.expr.	Definition
NMS	Ncms-	Common noun, masc.sing.

A table displaying the mapping between lexicon specifications and corpus tags is provided, along with an examplification.

5.1.1 Nouns (N)

5.1.1.1 Lexicon

Attribute	Value	Example	Code
Type	common	libro	c
	proper	Gianni	p
Gender	masculine	uomo	m
	feminine	donna	f
l-spec	common	insegnante	c
Number	singular	uomini	s
	plural	donne	p
l-spec	invariant	attivita'	n
Case	(n.a.)	(n.a.)	-

5.1.1.2 Corpus

Tag	Regular expression	Definition

NMS	Ncms-	Common noun, masc. sing.
NMP	Ncmp-	Common noun, masc. plur.
NMN	Ncmn-	Common noun, masc. invar.
NFS	Ncfs-	Common noun, fem. sing.
NFP	Ncfp-	Common noun, fem. plur.
NFN	Ncfn-	Common noun, fem. invar.
NNS	Nccs-	Common noun, comm. sing.
NNP	Nccp-	Common noun, comm. plur.
NNN	Nccn-	Common noun, comm. invar.
NP	Np..-	Proper noun

5.1.1.3 Combinations

Lexicon	Corpus	Example
Ncms-	NMS	libro
Ncmp-	NMP	libri
Ncmn-	NMN	re, caffe' (il/i)
Ncfs-	NFS	casa
Ncfp-	NFP	case
Ncfn-	NFN	attivita' (la/le)
Nccs-	NNS	insegnante (un/una)
Nccp-	NNP	insegnanti (gli/le)
Nccn-	NNN	sosia (il/la, i/le)
Np..-	NP	Mario, Maria, Borboni

5.1.1.4 Some obsevations for the corpus tagset

The idea of the French group to tag Proper Nouns simply with NP (collapsing the information on gender and number) seems the best solution.

5.1.2 Verb (V)

5.1.2.1 Lexicon

Attribute	Value	Example	Code

Type	main	amare	m
	auxiliary	avere	a
Mood/VForm	indicative	amo	i
	subjunctive	ami	s
	imperative	ama	m
	conditional	amerei	c
	infinitive	amare	n
	participle	amato	p
	gerund	amando	g
Tense	present	amo	p
	imperfect	amavo	i
	future	amero'	f
	past	amai	s
Person	first	amo	1
	second	ami	2
	third	ama	3
Number	singular	amo	s
	plural	amiamo	p
Gender	masculine	amato	m
	feminine	amata	f
l-spec	common	amante	c

5.1.2.2 Corpus

Tag	Regular Expression	Definition
VAS1IP	Vaip1s-	Aux. Verb, 1st pers.sing., pres.indic.
VAS2IP	Vaip2s-	Aux. Verb, 2nd pers.sing., pres.indic.
VAS3IP	Vaip3s-	Aux. Verb, 3rd pers.sing., pres.indic.
VAP2IP	Vaip2p-	Aux. Verb, 2nd pers.plur., pres.indic.
VAP3IP	Vaip3p-	Aux. Verb, 3rd pers.plur., pres.indic.
VAP1ICP	Va[is]p1p-	Aux. Verb, 1stpers.plur.,pres.indic/cong
VAY^2IP	Vaip(1s|3p)-	Aux. Verb, 1st sing./3rd plur., pres.indic
VAS1II	Vaii1s-	Aux. Verb, 1st pers.sing., impf.indic.
VAS2II	Vaii2s-	Aux. Verb, 2nd pers.sing., impf.indic.
VAS3II	Vaii3s-	Aux. Verb, 3rd pers.sing., impf.indic.

VAP1II	Vaii1p-	Aux. Verb, 1st pers.plur., impf.indic.
VAP2II	Vaii2p-	Aux. Verb, 2nd pers.plur., impf.indic.
VAP3II	Vaii3p-	Aux. Verb, 3rd pers.plur., impf.indic.
VAS1IF	Vaif1s-	Aux. Verb, 1st pers.sing., fut. indic.
VAS2IF	Vaif2s-	Aux. Verb, 2nd pers.sing., fut. indic.
VAS3IF	Vaif3s-	Aux. Verb, 3rd pers.sing., fut. indic.
VAP1IF	Vaif1p-	Aux. Verb, 1st pers.plur., fut. indic.
VAP2IF	Vaif2p-	Aux. Verb, 2nd pers.plur., fut. indic.
VAP3IF	Vaif3p-	Aux. Verb, 3rd pers.plur., fut. indic.
VAS1IR	Vais1s-	Aux. Verb, 1st pers.sing., past indic.
VAS2IR	Vais2s-	Aux. Verb, 2nd pers.sing., past indic.
VAS3IR	Vais3s-	Aux. Verb, 3rd pers.sing., past indic.
VAP1IR	Vais1p-	Aux. Verb, 1st pers.plur., past indic.
VAP3IR	Vais3p-	Aux. Verb, 3rd pers.plur., past indic.
VAP2ICR	Va(is)|(si)2p-	Aux. Verb, 2nd p.pl., past indic./pres.cong
VASXCP	Vacp.s-	Aux. Verb, 1/2/3 p. sing., pres.subj.
VAP2CMP	Va[sm]p2p-	Aux. Verb, 2nd pers.plur., pres.subj./imper.
VAP3CP	Vasp3p-	Aux. Verb, 3rd pers.plur., pres.subj.
VAS^3CI	Vasi^3s-	Aux. Verb, 1/2 pers.sing., impf.subj.
VAS3CI	Vasi3s-	Aux. Verb, 3rd pers.sing., impf.subj.
VAP1CI	Vasi1p-	Aux. Verb, 1st pers.plur., impf.subj.
VAP3CI	Vasi3p-	Aux. Verb, 3rd pers.plur., impf.subj.
VAS2MP	Vamp2s-	Aux. Verb, 2nd pers.sing., pres.impr.
VAS2MPE	Vamp2s-y	Aux. Verb, 2nd pers.sing., pres.impr. + clit.
VAP2MPE	Vamp2p-y	Aux. Verb, 2nd pers.plur., pres.impr. + clit.
VAS1DP	Vacp1s-	Aux. Verb, 1st pers.sing., pres.cond.
VAS2DP	Vacp2s-	Aux. Verb, 2nd pers.sing., pres.cond.
VAS3DP	Vacp3s-	Aux. Verb, 3rd pers.sing., pres.cond.
VAP1DP	Vacp1p-	Aux. Verb, 1st pers.plur., pres.cond.
VAP2DP	Vacp2p-	Aux. Verb, 2nd pers.plur., pres.cond.
VAP3DP	Vacp3p-	Aux. Verb, 3rd pers.plur., pres.cond.

VAF	Vanp---	Aux. Verb, infinitive
VAFE	Vanp--cy	Aux. Verb, infinitive + clitic
VANSPP	Vapp-sc	Aux. Verb, comm.sing., pres.part.
VANPPP	Vapp-pc	Aux. Verb, comm.plur., pres.part.
VAMSPR	Vaps-sm	Aux. Verb, masc.sing., past part.
VAMPPR	Vaps-pm	Aux. Verb, masc.plur., past part.
VAFSPR	Vaps-sf	Aux. Verb, femm.sing., past part.
VAFPPR	Vaps-pf	Aux. Verb, femm.plur., past part.
VAMSPRE	Vaps-smy	Aux. Verb, masc.sing., past part. + clitic
VAMPPRE	Vaps-pmy	Aux. Verb, masc.plur., past part. + clitic
VAFSPRE	Vaps-sfy	Aux. Verb, femm.sing., past part. + clitic
VAFPPRE	Vaps-pfy	Aux. Verb, femm.plur., past part. + clitic
VAG	Vagp---	Aux. Verb, gerund
VAGE	Vagp---y	Aux. Verb, gerund + clitic
VS1IP	Vmip1s-	Main Verb, 1st pers.sing., pres.indic
VS3IP	Vmip3s-	Main Verb, 3rd pers.sing., pres.indic
VP3IP	Vmip3p-	Main Verb, 3rd pers.plur., pres.indic
VP1ICP	Vm[is]p1p	Main Verb,1stpers.plur.,pres.indic/cong
VP2IMPP	Vm([im]p2p-)|(ps-pf)	M.V., 2nd pl., pres.indic/imper|pstprt f.pl.
VP2IMP	Vm([im]p2p)-	Main Verb, 2nd pl., pres.indic/imper
VSXICP	Vm(sp.s)|(ip2s)-	M.V., 1/2/3 sg.,pres.subj.|2ndsg. pres.indic.
VS^1IMP	Vm[im]^1s-	Main Verb, not 1stsg.,pres.indic./imper.
VS2IMP	Vm[im]p2s-	Main Verb, 2nd sg., pres.indic/imper
VP2IMCPP	Vm([ims]p2p-)|(ps-pf)	M.V., 2pl., pr.ind/imp/sub|pst.prt f.pl.
VS1II	Vmii1s-	Main Verb, 1st pers.sing., impf.indic.
VS2II	Vmii2s-	Main Verb, 2nd pers.sing., impf.indic.
VS3II	Vmii3s-	Main Verb, 3rd pers.sing., impf.indic.
VP1II	Vmii1p-	Main Verb, 1st pers.plur., impf.indic.
VP2II	Vmii2p-	Main Verb, 2nd pers.plur., impf.indic.
VP3II	Vmii3p-	Main Verb, 3rd pers.plur., impf.indic.
VS1IF	Vmif1s-	Main Verb, 1st pers.sing., fut. indic.
VS2IF	Vmif2s-	Main Verb, 2nd pers.sing., fut. indic.
VS3IF	Vmif3s-	Main Verb, 3rd pers.sing., fut. indic.
VP1IF	Vmif1p-	Main Verb, 1st pers.plur., fut. indic.

VP2IF	Vmif2p-	Main Verb, 2nd pers.plur., fut. indic.
VP3IF	Vmif3p-	Main Verb, 3rd pers.plur., fut. indic.
VS1IR	Vmis1s-	Main Verb, 1st pers.sing., past indic.
VS2IR	Vmis2s-	Main Verb, 2nd pers.sing., past indic.
VS3IR	Vmis3s-	Main Verb, 3rd pers.sing., past indic.
VP1IR	Vmis1p-	Main Verb, 1st pers.plur., past indic.
VP3IR	Vmis3p-	Main Verb, 3rd pers.plur., past indic.
VP2ICR	Vm(is)|(si)2p-	Main Verb, 2nd p.pl., past indic./pres.subj.
VP2CP	Vmsp2p-	Main Verb, 2nd pers.plur., pres.subj. amiate
VP3CP	Vmsp3p-	Main Verb, 3rd pers.plur., pres.subj. amino
VSXCP	Vmcp.s-	Main Verb, 1/2/3 p. sing., pres.subj.
VS^3CI	Vmsi^3s-	Main Verb, 1/2 pers.sing., impf.subj.
VS3CI	Vmsi3s-	Main Verb, 3rd pers.sing., impf.subj.
VP1CI	Vmsi1p-	Main Verb, 1st pers.plur., impf.subj.
VP3CI	Vmsi3p-	Main Verb, 3rd pers.plur., impf.subj.
VS2MPE	Vmmp2s-y	Main Verb, 2nd pers.sing., pres.impr. + clit.
VP2MPE	Vmmp2p-y	Main Verb, 2nd pers.plur., pres.impr. + clit.
VS1DP	Vmcp1s-	Main Verb, 1st pers.sing., pres.cond.
VS2DP	Vmcp2s-	Main Verb, 2nd pers.sing., pres.cond.
VS3DP	Vmcp3s-	Main Verb, 3rd pers.sing., pres.cond.
VP1DP	Vmcp1p-	Main Verb, 1st pers.plur., pres.cond.
VP2DP	Vmcp2p-	Main Verb, 2nd pers.plur., pres.cond.
VP3DP	Vmcp3p-	Main Verb, 3rd pers.plur., pres.cond.
VF	Vmnp---	Main Verb, infinitive
VFE	Vmnp---y	Main Verb, infinitive + clitic
VNSPP	Vmpp-sc	Main Verb, comm.sing., pres.part.
VNPPP	Vmpp-pc	Main Verb, comm.plur., pres.part.
VMSPR	Vmps-sm	Main Verb, masc.sing., past part.
VMPPR	Vmps-pm	Main Verb, masc.plur., past part.

```
VFSPR     Vmps-sf                   Main Verb, femm.sing., past part.
VFPPR     Vmps-pf                   Main Verb, femm.plur., past part.

VMSPRE    Vmps-smy                  Main Verb, masc.sing., past part. +c
VMPPRE    Vmps-pmy                  Main Verb, masc.plur., past part. +c
VFSPRE    Vmps-sfy                  Main Verb, femm.sing., past part. +c
VFPPRE    Vmps-pfy                  Main Verb, femm.plur., past part. +c

VG        Vmgp---                   Main Verb, gerund
VGE       Vmgp---y                  Main Verb, gerund + clitic


-------------------- more collapsed tagset -----------------------

VA1P   Va[iscm][pifs]1p--      Aux. verb, 1st person plur.
VA1S   Va[iscm][pifs]1s--      Aux. verb, 1st person sing.
VA2P   Va[iscm][pifs]2p--      Aux. verb, 2nd person plur.
VA2S   Va[iscm][pifs]2s--      Aux. verb, 2nd person sing.
VA3P   Va[iscm][pifs]3p--      Aux. verb, 3rd person plur.
VA3S   Va[iscm][pifs]3s--      Aux. verb, 3rd person sing.
VAFPPS Vaps-pf-                Aux. verb, fem. plur., past part.
VAFSPS Vaps-sf-                Aux. verb, fem. sing., past part.
VAMPPS Vaps-pm-                Aux. verb, masc. plur., past part.
VAMSPS Vaps-sm-                Aux. verb, masc. sing., past part.
VAN    Vanp----                Aux. verb, infinitive
VAFE   Vanp----                Aux. Verb, infinitive + enclitic
VAG    Vagp----                Aux. Verb, gerund
VAGE   Vagp----                Aux. Verb, gerund + enclitic
VAPP   Vapp-..-                Aux. verb, pres. participle

V1P    Vm[iscm][pifs]1p--      Main Verb, 1st person plur.
V1S    Vm[iscm][pifs]1s--      Main Verb, 1st person sing.
V2P    Vm[iscm][pifs]2p--      Main Verb, 2nd person plur.
V2S    Vm[iscm][pifs]2s--      Main Verb, 2nd person sing.
V3P    Vm[iscm][pifs]3p--      Main Verb, 3rd person plur.
V3S    Vm[iscm][pifs]3s--      Main Verb, 3rd person sing.
VFPPS  Vmps-pf-                Main Verb, fem. plur., past part.
VFSPS  Vmps-sf-                Main Verb, fem. sing., past part.
VMPPS  Vmps-pm-                Main Verb, masc. plur., past part.
VMSPS  Vmps-sm-                Main Verb, masc. plur., past part.
VF     Vmnp----                Main Verb, infinitive
VFE    Vmnp----y               Main Verb, infinitive + enclitic
VG     Vmgp----                Main Verb, gerund
VGE    Vmgp----y               Main Verb, gerund + enclitic
VPP    Vmpp-..-                Main Verb, pres. participle
--------------------- more collapsed end ----------------------
====== =================== ===================================
```

5.1.2.3 Combinations

Lexicon	Corpus	Example
Vaip1s- +++	VAS1IP	ho
Vaip2s-	VAS2IP	hai, sei
Vaip3s-	VAS3IP	ha, e’
Vaip2p-	VAP2IP	avete, siete
Vaip3p- +++	VAP3IP	hanno
Vaip1p-	VAP1ICP	abbiamo, siamo
Vaip1s- +++	VAY^2IP	sono
Vaip3p- +++	VAY^2IP	sono
Vaii1s-	VAS1II	avevo, ero
Vaii2s-	VAS2II	avevi, eri
Vaii3s-	VAS3II	aveva, era
Vaii1p-	VAP1II	avevamo, eravamo
Vaii2p-	VAP2II	avevate, eravate
Vaii3p-	VAP3II	avevano, erano
Vaif1s-	VAS1IF	avro’, saro’
Vaif2s-	VAS2IF	avrai, sarai
Vaif3s-	VAS3IF	avra’, sara’
Vaif1p-	VAP1IF	avremo, saremo
Vaif2p-	VAP2IF	avrete, sarete
Vaif3p-	VAP3IF	avranno, saranno
Vais1s-	VAS1IR	ebbi, fui
Vais2s-	VAS2IR	avesti, fosti
Vais3s-	VAS3IR	ebbe, fu
Vais1p-	VAP1IR	avemmo, fummo
Vais3p-	VAP3IR	ebbero, furono
Vais2s-	VAP2ICR	aveste, foste
Vasp1s-	VASXCP	abbia, sia
Vasp2s-	VASXCP	abbia, sia
Vasp3s-	VASXCP	abbia, sia
Vasp1p-	VAP1ICP	abbiamo, siamo

Vasp2p-	VAP2CMP	abbiate, siate
Vasp3p-	VAP3CP	abbiano, siano

Vasi1s-	VAS^3CI	avessi, fossi
Vasi2s-	VAS^3CI	avessi, fossi
Vasi3s-	VAS3CI	avesse, fosse
Vasi1p-	VAP1CI	avessimo, fossimo
Vasi2s-	VAP2ICR	aveste, foste
Vasi3p-	VAP3CI	avessero, fossero

Vamp2s-	VAS2MP	abbi, sii
Vamp2s-y	VAS2MPE	abbilo, siilo
Vamp2p-	VAP2CMP	abbiate, siate
Vamp2p-y	VAP2MPE	abbiatelo, siatelo

Vacp1s-	VAS1DP	avrei, sarei
Vacp2s-	VAS2DP	avresti, saresti
Vacp3s-	VAS3DP	avrebbe, sarebbe
Vacp1p-	VAP1DP	avremmo, saremmo
Vacp2p-	VAP2DP	avreste, sareste
Vacp3p-	VAP3DP	avrebbero, sarebbero

Vanp---	VAF	avere, essere
Vanp--cy	VAFE	averlo, esserlo

Vapp-sc	VANSPP	avente, essente
Vapp-pc	VANPPP	aventi, essenti

Vaps-sm	VAMSPR	avuto, stato
Vaps-pm	VAMPPR	avuti, stati
Vaps-sf	VAFSPR	avuta, stata
Vaps-pf	VAFPPR	avute, state

Vaps-smy	VAMSPRE	avutolo
Vaps-pmy	VAMPPRE	avutili
Vaps-sfy	VAFSPRE	avutala
Vaps-pfy	VAFPPRE	avuteli

Vagp---	VAG	avendo, essendo
Vagp---y	VAGE	avendolo, essendolo

```
Vmip1s-      VS1IP    amo, leggo, servo
Vmip2s- +++  VSXICP   ami
Vmip2s- +++  VS2IMP   leggi, servi
Vmip3s- ---  VS^1IMP  ama
Vmip3s- ---  VS3IP    legge, serve
Vmip1p-      VP1ICP   amiamo, leggiamo, serviamo
Vmip2p- ***  VP2IMPP  amate, servite
Vmip2p- ***  VP2IMP   leggete
Vmip2p- ***  VP2IMCPP premiate
Vmip3p-      VP3IP    amano, leggono, servono


Vmii1s-      VS1II    amavo,
Vmii2s-      VS2II    amavi,
Vmii3s-      VS3II    amava
Vmii1p-      VP1II    amavano
Vmii2p-      VP2II    amavate
Vmii3p-      VP3II    amavano


Vmif1s-      VS1IF    amero'
Vmif2s-      VS2IF    amerai
Vmif3s-      VS3IF    amera'
Vmif1p-      VP1IF    ameremo
Vmif2p-      VP2IF    amerete
Vmif3p-      VP3IF    ameranno


Vmis1s-      VS1IR    amai
Vmis2s-      VS2IR    amasti
Vmis3s-      VS3IR    amo'
Vmis1p-      VP1IR    amammo
Vmis2p-      VP2ICR   amaste, leggeste, serviste
Vmis3p-      VP3IR    amarono


Vmsp1s- +++  VSXCP    legga
Vmsp1s- +++  VSXICP   ami
Vmsp2s- ---  VSXCP    legga
Vmsp2s- ---  VSXICP   ami
Vmsp3s- ***  VSXCP    legga
Vmsp3s- ***  VSXICP   ami
Vmsp1p-      VP1ICP   amiamo, leggiamo, serviamo
Vmsp2p- """  VP2CP    amiate, leggiate, serviate
Vmsp2p- """  VP2ICMPP premiate
Vmsp3p-      VP3CP    amino, leggano, servano
```

```
Vmsi1s-       VS^3CI   amassi, leggessi, servissi
Vmsi2s-       VS^3CI   amassi, leggessi, servissi
Vmsi3s-       VS3CI    amasse, leggesse, servisse
Vmsi1p-       VP1CI    amassimo
Vmsi2p-       VP2ICR   amaste, leggeste, serviste
Vmsi3p-       VP3CI    amassero

Vmmp2s- +++   VS^1IMP  ama
Vmmp2s- +++   VS2IMP   leggi, servi
Vmmp2p- ---   VP2IMPP  amate
Vmmp2p- ---   VP2IMP   leggete, servite
Vmmp2p- ---   VP2IMCPP premiate
Vmmp2s-y      VS2MPe   amalo, leggilo, servilo
Vmmp2p-y      VP2MPe   amatelo, leggetelo, servitelo


Vmcp1s-       VS1DP    amerei
Vmcp2s-       VS2DP    ameresti
Vmcp3s-       VS3DP    amarebbe
Vmcp1p-       VP1DP    ameremmo
Vmcp2p-       VP2DP    amereste
Vmcp3p-       VP3DP    amerebero


Vmnp---       VF       amare
Vmnp---y      VFE      amarlo


Vmpp-sc       VNSPP    amante
Vmpp-pc       VNPPP    amanti


Vmps-sm       VMSPR    amato, letto, servito
Vmps-pm       VMPPR    amati, letti, serviti
Vmps-sf       VFSPR    amata, letta, servita
Vmps-pf +++   VP2IMCPP premiate
Vmps-pf +++   VP2IMPP  amate, servite
Vmps-pf +++   VFPPR    lette


Vmps-smy      VMSPRE   amatolo
Vmps-pmy      VMPPRE   amatili
Vmps-sfy      VFSPRE   amatala
Vmps-pfy      VFPPRE   amatele
```

```
Vmgp---       VG       amando
Vmgp---y      VGE      amandolo

----------------------- more collapsed tagset ------------------

Vaip1s-   VA1S         ho
Vaip2s-   VA2S         hai
Vaip3s-   VA3S         ha
Vaip1p-   VA1P         abbiamo
Vaip2p-   VA2P         avete
Vaip3p-   VA3P         hanno
Vaii1s-   VA1S         avevo
Vaii2s-   VA2S         avevi
Vaii3s-   VA3S         aveva
Vaii1p-   VA1P         avevamo
Vaii2p-   VA2P         avevate
Vaii3p-   VA3P         avevano
Vaif1s-   VA1S         avro'
Vaif2s-   VA2S         avrai
Vaif3s-   VA3S         avra'
Vaif1p-   VA1P         avremo
Vaif2p-   VA2P         avrete
Vaif3p-   VA3P         avranno
Vais1s-   VA1S         ebbi
Vais2s-   VA2S         avesti
Vais3s-   VA3S         ebbe
Vais1p-   VA1P         avemmo
Vais2p-   VA2P         aveste
Vais3p-   VA3P         ebbero
Vasp1s-   VA1S         abbia
Vasp2s-   VA2S         abbia
Vasp3s-   VA3S         abbia
Vasp1p-   VA1P         abbiamo
Vasp2p-   VA2P         abbiate
Vasp3p-   VA3P         abbiano
Vasi1s-   VA1S         avessi
Vasi2s-   VA2S         avessi
Vasi3s-   VA3S         avesse
Vasi1p-   VA1P         avessimo
Vasi2p-   VA2P         aveste
Vasi3p-   VA3P         avessero
Vamp2s-   VA2S         abbi
Vamp2p-   VA2P         abbiate
Vacp1s-   VA1S         avrei
Vacp2s-   VA2S         avresti
Vacp3s-   VA3S         avrebbe
Vacp1p-   VA1P         avremmo
```

Vacp2p-	VA2P	avreste
Vacp3p-	VA3P	avrebbero
Vanp---	VAF	avere
Vanp---y	VAFE	averlo
Va-cspp	VANSPP	avente
Va-cppp	VANPPP	aventi
Va-msps	VAMSPR	avuto
Va-mpps	VAMPPR	avuti
Va-fsps	VAFSPR	avuta
Va-fpps	VAFPPR	avute
Va-gp--	VAG	avendo
Va-gp--y	VAGE	avendolo

Vmip1s-	V1S	amo
Vmip2s-	V2S	ami
Vmip3s-	V3S	ama
Vmip1p-	V1P	amiamo
Vmip2p-	V2P	amate
Vmip3p-	V3P	amano
Vmii1s-	V1S	amavo
Vmii2s-	V2S	amavi
Vmii3s-	V3S	amava
Vmii1p-	V1P	amavamo
Vmii2p-	V2P	amavate
Vmii3p-	V3P	amavano
Vmif1s-	V1S	amero’
Vmif2s-	V2S	amerai
Vmif3s-	V3S	amera’
Vmif1p-	V1P	ameremo
Vmif2p-	V2P	amerete
Vmif3p-	V3P	ameranno
Vmis1s-	V1S	amai
Vmis2s-	V2S	amasti
Vmis3s-	V3S	amo’
Vmis1p-	V1P	amammo
Vmis2p-	V2P	amaste
Vmis3p-	V3P	amarono
Vmsp1s-	V1S	ami
Vmsp2s-	V2S	ami
Vmsp3s-	V3S	ami
Vmsp1p-	V1P	amiamo
Vmsp2p-	V2P	amiate
Vmsp3p-	V3P	amino
Vmsi1s-	V1S	amassi
Vmsi2s-	V2S	amassi
Vmsi3s-	V3S	amasse
Vmsi1p-	V1P	amassimo

Vmsi2p-	V2P	amaste
Vmsi3p-	V3P	amassero
Vmmp2s-	V2S	ama
Vmmp2p-	V2P	amate
Vmcp1s-	V1S	amerei
Vmcp2s-	V2S	ameresti
Vmcp3s-	V3S	amerebbe
Vmcp1p-	V1P	ameremmo
Vmcp2p-	V2P	amereste
Vmcp3p-	V3P	amerebbero
Vmnp---	VF	amare
Vmnp---y	VFE	amarlo
Vm-cspp	VNSPP	amante
Vm-cppp	VNPPP	amanti
Vm-msps	VMSPR	amato
Vm-mpps	VMPPR	amati
Vm-fsps	VFSPR	amata
Vm-fpps	VFPPR	amate
Vm-gp--	VG	amando
Vm-gp--y	VGE	amandolo

5.1.2.4 Some observations for corpus tagset

An observation concerns the special marking for the auxiliaries: the taggers are in general not able to disambiguate the cases in which the auxiliaries are used as full verbs ("io ho un cane" , "i bambini sono nel prato") from the cases when they are auxiliaries. The distinction of the auxiliaries is used only in order to isolate 'avere' and 'essere' from the other verbs.

