

**SEVENTH FRAMEWORK PROGRAMME**  
**THEME 3**  
**Information and communication Technologies**

# **PANACEA Project**

**Grant Agreement no.: 248064**

**Platform for Automatic, Normalized Annotation and  
Cost-Effective Acquisition**  
of Language Resources for Human Language Technologies

## **Travelling Object definition for multilevel lexicon in PANACEA platform**

<b>Dissemination Level:</b>	Public
<b>Delivery Date:</b>	13/11/12
<b>Status – Version:</b>	V3.0
<b>Author(s) and Affiliation:</b>	Laura Rimell (UCAM), Núria Bel, Muntsa Padró (UPF), Valeria Quochi, Francesca Frontini, Monica Monachini, (CNR-ILC)

### **Relevant Panacea Deliverables**

<b>D6.1</b>	Technologies and Tools for Lexical Acquisition
<b>D6.2</b>	Integrated Final Version of the Components for Lexical Acquisition

*Travelling Object definition for multilevel lexicon in PANACEA platform*

This document is part of technical documentation generated in the PANACEA Project, Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition (Grant Agreement no. 248064).



This document is licensed under a Creative Commons Attribution 3.0 Spain License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/es/>.

Please send feedback and questions on this document to: [iulatri@upf.edu](mailto:iulatri@upf.edu)

TRL Group (Tecnologies dels Recursos Lingüístics), Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra (IULA-UPF)

## Table of Contents

1	Introduction .....	1
2	General issues about LMF.....	1
3	General information about the resource/lexicon .....	2
3.1	Features added automatically by the lexicon acquisition component(s).....	2
3.1.1	GlobalInformation .....	2
3.1.2	Lexicon.....	3
3.2	Features to be added manually by the lexicon “curator”/publisher .....	4
3.2.1	GlobalInformation .....	4
3.2.2	Lexicon.....	5
3.3	Example .....	5
4	PANACEA SubCat lexicon format.....	6
4.1	Spanish SCF.....	8
5	PANACEA Multiword lexicon format.....	10
6	PANACEA Lexical Classes lexicon format.....	12
6.1	Scored LMF example: .....	13
6.2	Filtered LMF example: .....	14

## 1 Introduction

In this document, the format of the lexica acquired and delivered by PANACEA platform is specified. The targeted standard format chosen for these travelling objects is basic Lexical Markup Framework, LMF (Francopoulo et al. 2008).

Below we provide LMF examples for the levels of concern in PANACEA, all of them based on LMF using [DTD in revision 16](#).

## 2 General issues about LMF

(extracted from LMF specifications revision 16<sup>1</sup>)

“The LMF core package describes the basic hierarchy of information of a lexical entry, including information on the form. The core package is supplemented by various resources that are part of the definition of LMF. These resources include:

- Specific data categories used by the variety of resource types associated with LMF, both those data categories relevant to the metamodel itself, and those associated with the

---

<sup>1</sup> [http://www.tagmatica.fr/lmf/iso\\_tc37\\_sc4\\_n453\\_rev16\\_FDIS\\_24613\\_LMF.pdf](http://www.tagmatica.fr/lmf/iso_tc37_sc4_n453_rev16_FDIS_24613_LMF.pdf)

---

extensions to the core package (for data categories here we understand the names of the XML elements, that correspond to the main building blocks of a lexical resource (e.g. LexicalEntry, Lemma, Sense etc), and of the mandatory attributes (e.g. id, entry, targets...)).

- The constraints governing the relationship of these data categories to the meta-model and to its extensions;
- Standard procedures for expressing these categories and thus for anchoring them on the structural skeleton of LMF and relating them to the respective extension models;
- The vocabularies used by LMF to express related informational objects for describing how to extend LMF through linkage to a variety of specific resources (extensions) and methods for analyzing and designing such linked systems.

[...]

LMF extensions are expressed in a framework that describes the reuse of the LMF core components (such as structures, data categories, and vocabularies) in conjunction with the additional components required for a specific resource.

