

Deliverable 2.1

Evaluation of existing standards for NLP Lexica

Project reference number	e-Content-22236-LIRICS
Project acronym	LIRICS
Project full title	Linguistic Infrastructure for Interoperable Resource and Systems
Project contact point	Laurent Romary, INRIA-Loria 615, rue du jardin botanique BP101. 54602 Villers lès Nancy (France) romary@loria.fr
Project web site	http://lirics.loria.fr
EC project officer	Erwin Valentini
Document title	Evaluation of existing standard for NLP Lexica
Deliverable ID	2.1
Document type	Report
Dissemination level	Public
Contractual date of delivery	M9
Actual date of delivery	31.10.2005
Status & version	Draft 1.1
Work package, task & deliverable responsible	DFKI
Author(s) & affiliation(s)	Monica Monachini (CNR-ILC), Claudia Soria (CNR-ILC), Marisa Ulivieri (CNR-ILC)
Additional contributor(s)	Nicoletta Calzolari (CNR-ILC), Thierry Declerck (DFKI), Michele Mammini (CNR-ILC)
Keywords	NLP Lexicon Morpho-Syntax, Syntax, Semantics, Standards

Document evolution

Version	date	version	date
1.0	20 June 2005		
1.1	31st October 2005		

1 Objectives and Interrelationships between LIRICS and ISO activities

As stated in the Technical Annex of the project, LIRICS will develop, implement and demonstrate a set of standards for language technology ratified by ISO. LIRICS will do so by taking as a starting point the results achieved in past and on-going standardization initiatives and trying to address gaps emerged in these activities so far.

The present deliverable reports the activities within *Task 1*, concerned with the definition of the indispensable background for the formulation of a set of Data Categories needed for populating the different layers of the lexical data model, i.e. the abstract framework for lexicon representation to be designed in *Work Package 2*.

The **objective** of the present deliverable is to:

- gather linguistic information, reliable, harmonized enough and crucial for the description of a computational lexical entry;
- draft a unified inventory of lexical information, good candidate to become Data Categories, with unified descriptors and short descriptions/exemplifications.

The methodology for attaining such an objective consists in a comparative analysis and integration of major past or ongoing efforts towards standardization and best practices in the field.

This set of lexical information can be seen as a kind of *Pre-Data Categories*, that, once consolidated, will play the role of Data Categories that are crucial to decorate the lexicon layers of the meta-model and instantiate the abstract lexical components associated to it. Data Categories will constitute one of the key results of the LIRICS project: they represent the main *building blocks* which, in combination with the structures of the LIRICS lexicon data-model, will make it possible to design different possible lexical entries as instances of the abstract schema¹.

The identification and definition of Data Categories in LIRICS, represents one of the main contact points with objectives defined in **ISO/TC37**, where a total revision of the **ISO 12620 Data Category Registry** is underway. The LIRICS Data Categories will serve as a coherent input to this Registry and will directly enrich it. They will be formalized according to the ISO standards and will be represented and ruled in conformance to ISO procedures. This constitutes an enormous advantage which will ensure that, beyond the life-time of the project, Data Categories issued by LIRICS will be governed and maintained as an official ISO global resource.

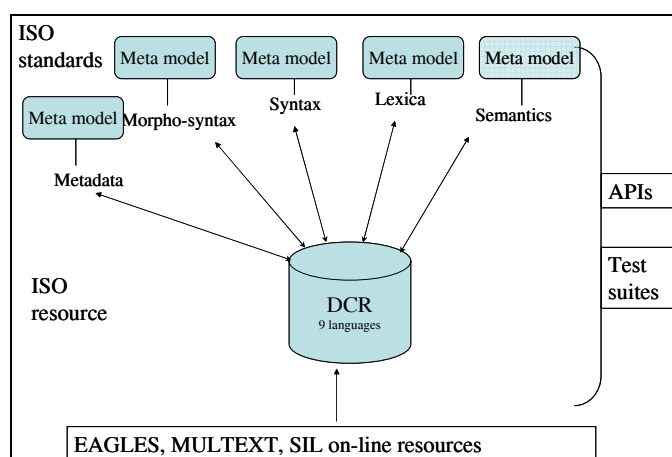
The Data Category Registry plays a central role in the development of flexible standards in the ISO/TC37/SC4 perspective and thus ensures the coherence of LIRICS activities from a vertical point of view, since the various standard proposals will be grounded on the common principle of combining a meta-model with data categories selected from the registry.

As concerns Data Categories to be identified for lexicons, it should be mentioned that, within the ISO/TC37/SC4, the working group devoted to the definition of a standard framework for lexical resources, the so-called LMF (Lexical Mark-up Framework), has recently established a **Data Category task force** explicitly dedicated to Data Category creation and management and formally approved with a plenary ISO Committee resolution. The work on data category will find assessment and support from cooperation with this group of experts, (where a number of LIRICS representatives are involved).

¹ Lexicon designers are allowed, in case they need to represent lexical information which is not covered in the Data Category Registry, to define a set of user-defined data categories, provided that this is represented according to ISO 12620 procedures and rules.

Data Categories are crucial also from a horizontal point of view, since APIs (WP5) and test suites will be designed according to a methodology based on the same principles, throughout all the project.

Moreover, data categories ensure transversal coherence in LIRICS: strict **interrelationships** exist between the work described here and activities being carried out in other LIRICS Work Packages towards the formulation of a set of agreed on data categories for linguistic annotation in corpora (WP3) and harmonized linguistic information for semantic representation in texts (WP4). The figure in the Technical Annex clearly displays such interrelationships.



The LIRICS Data Category registry

In the actual version of this deliverable we aim at formulating the nucleus of the proposal, i.e. the maximum set of information available subdivided along the different layers of linguistic representation. Moreover (as explained in more details below), as an added-value of this purely linguistic activity, we tried to offer an abstract model with an XML formalization and instantiation of the data. This effort is crucial in order to make this bulk of information more computationally manageable, thus helping and fostering next phases of Data Category creation and API definition.

These *Pre-Data Categories* will receive the necessary adjustments and modifications, after discussion among the LIRICS partners and feedback from connected LIRICS activities, revisions coming from ISO experts and, hopefully, suggestions issued from end-users which will test adequacy, appropriateness and coverage with respect to application and language-specific considerations.

In agreement with D3.1, we suggest to consider this deliverable as a document in-progress. The work of defining Data Categories for lexical description is, by its very nature, to be seen as an incremental one, since it will proceed step by step with the specification of the meta-model. Only the combination of the meta-model (its layers and components) with the Data Categories will concur to define the Lexical Model and the Lexical Mark-up Language necessary to instantiate it.

2 Driving Criteria

As already mentioned in the introductory section, the main input to this work comes from the previous experiences in the field of lexicons. The survey phase has been conducted in collaboration with other LIRICS partners involved in the past standardization initiatives, with the following **driving criteria** in mind:

- **Transversal coherence between lexicon and corpus.** Endorsing the fruitful and well-proven methodology proposed within EAGLES (Monachini and Calzolari 1999),

activities in this Work package proceeded in close cooperation with parallel work in the field of linguistic annotation in texts (WP3). A combined strategy between ILC and DFKI has been adopted in order to investigate interdependencies between linguistic categories, thus ensuring transversal compatibility between the two actual compilations of information and coherence between the two Deliverables². D2.1 and D3.1 should, hence, be seen as the “two sides of the same coin”, the first presenting linguistic information from the point of view of lexicon encoding, whereas the latter offers the perspective of text annotation. Reasons of this choice are many-fold. The first obvious consideration is that a lexicon and a corpus are strictly interrelated in NLP applications: morpho-syntactic taggers and syntactic parsers are systematically based on a lexicon to perform corpus analysis and the higher is the performance, the more compatible the content (and representation) of linguistic information is between the two resources. Corpora and lexica demand common representation and content which can only be obtained via standard converging formats.

- **Maximum coverage.** Linguistic information proposed in the present deliverable come from integration of major past or ongoing efforts in EU, outside EU and in accessing countries. Interaction with emerging standards in the speech community have been also taken into account, by considering outcomes of parallel initiatives, such as the specifications carried out by the LCStar project.
- **Maximum decomposition.** We aim “to reach a maximal decomposition into the minimal *basic information units* that reflect the phenomena we are dealing with” (Calzolari *et al.* 2002.) This principle is used to allow easier reusability or mappability into different theoretical or system approaches: small units can be assembled, in different frameworks, according to different (theory/application dependent) generalization principles. Lexica are built for different purposes/users and can be specialised to cover a few linguistic phenomena, only describe one category (verbs, nouns etc.) or apply to specific NLP systems and/or applications. All these differences can have a repercussion on linguistic information. Basic information must be established before considering any system-specific instantiation, otherwise their finding may be too conditioned by system-specific approaches. At the same time, whenever consensus can be found around a more complex linguistic object, we provide such shareable commonly agreed linguistic objects (e.g. *synsets* for WordNet-like lexicons and *qualia relations* for SIMPLE-like lexicons are proposed as *de-facto* standard, even if linked to specific theoretical approaches).
- **Acceptance and Impact.** By relying on past standardization activities and attested best-practices in the field, we have implicit guarantee that the obtained a set of information is used, accepted and widespread inside the community. Through the massive involvement of the pool of industrial partners, i.e. the LIRICS Industrial Advisory Board, the usability of proposed linguistic categories, their degree of acceptability and diffusion, the necessary sustainability, the adherence to the real end-user requirements will be checked once more.