For verbs, two different sets of tags are proposed, the first more fine-grained for more accurate distinctions and the latter more coarse-grained, which follows the approach proposed by the French group.

The collapsing proposed by the French group of Moods and Tenses, if considered wrt to the performances of our tagger, appears restrictive: for many unambiguous tenses and moods, the Italian tagger is able to formulate the correct analysis (e.g. conditional, subjunctive imperfect, indicative past etc.) and these distinctions are, in our opinion, worth being maintained. It has to be noticed that the ambiguities between verb forms depend also on different lexical verbs.

In Italian, the major ambiguities concerns the 2nd sing and plur of the present indicative and imperative, ama-amate; leggi-leggete.

However, this is again not a general rule.

Another very common ambiguity is between the 2nd pers. of the indicative and the 1st, 2nd, 3rd person of the present subjunctive. Therefore not always it is possible to decide unambiguosly on the person.

Some more frequent typical homographies in Italian are listed below:

```
VP1ICP   amiamo
VP2IMP   leggete
VP2IMPP  amate
VP2IMCPP premiate
VP2ICR   amaste
VS^3CI   amassi
VSXCP    legga
VAY^2IP  sono
VSXICP   ami
VS2IMP   leggi
VS^1IMP  ama
VS^1IMP  ama
```

In the design of corpus tagsets for verbs careful attention should be given to the enclitic phenomenon: at present our tagger is able to recognize the presence of the clitics which is signalled by the addition of the mark "+E" (plus clitic) to the regular verb tag.

5.1.3 Adjectives (A)

5.1.3.1 Lexicon

Attribute	Value	Example	Code
Type	-	-	-
Degree	positive	buono	p
	comparative	migliore	c
	superlative	buonissimo	s
Gender	masculine	buono	m
	feminine	buona	f
l-spec	common	dolce	c
Number	singular	buono	s
	plural	buoni	p

l-spec	invariant	pari	n
Case	(n.a.)	(n.a.)	-

5.1.3.2 Corpus

Tag	Regular expression	Definition
AFP	A-.fp-	Adjective fem. plur.
AFS	A-.fs-	Adjective fem. sing.
AFN	A-.fn-	Adjective fem. invar.
AMP	A-.mp-	Adjective masc. plur.
AMS	A-.ms-	Adjective masc. sing.
AMN	A-.mn-	Adjective masc. invar.
AMP	A-.mp-	Adjective comm. plur.
AMS	A-.ms-	Adjective comm. sing.
AMN	A-.mn-	Adjective comm. invar.

5.1.3.3 Combinations

Lexicon	Corpus	Example
A-pms-	AMS	vero
A-pmp-	AMP	veri
A-pmn-	AMN	oggetto (complemento/i oggetto: grammatical language)
A-pfs-	AFS	vera
A-pfp-	AFP	vere
A-pfn-	AFN	valore (clausola valore: juridical language)
A-pcs-	ANS	dolce (biscotto, torta)
A-pcp-	ANP	dolci (biscotti, dolci)
A-pcn-	ANN	pari (risultato/i, somma/e)
A-sms-	AMS	verissimo
A-smp-	AMP	verissimi
A-sfs-	AFS	verissima
A-sfp-	AFP	verissime

5.1.3.4 Observations

The comparative Degree applies only to a close set of adjectives (e.g.

maggiore, migliore, etc). All other adjectives form their comparatives with "piu'" + adjective (e.g., piu' forte). Superlative is also an analytical form (il piu' forte), but can be also synthetically formed: grandissimo, massimo.

5.1.4. Pronouns

5.1.4.1. Lexicon

Attribute	Value	Example	Code
Type	personal	io	p
	demonstrat.	quello	d
	indefinite	chiunque	i
	possessive	mio	s
	interrog.	chi	t
	relative	che	r
	exclamative	quanto	e
Person	first	io	1
	second	tu	2
	third	egli	3
Gender	masculine	questo	m
	feminine	questa	f
l-spec	common	io	c
Number	singular	questo	s
	plural	questi	p
l-spec	invariant	che	n
Case	(n.a.)	(n.a.)	-
Possessor	-	-	-

5.1.4.2 Corpus

Tag	Reg.Expr.	Definition
PDMS	Pd-ms--	Demonstrative pronoun masc.sing.
PDMP	Pd-mp--	Demonstrative pronoun masc.plur.
PDFS	Pd-fs--	Demonstrative pronoun femm.sing.

```
PDFP   Pd-fp--       Demonstrative pronoun femm.plur.
PDNS   Pd-cs--       Demonstrative pronoun comm.sing.
PDNP   Pd-cp--       Demonstrative pronoun comm.plur.

PIMS   Pi-ms--       Indefinite    pronoun masc.sing.
PIMP   Pi-mp--       Indefinite    pronoun masc.plur.
PIFS   Pi-fs--       Indefinite    pronoun femm.sing.
PIFP   Pi-fp--       Indefinite    pronoun femm.plur.
PINS   Pi-cs--       Indefinite    pronoun comm.sing.
PINP   Pi-cp--       Indefinite    pronoun comm.plur.

PPMS   Ps.ms--       Possessive pronoun, masc.sing.
PPMP   Ps.mp--       Possessive pronoun, masc.plur.
PPFS   Ps.fs--       Possessive pronoun, femm.sing.
PPFP   Ps.fp--       Possessive pronoun, femm.plur.
PPNP   Ps.cp--       Possessive pronoun, comm.plur.

PWNS   P[tre]-cs--   Interr./Rel./Escl. pronoun, comm.sing.
PWNP   P[tre]-cp--   Interr./Rel./Escl. pronoun, comm.plur.
PWNN   P[tre]-cn--   Interr./Rel./Escl. pronoun, comm.plur.
PWMS   P[tre]-ms--   Interr./Rel./Escl. pronoun, masc.sing.
PWMP   P[tre]-mp--   Interr./Rel./Escl. pronoun, masc.plur.
PWFS   P[tre]-fs--   Interr./Rel./Escl. pronoun, femm.sing.
PWFP   P[tre]-fp--   Interr./Rel./Escl. pronoun, femm.plur.

PQNS1  Pp1cs--       Personal   pronoun, 1st pers., comm.sing.
PQNS2  Pp2cs--       Personal   pronoun, 2nd pers., comm.sing.
PQMS3  Pp3ms--       Personal   pronoun, 3rd pers., masc.sing.
PQFS3  Pp3fs--       Personal   pronoun, 3rd pers., femm.sing.
PQNN3  Pp3cn--       Personal   pronoun, 3rd pers., comm.inv.
PQNP1  Pp1cp--       Personal   pronoun, 1st pers., comm.plur.
PQNP2  Pp2cp--       Personal   pronoun, 2nd pers., comm.plur.
PQNP3  Pp3cp--       Personal   pronoun, 3rd pers., comm.plur.
PQMP3  Pp3mp--       Personal   pronoun, 3rd pers., masc.plur.
PQFP3  Pp3fp--       Personal   pronoun, 3rd pers., femm.plur.

--------- more collapsed ----------------
PFP    P..fp---      Pronoun, fem. plur.
PFS    P..fs---      Pronoun, fem. plur.
PMP    P..mp---      Pronoun, masc. plur.
PMS    P..ms---      Pronoun, masc. sing.
PNS    P..cs---      Pronoun, comm. sing.
PNP    P..cp---      Pronoun, comm. plur.
PNN    P..cn---      Pronoun, comm. inv.
---------- more collapsed end -----------

====== ========== ==========================================
```

5.1.4.3 Combinations

Lexicon	Corpus	Example
Pd-ms--	PDMS	quello, costui
Pd-mp--	PDMP	quelli
Pd-fs--	PDFS	quella
Pd-fp--	PDFP	quelle
Pd-cs--	PDNS	cio'
Pd-cp--	PDNP	coloro
Pi-ms--	PIMS	ognuno
Pi-mp--	PIMP	alcuni
Pi-fs--	PIFS	ognuna
Pi-fp--	PIFP	alcune
Pi-cs--	PINS	chiunque, tale
Pi-cp--	PINP	tali
Ps1ms--	PPMS	mio, nostro
Ps1mp--	PPMP	miei
Ps1fs--	PPFS	mia
Ps1fp--	PPFP	mie
Ps2ms--	PPMS	tuo, vostro
Ps2mp--	PPMP	tuoi
Ps2fs--	PPFS	tua
Ps2fp--	PPFP	tue
Ps3ms--	PPMS	suo
Ps3mp--	PPMP	suoi
Ps3fs--	PPFS	sua
Ps3fp--	PPFP	sue
Ps3cp--	PPNP	loro
Pt-cs--	PWNS	chi? quale?
Pt-cp--	PWNP	quali?
Pt-cn--	PWNN	che?
Pt-ms--	PWMS	quanto?
Pt-mp--	PWMP	quanti?
Pt-fs--	PWFS	quanta?
Pt-fp--	PWFP	quante?
Pr-cn--	PWNN	cui
Pr-ms--	PWMS	quanto
Pr-mp--	PWMP	quanti
Pr-fs--	PWFS	quanta

```
Pr-fp--   PWFP      quante
Pr-cs--   PWNS      chi, quale
Pr-cp--   PWNP      quali

Pe-ms--   PWMS      quanto!
Pe-mp--   PWMP      quanti!
Pe-fs--   PWFS      quanta!
Pe-fp--   PWFP      quante!
Pe-cs--   PWNS      quale!
Pe-cp--   PWNP      quali!
Pe-cn--   PWNN      che!

Pp1cs--   PQNS1     io, me, mi
Pp2cs--   PQNS2     tu, te, ti,
Pp3ms--   PQMS3     egli, lui, esso, gli, lo
Pp3fs--   PQFS3     ella, lei, essa, le, la
Pp3cn--   PQNN3     si
Pp1cp--   PQNP1     noi, ci
Pp2cp--   PQNP2     voi, vi
Pp3cp--   PQNP3     loro,
Pp3mp--   PQMP3     essi, li
Pp3fp--   PQFP3     esse, le

--------------------- more collapsed -----------------------------
P..fp---  PFP        mie, queste, quante etc.
P..fs---  PFS        mia, questa, quanta etc.
P..mp---  PMP        miei, questi, quanti etc.
P..ms---  PMS        mio, questo, quanto etc.
P..cs---  PNS        quale
P..cp---  PNP        quali
P..cn---  PNN        che, cui, altrui
-------------------- more collapsed end --------------------
==============================
```

5.1.4.4 Observations

For pronouns, the strategy of proposing two different tagsets, the one more collapsed and the other more fine-grained is followed.

As far as the pronominal paradigm is concerned, Case is not encoded at present in our DMI (Calzolari et al. 1983).
Personal pronouns are not lemmatized: 'gli' is not considered the dative form of the base pronoun 'egli' (he), but constitutes a separate entry.

The Italian pronominal paradigm is the following:

```
'forme toniche' (strong forms): subj (io, egli), compl (me, lui)

              ama me / da' a me -- dir-obj/prep-obj --
              (he loves me / he gives to me)

              ama lui / da' a lui -- dir-obj/prep-obj --
              (she loves him / she gives to him)

'forme atone' (weak forms):        -           compl (mi, gli/lo)

              mi da' / mi ama    -- ind-obj/dir-obj --
              (he gives me / he loves me)
              gli da'            -- ind-obj --
              (he gives him)
              lo ama             -- dir-obj --
              (she loves him)
```

This paradigm can be mapped on the Case system proposed by the French group, in the following way:

```
io, egli = subj                      = nom
mi/me    = dir-obj/ind-obj/prep-obj = obj -] acc, dat, prep+obl
lui      = dir-obj/prep-obj         = obj -] acc, prep+obl
gli      = ind-obj                  = dat
lo       = dir-obj                  = acc
```

5.1.5 Determiners (Pronominal Adjectives) (D)

5.1.5.1. Lexicon

Attribute	Value	Example	Code	
Type	demonstrat.	questo	d	
	indefinite	ogni	i	
	possessive	mio	s	
	interrogat.	che	t	
	exclamative	quanto	e	this value has been added
	relative	quanto	r	this value has been added
Person	first	mio	1	
	second	tuo	2	
	third	suo	3	

Gender	masculine	questo	m
	feminine	questa	f
l-spec	common	ogni	c
Number	singular	quello	s
	plural	quelli	p
l-spec	invariant	altrui	n
Case	(n.a.)	(n.a.)	-
Possessor	-	-	-

5.1.5.2 Corpus

Tag	Regular exp.	Definition
DDNS	Dd-ns--	Demonstrative pron.adj. comm.inv.
DDNP	Dd-np--	Demonstrative pron.adj. comm.plur.
DDMS	Dd-ms--	Demonstrative pron.adj. masc.sing.
DDMP	Dd-mp--	Demonstrative pron.adj. masc.plur.
DDFS	Dd-fs--	Demonstrative pron.adj. femm.sing.
DDFP	Dd-fp--	Demonstrative pron.adj. femm.plur.
DIMS	Di-ms--	Indefinite pron.adj. masc.sing.
DIMP	Di-mp--	Indefinite pron.adj. masc.plur.
DIFS	Di-fs--	Indefinite pron.adj. femm.sing.
DIFP	Di-fp--	Indefinite pron.adj. femm.plur.
DINS	Di-cs--	Indefinite pron.adj. comm.sing.
DINP	Di-cp--	Indefinite pron.adj. comm.plur.
DPMS	Ds.ms--	Possessive pron.adj., masc.sing.
DPMP	Ds.mp--	Possessive pron.adj., masc.plur.
DPFS	Ds.fs--	Possessive pron.adj., femm.sing.
DPFP	Ds.fp--	Possessive pron.adj., femm.plur.
DPNN	Ds-cn--	Possessive pron.adj., comm.inv.
DWNN	D[tre]-cn--	Interr/Relat./escl. pron.adj., comm.inv.
DWMS	D[tre]-ms--	Interr/Relat./escl. pron.adj., masc.sing.
DWMP	D[tre]-mp--	Interr/Relat./escl. pron.adj., masc.plur.
DWFS	D[tre]-fs--	Interr/Relat./escl. pron.adj., femm.sing.
DWFP	D[tre]-fp--	Interr/Relat./escl. pron.adj., femm.plur.
DWNS	D[tre]-cs--	Interr/Relat./escl. pron.adj., comm.sing.
DWNP	D[tre]-cp--	Interr/Relat./escl. pron.adj., comm.plur.

```
--------- more collapsed ----------------
DFP     D..fp---     Determiner, fem. plur.
DFS     D..fs---     Determiner, fem. plur.
DMP     D..mp---     Determiner, masc. plur.
DMS     D..ms---     Determiner, masc. sing.
DNS     D..cs---     Determiner, comm. sing.
DNP     D..cp---     Determiner, comm. plur.
DNN     D..cn---     Determiner, comm. inv.
--------- more collapsed end --------------
======= ============ ================================================
```

5.1.5.3 Combinations

```
========= ========== =================================
Lexicon   Corpus     Example
========= ========== =================================
Dd-cs--   DDNS       tale
Dd-cp--   DDNP       tali
Dd-ms--   DDMS       quello
Dd-mp--   DDMP       quelli
Dd-fs--   DDFS       quella
Dd-fp--   DDFP       quelle

Di-ms--   DIMS       nessun
Di-mp--   DIMP       alcuni
Di-fs--   DIFS       nessuna
Di-fp--   DIFP       alcune
Di-cs--   DINS       ogni
Di-cp--   DINP       quali

Ds1ms--   DPMS       mio, nostro
Ds1mp--   DPMP       miei
Ds1fs--   DPFS       mia
Ds1fp--   DPFP       mie
Ds2ms--   DPMS       tuo, vostro
Ds2mp--   DPMP       tuoi
Ds2fs--   DPFS       tua
Ds2fp--   DPFP       tue
Ds3ms--   DPMS       suo
Ds3mp--   DPMP       suoi
Ds3fs--   DPFS       sua
Ds3fp--   DPFP       sue
Ds-cn--   DPNN       altrui

Dr-cn--   DWNN       cui
Dr-ms--   DWMS       quanto
```

```
Dr-mp--   DWMP       quanti
Dr-fs--   DWFS       quante
Dr-fp--   DWFP       quanti
Dr-cs--   DWNS       quale
Dr-cp--   DWNP       quale

Dt-cn--   DWNN       che
Dt-ms--   DWMS       quanto
Dt-mp--   DWMP       quanti
Dt-fs--   DWFS       quante
Dt-fp--   DWFP       quanti
Dt-cs--   DWNS       quale
Dt-cp--   DWNP       quale

De-cn--   DWNN       che
De-cp--   DWNP       quali
De-cs--   DWNS       quale
De-ms--   DWMS       quanto
De-mp--   DWMP       quanti
De-fs--   DWFS       quanta
De-fp--   DWFP       quante

----------------------- more collapsed ----------------------------
D..fp---  DFP        mie, queste, quante etc.
D..fs---  DFS        mia, questa, quanta etc.
D..mp---  DMP        miei, questi, quanti etc.
D..ms---  DMS        mio, questo, quanto etc.
D..cs---  DNS        quale
D..cp---  DNP        quali
D..cn---  DNN        altrui
----------------------- more collapsed end -----------------------
========= ========== =================================
```

5.1.5.4 Combinations

On the basis of the strategy adopted for Pronouns, also for Determiners two tagsets are proposed.

5.1.6 Articles (T)

5.1.6.1. Lexicon

```
============ =========== =========== ====
Attribute    Value       Example     Code
```

Type	definite	il	d
	indefinite	un	i
Gender	masculine	il	m
	feminine	la	f
l-spec	common	l'	c
Number	singular	la	s
	plural	le	p
Case	(n.a.)	(n.a.)	-

5.1.6.2. Corpus

Tag	Reg.Expr.	Definition
RMS	Tdms-	Article, definite, masc.sing.
RMP	Tdmp-	Article, definite, masc.plur.
RFS	Tdfs-	Article, definite, femm.sing.
RFP	Tdfp-	Article, definite, femm.plur.
RNS	Tdcs-	Article, definite, comm.sing.
RIMS	Tims-	Article, indefinite, masc.sing.
RIFS	Tifs-	Article, indefinite, femm.sing.

5.1.6.3. Combinations

Lexicon	Corpus	Example
Tdms-	RMS	il, lo
Tdmp-	RMP	i, gli
Tdfs-	RFS	la
Tdfp-	RFP	le
Tdcs-	RNS	l' (amico/a)
Tims-	RIMS	un, uno
Tifs-	RIFS	una, un'

5.1.7 Adverbs (R)

5.1.7.1 Lexicon

Attribute	Value	Example	Code
Type	-	-	-
Degree	positive	bene	p
	superlative	benissimo	s

5.1.7.2 Corpus

Tag	Regular Expression	Definition
B	R-p	Adverb positive
BS	R-s	Adverb superaltive

5.1.7.3 Combinations

Lexicon	Corpus	Example
R-p	B	fortemente
R-s	BS	fortissimamente

5.1.7.4. Observations

The feature Type is not encoded in the Italian lexicon.

5.1.8. Adposition (S)

5.1.8.1. Lexicon

Attribute	Value	Example	Code
Type	preposition	di, a, da	p

Formation	simple	di	s	
	compound	dello	c	
Gender	masculine	dello	m	This attribute and values
	feminine	alla	f	have been added
l-spec	common	dell'	c	
Number	singular	al	s	This attribute and values
	plural	ai	p	have been added

5.1.8.2 Corpus

Tag	Regular Expression	Definition
E	Sp-	Preposition simple
EA	Spc..	Preposition compound

5.1.8.3 Combinations

Lexicon	Corpus	Example
Sp	E	di
Spcfs	EA	della
Spcfp	EA	delle
Spcms	EA	del, dello
Spcmp	EA	dei, degli
Spccn	EA	dell'

5.1.8.4 Observations

The Italian policy for encoding fused prepositions foresees to attach the morphological information of the article to the preposition tag.

5.1.9 Conjunctions (C)

5.1.9.1. Lexicon

Attribute	Value	Example	Code
Type	coordinat.	e	c
	subordinat.	perche'	s

5.1.9.2 Corpus

Tag	Regular Expression	Definition
CC	Cc	Coordinative conjunction
CS	Cs	Subordinative conjunction

5.1.9.3 Combinations

Lexicon	Corpus	Example
Cc	CC	ma
Cs	CS	perche'

5.1.10 Numerals (M)

5.1.10.1. Lexicon

Attribute	Value	Example	Code
Type	cardinal	cento	c
	ordinal	primo	o
Gender	masculine	primo	m
	feminine	prima	f
Number	singular	secondo	s
	plural	secondi	p
Case	(n.a.)	(n.a.)	-

============ =========== =========== ====

6.4.10.2 Corpus

Tag	Regular Expression	Definition
NMS	M.ms-	Numeral, masc.sing.
NFS	M.fs-	Numeral, femm.sing.
NMP	M.mp-	Numeral, masc.plur.
NFP	M.fp-	Numeral, femm.plur.
N	Mc---	Numeral cardinal

5.1.10.3 Combinations

Lexicon	Corpus	
M.ms-	NMS	primo
M.fs-	NFS	prima
M.mp-	NMP	primi
M.fp-	NFP	prime
Mc---	N	zero, cento

5.1.11 Interjection (I)

5.1.11.1. Corpus

Tag	Reg. Expr.	Definition
I	I	Interjection

5.1.11.2. Combinations

Lexicon	Corpus	Example
I	I	oh

======= =========== =====================================

5.1.12 Unique membership class (U)

None

5.1.13. Residual (X)

5.1.13.2 Corpus

Tag	Regular Expression	Definition
NY	???	"Guessed" Noun
AY	???	"Guessed" Adjective

5.1.13.3 Combinations

Lexicon	Corpus	Example
???	NY	bit
???	AY	computerizzato

5.1.13.4 Observations

At corpus level, we have the tag SY which is used to mark symbols, letters, acronyms, foreign words, toponyms etc., in general unknown words, for which a "guess" is provided.

5.1.13 Punctuation

Tag	Example
punct	.,;:?! etc.

5.2 Application to German

The application of the MULTEXT encoding scheme to German has been carried out by the German group (Steiner and Lemnitzer 1994).

It has been attempted to keep as close as possible to the conventions. However, some deviations were unavoidable. This concerns:

a. The extension of value sets for some attributes
b. The addition of some minor classes, described in a separate section (see Add on classes)
c. The additon or deletion of an attribute

We will try to justify the changes, or mark them as language-specific. However, some features will be topics for further discussion.