[...]

LMF provides general structures and mechanisms for analyzing and designing new electronic lexical resources, but LMF does not specify the structures, data constraints, and vocabularies to be used in the design of specific electronic lexical resources. LMF also provides mechanisms for analyzing and describing existing resources using a common descriptive framework. For the purpose of both designing new lexical resources and describing existing lexical resources, LMF defines the conditions that allow the data expressed in any one lexical resource to be mapped to the LMF framework, and thus provides an intermediate format for lexical data exchange.”

### **3 General information about the resource/lexicon**

LMF requires two high level, general elements: Lexical Resource and Lexicon. They are supposed to be used to encode general and administrative information about the resource and the lexicons included in it such as size, date of creation, authors, availability, and so on. Here below we list the features that are/can be added automatically by the service and those that may be added manually if the final user wants to publish/distribute the resource.

The used features will be presented in tables, containing the name of the attribute, the kind of values it can have and whether it is mandatory or optional. The proposed features are already compliant to Metashare. Furthermore, most of them are traceable in IsoCat. In the tables below we give the link to these correspondences when available.

#### **3.1 Features added automatically by the lexicon acquisition component(s)**

##### **3.1.1 GlobalInformation**

In the table below we list the features that can be found under <GlobalInformation> in PANACEA generated lexica.

Attribute	Value	Status	IsoCat
resourceType	"lexicalConceptualResource"	mandatory	<a href="http://www.isocat.org/rest/dc/3806">http://www.isocat.org/rest/dc/3806</a>
lexicalConceptualResourceType	"lexicon"	mandatory	<a href="http://www.isocat.org/datcat/DC-2487">http://www.isocat.org/datcat/DC-2487</a>
conformanceToStandardsBestPractices	"LMF"	mandatory	
mediaType	"text"	mandatory	<a href="http://www.isocat.org/datcat/DC-2490">http://www.isocat.org/datcat/DC-2490</a>
mimeType	"text/xml"	mandatory	<a href="http://www.isocat.org/datcat/DC-2571">http://www.isocat.org/datcat/DC-2571</a>
characterEncoding	"UTF-8"	mandatory	<a href="http://www.isocat.org/datcat/DC-2564">http://www.isocat.org/datcat/DC-2564</a>
resourceName	open	optional	
modalityType	"written language"		<a href="http://www.isocat.org/datcat/DC-2490">http://www.isocat.org/datcat/DC-2490</a>

### 3.1.2 Lexicon

Those are the features that are related to <Lexicon> entry. Nevertheless, some PANACEA tools (such as the mergers) deliver this information in <GlobalInformation> in order to ease the extraction of the data to be used in the Metashare Metadata creator. The idea is to put all metadata in <GlobalInformation> and then convert automatically the LMF data inside this element to Metadata compliant with Metashare.

Attribute	Value	Status	IsoCat
originalSource	open (e.g. corpus name) *	optional	<a href="http://www.isocat.org/datcat/DC-2534">http://www.isocat.org/datcat/DC-2534</a>
domain	open *	optional	<a href="http://www.isocat.org/datcat/DC-2467">http://www.isocat.org/datcat/DC-2467</a>
size	open, value type: number.	mandatory	<a href="http://www.isocat.org/datcat/DC-2580">http://www.isocat.org/datcat/DC-2580</a>
sizeUnit	open suggested: "entry")	mandatory	<a href="http://www.isocat.org/datcat/DC-2583">http://www.isocat.org/datcat/DC-2583</a>
creationMode	open (suggested value list: "automatic manual mixed")	mandatory	<a href="http://www.isocat.org/datcat/DC-2516">http://www.isocat.org/datcat/DC-2516</a>
creationModeDetails	open	optional	<a href="http://www.isocat.org/datcat/DC-2511">http://www.isocat.org/datcat/DC-2511</a>
creationTool	open	optional	
encodingLevel	open	optional	
lingualityType	list: "monolingual bilingual multilingual"	mandatory	<a href="http://www.isocat.org/datcat/DC-2491">http://www.isocat.org/datcat/DC-2491</a>
languageID	open (the feat can be	mandatory	<a href="http://www.isocat.org">http://www.isocat.org</a>