3 State-of-the-Art Linguistic Information in Standardised Lexicons

3.1 Morpho-Syntactic Lexical Encoding

This is the information that is used to qualify a Lexical Entry with reference to the morpho-syntactic category (part of speech) and, where appropriate, a sub-category, its morphological features, and the combination of morphological features in inflected word-forms.

² See minutes of the DFKI– CNR-ILC bilateral meeting (lyrics.loria.fr/documents.html) for further details.

Previous activities related to the harmonization of Morpho-syntactic encoding for lexicons are all EAGLES-centric. The EAGLES activity was started in 1993 and culminated with the publication of the main document about lexicon encoding in 1996 (Monachini and Calzolari, 1996: <http://www.ilc.cnr.it/EAGLES96/annotate/morphsyn.html>) in combination with the document about morpho-syntactic annotation of corpora (Leech and Wilson, 1996: <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>). The EAGLES Morpho-syntactic specifications are the result of a bottom-up approach, consisting of a comparison of the main encoding practices in lexica and corpora and resulting in a consensual descriptive collection of morpho-syntactic lexical properties for twelve European languages. These recommendations – by the way they came into existence, i.e. based on commonly accepted practices – constitute a detailed agreed on platform, which had great influence in addressing later projects. Firstly, MULTEXT where morphological lexicons for corpus annotation were implemented in conformance with those guidelines (Bel, Calzolari and Monachini, 1995) and lexicon specifications were used as interface between different tagsets. Secondly, MULTEXT-East where the specifications were extended to cover Eastern languages and to allow the production of harmonized morphological lexica for those languages (Monachini 1995 and Erjavec and Monachini 1997). Moreover, in the second phase of EAGLES, efforts were made to foster the use of lexicon guidelines in application framework, by offering more detailed and formalized specifications in the form of language specific implementations for Italian, German, French and English lexicons: lexical specifications were arranged in type hierarchies, thus allowing a more formal representation and expression of constraints in the application of feature and values. Then, the PAROLE project implemented the EAGLES guidelines for the construction of pluri-lingual lexicons for twelve EU languages (Monachini 1995). More recently, in the ISLE framework, the EAGLES morpho-syntactic specifications have been adopted, reinterpreted in a multilingual perspective and reassessed in view of the requirements of Asian languages as well (Calzolari, Bertagna, Lenci and Monachini, 2001). ELRA, the major European agency active in the field of linguistic resource, grounded its standardization activities on EAGLES (Monachini *et al.* 2003).

The EAGLES Guidelines for Morpho-syntax, issued as a maximal set of linguistic specifications proposed for the description of a lexical entry, constitute the platform for drafting the set of harmonized LIRICS Pre Data Categories for this level of lexical description.

Moreover, in view of fulfilling the requirements of lexicons oriented to speech applications, as far as this descriptive layer is concerned, we also took into considerations the achievements obtained in an experiment of integration and unification of two Italian lexicons, the PAROLE morphological lexicon and the LCStar pronunciation lexicon. A *by-product* of this work has been, indeed, an agreed-on set of morpho-syntactic specifications which has been the base to the formulation of a nucleus of morpho-syntactic Pre Data Categories organized in sets of admitted combinations.

3.2 Syntactic Lexical Encoding

We deal here with all the information coming into play when describing the subcategorization frame. Synthetically, subcategorization corresponds to a set of possible syntactic structures (the head and its syntactic arguments, with their phrasal realization) associated with an entry (typically a verb, but also a so-called predicative noun, an adjective or an adverb).

A general approach similar to the one adopted for Morphosyntax was adopted for this layer of linguistic description. Here, the starting point are the Recommendations on Subcategorization (available at <http://www.ilc.pi.cnr.it/EAGLES96/synlex/synlex.html>, Sanfilippo *et al.* 1999). As with the morphosyntactic specifications, the EAGLES strategy for standardising subcategorisation in lexica was bottom-up, comparing a number of syntactic theories (GB, LFG, HPSG, Categorical Grammar and Dependency Grammar). In addition, the practices in 7 practical NLP lexica and 6 annotation schemata for tagging corpora were surveyed and used as input to a consensual definition of the specifications for subcategorisation information to be included in lexica. This constitutes the major contact point with the outcomes of the survey carried out in WP3. Moreover, the tracing of notions crucial in lexical entries to represent information that concurs to define and discriminate a syntactic structure draws inspiration from the PAROLE project, where the EAGLES guidelines have been concretely applied to a

set of twelve lexicons. In addition, the main input will be constituted by the experience accumulated in ISLE, where the syntactic basic notions have been investigated for the monolingual level, but also in view of developing a standardized infrastructure for multilingual lexical resources to be used in HLT applications.

3.3 Semantic Lexical Encoding

The nuclear building block of the whole semantic layer is the *word-sense*. All information types concurring to discriminate senses in a monolingual framework (or to direct towards a given translation in multilingual operations) are candidate Data Categories.

The starting point in the area of lexical semantics is the previous EAGLES guidelines.. They have been hence re-interpreted under the perspective of the underlying semantic model, trying to provide the set of information necessary to be dealt with at this level of representation.

In this light, the bulk of semantic information encoded in the SIMPLE lexicons is also re-examined and integrated with other realities. In consideration of their being built up on the EAGLES recommendations and with the flexible harmonized GENELEX model, the SIMPLE lexicons can be seen as pluri-lingual lexicons, thus constituting an example of *best practices*, Lenci *et al.* 2000a. Other approaches have been taken into account, since the notion of *word meaning*, central to semantics description, is not uncontroversial. In the lexicographic tradition, the word meaning is the *sense*, the unit resulting from the subdivision of the lemma in its readings. In lexicons *à la* GENELEX (or SIMPLE), the word meaning is represented by the SemU – the Semantic Unit – corresponding to the traditional notion of word sense and constituting the nuclear building block of the whole semantic description. It is the semantic unit that is linked to a given ontological type, it is the semantic unit that the semantic frame is associated to, and it is the semantic unit that, alternatively, works as the target and the source of all semantic relations. A different modality of representation resorts to the *synset*, the *set of synonyms* that constitutes the building block in WordNet and WordNet-like kind of resources (Vossen, 1999). During the years, WordNet has become an outstanding reality for the lexicon community, with WordNets dedicated to dozens of languages and used in a wide variety of applications. Thus, it is important to take WordNet and its basic structure into consideration, ensuring that the Registry cover all the data categories needed to easily map already encoded resources into the standard being designed.

All in all, the bootstrapping of semantic Data Categories has been done mainly on the basis of the ISLE Deliverable, carried out by the Computational Lexicon Working Group, where a compendium of the basic notions crucial to the creation of a standardized abstract schema for the construction of a lexical entry (the MILE) is provided (Calzolari *et al.* 2002).

4 Extracting possible Candidates for Lexical Data Categories

4.1 “Hooking” Data Categories to Meta-model Components and Classes

Morphology Extension

This is the information that in the meta-model is used to qualify the LexicalEntry class as concerns the morpho-syntactic category (*part_of_speech*) and, where appropriate, a sub-category, the MorphologicalFeature class, which is used to describe the entry and the inflected forms.

Syntactic Extension

We deal here with the set of Data Categories to be hooked to the NLP Syntax layer of the meta-model, i.e. all the information coming into play when describing the subcategorization frame. Synthetically, subcategorization corresponds to a set of possible syntactic structures (the head and its syntactic arguments, with their phrasal realization) associated with an entry

(typically a verb, but also a so-called predicative noun, an adjective or an adverb). This information is formalized under the SyntacticBehavior component. The central lexical element of the syntactic model is the SyntacticFrame and data categories are searched for to specifically decorate SyntacticPosition, SyntacticActant and SyntacticFeature elements.

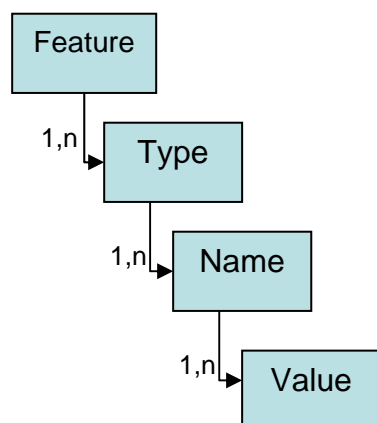
Semantic Extension

Data Categories here are intended to be attached to the SemanticFeature, Relation, SynsetRelation, PredicateRelation and Argument Classes. The central lexical element of the semantic extension is PredicateRepresentation, where the type of link between a Sense and its Predicate is made explicit.