5.2.1 Nouns (N)

5.2.1.1 Lexicon

Attribute	Value	Example	Code
Type	common	Buch	c
	proper	Peter	p
Gender	masculine	Mann	m
	feminine	Frau	f
	neuter	Kind	n
Number	singular	Mann	s
	plural	Frauen	p
Case	nominative	Kind	n
	genitive	Kindes	g
	dative	Kinde	d
	accusative	Kind	a

5.2.1.2 Corpus

Tag	Regular expression	Definition
NCMSN	Ncmsn	Common noun, masc. sing., nominative

NCMSG	Ncmsg	Common noun, masc. sing., genitive
NCMSD	Ncmsd	Common noun, masc. sing., dative
NCMSA	Ncmsa	Common noun, masc. sing., accusative
NCMPN	Ncmpn	Common noun, masc. plur., nominative
NCMPG	Ncmpg	Common noun, masc. plur., genitive
NCMPD	Ncmpd	Common noun, masc. plur., dative
NCMPA	Ncmpa	Common noun, masc. plur., accusative
NCFSN	Ncfsn	Common noun, fem. sing., nominative
NCFSG	Ncfsg	Common noun, fem. sing., genitive
NCFSD	Ncfsd	Common noun, fem. sing., dative
NCFSA	Ncfsa	Common noun, fem. sing., accusative
NCFPN	Ncfpn	Common noun, fem. plur., nominative
NCFPG	Ncfpg	Common noun, fem. plur., genitive
NCFPD	Ncfpd	Common noun, fem. plur., dative
NCFPA	Ncfpa	Common noun, fem. plur., accusative
NCNSN	Ncnsn	Common noun, neut. sing., nominative
NCNSG	Ncnsg	Common noun, neut. sing., genitive
NCNSD	Ncnsd	Common noun, neut. sing., dative
NCNSA	Ncnsa	Common noun, neut. sing., accusative
NCNPN	Ncnpn	Common noun, neut. plur., nominative
NCNPG	Ncnpg	Common noun, neut. plur., genitive
NCNPD	Ncnpd	Common noun, neut. plur., dative
NCNPA	Ncnpa	Common noun, neut. plur., accusative
NPMSN	Npmsn	Proper noun, masc. sing., nominative
NPMSG	Npmsg	Proper noun, masc. sing., genitive
NPMSD	Npmsd	Proper noun, masc. sing., dative
NPMSA	Npmsa	Proper noun, masc. sing., accusative
NPMPN	Npmpn	Proper noun, masc. plur., nominative
NPMPG	Npmpg	Proper noun, masc. plur., genitive
NPMPD	Npmpd	Proper noun, masc. plur., dative
NPMPA	Npmpa	Proper noun, masc. plur., accusative
NPFSN	Npfsn	Proper noun, fem. sing., nominative
NPFSG	Npfsg	Proper noun, fem. sing., genitive
NPFSD	Npfsd	Proper noun, fem. sing., dative
NPFSA	Npfsa	Proper noun, fem. sing., accusative
NPFPN	Npfpn	Proper noun, fem. plur., nominative
NPFPG	Npfpg	Proper noun, fem. plur., genitive
NPFPD	Npfpd	Proper noun, fem. plur., dative

NPFPA	Npfpa	Proper noun, fem. plur., accusative
NPNSN	Npnsn	Proper noun, neut. sing., nominative
NPNSG	Npnsg	Proper noun, neut. sing., genitive
NPNSD	Npnsd	Proper noun, neut. sing., dative
NPNSA	Npnsa	Proper noun, neut. sing., accusative
NPNPN	Npnpn	Proper noun, neut. plur., nominative
NPNPG	Npnpg	Proper noun, neut. plur., genitive
NPNPD	Npnpd	Proper noun, neut. plur., dative
NPNPA	Npnpa	Proper noun, neut. plur., accusative

5.2.1.3 Combinations

Lexique	Corpus	Example
Ncmsn	NCMSN	(der) Hund
Ncmsg	NCMSG	(des) Hundes
Ncmsd	NCMSD	(dem) Hunde
Ncmsa	NCMSA	(den) Hund
Ncmpn	NCMPN	(die) Hunde
Ncmpg	NCMPG	(der) Hunde
Ncmpd	NCMPD	(den) Hunden
Ncmpa	NCMPA	(die) Hunde
Ncfsn	NCFSN	(die) Frau
Ncfsn	NCFSG	(der) Frau
Ncfsd	NCFSD	(der) Frau
Ncfsa	NCFSA	(die) Frau
Ncfpn	NCFPN	(die) Frauen
Ncfpg	NCFPG	(der) Frauen
Ncfpd	NCFPD	(den) Frauen
Ncfpa	NCFPA	(die) Frauen
Ncnsn	NCNSN	(das) Kind
Ncnsg	NCNSG	(des) Kindes
Ncnsd	NCNSD	(dem) Kinde
Ncnsa	NCNSA	(das) Kind
Ncnpn	NCNPN	(die) Kinder
Ncnpg	NCNPG	(der) Kinder

Ncnpd	NCNPD	(den) Kindern
Ncnpa	NCNPA	(die) Kinder
Npmsn	NPMSN	Peter
Npmsg	NPMSG	Peters
Npmsd	NPMSD	Peter
Npmsa	NPMSA	Peter
Npmpn	NPMPN	Einsteins
Npmpg	NPMPG	Einsteins
Npmpd	NPMPD	Einsteins
Npmpa	NPMPA	Einsteins
Npfsn	NPFSN	Sabine
Npfsg	NPFSG	Sabines
Npfsd	NPFSD	Sabine
Npfsa	NPFSA	Sabine
Npfpn	NPFPN	Pyren"aen
Npfpg	NPFPG	Pyren"aen
Npfpd	NPFPD	Pyren"aen
Npfpa	NPFPA	Pyren"aen
Npnsn	NPNSN	Bayern
Npnsg	NPNSG	Bayerns
Npnsd	NPNSD	Bayern
Npnsa	NPNSA	Bayern
Npnpn	NPNPN	Bayerns
Npnpg	NPNPG	Bayerns
Npnpd	NPNPD	Bayerns
Npnpa	NPNPA	Bayerns

5.2.2 Verbs (V)

5.2.2.1 Lexicon

Attribute	Value	Example	Code	
Type	main	gehen	m	
	modal	sollen	o	this value has been added
	auxiliary	haben	a	

Mood	indicative	geht	i	
	subjunctive	gehe	s	
	imperative	geht	m	
	infinitive	gehen	n	
	inf. with inc. particle	wegzugehen	u	this value has been added
	participle	gehend	p	
Tense	present	geht	p	
	imperfect	ging	i	
Person	first	bin	1	
	second	bist	2	
	third	ist	3	
Number	singular	geht	s	
	plural	gehen	p	
Gender	///	///	-	
Clitic	no	hat	n	
	yes	hats	y	

Notes:

a. Gender

There is no distinction in gender for the third person singular.

5.2.2.2. Corpus

Tag	Regular expression	Definition
VAIP1PN	Vaip1p-n	Aux. verb, 1st pers. pl. ind. pres., nonclitic
VAIP1PY	Vaip1p-y	Aux. verb, 1st pers. pl. ind. pres., clitic
VAII1PN	Vaii1p-n	Aux. verb, 1st pers. pl. ind. imp., nonclitic
VAII1PY	Vaii1p-y	Aux. verb, 1st pers. pl. ind. imp., clitic
VASP1PN	Vasp1p-n	Aux. verb, 1st pers. pl. subj. pres., nonclit
VASP1PY	Vasp1p-y	Aux. verb, 1st pers. pl. subj. pres., clitic
VASI1PN	Vasi1p-n	Aux. verb, 1st pers. pl. subj. imp., nonclitic
VASI1PY	Vasi1p-y	Aux. verb, 1st pers. pl. subj. imp., clitic
VAIP1SN	Vaip1s-n	Aux. verb, 1st pers. sg. ind. pres., nonclitic

VAIP1SY	Vaip1s-y	Aux. verb, 1st pers. sg. ind. pres., clitic
VAII1SN	Vaii1s-n	Aux. verb, 1st pers. sg. ind. imp., nonclitic
VAII1SY	Vaii1s-y	Aux. verb, 1st pers. sg. ind. imp., clitic
VASP1SN	Vasp1s-n	Aux. verb, 1st pers. sg. subj. pres., nonclit
VASP1SY	Vasp1s-y	Aux. verb, 1st pers. sg. subj. pres., clitic
VASI1SN	Vasi1s-n	Aux. verb, 1st pers. sg. subj. imp., nonclitic
VASI1SY	Vasi1s-y	Aux. verb, 1st pers. sg. subj. imp., clitic
VAIP2PN	Vaip2p-n	Aux. verb, 2nd pers. pl. ind. pres., nonclitic
VAIP2PY	Vaip2p-y	Aux. verb, 2nd pers. pl. ind. pres., clitic
VAII2PN	Vaii2p-n	Aux. verb, 2nd pers. pl. ind. imp., nonclitic
VAII2PY	Vaii2p-y	Aux. verb, 2nd pers. pl. ind. imp., clitic
VASP2PN	Vasp2p-n	Aux. verb, 2nd pers. pl. subj. pres., nonclit
VASP2PY	Vasp2p-y	Aux. verb, 2nd pers. pl. subj. pres., clitic
VASI2PN	Vasi2p-n	Aux. verb, 2nd pers. pl. subj. imp., nonclitic
VASI2PY	Vasi2p-y	Aux. verb, 2nd pers. pl. subj. imp., clitic
VAM2PN	Vam-2p-n	Aux. verb, 2nd pers. pl. imperative, nonclitic
VAM2PY	Vam-2p-y	Aux. verb, 2nd pers. pl. imperative, clitic
VAIP2SN	Vaip2s-n	Aux. verb, 2nd pers. sg. ind. pres., nonclitic
VAIP2SY	Vaip2s-y	Aux. verb, 2nd pers. sg. ind. pres., clitic
VAII2SN	Vaii2s-n	Aux. verb, 2nd pers. sg. ind. imp., nonclitic
VAII2SY	Vaii2s-y	Aux. verb, 2nd pers. sg. ind. imp., clitic
VASP2SN	Vasp2s-n	Aux. verb, 2nd pers. sg. subj. pres., nonclit
VASP2SY	Vasp2s-y	Aux. verb, 2nd pers. sg. subj. pres., clitic
VASI2SN	Vasi2s-n	Aux. verb, 2nd pers. sg. subj. imp., nonclitic
VASI2SY	Vasi2s-y	Aux. verb, 2nd pers. sg. subj. imp., clitic
VAM2SN	Vam-2s-n	Aux. verb, 2nd pers. sg. imperative, nonclitic
VAM2SY	Vam-2s-y	Aux. verb, 2nd pers. sg. imperative, clitic
VAIP3PN	Vaip3p-n	Aux. verb, 3rd pers. pl. ind. pres., nonclitic
VAIP3PY	Vaip3p-y	Aux. verb, 3rd pers. pl. ind. pres., clitic
VAII3PN	Vaii3p-n	Aux. verb, 3rd pers. pl. ind. imp., nonclitic
VAII3PY	Vaii3p-y	Aux. verb, 3rd pers. pl. ind. imp., clitic
VASP3PN	Vasp3p-n	Aux. verb, 3rd pers. pl. subj. pres., nonclitic
VASP3PY	Vasp3p-y	Aux. verb, 3rd pers. pl. subj. pres., clitic
VASI3PN	Vasi3p-n	Aux. verb, 3rd pers. pl. subj. imp., nonclitic
VASI3PY	Vasi3p-y	Aux. verb, 3rd pers. pl. subj. imp., clitic
VAIS3SN	Vaip3s-n	Aux. verb, 3rd pers. sg. ind. pres., nonclitic
VAIS3SY	Vaip3s-y	Aux. verb, 3rd pers. sg. ind. pres., clitic
VAII3SN	Vaii3s-n	Aux. verb, 3rd pers. sg. ind. imp., nonclitic
VAII3SY	Vaii3s-y	Aux. verb, 3rd pers. sg. ind. imp., clitic
VASP3SN	Vasp3s-n	Aux. verb, 3rd pers. sg. subj. pres., nonclit
VASP3SY	Vasp3s-y	Aux. verb, 3rd pers. sg. subj. pres., clitic
VASI3SN	Vasi3s-n	Aux. verb, 3rd pers. sg. subj. imp., nonclitic
VASI3SY	Vasi3s-y	Aux. verb, 3rd pers. sg. subj. imp., clitic

VAPS	Vaps----	Aux. verb, past part.
VAN	Van-----	Aux. verb, infinitive
VAPP	Vapp----	Aux. verb, pres. participle
VOIP1PN	Voip1p-n	Mod. verb, 1st pers. pl. ind. pres., nonclitic
VOIP1PY	Voip1p-y	Mod. verb, 1st pers. pl. ind. pres., clitic
VOII1PN	Voii1p-n	Mod. verb, 1st pers. pl. ind. imp., nonclitic
VOII1PY	Voii1p-y	Mod. verb, 1st pers. pl. ind. imp., clitic
VOSP1PN	Vosp1p-n	Mod. verb, 1st pers. pl. subj. pres., nonclit
VOSP1PY	Vosp1p-y	Mod. verb, 1st pers. pl. subj. pres., clitic
VOSI1PN	Vosi1p-n	Mod. verb, 1st pers. pl. subj. imp., nonclitic
VOSI1PY	Vosi1p-y	Mod. verb, 1st pers. pl. subj. imp., clitic
VOIP1SN	Voip1s-n	Mod. verb, 1st pers. sg. ind. pres., nonclitic
VOIP1SY	Voip1s-y	Mod. verb, 1st pers. sg. ind. pres., clitic
VOII1SN	Voii1s-n	Mod. verb, 1st pers. sg. ind. imp., nonclitic
VOII1SY	Voii1s-y	Mod. verb, 1st pers. sg. ind. imp., clitic
VOSP1SN	Vosp1s-n	Mod. verb, 1st pers. sg. subj. pres., nonclit
VOSP1SY	Vosp1s-y	Mod. verb, 1st pers. sg. subj. pres., clitic
VOSI1SN	Vosi1s-n	Mod. verb, 1st pers. sg. subj. imp., nonclitic
VOSI1SY	Vosi1s-y	Mod. verb, 1st pers. sg. subj. imp.,clitic
VOIP2PN	Voip2p-n	Mod. verb, 2nd pers. pl. ind. pres., nonclitic
VOIP2PY	Voip2p-y	Mod. verb, 2nd pers. pl. ind. pres., clitic
VOII2PN	Voii2p-n	Mod. verb, 2nd pers. pl. ind. imp., nonclitic
VOII2PY	Voii2p-y	Mod. verb, 2nd pers. pl. ind. imp., clitic
VOSP2PN	Vosp2p-n	Mod. verb, 2nd pers. pl. subj. pres., nonclit
VOSP2PY	Vosp2p-y	Mod. verb, 2nd pers. pl. subj. pres., clitic
VOSI2PN	Vosi2p-n	Mod. verb, 2nd pers. pl. subj. imp., nonclitic
VOSI2PY	Vosi2p-y	Mod. verb, 2nd pers. pl. subj. imp., clitic
VOIP2SN	Voip2s-n	Mod. verb, 2nd pers. sg. ind. pres., nonclitic
VOIP2SY	Voip2s-y	Mod. verb, 2nd pers. sg. ind. pres., clitic
VOII2SN	Voii2s-n	Mod. verb, 2nd pers. sg. ind. imp., nonclitic
VOII2SY	Voii2s-y	Mod. verb, 2nd pers. sg. ind. imp., clitic
VOSP2SN	Vosp2s-n	Mod. verb, 2nd pers. sg. subj. pres., nonclit
VOSP2SY	Vosp2s-y	Mod. verb, 2nd pers. sg. subj. pres., clitic
VOSI2SN	Vosi2s-n	Mod. verb, 2nd pers. sg. subj. imp., nonclitic
VOSI2SY	Vosi2s-y	Mod. verb, 2nd pers. sg. subj. imp., clitic
VOIP3PN	Voip3p-n	Mod. verb, 3rd pers. pl. ind. pres., nonclitic
VOIP3PY	Voip3p-y	Mod. verb, 3rd pers. pl. ind. pres., clitic
VOII3PN	Voii3p-n	Mod. verb, 3rd pers. pl. ind. imp., nonclitic
VOII3PY	Voii3p-y	Mod. verb, 3rd pers. pl. ind. imp., clitic
VOSP3PN	Vosp3p-n	Mod. verb, 3rd pers. pl. subj. pres., nonclit
VOSP3PY	Vosp3p-y	Mod. verb, 3rd pers. pl. subj. pres., clitic

VOSI3PN	Vosi3p-n	Mod. verb, 3rd pers. pl. subj. imp., nonclitic
VOSI3PY	Vosi3p-y	Mod. verb, 3rd pers. pl. subj. imp., clitic
VOIS3SN	Voip3s-n	Mod. verb, 3rd pers. sg. ind. pres., nonclitic
VOIS3SY	Voip3s-y	Mod. verb, 3rd pers. sg. ind. pres., clitic
VOII3SN	Voii3s-n	Mod. verb, 3rd pers. sg. ind. imp., nonclitic
VOII3SY	Voii3s-y	Mod. verb, 3rd pers. sg. ind. imp., clitic
VOSP3SN	Vosp3s-n	Mod. verb, 3rd pers. sg. subj. pres., nonclit
VOSP3SY	Vosp3s-y	Mod. verb, 3rd pers. sg. subj. pres., clitic
VOSI3SN	Vosi3s-n	Mod. verb, 3rd pers. sg. subj. imp., nonclitic
VOSI3SY	Vosi3s-y	Mod. verb, 3rd pers. sg. subj. imp., clitic
VOPS	Vops----	Mod. verb, past part.
VON	Von-----	Mod. verb, infinitive
VOPP	Vopp----	Mod. verb, pres. participle
VMIP1PN	Vmip1p-n	Main verb, 1st pers. pl. ind. pres., nonclitic
VMIP1PY	Vmip1p-y	Main verb, 1st pers. pl. ind. pres., clitic
VMII1PN	Vmii1p-n	Main verb, 1st pers. pl. ind. imp., nonclitic
VMII1PY	Vmii1p-y	Main verb, 1st pers. pl. ind. imp., clitic
VMSP1PN	Vmsp1p-n	Main verb, 1st pers. pl. subj. pres., nonclit
VMSP1PY	Vmsp1p-y	Main verb, 1st pers. pl. subj. pres., clitic
VMSI1PN	Vmsi1p-n	Main verb, 1st pers. pl. subj. imp., nonclitic
VMSI1PY	Vmsi1p-y	Main verb, 1st pers. pl. subj. imp., clitic
VMIP1SN	Vmip1s-n	Main verb, 1st pers. sg. ind. pres., nonclitic
VMIP1SY	Vmip1s-y	Main verb, 1st pers. sg. ind. pres., clitic
VMII1SN	Vmii1s-n	Main verb, 1st pers. sg. ind. imp., nonclitic
VMII1SY	Vmii1s-y	Main verb, 1st pers. sg. ind. imp., clitic
VMSP1SN	Vmsp1s-n	Main verb, 1st pers. sg. subj. pres., nonclit
VMSP1SY	Vmsp1s-y	Main verb, 1st pers. sg. subj. pres., clitic
VMSI1SN	Vmsi1s-n	Main verb, 1st pers. sg. subj. imp., nonclitic
VMSI1SY	Vmsi1s-y	Main verb, 1st pers. sg. subj. imp., clitic
VMIP2PN	Vmip2p-n	Main verb, 2nd pers. pl. ind. pres., nonclitic
VMIP2PY	Vmip2p-y	Main verb, 2nd pers. pl. ind. pres., clitic
VMII2PN	Vmii2p-n	Main verb, 2nd pers. pl. ind. imp., nonclitic
VMII2PY	Vmii2p-y	Main verb, 2nd pers. pl. ind. imp., clitic
VMSP2PN	Vmsp2p-n	Main verb, 2nd pers. pl. subj. pres., nonclit
VMSP2PY	Vmsp2p-y	Main verb, 2nd pers. pl. subj. pres., clitic
VMSI2PN	Vmsi2p-n	Main verb, 2nd pers. pl. subj. imp., nonclitic
VMSI2PY	Vmsi2p-y	Main verb, 2nd pers. pl. subj. imp., clitic
VMM2PN	Vmm-2p-n	Main verb, 2nd pers. pl. imperative, nonclitic
VMM2PY	Vmm-2p-y	Main verb, 2nd pers. pl. imperative, clitic
VMIP2SN	Vmip2s-n	Main verb, 2nd pers. sg. ind. pres., nonclitic
VMIP2SY	Vmip2s-y	Main verb, 2nd pers. sg. ind. pres., clitic

```
VMII2SN  Vmii2s-n           Main verb, 2nd pers. sg. ind. imp., nonclitic
VMII2SY  Vmii2s-y           Main verb, 2nd pers. sg. ind. imp., clitic
VMSP2SN  Vmsp2s-n           Main verb, 2nd pers. sg. subj. pres., nonclit
VMSP2SY  Vmsp2s-y           Main verb, 2nd pers. sg. subj. pres., clitic
VMSI2SN  Vmsi2s-n           Main verb, 2nd pers. sg. subj. imp., nonclitic
VMSI2SY  Vmsi2s-y           Main verb, 2nd pers. sg. subj. imp., clitic
VMM2SN   Vmm-2s-n           Main verb, 2nd pers. sg. imperative, nonclitic
VMM2SY   Vmm-2s-y           Main verb, 2nd pers. sg. imperative, clitic

VMIP3PN  Vmip3p-n           Main verb, 3rd pers. pl. ind. pres., nonclitic
VMIP3PY  Vmip3p-y           Main verb, 3rd pers. pl. ind. pres., clitic
VMII3PN  Vmii3p-n           Main verb, 3rd pers. pl. ind. imp., nonclitic
VMII3PY  Vmii3p-y           Main verb, 3rd pers. pl. ind. imp., clitic
VMSP3PN  Vmsp3p-n           Main verb, 3rd pers. pl. subj. pres., nonclit
VMSP3PY  Vmsp3p-y           Main verb, 3rd pers. pl. subj. pres., clitic
VMSI3PN  Vmsi3p-n           Main verb, 3rd pers. pl. subj. imp., nonclitic
VMSI3PY  Vmsi3p-y           Main verb, 3rd pers. pl. subj. imp., clitic

VMIS3SN  Vmip3s-n           Main verb, 3rd pers. sg. ind. pres., nonclitic
VMIS3SY  Vmip3s-y           Main verb, 3rd pers. sg. ind. pres., clitic
VMII3SN  Vmii3s-n           Main verb, 3rd pers. sg. ind. imp., nonclitic
VMII3SY  Vmii3s-y           Main verb, 3rd pers. sg. ind. imp., clitic
VMSP3SN  Vmsp3s-n           Main verb, 3rd pers. sg. subj. pres., nonclit
VMSP3SY  Vmsp3s-y           Main verb, 3rd pers. sg. subj. pres., clitic
VMSI3SN  Vmsi3s-n           Main verb, 3rd pers. sg. subj. imp., nonclitic
VMSI3SY  Vmsi3s-y           Main verb, 3rd pers. sg. subj. imp., clitic

VMPS     Vmps----           Main verb, past part.
VMN      Vmn-----           Main verb, infinitive
VMU      Vmu-----           Main verb, infinitive with incorp. particle
VMPP     Vmpp----           Main verb, pres. participle
======= ================== ====================================
```

5.2.2.3 Combinations

```
========= ======= =============================================
Lexique   Corpus  Example
========= ======= =============================================

Vaip1p-n  VAIP1PN          sind
Vaip1p-y  VAIP1PY          sinds
Vaii1p-n  VAII1PN          waren
Vaii1p-y  VAII1PY          warens
Vasp1p-n  VASP1PN          seien
Vasp1p-y  VASP1PY          seiens
Vasi1p-n  VASI1PN          w"aren
```

Vasi1p-y	VASI1PY	w"arens
Vaip1s-n	VAIP1SN	bin
Vaip1s-y	VAIP1SY	bins
Vaii1s-n	VAII1SN	war
Vaii1s-y	VAII1SY	wars
Vasp1s-n	VASP1SN	sei
Vasp1s-y	VASP1SY	seis
Vasi1s-n	VASI1SN	w"are
Vasi1s-y	VASI1SY	w"ares
Vaip2p-n	VAIP2PN	seid
Vaip2p-y	VAIP2PY	seids
Vaii2p-n	VAII2PN	wart
Vaii2p-y	VAII2PY	warts
Vasp2p-n	VASP2PN	seiet
Vasp2p-y	VASP2PY	seiets
Vasi2p-n	VASI2PN	w"aret
Vasi2p-y	VASI2PY	w"arets
Vam-2p-n	VAM2PN	seid
Vam-2p-y	VAM2PY	seids
Vaip2s-n	VAIP2SN	bist
Vaip2s-y	VAIP2SY	bists
Vaii2s-n	VAII2SN	warst
Vaii2s-y	VAII2SY	warsts
Vasp2s-n	VASP2SN	seist
Vasp2s-y	VASP2SY	seists
Vasi2s-n	VASI2SN	w"arest
Vasi2s-y	VASI2SY	w"arests
Vam-2s-n	VAM2SN	sei
Vam-2s-y	VAM2SY	seis
Vaip3p-n	VAIP3PN	sind
Vaip3p-y	VAIP3PY	sinds
Vaii3p-n	VAII3PN	waren
Vaii3p-y	VAII3PY	warens
Vasp3p-n	VASP3PN	seien
Vasp3p-y	VASP3PY	seiens
Vasi3p-n	VASI3PN	w"aren
Vasi3p-y	VASI3PY	w"arens
Vaip3s-n	VAIS3SN	ist
Vaip3s-y	VAIS3SY	ists
Vaii3s-n	VAII3SN	war
Vaii3s-y	VAII3SY	wars
Vasp3s-n	VASP3SN	sei