	repeated)		<a href="http://www.isocat.org/datcat/DC-2482">g/datcat/DC-2482</a>
languageName	open (the feat can be repeated)	mandatory	<a href="http://www.isocat.org/rest/dc/2484">http://www.isocat.org/rest/dc/2484</a>
resourceName	open	optional	<a href="http://www.isocat.org/datcat/DC-2545">http://www.isocat.org/datcat/DC-2545</a>

\* Value should be passed to a wrapper from the Xces/Graf header.

### 3.2 Features to be added manually by the lexicon “curator”/publisher

Here we include only those recommended features, compliant to Metashare that may be added manually to those indicated above.

#### 3.2.1 GlobalInformation

Attribute	Value	Status	IsoCat
description	free text	optional	<a href="http://www.isocat.org/datcat/DC-2520">http://www.isocat.org/datcat/DC-2520</a>
availability	“available-restrictedUse”	optional	Similar to: <a href="http://www.isocat.org/datcat/DC-2453">http://www.isocat.org/datcat/DC-2453</a>
license	open (recommended “CC-BY-3.0”)	optional	<a href="http://www.isocat.org/datcat/DC-2457">http://www.isocat.org/datcat/DC-2457</a>
licenseurl	open	optional	<a href="http://www.isocat.org/datcat/DC-2457">http://www.isocat.org/datcat/DC-2457</a>
restrictionsOfUse	open (recommended “attribution”)	optional	
attributionText	open, free text	optional	
distributionAccessMedium	“downloadable”	optional	<a href="http://www.isocat.org/datcat/DC-2458">http://www.isocat.org/datcat/DC-2458</a>
foreseenUse	“nlpApplication”	optional	
owner	open, free text	optional	
email	email	recommended	<a href="http://www.isocat.org/datcat/DC-2521">http://www.isocat.org/datcat/DC-2521</a>
organisationName	open, free text	optional	<a href="http://www.isocat.org/datcat/DC-2459">http://www.isocat.org/datcat/DC-2459</a>
organisationShortName	open, acronym	optional	
departmentName	open, free text	optional	
projectName	open, free text	optional	<a href="http://www.isocat.org/datcat/DC-2537">http://www.isocat.org/datcat/DC-2537</a>
projectShortName	open	optional	<a href="http://www.isocat.org/datcat/DC-2536">http://www.isocat.org/datcat/DC-2536</a>
projectID	open	optional	<a href="http://www.isocat.org/datcat/DC-2535">http://www.isocat.org/datcat/DC-2535</a>
fundingType	open (see Metashare list)	optional	

### 3.2.2 Lexicon

Attribute	Value	Status	IsoCat
version	number	optional	<a href="http://www.isocat.org/datcat/DC-2547">http://www.isocat.org/datcat/DC-2547</a>
description	open, free text	optional	<a href="http://www.isocat.org/datcat/DC-2520">http://www.isocat.org/datcat/DC-2520</a>

### 3.3 Example

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM
"http://www.tagmatica.fr/lmf/DTD_LMF_REV_16.dtd" >
<LexicalResource dtdVersion="16">
  <!-- metadata for GlobalInfo are as far as possible compliant to MetaShare
  metadata and/or ISO Cat. Penny Labropoulou, Elina Desipri (eds)
  Documentation and User Manual of the META-SHARE Metadata Model. Date:
  06/03/2012. http://www.meta-net.eu/meta-share/metadata-schema/. However,
  metadata are given as flat lists; not organised into components
  Some of these metadata have been manually added for distributions. Other
  are automatically added by the acquisition tool.-->