4.2 Formalization of Candidate Data Categories

This section describes the modelling, formalization and instantiation of the collected lexical properties. This activity is deemed suitable to pave the way to the next step of Data Category design. Beyond the objectives designed in the Technical Annex for this Task, i.e. to deal only with linguistic content of lexical categories, we tried to define an abstract and unique meta-model capable of describing and representing information in a formal way, thus making it more computationally manageable.

Two basic expressive modalities are used in the lexical model to represent linguistic properties across the different layers of description: *features* and *relations*. A unique abstract meta-model has been identified to represent them with the following structural elements: the Element Feature or a Relation can be characterized by one to more Type(s), which on its turn is qualified by one to more attribute(s) (called Name in the model), further specified by one to more values (Value). The structural model can be depicted as follows.



This meta-model has constituted the base to develop four different XML schemas specialized to give formal representation of the Morphological, Syntactic and Semantic Features and Semantic Relations over the different lexical layer.

Below, the XML formalization, which follows the 3WC XML schema, is presented for the element MorphologicalFeature and Semantic Relation.

4.2.1 XML schema of Morphological Feature

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
  <xs:element name="MorphologicalFeature">
    <xs:complexType>
  
```

```

                <xs:sequence>
                    <xs:element ref="MorphologicalFeatureTypes"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="MorphologicalFeatureName">
            <xs:complexType>
                <xs:sequence>
                    <xs:element ref="name"/>
                    <xs:element ref="description"/>
                    <xs:element ref="MorphologicalFeatureValues"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="MorphologicalFeatureNames">
            <xs:complexType>
                <xs:sequence>
                    <xs:element ref="MorphologicalFeatureName"
maxOccurs="unbounded"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="MorphologicalFeatureType">
            <xs:complexType>
                <xs:sequence>
                    <xs:element ref="type"/>
                    <xs:element ref="description"/>
                    <xs:element ref="MorphologicalFeatureNames"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="MorphologicalFeatureTypes">
            <xs:complexType>
                <xs:sequence>
                    <xs:element ref="MorphologicalFeatureType" maxOccurs="unbounded"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="MorphologicalFeatureValue">
            <xs:complexType>
                <xs:sequence>
                    <xs:element ref="featval" minOccurs="0"/>
                    <xs:element ref="value" minOccurs="0"/>
                    <xs:element ref="description"/>
                    <xs:element ref="example"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="MorphologicalFeatureValues">
            <xs:complexType>
                <xs:sequence>
                    <xs:element ref="MorphologicalFeatureValue"
maxOccurs="unbounded"/>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
        <xs:element name="description">
            <xs:simpleType>
                <xs:restriction base="xs:string">
                </xs:restriction>
            </xs:simpleType>
        </xs:element>
        <xs:element name="example">
            <xs:simpleType>
                <xs:restriction base="xs:string">
                </xs:restriction>
            </xs:simpleType>

```



```

</xs:element>
<xs:element name="featval">
  <xs:simpleType>
    <xs:restriction base="xs:string">
    </xs:restriction>
  </xs:simpleType>
</xs:element>
<xs:element name="name">
  <xs:simpleType>
    <xs:restriction base="xs:string">
    </xs:restriction>
  </xs:simpleType>
</xs:element>
<xs:element name="type">
  <xs:simpleType>
    <xs:restriction base="xs:string">
    </xs:restriction>
  </xs:simpleType>
</xs:element>
</xs:schema>

```

4.2.2 XML Schema for Syntactic Feature

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
  <xs:element name="SyntacticFeature">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="SyntacticFeatureTypes"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="SyntacticFeatureName">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="name"/>
        <xs:element ref="description"/>
        <xs:element ref="SyntacticFeatureValues" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="SyntacticFeatureNames">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="SyntacticFeatureName" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="SyntacticFeatureType">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="type" minOccurs="0"/>
        <xs:element ref="description" minOccurs="0"/>
        <xs:element ref="SyntacticFeatureNames"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="SyntacticFeatureTypes">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="SyntacticFeatureType" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="SyntacticFeatureValue">
    <xs:complexType>

```

```

        <xs:sequence>
            <xs:element ref="featvalue"/>
            <xs:element ref="description"/>
            <xs:element ref="example"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="SyntacticFeatureValues">
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="SyntacticFeatureValue" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="description" type="xs:string"/>
<xs:element name="example" type="xs:string"/>
<xs:element name="featvalue">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            </xs:restriction>
        </xs:simpleType>
    </xs:element>
<xs:element name="name" type="xs:string"/>
<xs:element name="type" type="xs:string"/>
</xs:schema>

```

4.2.3 XML Schema for Semantic Relation

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
    <xs:element name="Relation">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="RelationTypes"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="RelationName">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="name"/>
                <xs:element ref="description"/>
                <xs:element ref="RelationValues" minOccurs="0"/>
                <xs:element ref="example" minOccurs="0"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="RelationNames">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="RelationName" maxOccurs="unbounded"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="RelationType">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="type"/>
                <xs:element ref="description"/>
                <xs:element ref="RelationNames"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="RelationTypes">
        <xs:complexType>

```