Vasp3s-y	VASP3SY	seis
Vasi3s-n	VASI3SN	w"are
Vasi3s-y	VASI3SY	w"ares
Vaps----	VAPS	gehabt
Van-----	VAN	haben
Vapp----	VAPP	habend
Voip1p-n	VOIP1PN	sollen
Voip1p-y	VOIP1PY	sollens
Voii1p-n	VOII1PN	sollten
Voii1p-y	VOII1PY	solltens
Vosp1p-n	VOSP1PN	sollen
Vosp1p-y	VOSP1PY	sollens
Vosi1p-n	VOSI1PN	sollten
Vosi1p-y	VOSI1PY	solltens
Voip1s-n	VOIP1SN	soll
Voip1s-y	VOIP1SY	solls
Voii1s-n	VOII1SN	sollte
Voii1s-y	VOII1SY	solltes
Vosp1s-n	VOSP1SN	solle
Vosp1s-y	VOSP1SY	solles
Vosi1s-n	VOSI1SN	sollte
Vosi1s-y	VOSI1SY	solltes
Voip2p-n	VOIP2PN	sollt
Voip2p-y	VOIP2PY	sollts
Voii2p-n	VOII2PN	solltet
Voii2p-y	VOII2PY	solltets
Vosp2p-n	VOSP2PN	sollet
Vosp2p-y	VOSP2PY	sollets
Vosi2p-n	VOSI2PN	solltet
Vosi2p-y	VOSI2PY	solltets
Vom-2p-n	VOM2PN	sollt
Vom-2p-y	VOM2PY	sollts
Voip2s-n	VOIP2SN	sollst
Voip2s-y	VOIP2SY	sollsts
Voii2s-n	VOII2SN	solltest
Voii2s-y	VOII2SY	solltests
Vosp2s-n	VOSP2SN	sollest
Vosp2s-y	VOSP2SY	sollests
Vosi2s-n	VOSI2SN	solltest
Vosi2s-y	VOSI2SY	solltests
Vom-2s-n	VOM2SN	soll

Vom-2s-y	VOM2SY	solls
Voip3p-n	VOIP3PN	sollen
Voip3p-y	VOIP3PY	sollens
Voii3p-n	VOII3PN	sollten
Voii3p-y	VOII3PY	solltens
Vosp3p-n	VOSP3PN	sollen
Vosp3p-y	VOSP3PY	sollens
Vosi3p-n	VOSI3PN	sollten
Vosi3p-y	VOSI3PY	solltens
Voip3s-n	VOIS3SN	soll
Voip3s-y	VOIS3SY	solls
Voii3s-n	VOII3SN	sollte
Voii3s-y	VOII3SY	solltes
Vosp3s-n	VOSP3SN	solle
Vosp3s-y	VOSP3SY	solles
Vosi3s-n	VOSI3SN	sollte
Vosi3s-y	VOSI3SY	solltes
Vops----	VOPS	gesollt
Von-----	VON	sollen
Vopp----	VOPP	sollend
Vmip1p-n	VMIP1PN	schreiben
Vmip1p-y	VMIP1PY	schreibens
Vmii1p-n	VMII1PN	schrieben
Vmii1p-y	VMII1PY	schriebens
Vmsp1p-n	VMSP1PN	schreiben
Vmsp1p-y	VMSP1PY	schreibens
Vmsi1p-n	VMSI1PN	schrieben
Vmsi1p-y	VMSI1PY	schriebens
Vmip1s-n	VMIP1SN	schreibe
Vmip1s-y	VMIP1SY	schreibes
Vmii1s-n	VMII1SN	schrieb
Vmii1s-y	VMII1SY	schriebs
Vmsp1s-n	VMSP1SN	schreibe
Vmsp1s-y	VMSP1SY	schreibes
Vmsi1s-n	VMSI1SN	schriebe
Vmsi1s-y	VMSI1SY	schriebes
Vmip2p-n	VMIP2PN	schreibt
Vmip2p-y	VMIP2PY	schreibts
Vmii2p-n	VMII2PN	schriebt
Vmii2p-y	VMII2PY	schriebts
Vmsp2p-n	VMSP2PN	schreibet

```
Vmsp2p-y  VMSP2PY           schreibets
Vmsi2p-n  VMSI2PN           schriebet
Vmsi2p-y  VMSI2PY           schriebets
Vmm-2p-n  VMM2PN            schreibt
Vmm-2p-y  VMM2PY            schreibts

Vmip2s-n  VMIP2SN           schreibst
Vmip2s-y  VMIP2SY           schreibsts
Vmii2s-n  VMII2SN           schriebst
Vmii2s-y  VMII2SY           schriebsts
Vmsp2s-n  VMSP2SN           schreibest
Vmsp2s-y  VMSP2SY           schreibests
Vmsi2s-n  VMSI2SN           schriebest
Vmsi2s-y  VMSI2SY           schriebests
Vmm-2s-n  VMM2SN            schreib
Vmm-2s-y  VMM2SY            schreibs

Vmip3p-n  VMIP3PN           schreiben
Vmip3p-y  VMIP3PY           schreibens
Vmii3p-n  VMII3PN           schrieben
Vmii3p-y  VMII3PY           schriebens
Vmsp3p-n  VMSP3PN           schreiben
Vmsp3p-y  VMSP3PY           schreibens
Vmsi3p-n  VMSI3PN           schrieben
Vmsi3p-y  VMSI3PY           schriebens

Vmip3s-n  VMIS3SN           schreibt
Vmip3s-y  VMIS3SY           schreibts
Vmii3s-n  VMII3SN           schrieb
Vmii3s-y  VMII3SY           schriebs
Vmsp3s-n  VMSP3SN           schreibe
Vmsp3s-y  VMSP3SY           schreibes
Vmsi3s-n  VMSI3SN           schriebe
Vmsi3s-y  VMSI3SY           schriebes

Vm---ps-  VMPS              gegangen
Vm---n--  VMN               schreiben
Vm---u--  VMU               wegzuschreiben
Vm---pp-  VMPP              schreibend
==========================  =============
```

Note:

Adding participles here would imply the addition of adjective features. Morphologically at least, participles behave like adjectives. A good - but not very elegant - solution would therefore be to handle participles as a type of adjective.

Secondly there are ambigous cases, e.g.

Er ist ger"uhrt.

where ger"uhrt is a participle, but might be tagged either as a verb or an adjective, depending on the context. This would be violating the assumption for applicativeness. The best solution at hand is to treat these forms as ambiguous with respect to membership in a word class (adjective and verb, respectively). However, this is a mixture of morphosyntactic with distributional criteria, and therefore unsatisfactory.

5.2.3 Adjectives (A)

5.2.3.1. Lexicon

Attribute	Value	Example	Code	
Type	qualificat.	gut	f	
	ordinal	zweites	o	
	cardinal	zwei	c	this value has been added
	possessive	mein	s	
	part1	lachende	1	this value has been added
	part2	gesungene	2	this value has been added
Degree	positive	gut	p	
	comparative	besser	c	
	superlative	beste	s	
Gender	masculine	guter	m	
	feminine	gute	f	
	neuter	gutes	n	
Number	singular	guter	s	
	plural	gute	p	
Case	nominative	guter	n	
	genitive	guten	g	
	dative	guten	d	
	accusative	guten	a	

Notes

a. We decided to include cardinal as well as ordinal numbers. Therefore there is no special class for numerals.

b. Although we have doubts concerning the "possessive adjectives" and would prefer to add the form

(Der Ball ist) mein

and

das ist meins

to the possessive Determiner or Pronouns because 'mein' originally is a possessive pronoun, we recognize that the present definition of these categories does not allow the addition of the value 'predicative'. This treatment is therefore a compromise.

c. German adjectives can be used in an attributive or a predicative mode. Predicative adjectives are not marked for gender, case or number with the exception of possessive and ordinal adjectives.

d. It would be necessary to specify the inflection type. The inflection type reflects the place of an adjective in an NP (following a determiner, an article, or none of these). Values are strong, mixed, and weak. However, we have left this feature in this generic version.

e. part1 refers to adjectives that are derived from present participles.
part2 refers to adjectives that are derived from past participles.

5.2.3.2 Corpus

Tag	Regular expression	Definition
AMSN	A..msn	Adjective masc. sing. nominative
AMSG	A..msg	Adjective masc. sing. genitive
AMSD	A..msd	Adjective masc. sing. dative
AMSA	A..msa	Adjective masc. sing. accusative
AMPN	A..mpn	Adjective masc. plur. nominative
AMPG	A..mpg	Adjective masc. plur. genitive
AMPD	A..mpd	Adjective masc. plur. dative
AMPA	A..mpa	Adjective masc. plur. accusative

AFSN	A..fsn	Adjective fem. sing. nominative
AFSG	A..fsg	Adjective fem. sing. genitive
AFSD	A..fsd	Adjective fem. sing. dative
AFSA	A..fsa	Adjective fem. sing. accusative
AFPN	A..fpn	Adjective fem. plur. nominative
AFPG	A..fpg	Adjective fem. plur. genitive
AFPD	A..fpd	Adjective fem. plur. dative
AFPA	A..fpa	Adjective fem. plur. accusative
ANSN	A..fsn	Adjective neut. sing. nominative
ANSG	A..fsg	Adjective neut. sing. genitive
ANSD	A..fsd	Adjective neut. sing. dative
ANSA	A..fsa	Adjective neut. sing. accusative
ANPN	A..fpn	Adjective neut. plur. nominative
ANPG	A..fpg	Adjective neut. plur. genitive
ANPD	A..fpd	Adjective neut. plur. dative
ANPA	A..fpa	Adjective neut. plur. accusative
A	A[q12][pc]---	Adjective, predic. without gender mark,comparable
AP	A[cp]---	Adjective, predic. without g.m., not comparable
AP.	A[op]p.--	Adjective, predic. with gender mark and num.mark
AS	A[q12]s---	Adjective, predicative superlative

5.2.3.3 Combinations

Lexique	Corpus	Example
Aqpmsn	AMSN	gute
Aqpmsg	AMSG	guten
Aqpmsd	AMSD	guten
Aqpmsa	AMSA	guten
Aqpmpn	AMPN	guten
Aqpmpg	AMPG	guten
Aqpmpd	AMPD	guten
Aqpmpa	AMPA	guten
Aqpfsn	AFSN	gute
Aqpfsg	AFSG	guten
Aqpfsd	AFSD	guten
Aqpfsa	AFSA	gute

Aqpfpn	AFPN	guten
Aqpfpg	AFPG	guten
Aqpfpd	AFPD	guten
Aqpfpa	AFPA	guten
Aqpnsn	ANSN	gute
Aqpnsg	ANSG	guten
Aqpnsd	ANSD	guten
Aqpnsa	ANSA	gute
Aqpnpn	ANPN	guten
Aqpnpg	ANPG	guten
Aqpnpd	ANPD	guten
Aqpnpa	ANPA	guten
Aqcmsn	AMSN	bessere
Aqcmsg	AMSG	besseren
Aqcmsd	AMSD	besseren
Aqcmsa	AMSA	besseren
Aqcmpn	AMPN	besseren
Aqcmpg	AMPG	besseren
Aqcmpd	AMPD	besseren
Aqcmpa	AMPA	besseren
Aqcfsn	AFSN	bessere
Aqcfsg	AFSG	besseren
Aqcfsd	AFSD	besseren
Aqcfsa	AFSA	bessere
Aqcfpn	AFPN	besseren
Aqcfpg	AFPG	besseren
Aqcfpd	AFPD	besseren
Aqcfpa	AFPA	besseren
Aqcnsn	ANSN	bessere
Aqcnsg	ANSG	besseren
Aqcnsd	ANSD	besseren
Aqcnsa	ANSA	bessere
Aqcnpn	ANPN	besseren
Aqcnpg	ANPG	besseren
Aqcnpd	ANPD	besseren
Aqcnpa	ANPA	besseren
Aqsmsn	AMSN	beste

Aqsmsg	AMSG	besten
Aqsmsd	AMSD	besten
Aqsmsa	AMSA	besten
Aqsmpn	AMPN	besten
Aqsmpg	AMPG	besten
Aqsmpd	AMPD	besten
Aqsmpa	AMPA	besten
Aqsfsn	AFSN	beste
Aqsfsg	AFSG	besten
Aqsfsd	AFSD	besten
Aqsfsa	AFSA	beste
Aqsfpn	AFPN	besten
Aqsfpg	AFPG	besten
Aqsfpd	AFPD	besten
Aqsfpa	AFPA	besten
Aqsnsn	ANSN	beste
Aqsnsg	ANSG	besten
Aqsnsd	ANSD	besten
Aqsnsa	ANSA	beste
Aqsnpn	ANPN	besten
Aqsnpg	ANPG	besten
Aqsnpd	ANPD	besten
Aqsnpa	ANPA	besten
Aopmsn	AMSN	zweite
Aopmsg	AMSG	zweiten
Aopmsd	AMSD	zweiten
Aopmsa	AMSA	zweiten
Aopmpn	AMPN	zweiten
Aopmpg	AMPG	zweiten
Aopmpd	AMPD	zweiten
Aopmpa	AMPA	zweiten
Aopfsn	AFSN	zweite
Aopfsg	AFSG	zweiten
Aopfsd	AFSD	zweiten
Aopfsa	AFSA	zweite
Aopfpn	AFPN	zweiten
Aopfpg	AFPG	zweiten
Aopfpd	AFPD	zweiten

Aopfpa	AFPA	zweiten
Aopnsn	ANSN	zweite
Aopnsg	ANSG	zweiten
Aopnsd	ANSD	zweiten
Aopnsa	ANSA	zweite
Aopnpn	ANPN	zweiten
Aopnpg	ANPG	zweiten
Aopnpd	ANPD	zweiten
Aopnpa	ANPA	zweiten
Acpmpn	AMPN	zwei
Acpmpg	AMPG	zwei
Acpmpd	AMPD	zwei
Acpmpa	AMPA	zwei
Acpfpn	AFPN	zwei
Acpfpg	AFPG	zwei
Acpfpd	AFPD	zwei
Acpfpa	AFPA	zwei
Acpnpn	ANPN	zwei
Acpnpg	ANPG	zwei
Acpnpd	ANPD	zwei
Acpnpa	ANPA	zwei
A1pmsn	AMSN	beruhigende
A1pmsg	AMSG	beruhigenden
A1pmsd	AMSD	beruhigenden
A1pmsa	AMSA	beruhigenden
A1pmpn	AMPN	beruhigenden
A1pmpg	AMPG	beruhigenden
A1pmpd	AMPD	beruhigenden
A1pmpa	AMPA	beruhigenden
A1pfsn	AFSN	beruhigende
A1pfsg	AFSG	beruhigenden
A1pfsd	AFSD	beruhigenden
A1pfsa	AFSA	beruhigende
A1pfpn	AFPN	beruhigenden
A1pfpg	AFPG	beruhigenden
A1pfpd	AFPD	beruhigenden
A1pfpa	AFPA	beruhigenden

A1pnsn	ANSN	beruhigende
A1pnsg	ANSG	beruhigenden
A1pnsd	ANSD	beruhigenden
A1pnsa	ANSA	beruhigende
A1pnpn	ANPN	beruhigenden
A1pnpg	ANPG	beruhigenden
A1pnpd	ANPD	beruhigenden
A1pnpa	ANPA	beruhigenden
A1cmsn	AMSN	beruhigendere
A1cmsg	AMSG	beruhigenderen
A1cmsd	AMSD	beruhigenderen
A1cmsa	AMSA	beruhigenderen
A1cmpn	AMPN	beruhigenderen
A1cmpg	AMPG	beruhigenderen
A1cmpd	AMPD	beruhigenderen
A1cmpa	AMPA	beruhigenderen
A1cfsn	AFSN	beruhigendere
A1cfsg	AFSG	beruhigenderen
A1cfsd	AFSD	beruhigenderen
A1cfsa	AFSA	beruhigendere
A1cfpn	AFPN	beruhigenderen
A1cfpg	AFPG	beruhigenderen
A1cfpd	AFPD	beruhigenderen
A1cfpa	AFPA	beruhigenderen
A1cnsn	ANSN	beruhigendere
A1cnsg	ANSG	beruhigenderen
A1cnsd	ANSD	beruhigenderen
A1cnsa	ANSA	beruhigendere
A1cnpn	ANPN	beruhigenderen
A1cnpg	ANPG	beruhigenderen
A1cnpd	ANPD	beruhigenderen
A1cnpa	ANPA	beruhigenderen
A1smsn	AMSN	beruhigendste
A1smsg	AMSG	beruhigendsten
A1smsd	AMSD	beruhigendsten
A1smsa	AMSA	beruhigendsten
A1smpn	AMPN	beruhigendsten
A1smpg	AMPG	beruhigendsten

A1smpd	AMPD	beruhigendsten
A1smpa	AMPA	beruhigendsten
A1sfsn	AFSN	beruhigendste
A1sfsg	AFSG	beruhigendsten
A1sfsd	AFSD	beruhigendsten
A1sfsa	AFSA	beruhigendste
A1sfpn	AFPN	beruhigendsten
A1sfpg	AFPG	beruhigendsten
A1sfpd	AFPD	beruhigendsten
A1sfpa	AFPA	beruhigendsten
A1snsn	ANSN	beruhigendste
A1snsg	ANSG	beruhigendsten
A1snsd	ANSD	beruhigendsten
A1snsa	ANSA	beruhigendste
A1snpn	ANPN	beruhigendsten
A1snpg	ANPG	beruhigendsten
A1snpd	ANPD	beruhigendsten
A1snpa	ANPA	beruhigendsten
A2pmsn	AMSN	geachtete
A2pmsg	AMSG	geachteten
A2pmsd	AMSD	geachteten
A2pmsa	AMSA	geachteten
A2pmpn	AMPN	geachteten
A2pmpg	AMPG	geachteten
A2pmpd	AMPD	geachteten
A2pmpa	AMPA	geachteten
A2pfsn	AFSN	geachtete
A2pfsg	AFSG	geachteten
A2pfsd	AFSD	geachteten
A2pfsa	AFSA	geachtete
A2pfpn	AFPN	geachteten
A2pfpg	AFPG	geachteten
A2pfpd	AFPD	geachteten
A2pfpa	AFPA	geachteten
A2pnsn	ANSN	geachtete
A2pnsg	ANSG	geachteten
A2pnsd	ANSD	geachteten
A2pnsa	ANSA	geachtete

A2pnpn	ANPN	geachteten
A2pnpg	ANPG	geachteten
A2pnpd	ANPD	geachteten
A2pnpa	ANPA	geachteten
A2cmsn	AMSN	geachtetere
A2cmsg	AMSG	geachteteren
A2cmsd	AMSD	geachteteren
A2cmsa	AMSA	geachteteren
A2cmpn	AMPN	geachteteren
A2cmpg	AMPG	geachteteren
A2cmpd	AMPD	geachteteren
A2cmpa	AMPA	geachteteren
A2cfsn	AFSN	geachtetere
A2cfsg	AFSG	geachteteren
A2cfsd	AFSD	geachteteren
A2cfsa	AFSA	geachtetere
A2cfpn	AFPN	geachteteren
A2cfpg	AFPG	geachteteren
A2cfpd	AFPD	geachteteren
A2cfpa	AFPA	geachteteren
A2cnsn	ANSN	geachtetere
A2cnsg	ANSG	geachteteren
A2cnsd	ANSD	geachteteren
A2cnsa	ANSA	geachtetere
A2cnpn	ANPN	geachteteren
A2cnpg	ANPG	geachteteren
A2cnpd	ANPD	geachteteren
A2cnpa	ANPA	geachteteren
A2smsn	AMSN	geachtetste
A2smsg	AMSG	geachtetsten
A2smsd	AMSD	geachtetsten
A2smsa	AMSA	geachtetsten
A2smpn	AMPN	geachtetsten
A2smpg	AMPG	geachtetsten
A2smpd	AMPD	geachtetsten
A2smpa	AMPA	geachtetsten
A2sfsn	AFSN	geachtetste

```
A2sfsg    AFSG    geachtetsten
A2sfsd    AFSD    geachtetsten
A2sfsa    AFSA    geachtetste

A2sfpn    AFPN    geachtetsten
A2sfpg    AFPG    geachtetsten
A2sfpd    AFPD    geachtetsten
A2sfpa    AFPA    geachtetsten

A2snsn    ANSN    geachtetste
A2snsg    ANSG    geachtetsten
A2snsd    ANSD    geachtetsten
A2snsa    ANSA    geachtetste

A2snpn    ANPN    geachtetsten
A2snpg    ANPG    geachtetsten
A2snpd    ANPD    geachtetsten
A2snpa    ANPA    geachtetsten

Aqp       A       gut
Aqc       A       besser
A1p       A       beruhigend
A1c       A       beruhigender
A2p       A       geachtet
A2c       A       geachteter

Acp       ACP     zwei
Asp       AP      mein

Aopms     AP.     zweiter
Aopfs     AP.     zweite
Aopns     AP.     zweites
Aop-p     AP.     zweite

Aspms     AP.     meiner
Aspfs     AP.     meine
Aspns     AP.     meines
Asp-p     AP.     meine

Aqs       AS      besten
A1s       AS      beruhigendsten
A2s       AS      geachtetesten

========= ======= =============================================
```

5.2.4 Pronouns (P)

5.2.4.1 Lexicon

Attribute	Value	Example	Code
Type	personal	ich	p
	demonstrat.	dieser	d
	indefinite	kein	i
	interrog.	wer	t
	relative	der	r
	reflexive	sich	x
Person	first	ich	1
	second	du	2
	third	es	3
Gender	masculine	dieser	m
	feminine	diese	f
	neutre	dieses	n
Number	singular	dieser	s
	plural	diese	p
Case	nominative	dieser	n
	genitive	dieses	g
	dative	diesem	d
	accusative	diesen	a
Possessor	-	-	-

Notes

a. Possessive. In German there are no possessives in pronominal use, so we do not need the attribute possessor here.

5.2.4.2 Corpus

Tag	Regular expression	Definition
PP1SN	Pp1-sn	Personal pron., 1st pers. sing., nomin.
PP1SG	Pp1-sg	Personal pron., 1st pers. sing., gen.

