Report on use of LMF for representing WordNets

June 2008

Monica Monachini Claudia Soria
CNR ILC

Knowledge Yielding Ontologies for Transition-based Organization

ICT 211423

1 Purpose

The purpose of this report is to investigate suitability of LMF for the representation of WordNets.

2 Approach

We have selected one exemplifying synset and have represented it in LMF according to the various WordNet dialects. The synset chosen is {fire_2, flame_1, flaming_1} corresponding to WordNet 2.1 no. 13307342. We have selected its counterparts in Italian, Spanish, and Chinese and made the corresponding representations.
The information codified in these example entries is taken either from an export from the database or as it appears from the web interface. Of course, we need to make sure that all relevant information is represented. To this end, we would encourage that all partners try to implement an LMF-compliant synset representation based on their resources so that they can get a feeling of the representational potential and limitations of LMF.

In addition to the monolingual synset representations, we have another one illustrating a scenario where all monolingual lexicons are represented in the same global resource, a picture that could be suitable for the Kyoto WordNet database.

3 Introduction of LMF

A complete description of LMF can be found in the document available on the Kyoto Twiki site: http://www2.let.vu.nl/twiki/bin/view/Kyoto/WP02:SystemDesign .

4 Description of KYOTO representation format

4.1 Representation of Synsets

Each synset in an LMF lexicon is represented via a `<Synset>` object. The attribute `id` codifies its original offset:

```
<Synset id="N_13307342">
```

A `<Synset>` element contains the following elements:

- a number of `<feat>` elements
- `<Definition>`
- A number of `<SynsetRelation>`
- `<MonolingualExternalRef>`

4.1.1 <feat>

An element `<feat>` is used to encode Part of Speech information:

```
<feat att="PartOfSpeech" val="N"/>
```

This is an attribute-value pair by means of the standardized Data Category.

4.1.2 <Definition>

It allows to represent the glossa associated with each synset. This element in turn contains an element <Statement> that allows to represent examples of use associated with the synset:

```
<Definition>
   <feat att="gloss" val="the process of combustion of inflammable materials
producing heat and light and (often) smoke"/>
   <Statement>
      <feat att="example" val="fire was one of our ancestors' first
discoveries"/>
   </Statement>
</Definition>
```

4.1.3 <SynsetRelation>

Relations between synsets are codified by means of <SynsetRelation> elements, one per relation. The attribute targets contains the ID values of the synsets that are target of the relation. The particular relation type (ex., hypernym, meronym, domain, etc.) is expressed as a value of the attribute type of a <feat> element:

```
<SynsetRelation targets="N_13277364">
    <feat att="type" val="has_hypernym"/>
</SynsetRelation>
```

4.1.4 <MonolingualExternalRef>

According to LMF, this object represents a relationship between a synset instance and an external system, be it a knowledge organisation system or a terminological repository. The externalSystem and externalReference recommended attributes of the <feat> element allow to encode, respectively, the name of the external system and the specific relevant nodes in the given external system. We have used this element to express information concerning linking to an ontology, where this info was available. This is the case, for instance, of the representations of Italian, Spanish and Chinese synsets.
The following is an example taken from Spanish WordNet:

```
<MonolingualExternalRef>
   <feat att="externalSystem" val="top_ontology"/>
   <feat                                              att="externalReference"
val="Dynamic_Experience_Phenomenal_Physical_UnboundedEvent"/>
   <feat att="externalSystem" val="SUMO_ontology"/>
   <feat att="externalReference" val="Combustion"/>
</MonolingualExternalRef>
```

For English WordNet 2.1, we have used it to encode the information about the lexicographic file.

```
<MonolingualExternalRef>
   <feat att="externalSystem" val="lexicographic_file"/>
   <feat att="externalReference" val="noun.process"/>
</MonolingualExternalRef>
```

4.2 Representation of variants

According to LMF format, synset variants are expressed outside the <synset> element, by means of the <LexicalEntry> object, one for each variant.

A <LexicalEntry> contains the following elements:

- <Lemma>
- <Sense>

4.2.1 <Lemma>

This can be left empty, but is an obligatory child of the parent <LexicalEntry>.