```

        <xs:sequence>
            <xs:element ref="RelationType" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="RelationValue">
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="featvalue"/>
            <xs:element ref="description"/>
            <xs:element ref="example"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="RelationValues">
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="RelationValue" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="description" type="xs:string"/>
<xs:element name="example" type="xs:string"/>
<xs:element name="featvalue">
    <xs:simpleType>
        <xs:restriction base="xs:string">
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="name">
    <xs:simpleType>
        <xs:restriction base="xs:string">
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="type">
    <xs:simpleType>
        <xs:restriction base="xs:string">
        </xs:restriction>
    </xs:simpleType>
</xs:element>
</xs:schema>

```

4.2.4 XML Instantiations

The above XML schemas provide the base for extracting the collection of candidate data categories in the form of (flat) lists, linked to the XML elements they pertain to: for example, for the MorphologicalName element, example of restricted values are typical morphological features, such as Gender, Number, Case, etc. A definition (taken from the surveyed practices, if available) and an example can be provided. The lists of candidate linguistic information are presented in the following sections, equipped with information regarding the class they are intended to fill in the lexical model.


Except for the restrictions of values pertaining to a given element, the XML schema does not provide constraints and restrictions of application of different values depending on different Types and Names, i.e. they do not allow to extract feature bundles such as, for example, (Noun(Gender(masculine,feminine))). In order to supply this kind of allowed data category combinations, for each layer of linguistic representation, we produced four different (but still conformant to the common meta-model) XML instantiations based on the designed XML schemas. The XML instantiations constitute Annexes of the present deliverable and are to be seen as the central outcome of the Task reported here.

5 List of values for morpho-syntactic features

5.1 Values

Below possible values of MorphosyntacticFeatureValue are listed. They correspond to possible data categories to express the **value** in the attribute-value pair characterizing the MorphosyntacticFeature Class of the Lexical Model.

element **featval**

diagram	
type	restriction of xs:string
properties	content simple
used by	element MorphologicalFeatureValue
source	<pre> <xs:element name="featval"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="1"/> <xs:enumeration value="2"/> <xs:enumeration value="3"/> <xs:enumeration value="Value"/> <xs:enumeration value="abbreviation"/> <xs:enumeration value="accusative"/> <xs:enumeration value="acronym"/> <xs:enumeration value="active"/> <xs:enumeration value="adject"/> <xs:enumeration value="adjectival"/> <xs:enumeration value="adjective"/> <xs:enumeration value="adverbial"/> <xs:enumeration value="affirmative"/> <xs:enumeration value="answer"/> <xs:enumeration value="aorist"/> <xs:enumeration value="approx"/> <xs:enumeration value="aspect"/> <xs:enumeration value="attributive"/> <xs:enumeration value="attributively"/> <xs:enumeration value="auxiliary"/> <xs:enumeration value="both"/> <xs:enumeration value="bound"/> <xs:enumeration value="cardinal"/> <xs:enumeration value="circumposition"/> <xs:enumeration value="collect"/> <xs:enumeration value="common"/> <xs:enumeration value="comp-m"/> <xs:enumeration value="compararative"/> <xs:enumeration value="comparartive"/> <xs:enumeration value="comparative"/> <xs:enumeration value="compound"/> <xs:enumeration value="conditional"/> <xs:enumeration value="conjunctive"/> <xs:enumeration value="coordinating"/> <xs:enumeration value="copulative"/> <xs:enumeration value="count"/> <xs:enumeration value="dative"/> <xs:enumeration value="dative (bei crater)"/> <xs:enumeration value="definite"/> <xs:enumeration value="demonstrative"/> <xs:enumeration value="determiner"/> <xs:enumeration value="digit"/> <xs:enumeration value="direct"/> <xs:enumeration value="dual"/> <xs:enumeration value="elative"/> <xs:enumeration value="emphatic"/> <xs:enumeration value="esclamative"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>


<xs:enumeration value="ex-there"/>
<xs:enumeration value="existential"/>
<xs:enumeration value="feminine"/>
<xs:enumeration value="finit"/>
<xs:enumeration value="finite"/>
<xs:enumeration value="first"/>
<xs:enumeration value="flection"/>
<xs:enumeration value="foreign-word"/>
<xs:enumeration value="formula"/>
<xs:enumeration value="fractal"/>
<xs:enumeration value="full-art"/>
<xs:enumeration value="fused"/>
<xs:enumeration value="future"/>
<xs:enumeration value="general"/>
<xs:enumeration value="genitive"/>
<xs:enumeration value="gerund"/>
<xs:enumeration value="imperative"/>
<xs:enumeration value="imperfect"/>
<xs:enumeration value="indefinite"/>
<xs:enumeration value="indicative"/>
<xs:enumeration value="indifferent"/>
<xs:enumeration value="infin-m"/>
<xs:enumeration value="infinitive"/>
<xs:enumeration value="infinitive-zu"/>
<xs:enumeration value="initial"/>
<xs:enumeration value="instrumental"/>
<xs:enumeration value="int./rel."/>
<xs:enumeration value="int.rel."/>
<xs:enumeration value="interrogative"/>
<xs:enumeration value="invariant"/>
<xs:enumeration value="letter"/>
<xs:enumeration value="locative"/>
<xs:enumeration value="m-form"/>
<xs:enumeration value="main"/>
<xs:enumeration value="masculine"/>
<xs:enumeration value="mixed"/>
<xs:enumeration value="modal"/>
<xs:enumeration value="modifier"/>
<xs:enumeration value="multiple"/>
<xs:enumeration value="neg-m"/>
<xs:enumeration value="negation"/>
<xs:enumeration value="negative"/>
<xs:enumeration value="neuter"/>
<xs:enumeration value="no"/>
<xs:enumeration value="no-flection"/>
<xs:enumeration value="nominal"/>
<xs:enumeration value="nominative"/>
<xs:enumeration value="non-attributively"/>
<xs:enumeration value="non-finite"/>
<xs:enumeration value="non-initial"/>
<xs:enumeration value="non-wh"/>
<xs:enumeration value="norm-infinitive"/>
<xs:enumeration value="oblique"/>
<xs:enumeration value="ordinal"/>
<xs:enumeration value="participle"/>
<xs:enumeration value="particle"/>
<xs:enumeration value="passive"/>
<xs:enumeration value="past"/>
<xs:enumeration value="personal"/>
<xs:enumeration value="pluperfect"/>
<xs:enumeration value="plural"/>
<xs:enumeration value="positive"/>
<xs:enumeration value="possessive"/>
<xs:enumeration value="postmodifier"/>
<xs:enumeration value="postnominal"/>
<xs:enumeration value="postposition"/>
<xs:enumeration value="premodifier"/>
<xs:enumeration value="pronominal"/>
<xs:enumeration value="preposition"/>
<xs:enumeration value="present"/>
<xs:enumeration value="pronominal"/>
<xs:enumeration value="pronoun"/>
<xs:enumeration value="proper"/>
<xs:enumeration value="qualif"/>

	<pre> <xs:enumeration value="qualificative"/> <xs:enumeration value="quantitative"/> <xs:enumeration value="question"/> <xs:enumeration value="reflexive"/> <xs:enumeration value="relative"/> <xs:enumeration value="roman"/> <xs:enumeration value="second"/> <xs:enumeration value="short-art"/> <xs:enumeration value="simple"/> <xs:enumeration value="singular"/> <xs:enumeration value="special"/> <xs:enumeration value="specifier"/> <xs:enumeration value="strong"/> <xs:enumeration value="subjunctive"/> <xs:enumeration value="subordinating"/> <xs:enumeration value="subordination"/> <xs:enumeration value="subunctive"/> <xs:enumeration value="superlative"/> <xs:enumeration value="supine"/> <xs:enumeration value="symbol"/> <xs:enumeration value="third"/> <xs:enumeration value="unclassified"/> <xs:enumeration value="underspecified"/> <xs:enumeration value="verbal"/> <xs:enumeration value="vocative"/> <xs:enumeration value="vpat"/> <xs:enumeration value="weak"/> <xs:enumeration value="wh"/> <xs:enumeration value="with-finite"/> <xs:enumeration value="with-infinitive"/> <xs:enumeration value="yes"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>
--	--

5.2 Attributes

Below possible values of MorphosyntacticFeatureName are listed. They correspond to possible data categories to express the **attribute** in the attribute-value pair of the MorphosyntacticFeature Class in the Morphological Component of the Lexical Meta-Model.

element name

diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>MorphologicalFeatureName</u>
source	<pre> <xs:element name="name"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="ANIMATED"/> <xs:enumeration value="Attributes"/> <xs:enumeration value="CASE"/> <xs:enumeration value="CLITIC"/> <xs:enumeration value="COORDINATION-TYPE"/> <xs:enumeration value="DECLINE"/> <xs:enumeration value="DEFINITNESS"/> <xs:enumeration value="DEGREE"/> <xs:enumeration value="FINITNESS"/> <xs:enumeration value="FLECTION"/> <xs:enumeration value="FORM"/> <xs:enumeration value="FORMATION"/> <xs:enumeration value="FUNCTION"/> <xs:enumeration value="GENDER"/> </pre>

	<pre> <xs:enumeration value="INFLECTION"/> <xs:enumeration value="MODE"/> <xs:enumeration value="MODIFICATION-TYPE"/> <xs:enumeration value="MOOD"/> <xs:enumeration value="NEGATIVE"/> <xs:enumeration value="NUMBER"/> <xs:enumeration value="OWNER-GENDER"/> <xs:enumeration value="OWNER-NUMBER"/> <xs:enumeration value="PERSON"/> <xs:enumeration value="POLARITY"/> <xs:enumeration value="POSSESSOR"/> <xs:enumeration value="PRONOUN-FORM"/> <xs:enumeration value="REFERENT-TYPE"/> <xs:enumeration value="SUBORDINATION-TYPE"/> <xs:enumeration value="SYNTACTIC-TYPE"/> <xs:enumeration value="TENSE"/> <xs:enumeration value="TYPE"/> <xs:enumeration value="USE"/> <xs:enumeration value="VERBFORM"/> <xs:enumeration value="VOICE"/> <xs:enumeration value="WH-TYPE"/> <xs:enumeration value="ZU-INCL"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>
--	---

5.3 Types

Below possible values of MorphosyntacticFeatureType are listed. They correspond to possible data categories to express the **PoS** of the LexicalEntry Class in the Core Lexical Meta-Model.


element type

diagram	
type	restriction of xs:string
properties	content simple
used by	element MorphologicalFeatureType
source	<pre> <xs:element name="type"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="ABBREVIATION"/> <xs:enumeration value="ADJECTIVE"/> <xs:enumeration value="ADPOSITION"/> <xs:enumeration value="ADVERB"/> <xs:enumeration value="ARTICLE"/> <xs:enumeration value="CONJUNCTION"/> <xs:enumeration value="DETERMINER"/> <xs:enumeration value="INTERJECTION"/> <xs:enumeration value="NOUN"/> <xs:enumeration value="NUMERAL"/> <xs:enumeration value="PARTICLE"/> <xs:enumeration value="PRONOUN"/> <xs:enumeration value="PoS"/> <xs:enumeration value="RESIDUAL"/> <xs:enumeration value="VERB"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>

6 List of Values for Syntactic Features

These are syntactic features which are used to encode SyntacticPosition and SyntacticActant Classes of the syntactic lexical model.

element **featvalue**


diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>SyntacticFeatureValue</u>