  <GlobalInformation>
    <feat att="resourceType" val="lexicalConceptualResource" />
    <feat att="lexicalConceptualResourceType" val="lexicon" />
    <feat att="resourceName" val="PANACEA_SCF_IT_ENV" />
    <feat att="description" val="This is the PANACEA acquired SCF lexica
    for Italian and Environment domain" />
    <feat att="conformanceToStandardsBestPractices" val="LMF" />
    <feat att="mediaType" val="text" />
    <feat att="modalityType" val="writtenLanguage" />
    <feat att="characterEncoding" val="UTF-8" />
    <feat att="availability" val="available-restrictedUse" />
    <feat att="license" val="CC-BY-3.0" />
    <feat att="licenseurl" val=http://creativecommons.org/licenses/by/3.0/ />
    <feat att="restrictionsOfUse" val="attribution" />
    <feat att="attributionText" val="The Language Resource Group. CNR-ILC.
    Caselli et al.(2012)" />
    <feat att="distributionAccessMedium" val="downloadable" />
    <feat att="foreseenUse" val="nlpApplication" />
    <feat att="owner" val="The Language Resources Group" />
    <feat att="email" val="risorse@ilc.cnr.it" />
    <feat att="organisationName" val="Consorzio Nazionale delle Ricerche"/>
    <feat att="organisationShortName" val="CNR-ILC" />
    <feat att="departmentName" val="Istituto di Linguistica Computazionale
    A. Zampolli" />
    <feat att="projectName" val="Platform for Automatic, Normalised
    Annotation and Cost-Effective Acquisition of Language Resources for
    Human Language Technologies" />
    <feat att="projectShortName" val="PANACEA" />
  </GlobalInformation>
</LexicalResource>

```

```
<feat att="projectID" val="FP7-ICT-2009-4-248064" />
<feat att="fundingType" val="euFunds" />
<feat att="description" val="This is an automatically acquired and
created lexicon for verb subcategorisation frames for the Environment
domain." />
</GlobalInformation>
<Lexicon>
  <feat att="domain" val="Environment" />
  <feat att="encodingLevel" val="syntax" />
  <feat att="linguisticInformation" val="syntax-SubcatFrame" />
  <feat att="creationMode" val="automatic" />
  <feat att="creationModeDetails" val="induction" />
  <feat att="creationTool" val="SCF_Extractor_IT" />
  <feat att="creationDate" val="20120715" />
  <feat att="originalSource" val="PANACEA_McV2_ENV_IT" />
  <feat att="version" val="1.0" />
  <feat att="lingualityType" val="monolingual" />
  <feat att="languageID" val="it" />
  <feat att="languageName" val="Italian" />
  <feat att="size" val="370" />
  <feat att="sizeUnit" val="SyntacticBehaviour" />
  <LexicalEntry id="le_1">
    (...)
  </LexicalEntry>
</Lexicon>
</LexicalResource>
```

## 4 PANACEA SubCat lexicon format

We present and discuss an example with one lexical entry for the verb *accusare*. Two syntactic frames for this verb are described here as an example:

@SUBJ@OBJ (i.e. a syntactic frame with two arguments/complements: a subject and a direct object)

@SUBJ@OBJ@COMP-DI (i.e. a syntactic frame with three arguments/complements: a subject, a direct object, and a prepositional phrase complement introduced by the preposition *di*)

Some comments are contained in the appropriate comment field. <!-- -->

```
<Lexicon>
  <LexicalResource>
    <!-- LexicalEntry represents the verb main entry -->
    <LexicalEntry id="le_1">
      <feat att="partOfSpeech" val="v" /> <!-- it is recommended to use feat
      "partOfSpeech" to set the partOfSpeech of the entry -->
      <!--Lemma is obligatory in LMF and should be used to encode the morphosyntactic
      information applicable to the whole lemma -->
```



---