4.2.2 <Sense>

It allows to encode the sense number and to which synset the variant belongs.

This is the representation of the variants of the English WN2.1 synset.

```
<LexicalEntry id="fire">
   <Lemma> </Lemma>
   <Sense id="fire_2" synset="N_13307342"/>
</LexicalEntry>
<LexicalEntry id="flame">
   <Lemma></Lemma>
   <Sense id="flame_1" synset="N_13307342"/>
</LexicalEntry>
<LexicalEntry id="flaming">
   <Lemma></Lemma>
   <Sense id="flaming_1" synset="N_13307342"/>
</LexicalEntry>
```

4.3 Representation of ILI linking

For WordNets other than Princeton WordNet, it is necessary to represent linking to the ILI.
In LMF, this is done by means of the Multilingual Notation Extension (p. 49). This package provides a means to encode multilingual information and it is designed as an independent package, in order not to overload the representation of monolingual lexicons. The model is based on the notion of "Axes" that link synsets pertaining to different languages. For the purposes of creating a grid of WordNets linked via ILI, the most appropriate device is the <SenseAxis>, since it is specifically designed to implement approaches based on an interlingual pivot.

4.3.1 <SenseAxis>

This element represents the relationships between different closely related senses in different languages. It may refer to an external knowledge representation system or to an external lexical resource. Any <SenseAxis> element groups together monolingual synsets that correspond one to another. It is possible to specify the type of this correspondence by means of a <feat> element indicating, for instance, whether the equivalence is of the type "eq_synonym", "eq_near_synonym", etc.

So, for instance, the equivalence relation between Spanish {fuego_3, llama_1} synset and English {fire_2, flame_1, flaming_1} synset is expressed as follows:

```
<SenseAxis synsets="N_09686541 r_8253345">
   <feat att="type" val="eq_syn"/>
</SenseAxis>
```

Where the attribute synsets contains the identifiers of the Spanish and English synsets, and the following element <feat> specifies that the equivalence relation is of the synonymy kind.

4.4 Representation of multilingual linking

The same package used to represent ILI linking from monolingual wordnets can also be applied in a scenario where all wordnets point to the same interlingual resource and from there to an ontology. Thus, in our example at hand, the Italian, Spanish and Chinese synsets all point to the same English WN3.0 synset and this, in turn, points to an external ontology. ID Prefixes were introduced in order to disambiguate among (possible) identical numberings of synsets across languages.

```
<SenseAxis synsets="I_N_1251 S_N_09686541 C_Na_05231501 E_N_13480848">
   <feat att="type" val="eq_syn"/>
   <InterlingualExternalRef>
      <feat att="ExternalSystem" val="SUMO_ontology"/>
      <feat att="ExternalReference" val="Combustion"/>
   </InterlingualExternalRef>
</SenseAxis>
```

The above example also shows how the linking to an external ontology can be instantiated, in a manner similar to the one explained for <MonolingualExternalRef>. In this case, the object <InterlingualExternalRef> represents a relationship between a SenseAxis instance and an external system, such as an ontology. The externalSystem and externalReference recommended attributes of the <feat> elements allow to encode, respectively, the name of the external system and the specific relevant nodes in the given external system.

4.4.1 Comments

The separation between the representation of language-specific information and interlingual axis makes the system more powerful and flexible. From a monolingual point of view, it makes it easier to handle synsets. From the point of view of inter-wordnet correspondence, it allows to compact together information that would be scattered around different synsets.