source	<pre><xs:element name="featvalue"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="AdjectivalPhrase"/> <xs:enumeration value="AdverbialPhrase"/> <xs:enumeration value="DeterminerPhrase"/> <xs:enumeration value="DirectObject"/> <xs:enumeration value="IndirectObject"/> <xs:enumeration value="Modifier"/> <xs:enumeration value="NounPhrase"/> <xs:enumeration value="Object"/> <xs:enumeration value="ObliqueObject"/> <xs:enumeration value="PrepositionalPhrase"/> <xs:enumeration value="Subject"/> <xs:enumeration value="VerbPhrase"/> </xs:restriction> </xs:simpleType> </xs:element></pre>

7 List of Values for Semantic Features

7.1 Values

Below possible values of `SemanticFeatureValue` element are listed. They correspond to possible data categories to fill in the `SemanticFeature` Class of the Lexical Model. They are used to encode domain information, the linking between sense and an ontological type-system, meaning components and other kind of semantic features which can be used in different best practices to refine the encoding of semantic content. From the XML instantiations provided below, the typing of values presented here as a flat list of values can be derived: e.g. the following features bundles can be extracted, `Constitutive(Age(old,young,adult))`, where *old young adult* are possible value of the attribute *age* which is of constitutive type.

element `featvalue`

diagram	
type	restriction of <code>xs:string</code>
properties	content simple
used by	element <code>SemanticFeatureValue</code>
source	<pre> <xs:element name="featvalue"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="1"/> <xs:enumeration value="2"/> <xs:enumeration value="3"/> <xs:enumeration value="Abstract"/> <xs:enumeration value="Accounting"/> <xs:enumeration value="Acoustics"/> <xs:enumeration value="Acquire_Knowledge"/> <xs:enumeration value="Administrative_Law"/> <xs:enumeration value="Advertising"/> <xs:enumeration value="Aerospace_Engineering"/> <xs:enumeration value="Against"/> <xs:enumeration value="Agriculture"/> <xs:enumeration value="Air"/> <xs:enumeration value="Air_Conditioning"/> <xs:enumeration value="Air_Transport"/> <xs:enumeration value="Airforce"/> <xs:enumeration value="Alchemy"/> <xs:enumeration value="American_Football"/> <xs:enumeration value="Anatomy"/> <xs:enumeration value="Anesthesiology"/> <xs:enumeration value="Angling"/> <xs:enumeration value="Animal"/> <xs:enumeration value="Arable_Farming"/> <xs:enumeration value="Archaeology"/> <xs:enumeration value="Archery"/> <xs:enumeration value="Architecture"/> <xs:enumeration value="Area"/> <xs:enumeration value="Army"/> <xs:enumeration value="Around"/> <xs:enumeration value="Artifact_Food"/> <xs:enumeration value="Artifactual_Area"/> <xs:enumeration value="Artifactual_Material"/> <xs:enumeration value="Artwork"/> <xs:enumeration value="Astrology"/> <xs:enumeration value="Astronomy"/> <xs:enumeration value="Athletics"/> <xs:enumeration value="Audiovisual"/> <xs:enumeration value="Automobile_Engineering"/> <xs:enumeration value="Baby_Care"/> <xs:enumeration value="Backward"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>

<xs:enumeration value="Bacteriology"/>
<xs:enumeration value="Badminton"/>
<xs:enumeration value="Bakery"/>
<xs:enumeration value="Ballet"/>
<xs:enumeration value="Banking"/>
<xs:enumeration value="Baseball"/>
<xs:enumeration value="Basketball"/>
<xs:enumeration value="Basketry"/>
<xs:enumeration value="Beekeeping"/>
<xs:enumeration value="Billiards"/>
<xs:enumeration value="Biochemistry"/>
<xs:enumeration value="Bookbinding"/>
<xs:enumeration value="Botany"/>
<xs:enumeration value="Boxing"/>
<xs:enumeration value="Brewing"/>
<xs:enumeration value="Buddhism"/>
<xs:enumeration value="Building"/>
<xs:enumeration value="Building_Crafts"/>
<xs:enumeration value="Bullfighting"/>
<xs:enumeration value="Bus_Transport"/>
<xs:enumeration value="Butchery"/>
<xs:enumeration value="Canon_Law"/>
<xs:enumeration value="Car_Transport"/>
<xs:enumeration value="Cardiology"/>
<xs:enumeration value="Cards"/>
<xs:enumeration value="Cartography"/>
<xs:enumeration value="Cattle_Farming"/>
<xs:enumeration value="Cause_Act"/>
<xs:enumeration value="Cause_Aspectual"/>
<xs:enumeration value="Cause_Change_Location"/>
<xs:enumeration value="Cause_Natural_Transition"/>
<xs:enumeration value="Cause_Relational_Change"/>
<xs:enumeration value="Ceramics"/>
<xs:enumeration value="Cereal_Farming"/>
<xs:enumeration value="Change_Possession"/>
<xs:enumeration value="Change_of_Location"/>
<xs:enumeration value="Chemistry"/>
<xs:enumeration value="Chess"/>
<xs:enumeration value="Christianity"/>
<xs:enumeration value="Circus"/>
<xs:enumeration value="City_Planning"/>
<xs:enumeration value="Civil_Engineering"/>
<xs:enumeration value="Civil_Law"/>
<xs:enumeration value="Cleaning"/>
<xs:enumeration value="Clear"/>
<xs:enumeration value="Climbing"/>
<xs:enumeration value="Clothing"/>
<xs:enumeration value="Clothing_Industry"/>
<xs:enumeration value="Cognitive_Event"/>
<xs:enumeration value="Coking_Industry"/>
<xs:enumeration value="Commerce"/>
<xs:enumeration value="Commercial_Law"/>
<xs:enumeration value="Computing"/>
<xs:enumeration value="Concrete"/>
<xs:enumeration value="Constitutional_Law"/>
<xs:enumeration value="Container"/>
<xs:enumeration value="Cosmetics"/>
<xs:enumeration value="Creation"/>
<xs:enumeration value="Creative_Writing"/>
<xs:enumeration value="Cricket"/>
<xs:enumeration value="Crime"/>
<xs:enumeration value="Criminal_Law"/>
<xs:enumeration value="Croquet"/>
<xs:enumeration value="Crystallography"/>
<xs:enumeration value="Cuisine"/>
<xs:enumeration value="Cycling"/>
<xs:enumeration value="Cytology"/>
<xs:enumeration value="Dance"/>
<xs:enumeration value="Death"/>
<xs:enumeration value="Demography"/>
<xs:enumeration value="Dentistry"/>
<xs:enumeration value="Dermatology"/>
<xs:enumeration value="Diplomacy"/>
<xs:enumeration value="Distilling"/>

<xs:enumeration value="Down"/>
<xs:enumeration value="Drink"/>
<xs:enumeration value="Drugs"/>
<xs:enumeration value="Durative"/>
<xs:enumeration value="Dyeing"/>
<xs:enumeration value="Ear_Nose_Throat"/>
<xs:enumeration value="Earth"/>
<xs:enumeration value="Earth_Sciences"/>
<xs:enumeration value="Ecology"/>
<xs:enumeration value="Economics"/>
<xs:enumeration value="Electrical_Engineering"/>
<xs:enumeration value="Electricity"/>
<xs:enumeration value="Electronic_Engineering"/>
<xs:enumeration value="Embryology"/>
<xs:enumeration value="Employment"/>
<xs:enumeration value="Enology"/>
<xs:enumeration value="Entomology"/>
<xs:enumeration value="Equestrian_Sport"/>
<xs:enumeration value="Ethnology"/>
<xs:enumeration value="Exist"/>
<xs:enumeration value="Experience_Event"/>
<xs:enumeration value="Explicit"/>
<xs:enumeration value="Family_Planning"/>
<xs:enumeration value="Fashion"/>
<xs:enumeration value="Female"/>
<xs:enumeration value="Fencing"/>
<xs:enumeration value="Feudalism"/>
<xs:enumeration value="Film"/>
<xs:enumeration value="Finance"/>
<xs:enumeration value="Fire"/>
<xs:enumeration value="Firefighting"/>
<xs:enumeration value="Fishing"/>
<xs:enumeration value="Flavouring"/>
<xs:enumeration value="Flower_Growing"/>
<xs:enumeration value="For"/>
<xs:enumeration value="Forestry"/>
<xs:enumeration value="Formal"/>
<xs:enumeration value="Fortification"/>
<xs:enumeration value="Forward"/>
<xs:enumeration value="Freshwater_Fishing"/>
<xs:enumeration value="Fruit_and_Vegetables"/>
<xs:enumeration value="Furnishing"/>
<xs:enumeration value="Furniture"/>
<xs:enumeration value="Future"/>
<xs:enumeration value="Games"/>
<xs:enumeration value="Gardening"/>
<xs:enumeration value="Gas"/>
<xs:enumeration value="Genetics"/>
<xs:enumeration value="Geography"/>
<xs:enumeration value="Geology"/>
<xs:enumeration value="Geometry"/>
<xs:enumeration value="Geopolitical_Location"/>
<xs:enumeration value="Geopolitics"/>
<xs:enumeration value="Give_Knowledge"/>
<xs:enumeration value="Glassmaking"/>
<xs:enumeration value="Glazing"/>
<xs:enumeration value="Golf"/>
<xs:enumeration value="Government-Administration"/>
<xs:enumeration value="Graphic_Arts"/>
<xs:enumeration value="Gymnastics"/>
<xs:enumeration value="Hair"/>
<xs:enumeration value="Health"/>
<xs:enumeration value="Heating"/>
<xs:enumeration value="Hematology"/>
<xs:enumeration value="Heraldry"/>
<xs:enumeration value="Herpetology"/>
<xs:enumeration value="High"/>
<xs:enumeration value="Higher_Education"/>
<xs:enumeration value="Hinduism"/>
<xs:enumeration value="Histology"/>
<xs:enumeration value="History"/>
<xs:enumeration value="Home_Laundry"/>
<xs:enumeration value="Horology"/>
<xs:enumeration value="Horse_Racing"/>

<xs:enumeration value="Horseshoeing"/>
<xs:enumeration value="Hotel_Business"/>
<xs:enumeration value="House_Painting"/>
<xs:enumeration value="Human"/>
<xs:enumeration value="Human_Sciences"/>
<xs:enumeration value="Hunting_and_Shooting"/>
<xs:enumeration value="Hydrology"/>
<xs:enumeration value="Hygiene"/>
<xs:enumeration value="Ichthyology"/>
<xs:enumeration value="Implicit"/>
<xs:enumeration value="Inchoative"/>
<xs:enumeration value="Informal"/>
<xs:enumeration value="Inland_Waterway_Transport"/>
<xs:enumeration value="Instrument"/>
<xs:enumeration value="Insurance"/>
<xs:enumeration value="Intelligence"/>
<xs:enumeration value="International_Affairs"/>
<xs:enumeration value="International_Law"/>
<xs:enumeration value="Inward"/>
<xs:enumeration value="Islam"/>
<xs:enumeration value="Jewelry"/>
<xs:enumeration value="Judaism"/>
<xs:enumeration value="Kitchen_Equipment"/>
<xs:enumeration value="Knitting"/>
<xs:enumeration value="Law_Enforcement"/>
<xs:enumeration value="Leisure"/>
<xs:enumeration value="Librarianship"/>
<xs:enumeration value="Life_Sciences"/>
<xs:enumeration value="Linguistics"/>
<xs:enumeration value="Liquid"/>
<xs:enumeration value="Livestock_Farming"/>
<xs:enumeration value="Locksmithing"/>
<xs:enumeration value="Logic"/>
<xs:enumeration value="Low"/>
<xs:enumeration value="Magic_and_Withcraft"/>
<xs:enumeration value="Mail"/>
<xs:enumeration value="Male"/>
<xs:enumeration value="Mammology"/>
<xs:enumeration value="Management"/>
<xs:enumeration value="Maritime_Law"/>
<xs:enumeration value="Marketing"/>
<xs:enumeration value="Marriage"/>
<xs:enumeration value="Martial_Arts"/>
<xs:enumeration value="Masonry"/>
<xs:enumeration value="Mathematics"/>
<xs:enumeration value="Mechanical_Engineering"/>
<xs:enumeration value="Media"/>
<xs:enumeration value="Medicine"/>
<xs:enumeration value="Meeting"/>
<xs:enumeration value="Metallurgy"/>
<xs:enumeration value="Metereology"/>
<xs:enumeration value="Metrology"/>
<xs:enumeration value="Micro-organism"/>
<xs:enumeration value="Microscopy"/>
<xs:enumeration value="Military_Law"/>
<xs:enumeration value="Mineralogy"/>
<xs:enumeration value="Mithology"/>
<xs:enumeration value="Modal_Event"/>
<xs:enumeration value="Monarchy"/>
<xs:enumeration value="Money"/>
<xs:enumeration value="Move"/>
<xs:enumeration value="Music"/>
<xs:enumeration value="Mycology"/>
<xs:enumeration value="Natural_Substance"/>
<xs:enumeration value="Natural_Transition"/>
<xs:enumeration value="Navy"/>
<xs:enumeration value="Negative"/>
<xs:enumeration value="Nespaper_Publishing"/>
<xs:enumeration value="Neuroanatomy"/>
<xs:enumeration value="Neurology"/>
<xs:enumeration value="New"/>
<xs:enumeration value="No"/>
<xs:enumeration value="Non_Relational_Act"/>
<xs:enumeration value="Nuclear_Engineering"/>