P1SD	P[px]1-sd	Personal pron., 1st pers. sing., dat.
P1SA	P[px]1-sa	Personal pron., 1st pers. sing., acc.
PP2SN	Pp2-sn	Personal pron., 2nd pers. sing., nomin.
PP2SG	Pp2-sg	Personal pron., 2nd pers. sing., gen.
P2SD	P[px]2-sd	Personal pron., 2nd pers. sing., dat.
P2SN	P[px]2-sa	Personal pron., 2nd pers. sing., acc.
PP3MSN	Pp3msn	Personal pron., 3rd pers., masc., sing.,nomin.
PP3MSG	Pp3msg	Personal pron., 3rd pers., masc., sing., gen.
PP3MSD	Pp3msd	Personal pron., 3rd pers., masc., sing., dat.
PP3MSA	Pp3msa	Personal pron., 3rd pers., masc., sing., acc.
PP3FSN	Pp3fsn	Personal pron., 3rd pers., fem., sing., nomin.
PP3FSG	Pp3fsg	Personal pron., 3rd pers., fem., sing., gen.
PP3FSD	Pp3fsd	Personal pron., 3rd pers., fem., sing., dat.
PP3FSA	Pp3fsa	Personal pron., 3rd pers., fem., sing., acc.
PP3NSN	Pp3nsn	Personal pron., 3rd pers., neut., sing., nomin.
PP3NSG	Pp3nsg	Personal pron., 3rd pers., neut., sing., gen.
PP3NSD	Pp3nsd	Personal pron., 3rd pers., neut., sing., dat.
PP3NSA	Pp3nsa	Personal pron., 3rd pers., neut., sing., acc.
PP1PN	Pp1-pn	Personal pron., 1st pers. plur., nomin.
PP1PG	Pp1-pg	Personal pron., 1st pers. plur., gen.
P1PD	P[px]1-pd	Personal/Refl. pron., 1st pers. plur., dat.
P1PA	P[px]1-pa	Personal/Refl. pron., 1st pers. plur., acc.
PP2PN	Pp2-pn	Personal pron., 2nd pers. plur., nomin.
PP2PG	Pp2-pg	Personal pron., 2nd pers. plur., gen.
P2PD	P[px]2-pd	Personal/Refl. pron., 2nd pers. plur., dat.
P2PA	P[px]2-pn	Personal/Refl. pron., 2nd pers. plur., acc.
PP3PN	Pp3-pn	Personal pron., 3rd pers. plur., nomin.
PP3PG	Pp3-pg	Personal pron., 3rd pers. plur., gen.
PP3PD	Pp3-pd	Personal pron., 3rd pers. plur., dat.
PP3PA	Pp3-pn	Personal pron., 3rd pers. plur., acc.
PDMSN	Pd-msn	Dem. pronoun, masc., sing., nominative
PDMSG	Pd-msg	Dem. pronoun, masc. sing. genitive
PDMSD	Pd-msd	Dem. pronoun, masc. sing. dative
PDMSA	Pd-msa	Dem. pronoun, masc. sing. accusative
PDFSN	Pd-fsn	Dem. pronoun, fem. sing. nominativ
PDFSG	Pd-fsg	Dem. pronoun, fem. sing. genitive

PDFSD	Pd-fsd	Dem. pronoun, fem. sing. dative
PDFSA	Pd-fsa	Dem. pronoun, fem. sing. accusative
PDNSN	Pd-nsn	Dem. pronoun, neut. sing. nominativ
PDNSG	Pd-nsg	Dem. pronoun, neut. sing. genitive
PDNSD	Pd-nsd	Dem. pronoun, neut. sing. dative
PDNSA	Pd-nsa	Dem. pronoun, neut. sing. accusative
PDPN	Pd--pn	Dem. pronoun, plur. nominative
PDPG	Pd--pg	Dem. pronoun, plur. genitive
PDPD	Pd--pd	Dem. pronoun, plur. dative
PDPA	Pd--pa	Dem. pronoun, plur. accusative
PIMSN	Pi-msn	Indef. pronoun, masc. sing. nominative
PIMSG	Pi-msg	Indef. pronoun, masc. sing. genitive
PIMSD	Pi-msd	Indef. pronoun, masc. sing. dative
PIMSA	Pi-msa	Indef. pronoun, masc. sing. accusative
PIFSN	Pi-fsn	Indef. pronoun, fem. sing. nominativ
PIFSG	Pi-fsg	Indef. pronoun, fem. sing. genitive
PIFSD	Pi-fsd	Indef. pronoun, fem. sing. dative
PIFSA	Pi-fsa	Indef. pronoun, fem. sing. accusative
PINSN	Pi-nsn	Indef. pronoun, neut. sing. nominativ
PINSG	Pi-nsg	Indef. pronoun, neut. sing. genitive
PINSD	Pi-nsd	Indef. pronoun, neut. sing. dative
PINSA	Pi-nsa	Indef. pronoun, neut. sing. accusative
PIPN	Pi--pn	Indef. pronoun, plur. nominative
PIPG	Pi--pg	Indef. pronoun, plur. genitive
PIPD	Pi--pd	Indef. pronoun, plur. dative
PIPA	Pi--pa	Indef. pronoun, plur. accusative
PTN	Pt--n	Interrogative pronoun, nom.
PTG	Pt--g	Interrogative pronoun, gen.
PTD	Pt--d	Interrogative pronoun, dat.
PTA	Pt--a	Interrogative pronoun, acc.
PRMSN	Pr-msn	Rel. pronoun, masc. sing. nominative
PRMSG	Pr-msg	Rel. pronoun, masc. sing. genitive
PRMSD	Pr-msd	Rel. pronoun, masc. sing. dative
PRMSA	Pr-msa	Rel. pronoun, masc. sing. accusative
PRFSN	Pr-fsn	Rel. pronoun, fem. sing. nominativ
PRFSG	Pr-fsg	Rel. pronoun, fem. sing. genitive
PRFSD	Pr-fsd	Rel. pronoun, fem. sing. dative
PRFSA	Pr-fsa	Rel. pronoun, fem. sing. accusative

PRNSN	Pr-nsn	Rel. pronoun, neut. sing. nominativ
PRNSG	Pr-nsg	Rel. pronoun, neut. sing. genitive
PRNSD	Pr-nsd	Rel. pronoun, neut. sing. dative
PRNSA	Pr-nsa	Rel. pronoun, neut. sing. accusative
PRPN	Pr--pn	Rel. pronoun, plur. nominative
PRPG	Pr--pg	Rel. pronoun, plur. genitive
PRPD	Pr--pd	Rel. pronoun, plur. dative
PRPA	Pr--pa	Rel. pronoun, plur. accusative
PX3	Px3-.[da]	Refl. pronoun, 3rd pers.

5.2.4.3 Combinations

Lexique	Corpus	Example
Pp1-sn	PP1SN	ich
Pp1-sg	PP1SG	meiner
Pp1-sd	P1SD	mir
Pp1-sa	P1SA	mich
Px1-sd	P1SD	mir
Px1-sa	P1SA	mich
Pp2-sn	PP2SN	du
Pp2-sg	PP2SG	deiner
Pp2-sd	P2SD	dich
Pp2-sa	P2SA	dir
Px2-sd	P2SD	dich
Px2-sa	P2SA	dir
Pp3msn	PP3MSN	er
Pp3msg	PP3MSG	seiner
Pp3msd	P3SD	ihm
Pp3msa	P3SA	ihn
Pp3fsn	PP3FSN	sie
Pp3fsg	PP3FSG	ihrer
Pp3fsd	P3SD	sie
Pp3fsa	P3SA	sie

Pp3nsn	PP3NSN	es
Pp3nsg	PP3NSG	seiner
Pp3nsd	P3SD	ihm
Pp3nsa	P3SA	es
Pp1-pn	PP1PN	wir
Pp1-pg	PP1PG	unser
Pp1-pd	P1PD	uns
Pp1-pa	P1PA	uns
Px1-pd	P1PD	uns
Px1-pa	P1PA	uns
Pp2-pn	PP2PN	ihr
Pp2-pg	PP2PG	eurer
Pp2-pd	P2PD	euch
Pp2-pn	P2PA	euch
Px2-pd	P2PD	euch
Px2-pn	P2PA	euch
Pp3-pn	PP3PN	sie
Pp3-pg	PP3PG	ihrer
Pp3-pd	PP3PD	ihnen
Pp3-pn	PP3PN	sie
Pd-ms	PDMSN	dieser
Pd-msg	PDMSG	dieses
Pd-msd	PDMSD	diesem
Pd-msa	PDMSA	diesen
Pd-fsn	PDFSN	diese
Pd-fsg	PDFSG	dieser
Pd-fs	PDFSD	dieser
Pd-fsa	PDFSA	diese
Pd-nsn	PDNSN	dieses
Pd-nsg	PDNSG	dieses
Pd-nsd	PDNSD	diesem
Pd-nsa	PDNSA	dieses
Pd--pn	PDPN	diese
Pd--pg	PDPG	dieser
Pd--pd	PDPD	diesen
Pd--pa	PDPA	diese
Pi-msn	PIMSN	keiner
Pi-msg	PIMSG	keines

```
Pi-msd    PIMSD    keinem
Pi-msa    PIMSA    keinen

Pi-fsn    PIFSN     keine
Pi-fsg    PIFSG     keiner
Pi-fsd    PIFSD     keiner
Pi-fsa    PIFSA     keine

Pi-nsn    PINSN     keines
Pi-nsg    PINSG     keines
Pi-nsd    PINSD     keinem
Pi-nsa    PINSA     keines

Pi--pn    PIPN     keine
Pi--pg    PIPG     keiner
Pi--pd    PIPD     keinem
Pi--pa    PIPA     keinen

Pt--n     PTN      wer
Pt--g     PTG      wessen
Pt--d     PTD      wem
Pt--a     PTA      was

Pr-msn    PRMSN    der
Pr-msg    PRMSG    dessen
Pr-msd    PRMSD    dem
Pr-msa    PRMSA    den

Pr-fsn    PRFSN    die
Pr-fsg    PRFSG    deren
Pr-fsd    PRFSD    der
Pr-fsa    PRFSA    die

Pr-nsn    PRNSN    das
Pr-nsg    PRNSG    dessen
Pr-nsd    PRNSD    dem
Pr-nsa    PRNSA    das

Pr--pn    PRPN     die
Pr--pg    PRPG     deren
Pr--pd    PRPD     denen
Pr--pa    PRPA     die

Px3-sa    P3SA     sich
Px3-sd    P3SD     sich

========= ======= ============================================
```

5.2.5. Determiners (D)

5.2.5.1 Lexicon

Attribute	Value	Example	Code
Type	demonstrat.	dieser	d
	indefinite	kein	i
	possessive	mein	s
	interrog.	welche	t
Person	first	mein	1
	second	dein	2
	third	sein	3
Gender	masculine	dieser	m
	feminine	diese	f
	neutre	dieses	n
Number	singular	dieser	s
	plural	diese	p
Case	nominative	dieser	n
	genitive	dieses	g
	dative	diesem	d
	accusative	diesen	a
Possessor	-	-	-

Note:

We included "person" as a feature in the lexical list, but we doubt whether this treatment is useful, at least seen from a German point of view. We prefer to treat the six possessive determiners which are derived from the personal pronouns as different lexemes.

Otherwise, the feature "Number" would have to be specified twice. First, to mark the proper attributes of the basic pronoun (e.g. 1. person plural for "unser") and second to mark the features of agreement with the head of the NP ("unser Leben" vs. "unsere Kinder").

Furthermore, the person information is of semantic character for the possessive determiner, in contrast to the underlying personal pronoun, where the person information is also a feature of agreement with the verb or VP of a sentence.

The attribute 'possessor' is not grammatically relevant.

5.2.5.2 Corpus

Tag	Regular expression	Definition
DDMSN	Dd-msn-	Dem. determiner, masc., sing., nominative
DDMSG	Dd-msg-	Dem. determiner, masc. sing. genitive
DDMSD	Dd-msd-	Dem. determiner, masc. sing. dative
DDMSA	Dd-msa-	Dem. determiner, masc. sing. accusative
DDFSN	Dd-fsn-	Dem. determiner, fem. sing. nominativ
DDFSG	Dd-fsg-	Dem. determiner, fem. sing. genitive
DDFSD	Dd-fsd-	Dem. determiner, fem. sing. dative
DDFSA	Dd-fsa-	Dem. determiner, fem. sing. accusative
DDNSN	Dd-nsn-	Dem. determiner, neut. sing. nominativ
DDNSG	Dd-nsg-	Dem. determiner, neut. sing. genitive
DDNSD	Dd-nsd-	Dem. determiner, neut. sing. dative
DDNSA	Dd-nsa-	Dem. determiner, neut. sing. accusative
DDDN	Dd--pn-	Dem. determiner, plur. nominative
DDDG	Dd--pg-	Dem. determiner, plur. genitive
DDDD	Dd--pd-	Dem. determiner, plur. dative
DDDA	Dd--pa-	Dem. determiner, plur. accusative
DIMSN	Di-msn-	Indef. determiner, masc. sing. nominative
DIMSG	Di-msg-	Indef. determiner, masc. sing. genitive
DIMSD	Di-msd-	Indef. determiner, masc. sing. dative
DIMSA	Di-msa-	Indef. determiner, masc. sing. accusative
DIFSN	Di-fsn-	Indef. determiner, fem. sing. nominativ
DIFSG	Di-fsg-	Indef. determiner, fem. sing. genitive
DIFSD	Di-fsd-	Indef. determiner, fem. sing. dative
DIFSA	Di-fsa-	Indef. determiner, fem. sing. accusative
DINSN	Di-nsn-	Indef. determiner, neut. sing. nominativ
DINSG	Di-nsg-	Indef. determiner, neut. sing. genitive
DINSD	Di-nsd-	Indef. determiner, neut. sing. dative
DINSA	Di-nsa-	Indef. determiner, neut. sing. accusative
DIPN	Di--pn-	Indef. determiner, plur. nominative

DIPG	Di--pg-	Indef. determiner, plur. genitive
DIPD	Di--pd-	Indef. determiner, plur. dative
DIPA	Di--pa-	Indef. determiner, plur. accusative
DPMSN	Dp-msn-	Poss. determiner, masc. sing. nominative
DPMSG	Dp-msg-	Poss. determiner, masc. sing. genitive
DPMSD	Dp-msd-	Poss. determiner, masc. sing. dative
DPMSA	Dp-msa-	Poss. determiner, masc. sing. accusative
DPFSN	Dp-fsn-	Poss. determiner, fem. sing. nominativ
DPFSG	Dp-fsg-	Poss. determiner, fem. sing. genitive
DPFSD	Dp-fsd-	Poss. determiner, fem. sing. dative
DPFSA	Dp-fsa-	Poss. determiner, fem. sing. accusative
DPNSN	Dp-nsn-	Poss. determiner, neut. sing. nominativ
DPNSG	Dp-nsg-	Poss. determiner, neut. sing. genitive
DPNSD	Dp-nsd-	Poss. determiner, neut. sing. dative
DPNSA	Dp-nsa-	Poss. determiner, neut. sing. accusative
DPDN	Dp--pn-	Poss. determiner, plur. nominative
DPDG	Dp--pg-	Poss. determiner, plur. genitive
DPDD	Dp--pd-	Poss. determiner, plur. dative
DPDA	Dp--pa-	Poss. determiner, plur. accusative
DTMSN	Dt-msn-	Interrog. determiner, masc. sing. nominative
DTMSG	Dt-msg-	Interrog. determiner, masc. sing. genitive
DTMSD	Dt-msd-	Interrog. determiner, masc. sing. dative
DTMSA	Dt-msa-	Interrog. determiner, masc. sing. accusative
DTFSN	Dt-fsn-	Interrog. determiner, fem. sing. nominativ
DTFSG	Dt-fsg-	Interrog. determiner, fem. sing. genitive
DTFSD	Dt-fsd-	Interrog. determiner, fem. sing. dative
DTFSA	Dt-fsa-	Interrog. determiner, fem. sing. accusative
DTNSN	Dt-nsn-	Interrog. determiner, neut. sing. nominativ
DTNSG	Dt-nsg-	Interrog. determiner, neut. sing. genitive
DTNSD	Dt-nsd-	Interrog. determiner, neut. sing. dative
DTNSA	Dt-nsa-	Interrog. determiner, neut. sing. accusative
DTDN	Dt--pn-	Interrog. determiner, plur. nominative
DTDG	Dt--pg-	Interrog. determiner, plur. genitive
DTDD	Dt--pd-	Interrog. determiner, plur. dative
DTDA	Dt--pa-	Interrog. determiner, plur. accusative

5.2.5.3 Combinations

```
======= ======= =============================================
Lexique Corpus  Example
======= ======= =============================================
Dd-msn- DDMSN    dieser
Dd-msg- DDMSG    dieses
Dd-msd- DDMSD    diesem
Dd-msa- DDMSA    diesen

Dd-fsn- DDFSN    diese
Dd-fsg- DDFSG    dieser
Dd-fs-  DDFSD    dieser
Dd-fsa- DDFSA    diese

Dd-nsn- DDNSN    dieses
Dd-nsg- DDNSG    dieses
Dd-nsd- DDNSD    diesem
Dd-nsa- DDNSA    dieses

Dd--pn- DDPN     diese
Dd--pg- DDPG     dieser
Dd--pd- DDPD     diesen
Dd--pa- DDPA     diese

Di-msn- DIMSN    keiner
Di-msg- DIMSG    keines
Di-msd- DIMSD    keinem
Di-msa- DIMSA    keinen

Di-fsn- DIFSN    keine
Di-fsg- DIFSG    keiner
Di-fsd- DIFSD    keiner
Di-fsa- DIFSA    keine

Di-nsn- DINSN    keines
Di-nsg- DINSG    keines
Di-nsd- DINSD    keinem
Di-nsa- DINSA    keines

Di--pn- DIPN     keine
Di--pg- DIPG     keiner
Di--pd- DIPD     keinen
Di--pa- DIPA     keine

Dp-msn- DPMSN    meiner
Dp-msg- DPMSG    meines
Dp-msd- DPMSD    meinem
Dp-msa- DPMSA    meinen
```

```
Dp-fsn- DPFSN    meine
Dp-fsg- DPFSG    meiner
Dp-fsd- DPFSD    meiner
Dp-fsa- DPFSA    meine

Dp-nsn- DPNSN    meines
Dp-nsg- DPNSG    meines
Dp-nsd- DPNSD    meinem
Dp-nsa- DPNSA    meines

Dp--pn- DPPN     meine
Dp--pg- DPPG     meiner
Dp--pd- DPPD     meinen
Dp--pa- DPPA     meine


Dt-msn- DTMSN    welcher
Dt-msg- DTMSG    welches
Dt-msd- DTMSD    welchem
Dt-msa- DTMSA    welchen

Dt-fsn- DTFSN    welche
Dt-fsg- DTFSG    welcher
Dt-fsd- DTFSD    welcher
Dt-fsa- DTFSA    welche

Dt-nsn- DTNSN    welches
Dt-nsg- DTNSG    welches
Dt-nsd- DTNSD    welchem
Dt-nsa- DTNSA    welches

Dt--pn- DTPN     welche
Dt--pg- DTPG     welcher
Dt--pd- DTPD     welchen
Dt--pa- DTPA     welche
```

Note:

We deliberately treat the Articles as an extra class, but related to the Determiners. The syntactic behaviour of Articles and Determiners in NPs is different, as can be seen by the adjective in the following examples:

```
      Ein kleines Haus
      Welches kleine Haus ?
```

This difference supports our attitude towards separating articles and determiners generally.

5.2.6 Articles (T)

5.2.6.1 Lexicon

Attribute	Value	Example	Code
Type	definite	der	d
	indefinite	ein	i
Gender	masculine	der	m
	feminine	die	f
	neuter	das	n
Number	singular	ein	s
	plural	die	p
Case	nominative	der	n
	genitive	dessen	g
	dative	dem	d
	accusative	den	a

5.2.6.2 Corpus

Tag	Regular expression	Definition
TD	Td..-	definite article
TI	Ti.s-	indefinite article

5.2.6.3 Combinations

Lexique	Corpus	Example

Tdmsn	TD	der
Tdmsg	TD	des
Tdmsd	TD	dem
Tdmsa	TD	den
Tdfsn	TD	die
Tdfsg	TD	der
Tdfsd	TD	der
Tdfsa	TD	die
Tdnsn	TD	das
Tdnsn	TD	des
Tdnsd	TD	dem
Tdnsa	TD	das
Tdpn	TD	die
Tdpg	TD	der
Tdpd	TD	den
Tdpa	TD	die
Timsn	TI	einer
Timsg	TI	eines
Timsd	TI	einem
Timsa	TI	einen
Tifsn	TI	eine
Tifsn	TI	einer
Tifsd	TI	einer
Tifsa	TI	eine
Tinpn	TI	ein
Tinpg	TI	eines
Tinpd	TI	einem
Tinpa	TI	ein

5.2.7 Adverbs (R)

5.2.7.1 Lexicon

Attribute	Value	Example	Code	
Type	general	frischweg	g	
	degree	sogar	d	this value has been added

	interrog	worum	i	this value has been added
	conjunction	mithin	c	this value has been added
	modal	scheinbar	m	this value has been added
	pronom	so	p	this value has been added
	temporal	heute	t	this value has been added
	place	hier	l	this value has been added
Degree	positive	hoch	p	
	comparative	h"oher	c	
	superlative	h"ochst	s	

5.2.7.2 Corpus

Tag	Regular expression	Definition
RG	R[gdcmtl].	General adverb
RP	Rp-	pronominal adverb
RI	Ri-	interrogative adverb

5.2.7.3 Combinations

Lexique	Corpus	Example
Rgp	RG	frischweg
Rp	RP	so
Rgs	RG	h"ochst
Rgc	RG	h"oher
Ri	RI	worum

5.2.8 Adpositions (S)

5.2.8.1 Lexicon

Attribute	Value	Example	Code
Type	pre	an	p

	post	wegen	t	this value has been added
	circum	von - an	c	this value has been added
	part1	von	a	this value has been added
	part2	an	z	this value has been added
Formation	clitic	ans	c	
	simple	an	s	

In German, most prepositions precede the NP. However, there is a good deal of them which follow the NP ("entlang") or enclose it ("von" NP "an"). This behaviour is unpredictable and must therefore be marked lexically. This is done by increasing the values of the "Type" attribute. This extension can be considered as language specific, as long as there is no evidence from other languages which supports this distinction. For practical reasons of text segmentation, the two parts of a circumposition have to be distinguished by different tags.

The attribute "formation" allows us to deal with clitic contraction of preposition and article of the following NP ("zum", "ans")

5.2.8.2 Corpus

Tag	Regular expression	Definition
SPS	Sps	pre-position, simple
STS	Sts	post-position, simple
SPC	Spc	pre-position, clitic
SC	Sas	circumposition, partI, simple
SC	Szs	circumposition, partII, simple

5.2.8.3 Combinations

Lexique	Corpus	Example
Sps	SPS	an
Sts	STS	wegen
Spc	SPC	ans
Sas	SC	von
Szs	SC	an

========= ======= ==

5.2.9. Conjunctions (C)

5.2.9.1 Lexicon

Attribute	Value	Example	Code	
Type	coordinat.	oder	c	
	subordinat.	als	s	
	compar	als	v	this value has been added
	infinitive	um	i	this value has been added
	part1	entweder	a	this value has been added
	part2	oder	z	this value has been added

Comparative conjunctions are restricted to constitutents as arguments. Subordinate conjunctions introducing an infinite clause lead to an infinite verb form in sentence final position. The infinitive particle may be incorporated, as in "wegzugehen". These morphosyntactic features lead to an extension of the feature-Type values, which may be considered as language specific.

There are also a few complex conjunctions which appear at the beginning of both phrases which are conjoined by it ("entweder - oder"). This feature has to be marked lexically, again by increasing the values of the "Type" attribute. For practical reason of text segmentation, the two parts of a complex conjunction have to be distinguished by different tags.

5.2.9.2 Corpus

Tag	Regular expression	Definition
CC	Cc	Cooordinative cunjunction
CS	Cs	Subordinative cunjunction
CI	Ci	Subord. conjunctions introd. an infinit. clause
CV	Cv	Comparative Conjunction
CA	Ca	Conjunction Part I
CZ	Cz	Conjunction Part II

5.2.9.3 Combinations

Lexique	Corpus	Example
Cc	CC	aber
Cs	CS	als
Ci	CI	um
Cv	CV	als

5.2.10 Numerals (M)

We do not support the class "Numerals". Possible elements of this class will be treated as nouns, adjectives, or adverbs.

5.2.11 Interjection (I)

5.2.11.1 Lexicon

Tag	Example
I	oh

5.2.11.2 Corpus

Tag	Regular expression	Definition
I	I	Interjection

5.2.11.3 Combinations

=========== ======= ==

Lexique	Corpus	Example
I	I	oh

5.2.12 Add on classes

5.2.12.1 Particle (Q)

5.2.12.1.1 Lexicon

Attribute	Value	Example	Code
Type	infinitive	zu	i
	superlative	am	s
	verbal pref.	hinzu	v

5.2.12.1.2 Corpus

Tag	Regular expression	Definition
QS	Qs	superlative particle
QI	Qi	infinitive particle
QV	Qv	verbal prefix

Superlative particles precede the superlative form of adjectives, if used predicatively (er ist am gr"o"sten), and adverbs (applies also to Dutch).

Infinitive particles are found in each of the Germanic languages to be described (Danish, Dutch, English, German). This should be accounted for somewhere in the generic set of classes.

Verbal prefixes are a particular German (and Dutch) phenomenon and can be considered to be language specific.

5.2.12.1.3 Combinations

Lexique	Corpus	Example
Qi	QI	zu
Qs	QS	am
Qv	QV	hinzu

5.2.12.2 Punctuation (F)

5.2.12.2.1 Corpus

Tag	Regular expression	Definition
FE	Fe	sentence final
FI	Fi	sentence internal
FA	Fa	Quot mark / parenthesis initial
FZ	Fz	Quot mark / parenthesis final
FB	Fb	Hyphen, underscore, dash

The use of Tags for punctuation sign is obvious for the stochastic modelling of utterances. However, it is not clear whether they should be treated lexically. Therefore, we can describe our corpus tags, but not a lexical treatment of these units.

5.2.12.3 Abbreviations (Y)

5.2.12.3.1 Lexicon

Tag	Example
Y	bzw.

5.2.12.3.2 Corpus

Lexical abbreviations do not have a particular corresponding Corpus Tag, but are tagged according to the morphosyntactic behaviour of the form written out in full.

5.2.12.3.3 Combinations

Lexique	Corpus	Example
Y	RG	bzw.
Y	NCFSN	AG

5.2.12.4 Others (X)

Other particular corpus tags can correspond to the common class of Residual.

5.2.12.4.1 Corpus

Tag	Regular expression	Definition
SYM	X	Symbols (%, ' etc.)
EQ	X	Formulae (5x + 3y)

5.3 Application to Spanish

The application of the MULTEXT morphosyntactic enconding for lexical descriptions and corpus tags to Spanish has been performed by the Spanish group (Bel and Aguilar 1994), and has been revised during phase B.

The proposed set of TAGS is not definitive. As repeteadly mentioned, we understand that this set will have to be refined depending on the results of application. These tags are the result of a comparison of two tagset sources:

a. information supplied by the tool SAC (a corpus analysis tool which includes a PoS rule based tagger and a lemmatizer) in form of attribute-value pairs.

b. tagsets proposed by the CRATER project.