5 Representation of English WordNet 2.1 {fire_2, flame_1, flaming_1} synset.

```
<!DOCTYPE LexicalResource SYSTEM "LMF_rev16.dtd">
<LexicalResource dtdVersion="16">
  <GlobalInformation>
    <feat att="label" val="ILC-CNR WordNet Representation of synset
N_13307342, WordNet2.1"></feat>
  </GlobalInformation>
  <Lexicon>
    <LexicalEntry id="fire">
      <Lemma></Lemma>
      <Sense id="fire_2" synset="N_13307342"/>
    </LexicalEntry>
    <LexicalEntry id="flame">
      <Lemma></Lemma>
      <Sense id="flame_1" synset="N_13307342"/>
    </LexicalEntry>
    <LexicalEntry id="flaming">
      <Lemma></Lemma>
      <Sense id="flaming_1" synset="N_13307342"/>
    </LexicalEntry>
    <LexicalEntry id="combustion">
      <Lemma></Lemma>
      <Sense id="combustion_1" synset="N_13277364"/>
    </LexicalEntry>
    <LexicalEntry id="burning">
      <Lemma></Lemma>
      <Sense id="burning_3" synset="N_13277364"/>
    </LexicalEntry>
    <Synset id="N_13307342">
      <feat att="PartOfSpeech" val="N"/>
      <Definition>
        <feat att="gloss" val="the process of combustion of inflammable
materials producing heat and light and (often) smoke"/>
        <Statement>
          <feat att="example" val="fire was one of our ancestors' first
discoveries"/>
        </Statement>
      </Definition>
      <SynsetRelation targets="N_13277364">
          <feat att="type" val="has_hypernym"/>
      </SynsetRelation>
      <MonolingualExternalRef>
      <!-- this is to encode a link to the ontology.  -->
          <feat att="externalSystem" val="lexicographic_file"/>
          <feat att="externalReference" val="noun.process"/>
      </MonolingualExternalRef>
    </Synset>
    <Synset id="N_13277364">
        <feat att="PartOfSpeech" val="N"/>
        <Definition>
            <feat att="gloss" val="a process in which a substance reacts
with oxygen to give heat and light"/>
        </Definition>
    </Synset>
 </Lexicon>
</LexicalResource>
```

6 Representation of Italian WordNet {fire, flame, flaming} synset.

```
<!DOCTYPE LexicalResource SYSTEM "LMF_rev16.dtd">
<LexicalResource dtdVersion="16">
<GlobalInformation>
<feat att="label" val="ILC-CNR WordNet Representation of ItalWordNet synset
N_1251"></feat>
</GlobalInformation>
<!-- beginning of Italian WordNet-->
<Lexicon>
   <LexicalEntry id="fuoco">
      <Lemma></Lemma>
      <Sense id="fuoco_1" synset="N_1251"/>
   </LexicalEntry>
   <LexicalEntry id="fiamma">
      <Lemma></Lemma>
      <Sense id="fiamma_1" synset="N_1251"/>
   </LexicalEntry>
   <LexicalEntry id="fenomeno">
      <Lemma></Lemma>
      <Sense id="fenomeno_1" synset="N_26410"/>
   </LexicalEntry>
   <Synset id="N_1251">
      <feat att="PartOfSpeech" val="N"/>
      <Definition>
         <feat att="gloss" val="il fenomeno visivo legato alla combustione;
il fenomeno legato alla combustione"/>
      </Definition>
      <SynsetRelation targets="N_26410">
         <feat att="type" val="has_hyperonym"/>
      </SynsetRelation>
      <MonolingualExternalRef>
      <!-- this is to encode a link to the ontology.  -->
         <feat att="externalSystem" val="top_ontology"/>
         <feat att="externalReference" val="Dynamic_Experience_Phenomenal"/>
         <feat att="externalSystem" val="SIMPLE_ontology"/>
         <feat att="externalReference" val="Phenomenon"/>
      </MonolingualExternalRef>
   </Synset>
   <Synset id="N_26410"></Synset>
</Lexicon>
<!-- end of Italian WordNet-->
<!-- beginning of the English WN 1.5-->
<Lexicon>
   <LexicalEntry>
      <Lemma></Lemma>
      <Sense id="fire_1" synset="r_8253345"/>
   </LexicalEntry>
   <LexicalEntry>
      <Lemma></Lemma>
      <Sense id="flame_1" synset="r_8253345"/>
   </LexicalEntry>
   <LexicalEntry>
      <Lemma></Lemma>
      <Sense id="flaming_1" synset="r_8253345"/>
   </LexicalEntry>
   <Synset id="r_8253345">
      <Definition>
         <feat att="gloss" val="combustion of inflammable materials producing
heat and light and (often) smoke"/>
      </Definition>
```

```
   </Synset>
</Lexicon>
<!-- end of English WordNet-->
<!-- beginning of the multilingual correspondence section -->
<SenseAxis id="sa_it_en_001" synsets="N_1251 r_8253345">
   <feat att="type" val="eq_syn"/>
</SenseAxis>
</LexicalResource>
```