```

<xs:enumeration value="Nuclear_Physics"/>
<xs:enumeration value="Obstetrics"/>
<xs:enumeration value="Oceanography"/>
<xs:enumeration value="Office_Equipment"/>
<xs:enumeration value="Oil_Industry"/>
<xs:enumeration value="Oncology"/>
<xs:enumeration value="Opening"/>
<xs:enumeration value="Opera"/>
<xs:enumeration value="Ophthalmology_Optomety"/>
<xs:enumeration value="Optics"/>
<xs:enumeration value="Ornithology"/>
<xs:enumeration value="Ortodox_Church"/>
<xs:enumeration value="Outward"/>
<xs:enumeration value="Packaging"/>
<xs:enumeration value="Paintmaking"/>
<xs:enumeration value="Paleobiology"/>
<xs:enumeration value="Palmistry"/>
<xs:enumeration value="Paperhanging"/>
<xs:enumeration value="Papermaking"/>
<xs:enumeration value="Parapsychology"/>
<xs:enumeration value="Past"/>
<xs:enumeration value="Pediatrics"/>
<xs:enumeration value="Pedology"/>
<xs:enumeration value="Penal_System"/>
<xs:enumeration value="Perception"/>
<xs:enumeration value="Perfumery"/>
<xs:enumeration value="Persistent"/>
<xs:enumeration value="Petrography"/>
<xs:enumeration value="Petrology"/>
<xs:enumeration value="Pets"/>
<xs:enumeration value="Pharmacy"/>
<xs:enumeration value="Philately"/>
<xs:enumeration value="Philosophy"/>
<xs:enumeration value="Photography"/>
<xs:enumeration value="Phrenology"/>
<xs:enumeration value="Physical_Science"/>
<xs:enumeration value="Physics"/>
<xs:enumeration value="Physiology"/>
<xs:enumeration value="Pig_Farming"/>
<xs:enumeration value="Plastering"/>
<xs:enumeration value="Plastics"/>
<xs:enumeration value="Plumbing"/>
<xs:enumeration value="Poetics"/>
<xs:enumeration value="Politics"/>
<xs:enumeration value="Polo"/>
<xs:enumeration value="Positive"/>
<xs:enumeration value="Pottery"/>
<xs:enumeration value="Poultry_Farming"/>
<xs:enumeration value="Present"/>
<xs:enumeration value="Presupposed"/>
<xs:enumeration value="Primary_and_secondary_Education"/>
<xs:enumeration value="Printing"/>
<xs:enumeration value="Private"/>
<xs:enumeration value="Protestantism"/>
<xs:enumeration value="Psychiatry"/>
<xs:enumeration value="Psychoanalysis"/>
<xs:enumeration value="Psychology"/>
<xs:enumeration value="Public"/>
<xs:enumeration value="Publishing"/>
<xs:enumeration value="Pyrotechnics"/>
<xs:enumeration value="Quarrying"/>
<xs:enumeration value="Radio-Television"/>
<xs:enumeration value="Radiology_and_Radiotherapy"/>
<xs:enumeration value="Rail_Transport"/>
<xs:enumeration value="Real_Estate"/>
<xs:enumeration value="Relational_Act"/>
<xs:enumeration value="Relational_Change"/>
<xs:enumeration value="Relational_State"/>
<xs:enumeration value="Restauration"/>
<xs:enumeration value="Retail"/>
<xs:enumeration value="Rhetoric"/>
<xs:enumeration value="Road_Transport"/>
<xs:enumeration value="Roman_Catholicism"/>
<xs:enumeration value="Roman_Law"/>

```

<xs:enumeration value="Roofing"/>
<xs:enumeration value="Rowing"/>
<xs:enumeration value="Rugby"/>
<xs:enumeration value="Sailing_Yachting_and_Boating"/>
<xs:enumeration value="Scouting"/>
<xs:enumeration value="Sculpture"/>
<xs:enumeration value="Sea_Fishing"/>
<xs:enumeration value="Sea_Transport"/>
<xs:enumeration value="Seismology"/>
<xs:enumeration value="Semiotic_Artifact"/>
<xs:enumeration value="Sewing"/>
<xs:enumeration value="Sex"/>
<xs:enumeration value="Shaving"/>
<xs:enumeration value="Sheep_Farming"/>
<xs:enumeration value="Ship_Building"/>
<xs:enumeration value="Shoemaking"/>
<xs:enumeration value="Shows"/>
<xs:enumeration value="Skiing"/>
<xs:enumeration value="Smoking"/>
<xs:enumeration value="Soapmaking"/>
<xs:enumeration value="Soccer"/>
<xs:enumeration value="Social_Action"/>
<xs:enumeration value="Social_Security"/>
<xs:enumeration value="Sociology"/>
<xs:enumeration value="Solid"/>
<xs:enumeration value="Speech_Act"/>
<xs:enumeration value="Sport"/>
<xs:enumeration value="Statistics"/>
<xs:enumeration value="Steel_Industry"/>
<xs:enumeration value="Strong"/>
<xs:enumeration value="Substance_Food"/>
<xs:enumeration value="Subway_Transport"/>
<xs:enumeration value="Surface_Treatment"/>
<xs:enumeration value="Surfing"/>
<xs:enumeration value="Surgery"/>
<xs:enumeration value="Surveying"/>
<xs:enumeration value="Swimming"/>
<xs:enumeration value="Tanning"/>
<xs:enumeration value="Taxation"/>
<xs:enumeration value="Telecommunications"/>
<xs:enumeration value="Temporary"/>
<xs:enumeration value="Tennis"/>
<xs:enumeration value="Terminative"/>
<xs:enumeration value="Textiles"/>
<xs:enumeration value="Theater"/>
<xs:enumeration value="Theology"/>
<xs:enumeration value="Tiling"/>
<xs:enumeration value="Tobacco_Industry"/>
<xs:enumeration value="Topography"/>
<xs:enumeration value="Town_and_Country_Planning"/>
<xs:enumeration value="Trucking"/>
<xs:enumeration value="Typography"/>
<xs:enumeration value="Unclear"/>
<xs:enumeration value="Underspecified"/>
<xs:enumeration value="Up"/>
<xs:enumeration value="Utilities"/>
<xs:enumeration value="Vegetal_Entity"/>
<xs:enumeration value="Vehicle"/>
<xs:enumeration value="Venery"/>
<xs:enumeration value="Versification"/>
<xs:enumeration value="Veterinary_Medicine"/>
<xs:enumeration value="Virology"/>
<xs:enumeration value="Viticulture"/>
<xs:enumeration value="Volcanology"/>
<xs:enumeration value="Washing"/>
<xs:enumeration value="Waste_Treatment"/>
<xs:enumeration value="Water"/>
<xs:enumeration value="Water_Sport"/>
<xs:enumeration value="Weak"/>
<xs:enumeration value="Wheelwrighting"/>
<xs:enumeration value="Woodworking"/>
<xs:enumeration value="Wrestling"/>
<xs:enumeration value="Yes"/>
<xs:enumeration value="Zoology"/>