```
<Lemma>
  <feat att="writtenForm" val="accusare"/>
</Lemma>

<!-- SyntacticBehaviour contains the link btw the verb and the
subcategorisation frame(s) relevant for the verb at hand; auxiliary and frequency
information can also be encoded here. Because of the nature of the extracted data and
especially because of frequency information that usually refers to the verb-subcat pair
, we would recommend to have each syntactic behaviour point to only one
subcategorisation frame. But this is not constrained by LMF, and infact may not be
true nor a good practice for other types of lexicons -->

<SyntacticBehaviour id="sb_1" subcategorizationFrames="scf_11">
  <feat att="aux" val="avere"/>
  <feat att="freq" val="0.3"/>
  <!--"domain" is a label used to identify the domain in the case of manually
developed lexica or the corpus from where the lexicon has been extracted, in case it
has been automatically acquiered -->
  <feat att="domain" val="general"/>
</SyntacticBehaviour>

<SyntacticBehaviour id="sb_2" subcategorizationFrames="scf_22">
  <feat att="aux" val="avere"/>
  <feat att="freq" val="0.04"/>
  <feat att="domain" val="general"/>
</SyntacticBehaviour>
</LexicalEntry>

<!-- SubcategorizationFrame contains the description of the syntactic structures in
terms of syntactic arguments -->
<SubcategorizationFrame id="scf_11">
  <feat att="scf-type" val="@obj"/> <!-- the attribute scf-type here is used
simply to assign a lable to the whole SCF, which may be useful for evaluation
purposes -->
  <!-- SyntacticArgument specifies the properties of each single argument: e.g.
information about position, function, its optionality, syntactic realization, etc may be
expressed here. In the PANACEA TO it is recommended/mandatory to use function
(to express the grammatical function of the argument) and/or realisation (to express
somehow the surface realization of the argument) as the key/obligatory features for
syntctic arguments. They will be used e.g. for merging purposes. Optionality and
position are optional.-->
  <SyntacticArgument>
    <feat att="position" val="0"/>
    <feat att="optionality" val="yes"/>
    <feat att="function" val="subj"/>
    <feat att="realization" val="NP"/>
  </SyntacticArgument>
```

```
<SyntacticArgument>
  <feat att="position" val="1"/>
  <feat att="optionality" val="no"/>
  <feat att="function" val="obj"/>
  <feat att="realization" val="NP"/>
</SyntacticArgument>
</SubcategorizationFrame>

<SubcategorizationFrame id="scf_22">
  <feat att="scf-type" val="@obj@comp-di"/>
  <SyntacticArgument>
    <feat att="position" val="0"/>
    <feat att="optionality" val="yes"/>
    <feat att="function" val="subj"/>
    <feat att="realization" val="NP"/>
  </SyntacticArgument>
  <SyntacticArgument>
    <feat att="position" val="1"/>
    <feat att="optionality" val="no"/>
    <feat att="function" val="obj"/>
    <feat att="realization" val="NP"/>
  </SyntacticArgument>
  <SyntacticArgument>
    <feat att="position" val="2"/>
    <feat att="optionality" val="no"/>
    <feat att="function" val="comp"/>
    <feat att="realization" val="PP_di"/>
  </SyntacticArgument>
</SubcategorizationFrame>
</Lexicon>
</LexicalResource>
```

## 4.1 Spanish SCF

The general LMF structure used in the PANACEA lexica is common for all languages. Nevertheless, there is some information that can change in the different languages depending on the kind of information available in each case. Regarding SCFs, the concrete realization of the `SyntacticArgument` for Spanish is different than Italian. Thus, here we present some examples of Spanish `SyntacticArgument` and their contents:

```
<SyntacticArgument id="syn_arg_43_1">
  <feat att="position" val="1"/>
  <!-- for Spanish there is only one kind of complement, named "comp". The realization
  states the different kind of complements>
  <feat att="function" val="comp"/>
```

```
<!-- np: noun phrase -->
<feat att="realization" val="np"/>
</SyntacticArgument>

<SyntacticArgument id="syn_arg_47_1">
  <feat att="position" val="1"/>
  <feat att="function" val="comp"/>
  <!-- ppa: indirect object -->
  <feat att="realization" val="ppa"/>
</SyntacticArgument>