Therefore these tags have to be taken as a starting point for refinement.

```
5.3.1  Nouns (N)
----------------


5.3.1.1  Lexicon

============ =========== =========== ====
Attribute    Value       Example     Code
============ =========== =========== ====
Type         common      libro       c
             proper      Juan        p
------------ ----------- ----------- ----
Gender       masculine   hombre      m
             feminine    mujer       f
------------ ----------- ----------- ----
Number       singular    hombre      s
             plural      mujeres     p
------------ ----------- ----------- ----
Case         ///         ///         -
============ =========== =========== ====


5.3.1.2  Corpus TAGS: comments

Crater  tags for nouns  also  include  semantic  information  such  as
"A"(anthroponymous)  or  "T"(toponymous)  for proper nouns,  aswell as
"LOC"(ative),  "MEA(sure)", etc.  for common nouns. We will drop these
values in order to get  our  proposal which intends to be like the one
recommended in EAG-L1.
```

5.3.1.3 Combinations

Lexique	Corpus	Example
Ncms-	NCMS	hombre
Ncmp-	NCMP	hombres
Ncfs-	NCFS	mujer
Ncfp-	NCFP	mujeres
Npms-	NPMS	Juan
Npmp-	NPMP	Paris
Npfs-	NPFS	Ana
Npfp-	NPFP	Pirineos

5.3.1.4 Conversion tables

Reg.exp	TAG
Ncmp-	NCMP
Ncms-	NCMS
Ncfp-	NCFP
Ncfs-	NCFS
Nc.p-	NCP
Ncf.-	NCF
Ncm.-	NCM
Nc.s-	NCS
Np	NP

5.3.2 Verbs (V)

5.3.2.1. Lexicon

Attribute	Value	Example	Code
Status	main	comer	m
	auxiliar	haber	a
	modal	poder	o

Mood	indicative	viene	i
	subjunctive	venga	s
	imperative	ven	m
	conditional	vendri'a	c
	infinitive	venir	n
	participle	venido	p
	gerund	viniendo	g
Tense	present	vengo	p
	imperfect	veni'as	i
	future	vendre'	f
	past	vino	s
Person	first	soy	1
	second	eres	2
	third	es	3
Number	singular	viene	s
	plural	venimos	p
Gender	masculine	cantado	m
	feminine	cantada	f
Clitic	both	darselo	t
	accusative	darlo	a
	dative	darle	d

5.3.2.2 Corpus TAGS: comments

a. CRATER tags as well as our inhouse tags classify verb types into: main/ser/estar/haber/modal. "Ser", "estar", "haber" (normally considered auxiliaries) are more informative with respect to the different constructions such as: perfect tenses, active/pasive distinctions and the disambiguation between adjectives and past participles.

Following French suggestion we could pass this information into a language specific attribute as Lexical Class.

Attribute	Value	Example	Code

Lex. class	ser	ser	s	obviously
	estar	estar	e	language-specific
	haber	haber	h	

5.3.2.3 Combinations

Lexique	Corpus	Example
Van----t	VANT	haberselo
Van----d	VAND	haberle
Van----a	VANA	haberlo
Van----	VAN	haber
Vap--pf	VAPPF	habidas
Vap--sf	VAPSF	habida
Vap--pm	VAPPM	habidos
Vap--sm	VAPSM	habido
Vag----t	VAGT	habiindoselo
Vag----d	VAGD	habiindole
Vag----a	VAGA	habiindolo
Vag----	VAG	habiendo

Example	Lexique	lemma	Corpus
llamarmeles	Vmn----t	llamar	VMNT
llamarmele	Vmn----t	llamar	VMNT
llamarmelas	Vmn----t	llamar	VMNT
llamarmela	Vmn----t	llamar	VMNT
llamarmelos	Vmn----t	llamar	VMNT
llamarmelo	Vmn----t	llamar	VMNT
llamarme	Vmn----d	llamar	VMND
llamarteles	Vmn----t	llamar	VMNT
llamartele	Vmn----t	llamar	VMNT
llamartelas	Vmn----t	llamar	VMNT
llamartela	Vmn----t	llamar	VMNT
llamartelos	Vmn----t	llamar	VMNT
llamartelo	Vmn----t	llamar	VMNT
llamarte	Vmn----d	llamar	VMND
llamarseles	Vmn----t	llamar	VMNT
llamarsele	Vmn----t	llamar	VMNT
llamarselas	Vmn----t	llamar	VMNT

llamarsela	Vmn----t	llamar	VMNT
llamarselos	Vmn----t	llamar	VMNT
llamarselo	Vmn----t	llamar	VMNT
llamarse	Vmn----d	llamar	VMND
llamarnosles	Vmn----t	llamar	VMNT
llamarnosle	Vmn----t	llamar	VMNT
llamarnoslas	Vmn----t	llamar	VMNT
llamarnosla	Vmn----t	llamar	VMNT
llamarnoslos	Vmn----t	llamar	VMNT
llamarnoslo	Vmn----t	llamar	VMNT
llamarnos	Vmn----d	llamar	VMND
llamarosles	Vmn----t	llamar	VMNT
llamarosle	Vmn----t	llamar	VMNT
llamaroslas	Vmn----t	llamar	VMNT
llamarosla	Vmn----t	llamar	VMNT
llamaroslos	Vmn----t	llamar	VMNT
llamaroslo	Vmn----t	llamar	VMNT
llamaros	Vmn----d	llamar	VMND
llamarme	Vmn----a	llamar	VMNA
llamarte	Vmn----a	llamar	VMNA
llamarles	Vmn----a	llamar	VMNA
llamarle	Vmn----a	llamar	VMNA
llamarlas	Vmn----a	llamar	VMNA
llamarla	Vmn----a	llamar	VMNA
llamarlos	Vmn----a	llamar	VMNA
llamarlo	Vmn----a	llamar	VMNA
llamarnos	Vmn----a	llamar	VMNA
llamaros	Vmn----a	llamar	VMNA
llamar	Vmn-----	llamar	VMN
llamadas	Vmp--pf-	llamar	VMPPF
llamada	Vmp--sf-	llamar	VMPSF
llamados	Vmp--pm-	llamar	VMPPM
llamado	Vmp--sm-	llamar	VMPSM
llamandomeles	Vmg----t	llamar	VMGT
llamandomele	Vmg----t	llamar	VMGT
llamandomelas	Vmg----t	llamar	VMGT
llamandomela	Vmg----t	llamar	VMGT
llamandomelos	Vmg----t	llamar	VMGT
llamandomelo	Vmg----t	llamar	VMGT
llamandome	Vmg----d	llamar	VMGD
llamandoteles	Vmg----t	llamar	VMGT
llamandotele	Vmg----t	llamar	VMGT
llamandotelas	Vmg----t	llamar	VMGT
llamandotela	Vmg----t	llamar	VMGT
llamandotelos	Vmg----t	llamar	VMGT
llamandotelo	Vmg----t	llamar	VMGT
llamandote	Vmg----d	llamar	VMGD

llamandoseles	Vmg----t	llamar	VMGT
llamandosele	Vmg----t	llamar	VMGT
llamandoselas	Vmg----t	llamar	VMGT
llamandosela	Vmg----t	llamar	VMGT
llamandoselos	Vmg----t	llamar	VMGT
llamandoselo	Vmg----t	llamar	VMGT
llamandose	Vmg----d	llamar	VMGD
llamandonosles	Vmg----t	llamar	VMGT
llamandonosle	Vmg----t	llamar	VMGT
llamandonoslas	Vmg----t	llamar	VMGT
llamandonosla	Vmg----t	llamar	VMGT
llamandonoslos	Vmg----t	llamar	VMGT
llamandonoslo	Vmg----t	llamar	VMGT
llamandonos	Vmg----d	llamar	VMGD
llamandonos	Vmg----d	llamar	VMGD
llamandoosles	Vmg----t	llamar	VMGT
llamandoosle	Vmg----t	llamar	VMGT
llamandooslas	Vmg----t	llamar	VMGT
llamandoosla	Vmg----t	llamar	VMGT
llamandooslos	Vmg----t	llamar	VMGT
llamandooslo	Vmg----t	llamar	VMGT
llamandoos	Vmg----d	llamar	VMGD
llamandome	Vmg----a	llamar	VMGA
llamandote	Vmg----a	llamar	VMGA
llamandoles	Vmg----a	llamar	VMGA
llamandole	Vmg----a	llamar	VMGA
llamandolas	Vmg----a	llamar	VMGA
llamandola	Vmg----a	llamar	VMGA
llamandolos	Vmg----a	llamar	VMGA
llamandolo	Vmg----a	llamar	VMGA
llamandonos	Vmg----a	llamar	VMGA
llamandonos	Vmg----a	llamar	VMGA
llamandoos	Vmg----a	llamar	VMGA
llamando	Vmg-----	llamar	VMG
llamo	Vmip1s-	llamar	VMIP1S
llamas	Vmip2s-	llamar	VMIP2S
llama	Vmip3s-	llamar	VMIP3S
llamais	Vmip2p-	llamar	VMIP2P
llaman	Vmip3p-	llamar	VMIP3P
llamamos	Vmip1p-	llamar	VMIP1P
llame	Vmsp[13]s-	llamar	VMSPS
llames	Vmsp2s-	llamar	VMSP2S
llamemos	Vmsp1p-	llamar	VMSP1P
llamiis	Vmsp2p-	llamar	VMSP2P
llamen	Vmsp3p-	llamar	VMSP3P
llami	Vmif1s-	llamar	VMIF1S
llamaste	Vmif2s-	llamar	VMIF2S

llams	Vmif3s-	llamar	VMIF3S
llamasteis	Vmif2p-	llamar	VMIF2P
llamaron	Vmif3p-	llamar	VMIF3P
llamamos	Vmif1p-	llamar	VMIF1P
llamaba	Vmii[13]s-	llamar	VMIIS
llamabas	Vmii2s-	llamar	VMII2S
llamabamos	Vmii1p-	llamar	VMII1P
llamabais	Vmii2p-	llamar	VMII2P
llamaban	Vmii3p-	llamar	VMII3P
llamara	Vmsi[13]s-	llamar	VMSIS
llamaras	Vmsi2s-	llamar	VMSI2S
llamaramos	Vmsi1p-	llamar	VMSI1P
llamarais	Vmsi2p-	llamar	VMSI2P
llamaran	Vmsi3p-	llamar	VMSI3P
llamase	Vmsi[13]s-	llamar	VMSIS
llamases	Vmsi2s-	llamar	VMSI2S
llamasemos	Vmsi1p-	llamar	VMSI1P
llamaseis	Vmsi2p-	llamar	VMSI2P
llamasen	Vmsi3p-	llamar	VMSI3P
llamari	Vmis1s-	llamar	VMIS1S
llamaras	Vmis2s-	llamar	VMIS2S
llamara	Vmis3s-	llamar	VMIS3S
llamaremos	Vmis1p-	llamar	VMIS1P
llamariis	Vmis2p-	llamar	VMIS2P
llamaran	Vmis3p-	llamar	VMIS3P
llamarma	Vmc-[13]s-	llamar	VMCS
llamarmas	Vmc-2s-	llamar	VMC2S
llamarmamos	Vmc-1p-	llamar	VMC1P
llamarmais	Vmc-2p-	llamar	VMC2P
llamarman	Vmc-3p-	llamar	VMC3P

5.3.2.4. Conversion tables

Reg.expr.	TAG
Van----t	VANT
Van----d	VAND
Van----a	VANA
Van----	VAN
Vap--pf	VAPPF
Vap--sf	VAPSF
Vap--pm	VAPPM
Vap--sm	VAPSM
Vag----t	VAGT
Vag----d	VAGD
Vag----a	VAGA

```
Vag----     VAG
Vaip1s-     VAIP1S
Vaip2s-     VAIP2S
Vaip3s-     VAIP3S
Vaip2p-     VAIP2P
Vaip3p-     VAIP3P
Vaip1p-     VAIP1P
Vasp[13]s-  VASPS
Vasp2s-     VASP2S
Vasp1p-     VASP1P
Vasp2p-     VASP2P
Vasp3p-     VASP3P
Vais1s-     VAIS1S
Vais2s-     VAIS2S
Vais3s-     VAIS3S
Vais2p-     VAIS2P
Vais3p-     VAIS3P
Vais1p-     VAIS1P
Vaii[13]s-  VAIIS
Vaii2s-     VAII2S
Vaii1p-     VAII1P
Vaii2p-     VAII2P
Vaii3p-     VAII3P
Vasi[13]s-  VASIS
Vasi2s-     VASI2S
Vasi1p-     VASI1P
Vasi2p-     VASI2P
Vasi3p-     VASI3P
Vasi[13]s-  VASIS
Vasi2s-     VASI2S
Vasi1p-     VASI1P
Vasi2p-     VASI2P
Vasi3p-     VASI3P
Vaif1s-     VAIF1S
Vaif2s-     VAIF2S
Vaif3s-     VAIF3S
Vaif1p-     VAIF1P
Vaif2p-     VAIF2P
Vaif3p-     VAIF3P
Vac[13]s-   VACS
Vac2s-      VAC2S
Vac1p-      VAC1P
Vac2p-      VAC2P
Vac3p-      VAC3P


Von----t    VONT
```

Von----d	VOND
Von----a	VONA
Von----	VON
Vop--pf	VOPPF
Vop--sf	VOPSF
Vop--pm	VOPPM
Vop--sm	VOPSM
Vog----t	VOGT
Vog----d	VOGD
Vog----a	VOGA
Vog----	VOG
Voip1s-	VOIP1S
Voip2s-	VOIP2S
Voip3s-	VOIP3S
Voip2p-	VOIP2P
Voip3p-	VOIP3P
Voip1p-	VOIP1P
Vosp[13]s-	VOSPS
Vosp2s-	VOSP2S
Vosp1p-	VOSP1P
Vosp2p-	VOSP2P
Vosp3p-	VOSP3P
Vois1s-	VOIS1S
Vois2s-	VOIS2S
Vois3s-	VOIS3S
Vois2p-	VOIS2P
Vois3p-	VOIS3P
Vois1p-	VOIS1P
Voii[13]s-	VOIIS
Voii2s-	VOII2S
Voii1p-	VOII1P
Voii2p-	VOII2P
Voii3p-	VOII3P
Vosi[13]s-	VOSIS
Vosi2s-	VOSI2S
Vosi1p-	VOSI1P
Vosi2p-	VOSI2P
Vosi3p-	VOSI3P
Vosi[13]s-	VOSIS
Vosi2s-	VOSI2S
Vosi1p-	VOSI1P
Vosi2p-	VOSI2P
Vosi3p-	VOSI3P
Voif1s-	VOIF1S
Voif2s-	VOIF2S
Voif3s-	VOIF3S
Voif1p-	VOIF1P

Voif2p-	VOIF2P
Voif3p-	VOIF3P
Voc[13]s-	VOCS
Voc2s-	VOC2P
Voc1p-	VOC1P
Voc2p-	VOC2P
Voc3p-	VOC3P

Vmn----t	VMNT
Vmn----d	VMND
Vmn----a	VMNA
Vmn----	VMN
Vmp--pf	VMPPF
Vmp--sf	VMPSF
Vmp--pm	VMPPM
Vmp--sm	VMPSM
Vmg----t	VMGT
Vmg----d	VMGD
Vmg----a	VMGA
Vmg----	VMG
Vmip1s-	VMIP1S
Vmip2s-	VMIP2S
Vmip3s-	VMIP3S
Vmip2p-	VMIP2P
Vmip3p-	VMIP3P
Vmip1p-	VMIP1P
Vmsp[13]s-	VMSPS
Vmsp2s-	VMSP2S
Vmsp1p-	VMSP1P
Vmsp2p-	VMSP2P
Vmsp3p-	VMSP3P
Vmis1s-	VMIS1S
Vmis2s-	VMIS2S
Vmis3s-	VMIS3S
Vmis2p-	VMIS2P
Vmis3p-	VMIS3P
Vmis1p-	VMIS1P
Vmii[13]s-	VMIIS
Vmii2s-	VMII2S
Vmii1p-	VMII1P
Vmii2p-	VMII2P
Vmii3p-	VMII3P
Vmsi[13]s-	VMSIS
Vmsi2s-	VMSI2S
Vmsi1p-	VMSI1P
Vmsi2p-	VMSI2P

Vmsi3p-	VMSI3P
Vmsi[13]s-	VMSIS
Vmsi2s-	VMSI2S
Vmsi1p-	VMSI1P
Vmsi2p-	VMSI2P
Vmsi3p-	VMSI3P
Vmif1s-	VMIF1S
Vmif2s-	VMIF2S
Vmif3s-	VMIF3S
Vmif1p-	VMIF1P
Vmif2p-	VMIF2P
Vmif3p-	VMIF3P
Vmc[13]s-	VMCS
Vmc2s-	VMC2S
Vmc1p-	VMC1P
Vmc2p-	VMC2P
Vmc3p-	VMC3P

5.3.3 Adjectives (A)

5.3.3.1 Lexicon

Attribute	Value	Example	Code
Type	qualificat.	bueno	f
	possessive	vuestro	s
Degree	positive	bueno	p
	comparative	mejor	c
	superlative	bueni'simo	s
Gender	masculine	bueno	m
	feminine	buena	f
Number	singular	bueno	s
	plural	buenas	p
Case	///	///	-

Comments.

Possesive adjectives codification is still being considered.

5.3.3.2 Combinations

Lexique	Corpus	Example
Afpms-	AMS	bueno
Afpmp-	AMP	buenos
Afpfs-	AFS	buena
Afpfp-	AFP	buenas
Afc.s-	AS	(el/la) mejor
Afc.p-	AP	(los/las) mejores
Afsms-	AMS	interesanti'simo
Afsmp-	AMP	interesanti'simos
Afsfs-	AFS	interesanti'sima
Afsfp-	AFP	interesanti'simas

5.3.3.3. Conversion tables

Reg.Expr.	Corpus
Afp.p-	APP
Afp.s-	APS
Afpfp-	APFP
Afpfs-	APFS
Afpmp-	APMP
Afpms-	APMS
Afsfp-	ASFP
Afsfs-	ASFS
Afsmp-	ASMP
Afsms-	ASMS
Afc.p-	ACP
Afc.s-	ACS

5.3.4 Pronouns (P)

5.3.4.1 Lexicon

Attribute	Value	Example	Code
Type	personal	yo	p
	demonstrat.	este	d
	indefinite	alguno	i
	possessive	(el) tuyo	s
	interrog.	que'	t
	relative	que	r
	reflexive	se	x
Person	first	yo	1
	second	tu'	2
	third	e'l	3
Gender	masculine	esto, el	m
	feminine	esta, ella	f
Number	singular	alguno	s
	plural	algunos	p
Case	nominative	el	n
	dative	le	d
	accusative	lo	a
	oblique	mi',conmigo	o
Possessor	singular	mi'o	s
	plural	nuestro	p

5.3.4.2. Corpus TAGS: comments

a. CRATER TAGs make a distinction when proximal or remote deixis exists. We do not consider these values.

b. "se", which can be both dative and accusative depending on the existence of an 'external' accusative:

    Ella se lava (She washes herself)
    Ella se lava las manos (She washes her hands)

There is also an oblique reflexive pronoun "si'".

TAGs would be:

Px3..-a	PX3SA	se
Px3..-d	PX3SO	se
Px3.s-o	PX3SO	si’

c. Relative pronouns are a good example of rearranging of linguistic information when comparing with French. For example, due to in Spanish there exists a relative pronoun which is also a possessive one (cuyo, cuya, cuyos, cuyas, eng. "whose") possessor values have been marked.

5.3.4.3 Combinations

Example	Lexique	lemma	Corpus
yo	Pp1-sn-	yo	PP1SN
tu’	Pp2-sn-	tu’	PP2SN
ustedes	Pp3-p[no]-	usted	PP3PN
usted	Pp3-s[no]-	usted	PP3SN
il	Pp3ms[no]-	il	PP3MS
ellas	Pp3fp[no]-	il	PP3FP
ella	Pp3fs[no]-	il	PP3FS
ellos	Pp3mp[no]-	il	PP3MP
ello	Pp3ms[no]-	il	PP3MS
nosotras	Pp1fp[no]-	nosotros	PP1FP
nosotros	Pp1mp[no]-	nosotros	PP1MP
vosotras	Pp2fp[no]-	vosotros	PP2FP
vosotros	Pp2mp[no]-	vosotros	PP2MP
conmigo	Pp1-so-	conmigo	PP1SO
mi’	Pp1-so-	mi’	PP1SO
contigo	Pp2-so-	contigo	PP2SO
ti	Pp2-so-	ti’	PP2SO
si’	Pp3-so-	si’	PP3SO
mismas	Px.fpo-	mismo	PXFPO
misma	Px.fso-	mismo	PXFSO
mismos	Px.mpo-	mismo	PXMPO
mismo	Px.mso-	mismo	PXMSO
me	P[px]1.s[ad]-	me	P1S
te	P[px]2.s[ad]-	te	P2S
se	P[pxl]3..[ad]-se		P3
les	Pp3.pd-	le	PP3PD
le	Pp3.	sd-le	PP3SD
las	Pp3fpa-	lo	PP3FPA
la	Pp3fsa-	lo	PP3FSA
los	Pp3mpa-	lo	PP3MPA

```
lo            Pp3msa-       lo             PP3MSA
nos           P[pxl]1.p[ad]-nos            P1P
os            P[pxl]2-p[ad]-os             P2P

istas         Pd-fp--       iste           PDFP
ista          Pd-fs--       iste           PDFS
istos         Pd-mp--       iste           PDMP
isto          Pd-ms--       iste           PDMS
iste          Pd-ms--       iste           PDMS
estas         Pd-fp--       este           PDFP
esta          Pd-fs--       este           PDFS
estos         Pd-mp--       este           PDMP
esto          Pd-ms--       este           PDMS
este          Pd-ms--       este           PDMS

isas          Pd-fp--       ise            PDFP
isa           Pd-fs--       ise            PDFS
isos          Pd-mp--       ise            PDMP
iso           Pd-ms--       ise            PDMS
ise           Pd-ms--       ise            PDMS
esas          Pd-fp--       ese            PDFP
esa           Pd-fs--       ese            PDFS
esos          Pd-mp--       ese            PDMP
eso           Pd-ms--       ese            PDMS
ese           Pd-ms--       ese            PDMS

aquillas      Pd-fp--       aquil          PDFP
aquilla       Pd-fs--       aquil          PDFS
aquillos      Pd-mp--       aquil          PDMP
aquillo       Pd-ms--       aquil          PDMS
aquellas      Pd-fp--       aquel          PDFP
aquella       Pd-fs--       aquel          PDFS
aquellos      Pd-mp--       aquel          PDMP
aquello       Pd-ms--       aquel          PDMS
aquel         Pd-ms--       aquel          PDS

cuales        Pr--p--       cual           PRP
cual          Pr--s--       cual           PRS
cuales        Pt--p--       cual           PTP
cual          Pt--s--       cual           PTS
cuyas         Pr-f.-p       cuyo           PRFP
cuya          Pr-f.-s       cuyo           PRFS
cuyos         Pr-m.-p       cuyo           PRMP
cuyo          Pr-m.-s       cuyo           PRMS
quienes       Pr--p--       quien          PRP
quien         Pr--s--       quien          PRS
quiines       Pt--p--       quiin          PTP
```

quiin	Pt -s--	quiin	PTS
que	Pr--.--	que	PR
qui	Pt--.--	qui	PI
suyas	Ps3fp-.	suyo	PS3FP
suya	Ps3fs-.	suyo	PS3FS
suyos	Ps3mp-.	suyo	PS3MP
suyo	Ps3ms-.	suyo	PS3MS
tuyas	Ps2fp-s	tuyo	PS2FPS
tuya	Ps2fs-s	tuyo	PS2FSS
tuyos	Ps2mp-s	tuyo	PS2MPS
tuyo	Ps2ms-s	tuyo	PS2MSS
mmas	Ps1fp-s	mmo	PS1FPS
mma	Ps1fs-s	mmo	PS1FSS
mmos	Ps1mp-s	mmo	PS1MPS
mmo	Ps1ms-s	mmo	PS1MSS
nuestras	Ps1fp-p	nuestro	PS1FPP
nuestra	Ps1fs-p	nuestro	PS1FSP
nuestros	Ps1mp-p	nuestro	PS1MPP
nuestro	Ps1ms-p	nuestro	PS1MSP
vuestras	Ps2fp-p	vuestro	PS2FPP
vuestra	Ps2fs-p	vuestro	PS2FSP
vuestros	Ps2mp-p	vuestro	PS2MPP
vuestro	Ps2ms-p	vuestro	PS2MSP
algunas	Pi-fp--	algzn	PIFP
alguna	Pi-fs--	algzn	PIFS
algunos	Pi-mp--	algzn	PIMP
alguno	Pi-ms--	algzn	PIMS
ningunas	Pi-fp--	ningzn	PIFP
ninguna	Pi-fs--	ningzn	PIFS
ningunos	Pi-mp--	ningzn	PIMP
ninguno	Pi-ms--	ningzn	PIMS
ambos	Pi-mp--	ambos	PIMP
ambas	Pi-fp--	ambas	PIFP
muchas	Pi-fp--	mucho	PIFP
mucha	Pi-fs--	mucho	PIFS
muchos	Pi-mp--	mucho	PIMP
mucho	Pi-ms--	mucho	PIMS
nada	Pi-----	nada	PI
nadie	Pi-----	nadie	PI
otras	Pi-fp--	otro	PIFP
otra	Pi-fs--	otro	PIFS
otros	Pi-mp--	otro	PIMP
otro	Pi-ms--	otro	PIMS
pocas	Pi-fp--	poco	PIFP
poca	Pi-fs--	poco	PIFS

pocos	Pi-mp--	poco	PIMP
poco	Pi-ms--	poco	PIMS
todas	Pi-fp--	todo	PIFP
toda	Pi-fs--	todo	PIFS
todos	Pi-mp--	todo	PIMP
todo	Pi-ms--	todo	PIMS
varios	Pi-mp--	varios	PIMP
varias	Pi-mp--	varias	PIMP