```


<xs:enumeration value="adult"/>
<xs:enumeration value="electrical_Work"/>
<xs:enumeration value="mining-General"/>
<xs:enumeration value="old"/>
<xs:enumeration value="three_D_Location"/>
<xs:enumeration value="young"/>
</xs:restriction>
</xs:simpleType>
</xs:element>

```

7.2 Attributes

Below, the list presents all possible data categories to be used for expressing ontological upper nodes in a type-system, or to characterize the type of information provided by values. As specified above, the xml instantiations are crucial to make these flat lists meaningful since they qualify the type of information by means of features bundles.

element name

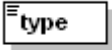
diagram	
type	restriction of xs:string
properties	content simple
used by	element SemanticFeatureName
source	<pre> <xs:element name="name"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Abstract_Entity"/> <xs:enumeration value="Act"/> <xs:enumeration value="Administrative"/> <xs:enumeration value="Affectedness"/> <xs:enumeration value="Age"/> <xs:enumeration value="Agriculture-Fishing-Forestry"/> <xs:enumeration value="Amount"/> <xs:enumeration value="Animal"/> <xs:enumeration value="Artifact"/> <xs:enumeration value="Arts"/> <xs:enumeration value="Aspect"/> <xs:enumeration value="Aspectual"/> <xs:enumeration value="Attitude"/> <xs:enumeration value="Audience"/> <xs:enumeration value="Body_Part"/> <xs:enumeration value="Business"/> <xs:enumeration value="Cause"/> <xs:enumeration value="Cause_Change"/> <xs:enumeration value="Change"/> <xs:enumeration value="Chemical"/> <xs:enumeration value="Cognitive_Fact"/> <xs:enumeration value="Collective"/> <xs:enumeration value="Color"/> <xs:enumeration value="Concrete_Entity"/> <xs:enumeration value="Connotation"/> <xs:enumeration value="Construction"/> <xs:enumeration value="Contact"/> <xs:enumeration value="Convention"/> <xs:enumeration value="Craft_Industry"/> <xs:enumeration value="Dermic"/> <xs:enumeration value="Dimension"/> <xs:enumeration value="Direction"/> <xs:enumeration value="Domain"/> <xs:enumeration value="Duration"/> <xs:enumeration value="Edible"/> <xs:enumeration value="Education"/> <xs:enumeration value="Elaborate"/> <xs:enumeration value="Emotive"/> </pre>

<xs:enumeration value="Emphasizer"/>
<xs:enumeration value="Event"/>
<xs:enumeration value="Explicitness"/>
<xs:enumeration value="Fictive"/>
<xs:enumeration value="Food"/>
<xs:enumeration value="Formality"/>
<xs:enumeration value="Geometric"/>
<xs:enumeration value="Group"/>
<xs:enumeration value="Habitat"/>
<xs:enumeration value="Health_and_Medicine"/>
<xs:enumeration value="Home_and_Garden"/>
<xs:enumeration value="Human"/>
<xs:enumeration value="Human_Group"/>
<xs:enumeration value="Iconic"/>
<xs:enumeration value="Information"/>
<xs:enumeration value="Insect"/>
<xs:enumeration value="Institution"/>
<xs:enumeration value="Intensifying_Property"/>
<xs:enumeration value="Intentionality"/>
<xs:enumeration value="Iterative"/>
<xs:enumeration value="Language"/>
<xs:enumeration value="Law"/>
<xs:enumeration value="Legal"/>
<xs:enumeration value="Liquid"/>
<xs:enumeration value="Living_Entity"/>
<xs:enumeration value="Location"/>
<xs:enumeration value="Locative"/>
<xs:enumeration value="Mammal"/>
<xs:enumeration value="Manner"/>
<xs:enumeration value="Manufacturing_Industry"/>
<xs:enumeration value="Material"/>
<xs:enumeration value="Measurable"/>
<xs:enumeration value="Meronym"/>
<xs:enumeration value="Military"/>
<xs:enumeration value="Mimetic"/>
<xs:enumeration value="Mineral"/>
<xs:enumeration value="Modal"/>
<xs:enumeration value="Moral_Standards"/>
<xs:enumeration value="Movement_of_Thought"/>
<xs:enumeration value="Natural"/>
<xs:enumeration value="Number"/>
<xs:enumeration value="Object-related"/>
<xs:enumeration value="Obligation"/>
<xs:enumeration value="Organic"/>
<xs:enumeration value="Organic_Object"/>
<xs:enumeration value="Part"/>
<xs:enumeration value="Partitive"/>
<xs:enumeration value="Pathological"/>
<xs:enumeration value="Phenomenon"/>
<xs:enumeration value="Physical_Object"/>
<xs:enumeration value="Physical_Power"/>
<xs:enumeration value="Physical_Property"/>
<xs:enumeration value="Plant"/>
<xs:enumeration value="Polarity"/>
<xs:enumeration value="Politics_and_Government"/>
<xs:enumeration value="Possible"/>
<xs:enumeration value="Presupposed"/>
<xs:enumeration value="Presupposition"/>
<xs:enumeration value="Professional"/>
<xs:enumeration value="Property"/>
<xs:enumeration value="Psychological_Property"/>
<xs:enumeration value="Psychological_Event"/>
<xs:enumeration value="Punctual"/>
<xs:enumeration value="Quality"/>
<xs:enumeration value="Reality"/>
<xs:enumeration value="Reciprocal"/>
<xs:enumeration value="Relational_Property"/>
<xs:enumeration value="Religion"/>
<xs:enumeration value="Representation"/>
<xs:enumeration value="Scalar"/>
<xs:enumeration value="Sciences"/>
<xs:enumeration value="Semiotic"/>
<xs:enumeration value="Service_Industry"/>
<xs:enumeration value="Sex"/>

	<pre> <xs:enumeration value="Shape"/> <xs:enumeration value="Sign"/> <xs:enumeration value="Social_Property"/> <xs:enumeration value="Sound"/> <xs:enumeration value="Sport_and_Leisure"/> <xs:enumeration value="State"/> <xs:enumeration value="Strength"/> <xs:enumeration value="Substance"/> <xs:enumeration value="Symbolic"/> <xs:enumeration value="Symptomatic"/> <xs:enumeration value="Temporal"/> <xs:enumeration value="Temporal_Property"/> <xs:enumeration value="Temporality_type"/> <xs:enumeration value="Themes"/> <xs:enumeration value="Time"/> <xs:enumeration value="Transport"/> <xs:enumeration value="Unit_of_Measurement"/> <xs:enumeration value="Voice_Quality"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>
--	--

7.3 Type


element type

diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>SemanticFeatureType</u>
source	<pre> <xs:element name="type"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Agentive"/> <xs:enumeration value="Constitutive"/> <xs:enumeration value="Distinctive"/> <xs:enumeration value="Domain"/> <xs:enumeration value="Entity"/> <xs:enumeration value="Extensional"/> <xs:enumeration value="Intensional"/> <xs:enumeration value="Telic"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>

8 List of Values of Predicate Features

Here are the features that are needed to qualify the Predicate Class.

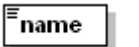
element **featvalue**

diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>PredicateValue</u>
source	<pre><xs:element name="featvalue"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Lexical"/> <xs:enumeration value="Primitive"/> </xs:restriction> </xs:simpleType> </xs:element></pre>

9 List of Values of Predicative Representation

Here are the features used to encode the relationships holding between a the Predicate Class and the Sense Class.


element **name**

diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>PredRepName</u>
source	<pre><xs:element name="name"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Adjective_Nominalization"/> <xs:enumeration value="Agent_Nominalization"/> <xs:enumeration value="Event_Nominalization"/> <xs:enumeration value="Instrument_Nominalization"/> <xs:enumeration value="Master"/> <xs:enumeration value="Patient_Nominalization"/> <xs:enumeration value="Process_Nominalization"/> <xs:enumeration value="State_Nominalization"/> <xs:enumeration value="Verb_Nominalization"/> </xs:restriction> </xs:simpleType> </xs:element></pre>

10 List of values of Semantic Role

This is a provisional list of possible semantic features to express semantic roles which is information used to qualify the Argument Class in the semantic model. The list here can be improved and augmented in the next release with Roles coming from FrameNet like lexicons.

element **featname**

diagram	
type	restriction of xs:string
properties	content simple
used by	element SemanticRoleValue
source	<pre> <xs:element name="featname"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="2_Participant"/> <xs:enumeration value="Agent"/> <xs:enumeration value="Head_Quantified"/> <xs:enumeration value="Kinship"/> <xs:enumeration value="Location/Direction/Origin"/> <xs:enumeration value="Patient"/> <xs:enumeration value="SoA_ARG"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>


11 List of Values for Semantic Relation

Below, possible values of semantic relations are presented. As already specified, the xml instantiations, by combining together information represented here under Value, Name and Type XML elements, will offer the correct interpretation and typing of the relations. This information is intended to fill the Relation Class of the lexical semantic meta-model.

In this version we only presented relations coming from the twelve SIMPLE lexicons. In the next release we will add relations coming from the WordNet lexicon family.

11.1 Relation Values

element **featvalue**

diagram	
type	restriction of xs:string
properties	content simple
used by	element RelationValue
source	<pre> <xs:element name="featvalue"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Agentive_Cause"/> <xs:enumeration value="Agentive_Experience"/> <xs:enumeration value="Agentive_prog"/> <xs:enumeration value="Antonym_comp"/> <xs:enumeration value="Antonym_grad"/> <xs:enumeration value="Antonym_mult"/> <xs:enumeration value="Caused_by"/> <xs:enumeration value="Causes"/> <xs:enumeration value="Concerns"/> <xs:enumeration value="Constitutive_activity"/> <xs:enumeration value="Contains"/> <xs:enumeration value="Derived_from"/> <xs:enumeration value="Has_as_colour"/> <xs:enumeration value="Has_as_effect"/> <xs:enumeration value="Has_as_member"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>