7 Representation of Spanish WordNet 2.1 {fire, flame, flaming} synset.

```
<!DOCTYPE LexicalResource SYSTEM "LMF_rev16.dtd">
<LexicalResource dtdVersion="16">
<GlobalInformation>
<feat att="label" val="ILC-CNR Spanish WordNet 1.6 Representation"></feat>
</GlobalInformation>
<!-- warning: this representation tries to stay as close as possible to the
WN model -->
<!-- beginning of Spanish WordNet-->
   <Lexicon>
      <LexicalEntry id="fuego">
         <Lemma></Lemma>
         <Sense id="fuego_3" synset="N_09686541"/>
      </LexicalEntry>
      <LexicalEntry id="llama">
         <Lemma> </Lemma>
         <Sense id="llama_1" synset="N_09686541"/>
      </LexicalEntry>
      <LexicalEntry id="combustion">
         <Lemma></Lemma>
         <Sense id="combustion_1" synset="N_09686366"/>
      </LexicalEntry>
      <Synset id="N_09686541">
         <feat att="PartOfSpeech" val="N"/>
         <feat att="domain" val="factotum"/>
         <Definition>
             <feat att="gloss" val=" Combustión de materiales inflamables que
produce calor, luz y muy a menudo humo"/>
         </Definition>
         <SynsetRelation targets="N_09686366">
            <feat att="type" val="has_hyperonym"/>
         </SynsetRelation>
         <MonolingualExternalRef>
         <!-- this is to encode a link to the ontology.  -->
            <feat att="externalSystem" val="top_ontology"/>
            <feat                                        att="externalReference"
val="Dynamic_Experience_Phenomenal_Physical_UnboundedEvent"/>
            <feat att="externalSystem" val="SUMO_ontology"/>
            <feat att="externalReference" val="Combustion"/>
         </MonolingualExternalRef>
      </Synset>
      <Synset id="N_09686366"></Synset>
   </Lexicon>
   <!-- end of Spanish WordNet-->
   <!-- beginning of the English WN 1.5-->
   <Lexicon>
      <LexicalEntry>
         <Lemma></Lemma>
         <Sense id="fire_1" synset="r_8253345"/>
      </LexicalEntry>
      <LexicalEntry>
         <Lemma></Lemma>
         <Sense id="flame_1" synset="r_8253345"/>
      </LexicalEntry>
      <LexicalEntry>
         <Lemma></Lemma>
         <Sense id="flaming_1" synset="r_8253345"/>
      </LexicalEntry>
```

```
       <Synset id="r_8253345">
             <Definition>
                <feat att="gloss" val="combustion of inflammable materials
producing heat and light and (often) smoke"/>
             </Definition>
       </Synset>
   </Lexicon>
   <!-- end of English WordNet-->
<!-- beginning of the multilingual correspondence section -->
<SenseAxis id="sa_sp_en_001" synsets="N_09686541 r_8253345">
   <feat att="type" val="eq_syn"/>
</SenseAxis>
</LexicalResource>
```