<SyntacticArgument id="syn_arg_43_2">
  <feat att="position" val="2"/>
  <feat att="function" val="comp"/>
  <!-- cp: object is a clause phrase, state also the kind of clause.-->
  <feat att="realization" val="cp"/>
  <!-- inf: infinitive clause -->
  <feat att="type" val="inf"/>
</SyntacticArgument>

<SyntacticArgument id="syn_arg_43_2">
  <feat att="position" val="2"/>
  <feat att="function" val="comp"/>
  <!-- cp: object is a clause phrase, state also the kind of clause -->
  <feat att="realization" val="cp"/>
  <!-- fin: finite clause -->
  <feat att="type+cl_type" val="fin"/>
</SyntacticArgument>

<SyntacticArgument id="syn_arg_43_3">
  <feat att="position" val="3"/>
  <feat att="function" val="comp"/>
  <!-- pp: prepositional complement, state which kind of object it has -->
  <feat att="realization" val="pp"/>
  <!-- concrete preposition that introduces the pp: -->
  <feat att="prep" val="a"/>
  <!-- the object of the pp can be "np" or "cp". If it is a "cp" the type of the "cp" is also
  stated -->
  <feat att="pp_object" val="cp"/>
  <feat att="pp_object+type" val="inf"/>
</SyntacticArgument>

<SyntacticArgument id="syn_arg_44_0">
```

```

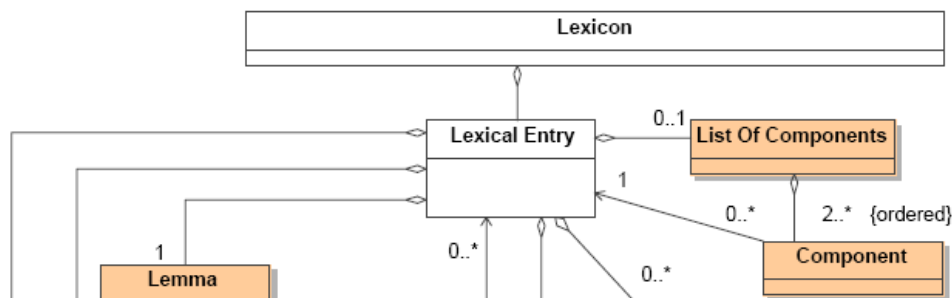
<!-- the subject is essentially equal to Italian, the realization can be "np" or "cp". If it
is "cp", it can have the same options than complement "cp" -->
<feat att="position" val="0"/>
<feat att="function" val="subj"/>
<feat att="optionality" val="yes"/>
<feat att="realization" val="np"/>
</SyntacticArgument>

```

## 5 PANACEA Multiword lexicon format

For the TO for MW lexica we propose the simplest representational means offered by LMF.

LMF has 3 possible extensions for representing MWE: the “Morphology”, the “Morphological patterns” and the “NLP multiword expression patterns” extensions, but the main components for their representations, available in all extensions, are List Of Components (aggregated to Lexical Entry) and Component (aggregated to List Of Components and pointing to Lexical Entry).



Already with these representational objects we may be able to describe also the internal composition and properties of MWE in a relatively simple way.

Lexical Entry may contain data categories that specify that the entry is a multiword and a data category specifying the POS (or MWE) pattern it instantiates.

See an example below:

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM
"http://www.tagmatica.fr/lmf/DTD_LMF_REV_16.dtd">
<LexicalResource dtdVersion="16">
  <!-- metadata for general info, to make compliant to MetaShare
  metadata -->
  <GlobalInformation>
    <feat att="originalSource" val="panacea_corpus_20111023"/>
    <feat att="crawlDate" val="2011"/>
    <feat att="size" val="20"/>
    <feat att="sizeUnit" val="words"/>
    <feat att="sizeUnitMultiplier" val="million"/>
    <feat att="author" val="CNR"/>