5.3.5 Determiners (D)

Attribute	Value	Example	Code
Type	demonstrat.	este	d
	indefinite	cierto	i
	possessive	mi	s
	interrog.	que'	t
Person	first	mi	1
	second	tu	2
	third	su	3
Gender	masculine	el	m
	feminine	la	f
Number	singular	el	s
	plural	los	p
Possessor	singular	mi	s
	plural	nuestro	p

5.3.5.2 Combinations

Lexique	Corpus	Example
Ds1.ss--	DS1SS	mi (taza - libro)
Ds1.ps--	DS1PS	mis (tazas - libros)
Ds1fsp--	DS1FSP	nuestra (taza)

```
Ds1fpp--  DS1FPP  nuestras (tazas)
Ds1msp--  DS1MSP  nuestro (libro)
Ds1mpp--  DS1MPP  nuestros (libros)

Ds2.ss--  DS2SS   tu (taza -libro)
Ds2.ps--  DS2PS   tus (tazas -libros)
Ds2fsp--  DS2FSP  vuestra (taza)
Ds2fpp--  DS2FPP  vuestras (tazas)
Ds2msp--  DS2MSP  vuestro (libro)
Ds2mpp--  DS2MPP  vuestros (libros)

Ds3.s.--  DS3S    su (taza - libro)
Ds3.p.--  DS3P    sus (tazas -libros)

Dd-fs---  DDFS     esta, esa, aquella
Dd-ms---  DDMS     este, ese, aquel
Dd-fp---  DDFP     estas, esas, aquellas
Dd-mp---  DDMP     estos, esos, aquellos

Di------  DI       cada, cualquier
Di-fs---  DIFS     alguna, ninguna, cierta
Di-ms---  DIMS     algu'n, ningu'n, cierto
Di-fp---  DIFP     algunas, ningunas, ciertas
Di-mp---  DIMP     algunos, ningunos,  ciertos

Dt-fs---  DTFS     cua'nta
Dt-ms---  DTMS     cua'nto
Dt-fp---  DTFP     cua'ntas
Dt-mp---  DTMP     cua'ntos
========= ======= ============================================

5.3.5.3. Conversion tables

=========    ============
Reg.exp        TAG
=========    ============

Dd-fp--      DDFP
Dd-fs--      DDFS
Dd-mp--      DDMP
Dd-ms--      DDMS
Di-----      DI
Di-.s--      DIS
Di-fp--      DIFP
Di-fs--      DIFS
Di-mp--      DIMP
Di-ms--      DIMS
```

Ds1.p-s	DS1PS
Ds1.s-s	DS1SS
Ds1fp-p	DS1FPP
Ds1fs-p	DS1FSP
Ds1mp-p	DS1MPP
Ds1ms-p	DS1MSP
Ds2.p-s	DS2PS
Ds2.s-s	DS2SS
Ds2fp-p	DS2FPP
Ds2fs-p	DS2FSP
Ds2mp-p	DS2MPP
Ds2ms-p	DS2MSP
Ds3.p-.	DS3P
Ds3.s-.	DS3S
Dt-fp--	DTFP
Dt-fs--	DTFS
Dt-mp--	DTMP
Dt-ms--	DTMS

5.3.6 Articles (T)

5.3.6.1. Lexicon

Attribute	Value	Example	Code
Type	definite	el	d
	indefinite	un	i
Gender	masculine	el	m
	feminine	la	f
Number	singular	el	s
	plural	los	p
Case	///	///	///

Reg.exp	TAG
Tifp-	TIFP

Tifs-	TIFS
Timp-	TIMP
Tims-	TIMS
Tdms-	TDMS
Tdfp-	TDFP
Tdfs-	TDFS
Tdmp-	TDMP

5.3.7 Adverbs (R)

5.3.7.1. Lexicon

Attribute	Value	Example	Code
Type	general	muy	g
	particle	no	p
Degree	positive	muy	p
	comparative	ma's	c
	superlative	muchi'simo	s

5.3.7.2 Combinations

Lexique	Corpus	Example
Rgp	RG	mucho
Rgc	RG	ma's
Rgn	RG	nunca
Rpn	RP	no

Reg.exp	TAG
R	R
Rg	RG
Rgp	RG

Rgc	RG

5.3.8 Adpositions (S)

5.3.8.1. Lexicon

Attribute	Value	Example	Code
Type	preposition	en, de	p
Formation	compound		y
	simple		n

5.3.7.2 Combinations

Lexique	Corpus	Example
Sp	SP	en

5.3.9 Conjunctions (C)

5.3.9.1 Lexicon

Attribute	Value	Example	Code
Type	coordinat.	y	c
	subordinat.	que	s

5.3.8.2 Combinations

Lexique	Corpus	Example
Cc	C	pero, o, y
Cs	C	que

5.3.10 Numerals (M)

5.3.10.1 Lexicon

Attribute	Value	Example	Code
Type	cardinal	dos	c
	ordinal	segundo	o
Gender	masculine	un	m
	feminine	una	f
Number	singular	un	s
	plural	dos	p
Case	///	///	-

5.3.9.2 Combinations

Example	Lexique	lema	Corpus
primeras	Mofp-	primero	MOFP
primera	Mofs-	primero	MOFS
primeros	Momp-	primero	MOMP
primero	Moms-	primero	MOMS
uno	Mcms-	uno	MCMS
una	Mcfs-	uno	MCFS
dos	Mc.p-	dos	MCP
doscientas	Mcfp-	doscientos	MCFP
doscientos	Mcmp-	doscientos	MCMP

```
========      ========        =========    ==========

5.3.11  Interjection (I)
------------------------


5.3.11.1  Lexicon

========= ===========
Tag       Example
========= ===========
I         eh
========= ===========


5.3.11.2  Combinations

========= ======= =============================================
Lexique   Corpus  Example
========= ======= =============================================
I         I       eh, ah, oh
========= ======= =============================================


5.3.12  Unique membership class (U)
-----------------------------------

None.



5.3.13  Residual (X)
--------------------


5.3.13.1  Lexicon

============ ==============
Tag          Example
============ ==============
X            symbols, etc.
============ ==============


5.3.12.2  Combinations

========= ======= =============================================
```

```
Lexique   Corpus  Example
========= ======= ==============================================
X         X       symbols, etc.
========= ======= ==============================================
```

5.4 Application to French

In this section the proposed encoding is applied to French (Veronis et al. 1994).

5.4.1 Nouns (N)

5.4.1.1 Lexicon

Attribute	Value	Example	Code
Type	common	livre	c
	proper	Jean	p
Gender	masculine	homme	m
	feminine	femme	f
Number	singular	hommes	s
	plural	femmes	p
Case	///	///	-

5.4.1.2 Corpus

Tag	Regular expression	Definition
NCMS	Ncms-	Common noun, masc. sing.
NCMP	Ncmp-	Common noun, masc. plur.
NCFS	Ncfs-	Common noun, fem. sing.
NCFP	Ncfp-	Common noun, fem. plur.
NPMS	Npms-	Proper noun, masc. sing.
NPMP	Npmp-	Proper noun, masc. plur.
NPFS	Npfs-	Proper noun, fem. sing.
NPFP	Npfp-	Proper noun, fem. plur.

5.4.1.3 Combinations

Lexique	Corpus	Example
Ncms-	NCMS	homme
Ncmp-	NCMP	hommes
Ncfs-	NCFS	femme
Ncfp-	NCFP	femmes
Npms-	NPMS	Jean

Npmp-	NPMP	Pays-bas
Npfs-	NPFS	Anne
Npfp-	NPFP	Pyrenees

Note

It is not clear that all proper nouns should receive gender and number information. In addition, even if this information exists, it might be difficult to find it automatically in corpora for unknown proper nouns. We must experiment to see if a single tag NP would be better.

5.4.2 Verbs (V)

5.4.2.1 Lexicon

Attribute	Value	Example	Code
Type	main	partir	m
	auxiliary	avoir	a
Mood/Vform	indicative	viens	i
	subjunctive	vienne	s
	imperative	viens	m
	conditional	viendrais	c
	infinitive	venir	n
	participle	venu	p
Tense	present	viens	p
	imperfect	venais	i
	future	viendrai	f
	past	vins	s
Person	first	suis	1
	second	es	2
	third	est	3
Number	singular	viens	s
	plural	venons	p
Gender	masculine	venu	m
	feminine	venue	f
Clitics	///	///	-

Notes

a. Conditional

Conditional is often considered as a tense rather than a mood. Encoding decision may change on this.

b. Auxiliaries
It is possible that we need to discriminate between "avoir" and "e^tre" auxiliaries. In that case, we could add a Lexical Class attribute:

Attribute	Value	Example	Code
Lex. class	e^tre	e^tre	e
	avoir	avoir	a

Note : obviously language-specific

c. Past participle
We decided to encode the past participle of the auxiliary verb "e^tre", the copulative verbs (e.g. "sembler") and impersonal verbs (e.g."falloir"), which do not agree in gender, with the value not applicable (-) :

e'te' : Va---ps--, semble' : Vm---ps--

5.4.2.2 Corpus

Tag	Regular expression	Definition
VA1P	Va[iscm][pifs]1p--	Aux. verb 1st person plur.
VA1S	Va[iscm][pifs]1s--	Aux. verb 1st person sing.
VA2P	Va[iscm][pifs]2p--	Aux. verb 2nd person plur.
VA2S	Va[iscm][pifs]2s--	Aux. verb 2nd person sing.
VA3P	Va[iscm][pifs]3p--	Aux. verb 3rd person plur.
VA3S	Va[iscm][pifs]3s--	Aux. verb 3rd person sing.
VAPSPF	Vaps-pf-	Aux. verb past part. plur. fem.
VAPSSF	Vaps-sf-	Aux. verb past part. sing. fem.
VAPSPM	Vaps-pm-	Aux. verb past part. plur. masc.
VAPSSM	Vaps-sm-	Aux. verb past part. sing. masc.
VAPS	Vaps----	Aux. verb past part.
VAPP	Vapp----	Aux. verb pres. part.
VAN	Van-----	Aux. verb infinitive
VM1P	Vm[iscm][pifs]1p--	Main.verb 1st person plur.
VM1S	Vm[iscm][pifs]1s--	Main.verb 1st person sing.
VM2P	Vm[iscm][pifs]2p--	Main.verb 2nd person plur.

```
VM2S    Vm[iscm][pifs]2s--  Main.verb 2nd person sing.
VM3P    Vm[iscm][pifs]3p--  Main.verb 3rd person plur.
VM3S    Vm[iscm][pifs]3s--  Main.verb 3rd person sing.
VMPSPF  Vmps-pf-            Main.verb past part. plur. fem.
VMPSSF  Vmps-sf-            Main.verb past part. sing. fem.
VMPSPM  Vmps-pm-            Main.verb past part. plur. masc.
VMPSSM  Vmps-sm-            Main.verb past part. sing. masc.
VMPS    Vmps----            Main.verb past part.
VMPP    Vmpp----            Main.verb pres. part.
VMN     Vmn-----            Main.verb infinitive

================== ====================================
```

5.4.2.3 Combinations

```
======= ==============================================
Lexique   Corpus  Example
======= ==============================================
INFINITIVE
Vmn-----  VMN      venir
Van-----  VAN      e^tre, avoir
PRESENT PARTICIPLE
Vmpp----  VMPP     venant
Vapp----  VAPP     e'tant, ayant
PAST PARTICIPLE
Vmps----  VM??PS   semble'
Vmps-pf-  VMFPPS   venues
Vmps-sf-  VMFSPS   venue
Vmps-pm-  VMMPPS   venus
Vmps-sm-  VMMSPS   venu

Vaps----  VA??PS   e'te'
Vaps-pf-  VAFPPS   eues
Vaps-sf-  VAFSPS   eue
Vaps-pm-  VAMPPS   eus
Vaps-sm-  VAMSPS   eu
INDICATIVE, PRESENT
Vmip1s-   VM1S     viens
Vmip2s-   VM2S     viens
Vmip3s-   VM3S     vient
Vmip1p-   VM1P     venons
Vmip2p-   VM2P     venez
Vmip3p-   VM3P     viennent

Vaip1s-   VA1S     suis, ai
aip2s-    VA2S     es, as
Vaip3s-   VA3S3    est, a
Vaip1p-   VA1P     sommes, avons
```

```
Vaip2p-   VA2P     e^tes, avez
Vaip3p-   VA3P     sont, ont


INDICATIVE, IMPERFECT
Vaii1s-   VA1S     e'tais, avais
Vaii2s-   VA2S     e'tais, avais
Vaii3s-   VA3S3    e'tait, avait
Vaii1p-   VA1P     e'tions, avions
Vaii2p-   VA2P     e'tiez, aviez
Vaii3p-   VA3P     e'taient, avaient

Vmii1s-   VM1S     venais
Vmii2s-   VM2S     venais
Vmii3s-   VM3S     venait
Vmii1p-   VM1P     venions
Vmii2p-   VM2P     veniez
Vmii3p-   VM3P     venaient
INDICATIVE, FUTURE
Vaif1s-   VA1S     serai, aurai
Vaif2s-   VA2S     seras, auras
Vaif3s-   VA3S3    sera, aura
Vaif1p-   VA1P     serons, aurons
Vaif2p-   VA2P     serez, aurez
Vaif3p-   VA3P     seront, auront

Vmif1s-   VM1S     viendrai
Vmif2s-   VM2SS    viendras
Vmif3s-   VM3S3    viendra
Vmif1p-   VM1PS    viendrons
Vmif2p-   VM2P     viendrez
Vmif3p-   VM3PS    viendront
INDICATIVE, PERFECT
Vais1s-   VA1S     fus, eus
Vais2s-   VA2S     fus, eus
Vais3s-   VA3S3    fut, eut
Vais1p-   VA1P     fu^mes, eu^mes
Vais2p-   VA2P     fu^tes, eu^tes
Vais3p-   VA3P     furent, eurent

Vmis1s-   VM1SS    vins
Vmis2s-   VM2S     vins
Vmis3s-   VM3S3    vint
Vmis1p-   VM1P     vinmes
Vmis2p-   VM2P     vintes
Vmis3p-   VM3P     vinrent
SUBJONCTIVE, PRESENT
```

```
Vasp1s-   VA1S     sois, aie
Vasp2s-   VA2S     sois, aies
Vasp3s-   VA3S     soit, ait
Vasp1p-   VA1P     soyons, ayons
Vasp2p-   VA2P     soyez, ayez
Vasp3p-   VA3P     soient, avaient, e'taient

Vmsp1s-   VM1S     finisse
Vmsp2s-   VM2S     finisse
Vmsp3s-   VM3S     finisse
Vmsp1p-   VM1P     finissions
Vmsp2p-   VM2P     finissiez
Vmsp3p-   VM3P     finissent
SUBJONCTIVE, IMPERFECT
Vasi1s-   VA1S     fusse, eusse
Vasi2s-   VA2S     fusses, eusses
Vasi3s-   VA3S3    fu^t, eu^t
Vasi1p-   VA1P     fussions, eussions
Vasi2p-   VA2P     fussiez, eussiez
Vasi3p-   VA3P     fussent, eussent

Vmsi1s-   VM1S     finisse
Vmsi2s-   VM2S     finisse
Vmsi3s-   VM3S     finit
Vmsi1p-   VM1P     finissions
Vmsi2p-   VM2P     finissiez
Vmsi3p-   VM3P     finissent
CONDITIONAL, PRESENT
Vacp1s-   VA1S     serais, aurais
Vacp2s-   VA2S     serais, aurais
Vacp3s-   VA3S     serait, aurait
Vacp1p-   VA1P     serions, aurions
Vacp2p-   VA2P     seriez, auriez
Vacp3p-   VA3P     seraient, auraient

Vmcp1s-   VM1S     viendrais
Vmcp2s-   VM2SS    viendrais
Vmcp3s-   VM3SS    viendrait
Vmcp1p-   VM1P     viendrions
Vmcp2p-   VM2PS    viendriez
Vmcp3p-   VM3P     viendraient
IMPERATIVE, PRESENT
Vamp2s-   VA2S     sois, aie
Vamp1p-   VA1P     soyons, ayons
Vamp2p-   VA2P     soyez, ayez
Vmmp2s-   VM2S     viens
Vmmp1p-   VM1P     venons
```

Vmmp2p-	VM2P	venez

5.4.3 Adjectives (A)

5.4.3.1 Lexicon

Attribute	Value	Example	Code
Type	qualificat.	bon	f
	ordinal	deuxie^me	o
	cardinal	deux	c
	indefinite	quelconque	i
	possessive	mien	s
Degree	positive	bon	p
	comparative	meilleur	c
Gender	masculine	bon	m
	feminine	bonne	f
Number	singular	bon	s
	plural	bons	p
Case	///	///	-

Notes

a. Degree
We encode Degree for compatibility with other languages, but the distinction positive/comparative applies only to two adjectives in French: "bon" and "mauvais". All other adjectives form their comparatives with "plus" + adjective (e.g., "plus grand"). Superlative is also a compound form ("le" + comparative, e.g. "le plus grand").

b. Possessor
We could add attributes for person and number of possessor.

c. Cardinal
The use of this value in french is still being considered as it seems perfectly redundant with the category numeral.

5.4.3.2 Corpus

Tag	Regular expression	Definition

AFP	A..fp-	Adjective fem. plur.
AFS	A..fs-	Adjective fem. sing.
AMP	A..mp-	Adjective masc. plur.
AMS	A..ms-	Adjective masc. sing.

5.4.3.3 Combinations

Lexique	Corpus	Example
Afcfp-	AFP	meilleures
Afcfs-	AFS	meilleure
Afcmp-	AMP	meilleurs
Afcms-	AMS	meilleur
Afpfp-	AFP	bonnes
Afpfs-	AFS	bonne
Afpmp-	AMP	bons
Afpms-	AMS	bon
Ai-fp-	AFP	certaines, me^mes, quelconques
Ai-fs-	AFS	certane, me^me, quelconque
Ai-mp-	AMP	certain, me^mes, quelconques
Ai-ms-	AMS	certain, me^me, quelconque
Ac-fp-	AFP	deux
Ac-fs-	AFS	une
Ac-mp-	AMP	deux
Ac-ms-	AMS	un
Ao-fp-	AFP	premi'eres
Ao-fs-	AFS	premi'ere
Ao-mp-	AMP	premiers
Ao-ms-	AMS	premier

Lexique	Corpus	Example
As-fp-	AFP	leurs, miennes,tiennes,siennes, no^tres, vo^tres
As-fs-	AFS	leur, mienne, tienne, sienne, no^tre,vo^tre
As-mp-	AMP	leurs, miens, tiens, siens, no^tres, vo^tres
As-ms-	AMS	leur, mien, tien, sien, no^tre, vo^tre

5.4.4 Pronouns (P)

5.4.4.1 Lexicon

Attribute	Value	Example	Code
Type	personal	je	p
	demonstrat.	celui	d
	indefinite	certain	i
	possessive	le_mien	s
	interrog.	lequel	t
	relative	quel	r
Person	first	je	1
	second	tu	2
	third	il	3
Gender	masculine	cet,il	m
	feminine	cette,elle	f
	neutre	ce	n
Number	singular	certain	s
	plural	certains	p
Case	nominative	il	n
	object	le, lui	j
	oblique	moi	o
Possessor	singular	mon	s
	plural	nos	p

Notes

a. Possessive

Possessive pronouns are compound forms only ("le mien"). The form "mien" is an adjective (see note supra).

b. Case

The case system proposed by EAGLES (nominative, accusative, dative, oblique, etc.) does not map readily to French personal pronouns. The usual typology is the following:

subject   je, tu, il, elle, nous, vous, ils, elles

object	me, te, le, la, lui, se, nous, vous, les, leur, se
other	moi, toi, lui, elle, soi, nous, vous, eux, elles, soi

The category "other" corresponds to reinforcement of subject or object ("Moi, je le dis"), attribute ("C'est moi"), etc.

We could use the following mapping:

```
Nominative  --> subject
Accusative  --> direct object
Dative      --> indirect object
Oblique     --> other
```

However, this solution splits "object" in "direct" and "indirect", and this distintion is valid only for the 3rd person pronouns in French (direct: le, la, les; indirect: lui, leur). Encoding this distinction would duplicate all other forms (direct:me, te, etc.; indirect: me, te, etc.).

We have therefore added one value to the case system proposed: the value "Object".

c. New values exclamative, reflexive, reciprocal

The addition of those new values for the attribute type has not yet been considered in French. It is clear that the value "exclamative" would be more useful for the Determiner category (where it is merged with the interrogative value). As for reflexive and reciprocal values, they may be redundant with the codes using the case-value "j" (object), applied to personal pronouns.

d. Agglutination

The presence of disjuncted lexical units among pronouns has led to their lexicalisation for the sake of consistency of some paradigms : for instance the paradigm "auquel", "auxquels" and "auxquelles" is completed with the unit _laquelle", which is not completely satisfactory.

5.4.4.2 Corpus

Tag	Regular expression	Definition
PDFP	Pd-fp--	Demonstrative pronoun fem. plur.

PDFS	Pd-fs--	Demonstrative pronoun fem. plur.
PDMP	Pd-mp--	Demonstrative pronoun masc. plur.
PDMS	Pd-[mn]s--	Demonstrative pronoun masc. sing.
PNFP	Pn-fp--	Indefinite pronoun fem. plur.
PNFS	Pn-fs--	Indefinite pronoun fem. plur.
PNMP	Pn-mp--	Indefinite pronoun masc. plur.
PNMS	Pn-ms--	Indefinite pronoun masc. sing.
PP1SN	Pp1-sn-	Personal pron.,1st pers.sing., nomin.

Tag	Regular expression	Definition
PP2SN	Pp2-sn-	Personal pron.,2nd pers. sing., nomin.
PP3SN	Pp3.sn-	Personal pron.,3rd pers.sing., nomin.
PP1PN	Pp1-sn-	Personal pron.,1st pers. plur., nomin.
PP2PN	Pp2-sn-	Personal pron., 2nd pers. plur.,nomin.
PP3PN	Pp3.sn-	Personal pron., 3rd pers. plur.,nomin.
PPJ	Pp...j-	Personal pron., object
PPO	Pp...o-	Personal pron., oblique
PQFP	P[rt]fp--	Interr. or relat. pronoun,fem. plur.
PQFS	P[rt]fs--	Interr. or relat. pronoun, fem. plur.
PQMP	P[rt]mp--	Interr. or relat. pronoun, masc. plur.
PQMS	P[rt]ms--	Interr. or relat. pronoun, masc. sing.
PSFP	Ps.fp.-	Possessive pronoun, fem. plur.
PSFS	Ps.fs.-	Possessive pronoun, fem. plur.
PSMP	Ps.mp.-	Possessive pronoun, masc. plur.
PSMS	Ps.ms.-	Possessive pronoun, masc. sing.