8 Representation of Chinese WordNet 2.1 {fire, flame, flaming} synset.

```
<!DOCTYPE LexicalResource SYSTEM "LMF_rev16.dtd">
<LexicalResource dtdVersion="16">
<GlobalInformation>
<feat att="label" val="ILC-CNR WordNet Representation"></feat>
</GlobalInformation>
<!-- beginning of Chinese WordNet-->
<Lexicon>
   <LexicalEntry id="火_huo3">
      <Lemma></Lemma>
      <Sense id="huo3_1" synset="Na_05231501"/>
   </LexicalEntry>
   <LexicalEntry id="能 2_neng2">
      <Lemma></Lemma>
      <Sense id="能 2_neng2_1" synset="Na_05001001"/>
   </LexicalEntry>
   <Synset id="Na_05231501">
      <feat att="PartOfSpeech" val="Na"/>
      <Definition>
         <feat att="gloss" val="物體燃燒時‧生的光和熱。"/>
             <Statement>
                <feat att="example" val="人站在旁邊，就覺得熱，離遠些，就沒感覺。"/>
             </Statement>
      </Definition>
      <SynsetRelation targets="Na_05001001">
         <feat att="type" val="has_hyperonym"/>
      </SynsetRelation>
      <MonolingualExternalRef>
        <!-- this is to encode a link to the ontology.  Copied from Spanish;
not checked-->
        <feat att="externalSystem" val="SUMO_ontology"/>
        <feat att="externalReference" val="Combustion"/>
      </MonolingualExternalRef>
   </Synset>
   <Synset id="Na_05001001"></Synset>
</Lexicon>
<!-- end of Chinese WordNet-->
<!-- beginning of the English WN 1.6-->
<Lexicon>
   <LexicalEntry>
      <Lemma></Lemma>
      <Sense id="fire_1" synset="N_09686541"/>
   </LexicalEntry>
   <LexicalEntry>
      <Lemma></Lemma>
      <Sense id="flame_1" synset="N_09686541"/>
   </LexicalEntry>
   <LexicalEntry>
      <Lemma></Lemma>
      <Sense id="flaming_1" synset="N_09686541"/>
   </LexicalEntry>
   <Synset id="N_09686541">
      <Definition>
         <feat att="gloss" val="combustion of inflammable materials producing
heat and light and (often) smoke"/>
      </Definition>
   </Synset>
```

```
</Lexicon>
<!-- end of English  WordNet 1.6-->
<!-- beginning of the multilingual correspondence section -->
<SenseAxis synsets="Na_05231501 N_09686541">
   <feat att="type" val="eq_syn"/>
</SenseAxis>
</LexicalResource>
```