```

---

```
<feat att="creationMode" val="automatic"/>
  <feat att="creationModeDetails" val="acquisition"/>
</GlobalInformation>

<Lexicon>
  <feat att="type" val="Panacea_MWE_Lexicon"/>
  <feat att="language" val="Italian"/>

  <--here follows a list of single words that are used in the Multiword lexicon -->
  <LexicalEntry id="le_ea38d68660cd14356bbc858586790d1e">
    <feat att="entryType" val="Singleword"/>
    <feat att="absoluteFrequency" val="33675"/>
    <feat att="pos" val="s"/>
    <Lemma>
      <feat att="writtenForm" val="datore"/>
    </Lemma>
  </LexicalEntry>

  <LexicalEntry id="le_fbc2154ed38299eea3458847ababafe3">
    <feat att="entryType" val="Singleword"/>
    <feat att="absoluteFrequency" val="295032"/>
    <feat att="pos" val="s"/>
    <Lemma>
      <feat att="writtenForm" val="lavoro"/>
    </Lemma>
  </LexicalEntry>

  <LexicalEntry id="le_ad72734656bb0f51bdd5dfcfc35607f">
    <feat att="entryType" val="Singleword"/>
    <feat att="pos" val="e"/>
    <Lemma>
      <feat att="writtenForm" val="di"/>
    </Lemma>
  </LexicalEntry>

  <--here is the list of actual MWEs, with their features and list of components. Each
  component points to the single Lexical Entry as referred above; the MWEs contain the
  feature Domain to mark the fact that they belong to a special domain. -->
  <LexicalEntry id="le_254b8f8a92b5d4efdd22e057edae1874">
    <feat att="entryType" val="Multiword"/>
    <feat att="MWEPattern" val="s+e+s"/>
    <feat att="absoluteFrequency" val="32149"/>
    <feat att="logLikelihood" val="0.002367902295912881"/>
    <feat att="writtenform" val="datore di lavoro"/>
    <feat att="lemmaPair" val="datore-lavoro"/>
    <feat att="domain" val="labour"/>
```

```
<Lemma></Lemma>
<ListOfComponents>
  <Component entry="le_ea38d68660cd14356bbc858586790d1e">
    <feat att="rank" val="0"/>
    <feat att="pos" val="s"/>
    <feat att="lemma" val="datore"/>
    <feat att="writtenform" val="datore"/>
    <feat att="function" val="head"/>
  </Component>
  <Component entry="le_ad72734656bb0f51bdd5dfcfc35607f">
    <feat att="rank" val="1"/>
    <feat att="pos" val="e"/>
    <feat att="lemma" val="di"/>
    <feat att="writtenform" val="di"/>
  </Component>
  <Component entry="le_fbc2154ed38299eea3458847ababafe3">
    <feat att="rank" val="2"/>
    <feat att="pos" val="s"/>
    <feat att="lemma" val="lavoro"/>
    <feat att="writtenform" val="lavoro"/>
  </Component>
</ListOfComponents>
</LexicalEntry>
</Lexicon>
</LexicalResource>
```

## 6 PANACEA Lexical Classes lexicon format

We present two LMF examples for lexical semantic classes. Our proposal is to include the information regarding the semantic class under `<Sense>` entry. For the given examples, we will assume that the nouns are classified in three classes: eventive, human and location. The two different LMF samples that we present differ only on how the information of belonging or not belonging to the class is encoded. This depends on how the classifier is used:

- 1. Scored LMF:** each noun in the lexicon receives a score (between -1 and 1) for each class indicating the confidence of the classifier. If the score is higher than 0, the noun is considered a member of the class, scores close to 1 indicate high confidence of the classifier. If the score is below zero, it is considered a non-member of the class (with more confidence as closer to -1 is the score).
- 2. Filtered LMF:** instead of giving a score for the classification, the nouns receive a ternary classification: yes/no/unknown. The unknown elements are those that have been classified with small confidence by the classifier.