5.4.4.3 Combinations

Lexique	Corpus	Example
Ps1fs-s	PSFS	la_mienne [mienne is not a pronoun]
Ps1fs-p	PSFS	la_no^tre
Ps1fp-s	PSFP	les_miennes
Ps1fp-p	PSFP	les_no^tres
Ps1ms-s	PSMS	le_mien
Ps1ms-p	PSMS	le_no^tre
Ps1mp-s	PSMP	les_miens
Ps1mp-p	PSMP	les_no^tres
Ps2fs-s	PSFS	la_tienne
Ps2fs-p	PSFS	la_vo^tre
Ps2fp-s	PSFP	les_tiennes
Ps2mp-p	PSMP	les_vo^tres

Ps2ms-s	PSMS	le_tien
Ps2ms-p	PSMS	le_vo^tre
Ps2mp-s	PSMP	les_tiens
Ps2fp-p	PSFP	les_vo^tres
Ps3fs-s	PSFS	la_sienne
Ps3fs-p	PSFS	la_leur
Ps3fp-s	PSFP	les_siennes
Ps3fp-p	PSFP	les_leurs
Ps3ms-s	PSMS	le_sien
Ps3ms-p	PSMS	le_leur
Ps3mp-s	PSMP	les_siens
Ps3mp-p	PSMP	les_leurs
Pp1-sn-	PP1SN	je
Pp2-sn-	PP2SN	tu
Pp3msn-	PP3SN	il, on
Pp3fsn-	PP3SN	elle
Pp1-pn-	PP1PN	nous
Pp2-pn-	PP2PN	vous
Pp3mpn-	PP3PN	ils
Pp3fpn-	PP3PN	elles
Pp1-sj-	PPJ	me (-moi after imperative)
Pp2-sj-	PPJ	te (-toi after imperative)
Pp3msj-	PPJ	le, se, lui
Pp3fsj-	PPJ	la, se, lui
Pp3n-j-	PPJ	en, y
Pp1-pj-	PPJ	nous
Pp2-pj-	PPJ	vous
Pp3mpj-	PPJ	les, se, leur
Pp3fpj-	PPJ	les, se, leur
Pp1-so-	PPO	moi
Pp2-so-	PPO	toi
Pp3mso-	PPO	lui, soi
Pp3fso-	PPO	elle, soi
Pp1-po-	PPO	nous
Pp2-po-	PPO	vous
Pp3mpo-	PPO	eux, soi
Pp3fpo-	PPO	elles, soi
Pd-fp--	PDFP	celles, celles-ci, celles-la’
Pd-fs--	PDFS	celle, celle-ci, celle-la’
Pd-mp--	PDMP	ceux, ceux-ci, ceux-la’
Pd-ms--	PDMS	celui, celui-ci, celui-la’
Pd-n---	PDMS	ce, ceci, cela, ca

Pi-fp--	PNFP	quelques-unes, certaines...
Pi-fs--	PNFS	aucune, nulle, certaine...
Pi-mp--	PNMP	quelques-uns, certains...
Pi-ms--	PNMS	aucun, nul, quelqu'un, certain...
Pr-fp--	PQFP	lesquelles,desquelles,auxquelles,qui, que, quoi, dont,
Pr-fs--	PQFS	laquelle, qui, que, quoi, dont, ou^
Pr-mp--	PQMP	lesquels, desquels, auxquels, qui, que, quoi, dont, ou^
Pr-ms--	PQMS	lequel, duquel, auquel, qui, que, quoi, dont, ou^
Pt-----	PQ??	quoi
Pt-fp--	PQFP	lesquelles, desquelles, auxquelles, qui, que
Pt-fs--	PQFS	laquelle, qui, que
Pt-mp--	PQMP	lesquels, desquels, auxquels, qui
Pt-ms--	PQMS	lequel, duquel, auquel, qui, que

5.4.5 Determiners (D)

5.4.5.1 Lexicon

Attribute	Value	Example	Code
Type	article	le	a
	demonstrat.	ce	d
	indefinite	certain	i
	possessive	mon	s
	interrog.	quel	t
Person	first	ma	1
	second	ta	2
	third	sa	3
Gender	masculine	le	m
	feminine	la	f
Number	singular	le	s
	plural	les	p
Case	///	///	-

Possessor	singular	mon	s
	plural	nos	p
Quantif.	definite	le	d
	indefinite	un	i

5.4.5.2 Corpus

Tag	Regular expression	Definition
DFP	D..fp--.	Determiner, fem. plur.
DFS	D..fs--.	Determiner, fem. plur.
DMP	D..mp--.	Determiner, masc. plur.
DMS	D..ms--.	Determiner, masc. sing.

5.4.5.3 Combinations

Lexique	Corpus	Example
Ds1fss--	DFS	ma (tasse)
Ds1mss--	DMS	mon (livre)
Ds1fps--	DFP	mes (tasses)
Ds1mps--	DMP	mes (livres)
Ds2fss--	DFS	ta (tasse)
Ds2mss--	DMS	ton (livre)
Ds2fps--	DFP	tes (tasses)
Ds2mps--	DMP	tes (livres)
Ds3fss--	DFS	sa (tasse)
Ds3mss--	DMS	son (livre)
Ds3fps--	DFP	ses (tasses)
Ds3mps--	DMP	ses (livres)
Ds1fsp--	DFS	notre (tasse)
Ds1msp--	DMS	notre (livre)
Ds1fpp--	DFP	nos (tasses)
Ds1mpp--	DMP	nos (livres)
Ds2fsp--	DFS	votre (tasse)
Ds2msp--	DMS	votre (livre)
Ds2fpp--	DFP	vos (tasses)
Ds2mpp--	DMP	vos (livres)
Ds3fsp--	DFS	leur (tasse)
Ds3msp--	DMS	leur (livre)
Ds3fpp--	DFP	leurs (tasses)
Ds3mpp--	DMP	leurs (livres)
Dd-fs---	DFS	cette
Dd-ms---	DMS	cet, ce

Dd-fp---	DFP	ces
Dd-mp---	DMP	ces
Di-fs---	DFS	aucune, nulle, certaine, toute, chacune...
Di-ms---	DMS	aucun, nul, certain, tout, chacun...
Di-fp---	DFP	certaines, toutes...
Di-mp---	DMP	certains, tous...
Dt-fs---	DFS	quelle
Dt-ms---	DMS	quel
Dt-fp---	DFP	quelles
Dt-mp---	DMP	quels
Da-fs--d	DFS	la
Da-ms--d	DMS	le
Da-fp--d	DFP	les
Da-mp--d	DMP	les
Da-fs--i	DFS	une
Da-ms--i	DMS	un
Da-fp--i	DFP	des
Da-mp--i	DMP	des

5.4.6 Adverbs (R)

5.4.6.1 Lexicon

Attribute	Value	Example	Code
Type	general	fortement	g
	particle	ne, pas	p
Degree	positive	fortement	p
	comparative	davantage	c
	negative	ne, pas	n

Note

We encode degree for compatibility with other languages, but as for adjectives, the comparative feature is not very productive in French. It applies only to "beaucoup" (comp.= "davantage"), "bien" (comp.= "mieux"), "mal" (comp.= "pis") and "peu" (comp. = "moins"). The comparative for other adverbs is marked by "plus" + adverb (e.g. "plus fortement"). The superlative is usually marked by "le" + comparative (e.g. le plus fortement).

5.4.6.2 Corpus

Tag	Regular expression	Definition

R	Rg.-	General adverb
R-NE	Rpn	ne
R-PAS	Rpn	pas

Note

It seems necessary in French to distinguish the two parts of the negation ("ne ... pas"), because they play an important role in disambiguation. However, this violates the applicative principle (the categories in the corpus should be broader than the categories in the lexicon). Here the same lexical category (Rpn) would split in two corpus tags (R-NE, R-PAS). As a result, the regular expression Rpn cannot be used to define the corpus tags unambiguously.

We could add an attribute "Lexical Class" to discriminate between the two particles, as, if needed, for the distinction between the "e^tre" and "avoir" auxiliaries (see note supra).

Attribute	Value	Example	Code
Lex. class	ne	ne, n'	n
	pas	pas, plus	p

However, the auxiliary distinction applied to many languages (English: be/have; Italian: essere/avere, etc.), whereas the negation problem seems specific to French. It seems therefore heavy to impose an attribute "Lexical Class" for all languages in the Adverb category.

Another point of view would be to consider that, in fact, some lexical subcategorization will be needed for one category or another in each language, and add a "Lexical Class" attribute to all the part-of-speech categories in a systematic way.

5.4.6.3 Combinations

Lexique	Corpus	Example
Rgp	R	beaucoup
Rgc	R	davantage
Rpn	R-NE	ne

```
Rpn       R-PAS   pas, plus
======= ============================================

5.4.7  Adpositions (S)
----------------------

5.4.7.1  Lexicon

============ =========== =========== ====
Attribute    Value       Example     Code
============ =========== =========== ====
Type         preposition en, de      p
============ =========== =========== ====

5.4.7.2  Corpus

================== ====================================
Tag     Regular expression Definition
================== ====================================
SP      Sp                 Preposition
================== ====================================

5.4.7.3  Combinations

======= ============================================
Lexique   Corpus  Example
======= ============================================
Sp        SP      en
======= ============================================

5.4.8  Conjunctions (C)
-----------------------

5.4.8.1  Lexicon

============ =========== =========== ====
Attribute    Value       Example     Code
============ =========== =========== ====
Type         coordinat.  et          c
             subordinat. que         s
============ =========== =========== ====

5.4.8.2  Corpus

================== ====================================
Tag     Regular expression Definition
================== ====================================
```

CC	Cc	Cooordinative conjunction
CS	Cs	Subordinative conjunction

5.4.8.3 Combinations

Lexique	Corpus	Example
Cc	CC	mais, ou
Cs	CS	que

5.4.9 Numerals (M)

5.4.9.1 Lexicon

Attribute	Value	Example	Code
Type	cardinal	deux	c
Gender	masculine	un	m
	feminine	une	f
Number	singular	un	s
	plural	deux	p
Case	///	///	-

Note:

Ordinals are simple adjectives in French. They can never be determiners (e.g. *premie're fois e'tait la bonne).

Traditionnal grammars usually distinguish un/article and un/numeral. However, it is very difficult to find linguistic tests that enable to discriminate between the two.
cf. J'ai vu un chat (article)
    J'ai vu un chat et deux chiens (numeral?)

We will not keep this distinction in the corpus tags.

Only un/une have a gender. All other numerals are invariant in gender. We will encode gender as not applicable rather than introduce a systematic homography.

5.4.9.2 Corpus

Tag	Regular expression	Definition
M	Mc-s-	Cardinal numeral, sing.
M	Mc-p-	Cardinal numeral, plur.

5.4.9.3 Combinations

Lexique	Corpus	Example
Mcms-	DMS	un
Mcfs-	DFS	une
Mc-s-	M	zero
Mc-p-	M	deux, trois

5.4.10 Interjection (I)

5.4.10.1 Lexicon

Tag	Example
I	eh

5.4.10.2 Corpus

Tag	Regular expression	Definition
I	I	Interjection

5.4.10.3 Combinations

Lexique	Corpus	Example
I	I	eh

5.4.11 Unique membership class (U)

None.

5.4.12 Residual (X)

5.4.12.1 Lexicon

Tag	c	Example
X	c	symbols, etc.

5.4.12.2 Corpus

Tag	Regular expression	Definition
X	X	Residual

5.4.12.3 Combinations

Lexique	Corpus	Example
X	X	symbols, etc.

5.5 Application to English

The application to English has been carried out by the MULTEXT Group at ISSCO (ISSCO 1994). Note that this application has been carried out on the basis of the attributes and the values as presented in the preceding version of this deliverable (MULTEXT WP1.6 A2 version).

Notation:

Trailing place-holders have been omitted.

It should be borne in mind that the need for place-holders is an artefact of the linear representation of lexical descriptions. A number of extensions have been made to the various proposals circulated. It is not clear how language-specific they are, but they represent phenomena that are plausibly relevant for various text-processing tasks.

```
5.5.1 Nouns (N)
---------------

= ============== ============== =
P ATT            VAL            C
= ============== ============== =
1 Type           Common         c
                 Proper         p
- -------------- -------------- -
2 Number         Singular       s
                 Plural         p
- -------------- -------------- -
3 Gender         Masculine      m
                 Feminine       f
                 Neuter         n
= ============== ============== =

Notes:

Case is not relevant for English.

Gender is probably unnecessary for most purposes. We can assume it may
be of interest in constructions with pronouns.  Many nouns are (or may
be) unmarked for number: fish, sheep, aircraft.

Examples:

    Ncsn   house
    Ncpn   houses
    Npsn   Thames
    Nppn   Alps
```

```
    Ncpf   women
    Ncsm   man
    Nc=n   sheep
```

5.5.2 Verbs (V)

```
= ============== ============== =
P ATT            VAL            C
= ============== ============== =
1 Type           Main           v
                 Auxiliary      a
                 Modal          m
- -------------- -------------- -
2 Form           Indicative     i   Form, not Mood
                 Imperative     m
                 Subjunctive    s
                 Base           b   base, not infinitive
                 Past Prt       p   Past Prt, not participle
                 Present Prt    g   Present Prt added (not gerund)
- -------------- -------------- -
3 Tense          Present        s
                 Past           d
- -------------- -------------- -
4 Number         Singular       s
                 Plural         p
- -------------- -------------- -
5 Person         First          1
                 Second         2
                 Third          3
= ============== ============== =
```

Notes:
Voice is not lexical.

Attributes have been reordered to minimize sequence lengths (assuming the   proposal about trailing "-"s) - Tense only applies to Finite verbs,
Number only to Present, and Person only to Singular. Modals have been included as a distinct subcategory.

Finiteness as an attribute is redundant - predictable from verb-form and past/present.

These tags do not attempt to represent distinctions found in the various compound verb-forms. These are composed of a sequence of auxiliary and non-finite verb as follows:

future	will/shall + base
conditional	would/should + base
passive	be + past participle
perfect	have + past participle
past perfect	have-past + past participle
present continuous	be + present participle
infinitive	to + base

So there is no "aspect" attribute or "future" value, for example.

Examples:

```
go        Vvm, Vvb, Vvs, Vvisp, Vviss1, Vviss2
goes      Vviss3
going     Vvg
gone      Vvp
went      Vvid

have      Vab, Vas, Vaip, Vaiss1, Vaiss2
has       Vaiss3
had       Vap Vaid
having    Vag

be        Vab, Vam
am        Vaiss1
are       Vaiss2, Vaisp
is        Vaiss3
was       Vaids1, Vaids3
were      Vaids2, Vaidp
been      Vap
being     Vag

will      Vmi
would     Vmi
```

5.5.3 Adjectives (A)

P	ATT	VAL	C
1	Degree	Positive	p
		Comparative	c

		Superlative	s	
2	Position	Attributive	a	Position added
		Predicative	p	

Notes:

Gender, number and case irrelevant for English.

Attributive/predicative distinction reflects positional constraints - some adjectives ('mere', 'utter', etc.) only appear in prenominal position, while others ('awake', 'devoid', etc.) only appear in predicative position. Most can appear in either.

Since many English comparatives & superlatives are formed with more/most, "positive" cannot be interpreted as "neither comparative nor superlative". See "Adverbs".

Examples:

big	Ap
bigger	Ac
biggest	As
more peculiar	Dscn+Ap
most remarkable	Dssn+A
awake	App
mere	Apa

5.5.4 Pronoun (P)

P	ATT	VAL	C	
1	Pron.-Type	General	g	General added
		Demonstrative	d	
		Possessive	s	
		Personal	p	
		Reflexive	x	
2	WH	Not-WH	n	WH added
		Relative	r	
		Int	q	

3 Number	Singular	s	
	Plural	p	
4 Person	First	1	
	Second	2	
	Third	3	
5 Gender	Masculine	m	
	Feminine	f	
	Neuter	n	
6 Case	Nominative	n	
	Accusative	a	
7 Poss-Number	Singular	s	
	Plural	p	
8 Poss-Person	First	1	Poss-Person added
	Second	2	
	Third	3	
9 Poss-Gender	Masculine	m	Poss-Gender added
	Feminine	f	
	Neuter	n	

Notes:

"General" pronouns are those which are not personal, possessive, demonstrative or reflexive. The choice of these four categories is based on distributional facts, though at a rather high level of abstraction. They enter into anaphoric dependencies which are signalled morphosyntactically and are therefore (in principle) more amenable to automatic detection. Most general pronouns do not, although they too sometimes encode number information.

"WH" attribute added to allow for combination of possessive and WH in "whose".

Examples:

Pgn	some, all, ...
Pgns	each, something, nothing, everything, -body, ...
Pgnp	both, ...
Pdns	this, that
Pdnp	these, those
Psn---s1	mine

Psn----2	yours
Psn---s3m	his
Psn---s3f	hers
Psn---s3n	its
Psn---p1	ours
Psn---p3	theirs
Psq	whose
Ppns1-n	I
Ppn-2	you
Ppns3mn	he
Ppns3fn	she
Ppns3n	it
Ppnp1-n	we
Ppnp3-n	they
Ppns1-a	me
Ppns3ma	him
Ppns3fa	her
Ppnp1-a	us
Ppnp3-a	them
Prns1	myself
Prns2	yourself
Prns3m	himself
Prns3f	herself
Prns3n	itself
Prnp1	ourselves
Prnp1	yourselves
Prnp1	themselves
Ppr	which
Ppq	which, what

5.5.5 Articles/Determiners (R)

P	ATT	VAL	C	
1	Type	Def-article	t	
		Indef-article	a	
		Demonstrative	d	
		Possessive	s	
		General	g	General added
2	WH	Not-WH	n	WH added
		Relative	r	
		Int/Excl	q	

3	Number	Singular	s	
		Plural	p	
4	Person	First	1	
		Second	2	
		Third	3	
5	Gender	Masculine	m	
		Feminine	f	
		Neuter	n	
6	Poss-Number	Singular	s	
		Plural	p	
7	Poss-Person	First	1	Poss-Person added
		Second	2	
		Third	3	
8	Poss-Gender	Masculine	m	Poss-Gender added
		Feminine	f	
		Neuter	n	

Notes:

Case not relevant to English.

Definite and indefinite articles represented as values of "Type". All of these have been marked as 3rd person. This is redundant, since determiners are all 3rd person, if anything.

Examples:

Rtn-3	the
Rans3	a/an
Rdns3-ns	this, that
Rdnp3-np	these, those
Rsn-3-s1	my
Rsn-3--2	your
Rsn-3-s3m	his
Rsn-3-s3f	her
Rsn-3-s3n	its
Rsn-3-p1	our
Rsn-3-p3	their
Rsr-3	whose
Rsq-3	whose

Rgr-3 which
Rgq-3 which, what,
Rgns3 each, ...
Rgnp3 all, both, certain, many, ...
Rgn-3 some, ...

5.5.6 Adverbs (D)

P	ATT	VAL	C	
1	Degree	Positive	p	
		Comparative	c	
		Superlative	s	
2	Function	Specifier	s	Function added
		Modifier	m	
3	WH	Yes	q	WH added
		No	n	

Notes:

No distinction has been made between different types of "modifier" adverbs ("sentence-modifying", "VP-modifying", etc.), since their distributions overlap considerably.

Examples:

Dbsn so, too, very, as
Dbsq how
Dcsn more
Dssn most
Dbmn quickly, soon, here, then, now
Dcmn better, worse
Dsmn best, worst
Dbmq where, when, how, why

5.5.7 Adpositions (S)

```
= ============== ============== =
```

P	ATT	VAL	C
1	Type	Preposition	e
		Postposition	o

Notes:

Postpositions are rare in English.

"possessive" 's and ' might be considered postpositions, especially if the alternative is to assign them to the unique membership class (where by definition they would be unrelated).

Examples:

Se	in, near, behind,...
So	notwithstanding, ago

5.5.8 Conjunctions (C)

P	ATT	VAL	C	
1	Type	Coordinating	c	
		Subordinating	s	
2	Comp-Type	Infinitive	i	Comp-Type added
		Finite	f	
3	Coord-Position	Initial	i	Coord-Position added
		Non-initial	n	

Notes:

Subordinating conjunctions are often identical to prepositions (before, after, since, ...).

"Comp-Type" encodes information about the complement of subordinating conjunctions. Present-participle complements have not been allowed for here; they consist of bare VPs and are often treated as nominal constructions, the words that introduce them (by, after, etc.) being classified as prepositions.

"Coord-Position" encodes the distinction between elements of a discontinuous coordination. "Initial" conjunctions are those that appear before the first conjunct, and "Non-initial" conjunctions are those that appear elsewhere. There is a dependency between initial and non-initial conjunctions ('both...and' but not 'either...and') which is not expressed in these attributes.

Examples:

```
Ccn  and, or, but
Cci  either, neither, both
Csi  for
Csf  that, because, ...
```

5.5.9 Numerals (M)

P	ATT	VAL	C
1	Type	Cardinal	c
		Ordinal	o

Notes:

"Function" depends on syntactic context.

These have not been subsumed under adjectives, pronouns, determiners, etc. because the internal structure of complex numerals is idiosyncratic.

Examples:

```
Mc    six
Mo    sixth
```

5.6 Application to Dutch

5.6.1 Multext morphology formalism

The main reason for having morphology is to facilitate maintenance of lexical lists, both during the project and afterwards. It follows that the rule formalism for morphology itself should not be a source of complexity. The designers of the Multext morphology tool have therefore decided to use fairly common, well-known methods, avoiding any adventurous modernism.

The tool has two parts: morphosyntax and morphographemics.

For morphosyntax, a user-friendly version of context-free grammar is used. The friendlyness comes with the possibility to annotate rules with features, and to set features to the same values through variables.

For morpho-graphemics, a version of two-level morphology is used.

The system can be used to generate a word list from an input word list, as well as to look words up in a given word list.

The full description and detail of the rule formalism are given in the report on the Multext morphology tool (Report nr. 2.3.1B) and the manuals accompanying the tool. Extensive exemplification can be found in the report on Multext morphology resources, Report nr. 5.3.1B.

5.6.2 Dutch word classes

The Dutch word classification used for Multext approximates as closely as possible the proposal in Deliverable A, section 1.6.1.

It can be summarized best in terms of (the relevant selection from) the types and attributes used in the actual Dutch description (Report 5.3.1B, section on Dutch).

There are 10 word class types:

```
V               : Vtype Vform  Person Number Tense
N               : Ntype Semgender Gender Number
A               : Inflected Degree
Adp             : AdpType
Det             : DetType Number Gender Defness
```

```
Pron           : PronType Number Defness Person Semgender Case
Adv            : Nil
Num            : Nil
Conj           : ConjType
Interj         : Nil

where:

Vtype          : Main Aux Copula Impersonal
Tense          : Pres Past
Person         : 1 2 3
Number         : Sg Pl
Vform          : Inf ImPart PerfPart Fin
Ntype          : Common Proper
Semgender      : M F N
Gender         : De Het none
Degree         : Pos Compar Super
Inflected      : 0 1
AdpType        : Post Pre
DetType        : Article Quantificational Possessive Demonstrative
Defness        : Def Indef
PronType       : Reciprocal Reflexive Personal Relative Demonstrative
                 Quantificational Interrogative
Case           : 1 4
ConjType       : Coord Subord
```

Dutch is not a morphologically rich language. A distinction that it makes which is different from most other Multext languages is that between ‘syntactic’ and ‘semantic’ gender. A good example is ‘meisje’ (girl). The syntactic gender is ‘het’ (‘het meisje’ *‘de meisje’) but the semantic gender is female (‘het meisje(i) dacht dat ze(i)/*hij(i)/??het een jongetje was’). It can be seen in the example that articles agree with their nouns in syntactic gender and that pronouns (usually) agree with their antecedents in semantic gender.

For the rest, the distinctions given differ from other languages mainly in what Dutch does not express.

The distinction between pronouns and determiners has been implemented as follows:

for any X that could be either a det or a pron:
if X distributes like NP, then X is a pronoun;
if X distributes like Det (i.e. NP-initial), then X is a determiner

It is hoped that this is a good starting point for tagging but this remains to be seen. As a consequence, a word like ‘mijn’ which is often called a pronoun is analysed as a determiner here.

Decisions like this one on function words can easily be changed; e.g. in the lexicon supplied for Dutch (Report nr. 5.4.1B), 59 words are classified as pronouns and 27 as determiners.

6 References

Bel N. and A. Aguilar (1994): "Proposal for Morphosyntactic encoding: Application to Spanish", Barcelona.

Calzolari N., Ceccotti M.L., Roventini A. (1983): "Documentazione sui tre nastri contenenti il DMI", ILC Technical Report, Pisa.

Leech J. and A. Wilson (1993): "Invitation Draft", EAGLES Document, Lancaster.

Leech J. and A. Wilson (1994): "Morphosyntactic Annotation", EAGLES Interim Report, Pisa.

Calzolari N. and M. Monachini (1994): "MULTEXT morphosyntactic encoding: Application to Italian", Pisa.

Heylen D. (1994): "Eagles Tagset for Dutch", Utrecht.

ISSCO (1994): "MULTEXT Morphosyntactic Encoding: Application to English", ISSCO, Geneva.

Lyons J. (1981): *Language and Linguistics*, Cambridge University Press.

Monachini M. and N. Calzolari (1994): "Synopsis and Comparison of Morphosyntactic Phenomena encoded in Lexicons and in Corpora and Application to European Languages, EAGLES document EAG-LSG-T4.6/CSG-T3.2, Pisa.

MULTEXT (1993): "MULTEXT Technical Annex", Aix-en-Provence.

MULTEXT WP1.6 Report A2 (1994): "Common Specifications and Notation for Lexicon Encoding", MULTEXT Report Milestone A2.

Steiner P.and L. Lemnitzer (1994): "An adaptation of the proposal for morphosyntactic encoding in MULTEXT for German", Muenster.

Veronis J. (1994): "Intermediary tagset: proposal for revision" Aix-en-provence.

Veronis J., L. Khouri and C. Meunier (1994): "Proposal for morphosyntactic encoding in MULTEXT", Aix-en-Provence.