9 Multilingual Representation of English, Italian, Spanish and Chinese synsets

```
<!DOCTYPE LexicalResource SYSTEM "LMF_rev16.dtd">
<LexicalResource dtdVersion="16">
        <GlobalInformation>
                <feat att="label" val="ILC-CNR WordNet Representation"/>
        </GlobalInformation>
        <!-- warning: this representation tries to stay as close as possible to the WN model -->
        <!-- beginning of Italian WordNet-->
        <Lexicon>
                <LexicalEntry id="fuoco">
                        <Lemma>
  </Lemma>
                        <Sense id="fuoco_1" synset="I_N_1251"/>
                </LexicalEntry>
                <LexicalEntry id="fiamma">
                        <Lemma>
  </Lemma>
                        <Sense id="fiamma_1" synset="I_N_1251"/>
                </LexicalEntry>
                <LexicalEntry id="fenomeno">
                        <Lemma>
  </Lemma>
                        <Sense id="fenomeno_1" synset="I_N_26410"/>
                </LexicalEntry>
                <Synset id="I_N_1251">
                        <feat att="PartOfSpeech" val="N"/>
                        <Definition>
                                <feat att="gloss" val="il fenomeno visivo legato alla combustione; il
fenomeno legato alla combustione"/>
                                <Statement>
                                </Statement>
                        </Definition>
                        <SynsetRelation targets="I_N_26410">
                                <feat att="type" val="has_hyperonym"/>
                        </SynsetRelation>
                        <MonolingualExternalRef>
                                <!-- this is to encode a link to the ontology.  -->
                                <feat att="externalSystem" val="top_ontology"/>
                                <feat att="externalReference"
val="Dynamic_Experience_Phenomenal"/>
                                <feat att="externalSystem" val="SIMPLE_ontology"/>
                                <feat att="externalReference" val="Phenomenon"/>
                                <!-- the following expresses a link from the synset to the term node in
the raw term list-->
                                <feat att="externalSystem" val="TMF_terminology"/>
                                <feat att="externalReference" val="t_001"/>
                        </MonolingualExternalRef>
                </Synset>
```

```
		<Synset id="I_N_26410"/>
	</Lexicon>
	<!-- end of Italian WordNet-->
	<!-- beginning of the English WN 1.5-->
	<Lexicon>
		<LexicalEntry>
			<Lemma/>
			<Sense id="fire_1" synset="E_N_13480848"/>
		</LexicalEntry>
		<LexicalEntry>
			<Lemma/>
			<Sense id="flame_1" synset="E_N_13480848"/>
		</LexicalEntry>
		<LexicalEntry>
			<Lemma/>
			<Sense id="flaming_1" synset="E_N_13480848"/>
		</LexicalEntry>
		<Synset id="E_N_13480848">
			<Definition>
				<feat att="gloss" val="combustion of inflammable materials
producing heat and light and (often) smoke"/>
			</Definition>
		</Synset>
	</Lexicon>
	<!-- end of English WordNet-->
	<!-- beginning of Spanish  WN -->
	<Lexicon>
		<LexicalEntry id="fuego">
			<Lemma/>
			<Sense id="fuego_3" synset="S_N_09686541"/>
		</LexicalEntry>
		<LexicalEntry id="llama">
			<Lemma> </Lemma>
			<Sense id="llama_1" synset="S_N_09686541"/>
		</LexicalEntry>
		<LexicalEntry id="combustion">
			<Lemma/>
			<Sense id="combustion_1" synset="S_N_09686366"/>
		</LexicalEntry>
		<Synset id="S_N_09686541">
			<feat att="PartOfSpeech" val="N"/>
			<feat att="domain" val="factotum"/>
			<Definition>
				<feat att="gloss" val=" Combustión de materiales inflamables que
produce calor, luz y muy a menudo humo"/>
				<Statement>
				</Statement>
			</Definition>
			<SynsetRelation targets="S_N_09686366">
				<feat att="type" val="has_hyperonym"/>
			</SynsetRelation>
```

```
                        <MonolingualExternalRef>
                                <!-- this is to encode a link to the ontology.  -->
                                <feat att="externalSystem" val="top_ontology"/>
                                <feat att="externalReference"
val="Dynamic_Experience_Phenomenal_Physical_UnboundedEvent"/>
                                <feat att="externalSystem" val="SUMO_ontology"/>
                                <feat att="externalReference" val="Combustion"/>
                        </MonolingualExternalRef>
                </Synset>
                <Synset id="S_N_09686366"/>
        </Lexicon>
        <!-- end of Spanish  WN -->
        <!-- beginning of Chinese WordNet -->
        <Lexicon>
                <LexicalEntry id="火_huo3">
                        <Lemma>
  </Lemma>
                        <Sense id="huo3_1" synset="C_Na_05231501"/>
                </LexicalEntry>
                <LexicalEntry id="能2_neng2">
                        <Lemma>
  </Lemma>
                        <Sense id="能2_neng2_1" synset="C_Na_05001001"/>
                </LexicalEntry>
                <Synset id="C_Na_05231501">
                        <feat att="PartOfSpeech" val="Na"/>
                        <Definition>
                                <feat att="gloss" val="物體燃燒時﹒生的光和熱。"/>
                                <Statement>
                                     <feat  att="example"  val="人站在旁邊，就覺得熱，離遠些，
就沒感覺。"/>
                                </Statement>
                        </Definition>
                        <SynsetRelation targets="C_Na_05001001">
                                <feat att="type" val="has_hyperonym"/>
                        </SynsetRelation>
                        <MonolingualExternalRef>
                                <!-- this is to encode a link to the ontology.  Copied from Spanish; not
checked-->
                                <feat att="externalSystem" val="SUMO_ontology"/>
                                <feat att="externalReference" val="Combustion"/>
                        </MonolingualExternalRef>
                </Synset>
                <Synset id="C_Na_05001001"/>
        </Lexicon>
        <!-- end of Chinese WordNet -->
        <!-- beginning of the multilingual correspondence section -->
        <!-- original IDs need to be prefixed in order to disambiguate between possible identical
numbers among different->
        <SenseAxis synsets="I_N_1251 S_N_09686541 C_Na_05231501 E_N_13480848">
                <feat att="type" val="eq_syn"/>
```

```
            <InterlingualExternalRef>
                <!-- the following element can be used to encode a link to a shared ontology,
where available; for instance-->
                <feat att="ExternalSystem" val="SUMO_ontology"/>
                <feat att="ExternalReference" val="Combustion"/>
            </InterlingualExternalRef>
        </SenseAxis>
</LexicalResource>
```