## 6.1 Scored LMF example:

```
<Lexicon>
  <LexicalEntry id="le_1">
    <!--Lemma is obligatory in LMF and should be used to encode the morphosyntactic
    information applicable to the whole lemma -->
    <Lemma>
      <feat att="writtenForm" val="boy"/>
    </Lemma>
    <!--use feat "partOfSpeech" to set the PoS of the entry -->
    <feat att="partOfSpeech" val="noun"/>
    <Sense>
      <!-- add one feat for each class and its assigned score -->
      <!-- "boy" belongs to the class human but not to location or eventive -->
      <feat att="event" val="-0.85"/>
      <feat att="hum" val="0.95"/>
      <feat att="loc" val="-0.75"/>

      <!-- "domain" is a label used to identify the corpus from where the lexicon has
      been extracted, in case it has been automatically acquired -->
      <feat att="domain" val="labour"/>
    </Sense>
  </LexicalEntry>
  <LexicalEntry id="le_2">
    <feat att="partOfSpeech" val="noun"/>
    <Lemma>
      <feat att="writtenForm" val="car"/>
    </Lemma>
    <Sense>
      <feat att="domain" val="labour"/>
      <!-- "car" does not belong to the class human nor eventive, but it has small score to
      belong to class location -->
      <feat att="event" val="-0.65"/>
      <feat att="hum" val="-0.75"/>
      <feat att="loc" val="0.25"/>
    </Sense>
  </LexicalEntry>
  <LexicalEntry id="le_3">
    <feat att="partOfSpeech" val="noun"/>
    <Lemma>
      <feat att="writtenForm" val="storm"/>
    </Lemma>
```

```
<Sense>
  <feat att="domain" val="labour"/>
  <!-- "storm" belongs to the class eventive but not to location or human -->
  <feat att="event" val="0.95"/>
  <feat att="hum" val="-0.75"/>
  <feat att="loc" val="-0.80"/>
</Sense>
</LexicalEntry>
</Lexicon>
```

## 6.2 Filtered LMF example:

```
<Lexicon>
  <LexicalEntry id="le_1">
    <feat att="partOfSpeech" val="noun"/>
    <!--Lemma is obligatory in LMF and should be used to encode the morphosyntactic
    information applicable to the whole lemma -->
    <Lemma>
      <feat att="writtenForm" val="boy"/>
    </Lemma>
    <!--use feat "partOfSpeech" to set the PoS of the entry -->
    <feat att="partOfSpeech" val="noun"/>
    <Sense>
      <!-- add one feat for each class and its assigned score -->
      <!-- "boy" belongs to the class human but not to location or eventive -->
      <feat att="event" val="no"/>
      <feat att="hum" val="yes"/>
      <feat att="loc" val="no"/>
      <!-- "domain" is a label used to identify the corpus from where the lexicon has
      been extracted, in case it has been automatically acquired -->
      <feat att="domain" val="labour"/>
    </Sense>
  </LexicalEntry>
  <LexicalEntry id="le_2">
    <feat att="partOfSpeech" val="noun"/>
    <Lemma>
      <feat att="writtenForm" val="car"/>
    </Lemma>
    <Sense>
      <feat att="domain" val="labour"/>
      <!-- "car" does not belong to the class human nor eventive, but the classifier is not
      sure about class location -->
      <feat att="event" val="no"/>
    </Sense>
  </LexicalEntry>
</Lexicon>
```



```
<feat att="hum" val="no"/>
  <feat att="loc" val="unknown"/>
</Sense>
</LexicalEntry>
<LexicalEntry id="le_3">
  <feat att="partOfSpeech" val="noun"/>
  <Lemma>
    <feat att="writtenForm" val="storm"/>
  </Lemma>
  <Sense>
    <feat att="domain" val="labour"/>
    <!-- "storm" belongs to the class eventive but not to location or human -->
    <feat att="event" val="yes"/>
    <feat att="hum" val="no"/>
    <feat att="loc" val="no"/>
  </Sense>
</LexicalEntry>
</Lexicon>
```