```

<xs:enumeration value="Has_as_part"/>
<xs:enumeration value="Has_as_property"/>
<xs:enumeration value="Indirect_telic"/>
<xs:enumeration value="Instrument"/>
<xs:enumeration value="Is_a_follower_of"/>
<xs:enumeration value="Is_a_member_of"/>
<xs:enumeration value="Is_a_part_of"/>
<xs:enumeration value="Is_in"/>
<xs:enumeration value="Is_the_ability_of"/>
<xs:enumeration value="Is_the_activity_of"/>
<xs:enumeration value="Is_the_habit_of"/>
<xs:enumeration value="Isa_relation"/>
<xs:enumeration value="Kinship"/>
<xs:enumeration value="Lives_in"/>
<xs:enumeration value="Made_of"/>
<xs:enumeration value="Measured_by"/>
<xs:enumeration value="Measures"/>
<xs:enumeration value="Object_of_the_activity"/>
<xs:enumeration value="Produced_by"/>
<xs:enumeration value="Produces"/>
<xs:enumeration value="Property_of"/>
<xs:enumeration value="Purpose"/>
<xs:enumeration value="Quantifies"/>
<xs:enumeration value="Related_to"/>
<xs:enumeration value="Relates"/>
<xs:enumeration value="Result_of"/>
<xs:enumeration value="Resulting_State"/>
<xs:enumeration value="Source"/>
<xs:enumeration value="Successor_of"/>
<xs:enumeration value="Typical_Location"/>
<xs:enumeration value="Typical_of"/>
<xs:enumeration value="Used_against"/>
<xs:enumeration value="Used_as"/>
<xs:enumeration value="Used_by"/>
<xs:enumeration value="Used_for"/>
<xs:enumeration value="Uses"/>
</xs:restriction>
</xs:simpleType>
</xs:element>

```

11.2 Names

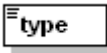
element name

diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>RelationName</u>
source	<pre> <xs:element name="name"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Activity"/> <xs:enumeration value="Agentive"/> <xs:enumeration value="Agentverb"/> <xs:enumeration value="Artifactual_agentive"/> <xs:enumeration value="Constitutive"/> <xs:enumeration value="DenominalVerbNoun"/> <xs:enumeration value="Direct_telic"/> <xs:enumeration value="Eventverb"/> <xs:enumeration value="Formal"/> <xs:enumeration value="Instrumental"/> <xs:enumeration value="Location"/> <xs:enumeration value="NounNoun"/> <xs:enumeration value="NounPropemoun"/> <xs:enumeration value="Nounadjective"/> <xs:enumeration value="Patientverb"/> </pre>

	<pre> <xs:enumeration value="Processverb"/> <xs:enumeration value="Property"/> <xs:enumeration value="Stateverb"/> <xs:enumeration value="Telic"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>
--	--

11.3 Types

element type

diagram	
type	restriction of xs:string
properties	content simple
used by	element <u>RelationType</u>
source	<pre> <xs:element name="type"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="Agentive"/> <xs:enumeration value="Constitutive"/> <xs:enumeration value="Derivational Relations"/> <xs:enumeration value="Formal"/> <xs:enumeration value="Telic"/> </xs:restriction> </xs:simpleType> </xs:element> </pre>

12 First Conclusions

In this version of the deliverable we presented the activities undertaken in LIRICS that seem to offer the most promising basis for the ISO standardisation work, i.e. preparing a nucleus of LIRICS data categories as input to the Data Category Registry under revision in ISO, in view of *decorating* the components of the lexical meta-model being created.

In the next version of the Deliverable, our main objective would be to offer language-specific instantiations of features, especially as far as the morphological layer is concerned, in the form of admitted combinations of Pos information and morphological features.

On this line, ILC is being implementing a tool that, starting from the morphological information presented in this Deliverable and in D3.1, allows to create a database of PoS with relevant morphological features for a specified language and/or a specified project. This tool is a perfect mean for describing the (constrained) relations between PoS of a given language and the morphological features associated with it. More specifically, the tool allows to formulate declarative rules that combine POS and morphological features for a specified language, thus defining constraints and dependencies between features and values in presence of a given feature or value. The resulting admitted combination(s) and constraints are stored in the database and can be made available to the community. An export function has been designed to map the content of the tool to XML, using at the beginning the XML DTD (the same that is presented in D3.1) and automatically extending it. The available information will be very valuable for every morpho-syntactic processing tool, since this information is supporting agreement check.

At present the database contains combinations and constraints for two Italian morphosyntactic lexical specifications, the IT-PAROLE and the IT-LCStar lexicons

The outcome of the tool is an addition to the kind of data and information that is foreseen as the outcome of WP2 and WP3, since it also addresses constraints that apply to the combination of PoS and morphological features.

The kind of constraints applying to the combination of POS and morphological features could be formulated according the ISO standard for feature structures

It should be evaluated if it is worthwhile to integrate this tool within the LORIA syntax tool for describing Data Category and/or in the framework of the LEXUS tool, which is being developed jointly with the LMF meta-model, in order to support the definition of hierarchies between attributes and values at the morpho-syntactic level.

13 Annex A – XML Instantiation of MorphologicalFeature Combinations

See file MorphologicalFeature.xml in the folder related to the deliverables

14 Annex B – XML Instantiation of Syntactic Features

See file syntacticfeature.xml in the folder related to the deliverables

15 Annex C – XML Instantiation of Semantic Features

See files predicate.xml, predicative_representation.xml, semanticrole.xml, semFeature.xml in the folder related to the deliverables

16 Annex D – XML Instantiation of Semantic Relations

See file relation.xml in the folder related to the deliverables

Bibliography (to be completed)

Bel, Calzolari and Monachini, 1995

Monachini 1995

Erjavec and Monachini 1997

Calzolari N., Grishman R., Palmer M., Bertagna F., Lenci A., Monachini M. 2002. **Introduction to the Final CLWG Guidelines**. ISLE Deliverable D1.1 - ISLE Computational Lexicon Working Group, draft version, Dec. 2002, Pisa.

Monachini M., Calzolari N. 1999. **Standardization in the lexicon**. In H. van Halteren (ed.), *Syntactic Wordclass Tagging*, London, Kluwer Publishers. 149-174.

Monachini *et al.* 2003

EAGLES Project: <http://www.ilc.cnr.it/EAGLES96/home.html>