

Multilingual Lexical Support for the SEMbySEM Project

Ingrid Falk¹³ Samuel Cruz-Lara²³ Nadia Bellalem²³ Tarik
Osswald¹³ Vincent Herrmann²³

¹Centre de Recherche INRIA Grand Est

²Nancy Université

³LORIA

Language Resources and Technology standards @ LREC 2010

Summary

- 1 Overview
- 2 The SEMbySEM project.
 - Linguistic needs in SEMbySEM
- 3 Linguistic information on the conceptual level.
 - Structure of the lexical ontology.
 - LIR in action: examples.
 - Implementation
- 4 Linguistic information on the visualisation level.
- 5 Conclusion

What this talk is about

An experience . . .

- ▶ in using standards for interoperability in the context of a project for managing complex systems,
- ▶ attaching lexical and linguistic information to a conceptual representation.

- 1 Overview
- 2 The SEMbySEM project.
 - Linguistic needs in SEMbySEM
- 3 Linguistic information on the conceptual level.
 - Structure of the lexical ontology.
 - LIR in action: examples.
 - Implementation
- 4 Linguistic information on the visualisation level.
- 5 Conclusion

Summary

- 1 Overview
- 2 The SEMbySEM project.**
 - Linguistic needs in SEMbySEM
- 3 Linguistic information on the conceptual level.
 - Structure of the lexical ontology.
 - LIR in action: examples.
 - Implementation
- 4 Linguistic information on the visualisation level.
- 5 Conclusion

SEMbySEM, European ITEA2 project, aim:

provide a framework for universal sensors management using **semantic representations**.

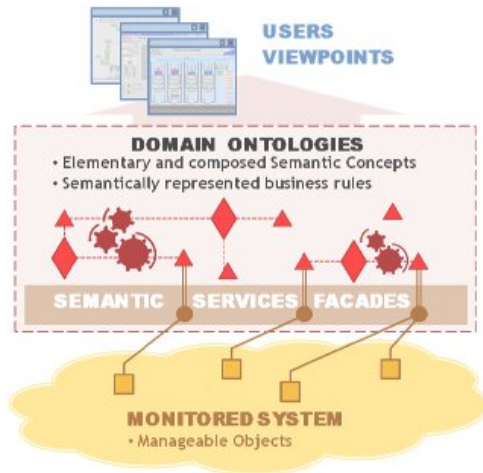
What is a sensor management system? Example: tramway network use case.

- ▶ supervise equipment of a tramway network.
- ▶ equipment includes sensors on/in stations: ticketing, traveller information terminals, intercom, CCTV, ...
- ▶ in praxis: ticket machine is faulty \rightsquigarrow SW sends alarm to operator \rightsquigarrow appropriate corrective action.



Figure: Sensor supervision and management system for transportation from ARC Informatique – a SEMbySEM partner.

System Architecture.



SEMbySEM needs multilingual linguistic information

- ▶ on the conceptual level
 - ▶ the conceptual representation must be visible in the designer's and in the user's language.
- ▶ on the GUI or visualization level
 - ▶ the HMI views must be designed in the end user's language.

Summary

- 1 Overview
- 2 The SEMbySEM project.
 - Linguistic needs in SEMbySEM
- 3 Linguistic information on the conceptual level.**
 - Structure of the lexical ontology.
 - LIR in action: examples.
 - Implementation
- 4 Linguistic information on the visualisation level.
- 5 Conclusion

Representing linguistic information for conceptual representations (ontologies)

Most common:

using `rdfs:label` and `rdfs:comment`.

But:

- ▶ restricted expressiveness,
- ▶ not user friendly,
- ▶ hardly reusable.

Other models:

- ▶ LIR, Linguistic Information Repository (Peters et al., 2009)
- ▶ LexInfo (Buitelaar et al., 2009)

Models for representing linguistic information: LIR vs. LexInfo

Similarities:

- ▶ linguistic information is stored in a lexical ontology,
- ▶ both models use LMF, the Lexical Markup Framework as building blocks.

Differences:

- ▶ lexical ontologies have different structure:

LexInfo emphasis on representation of properties, syntax ↔ semantics interface: subcategorization frames, predicate argument structure, mapping of syntactic and semantic arguments . . . ,

LIR more traditional lexicographic position: synonymy, translation (partial) equivalents

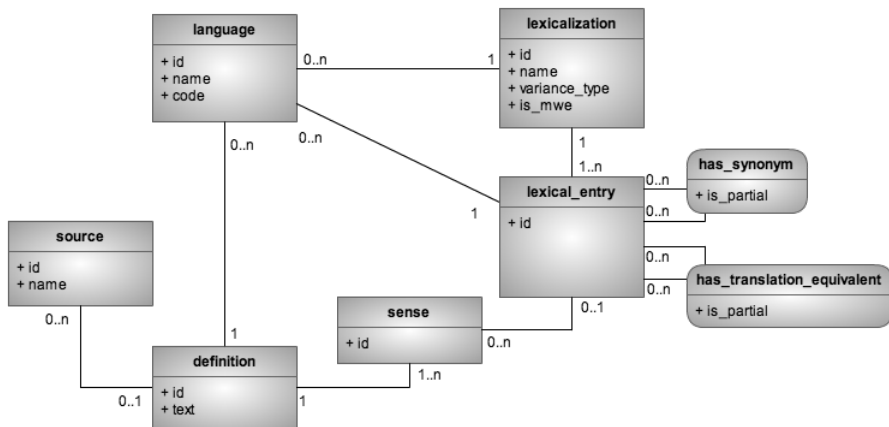
Lexical representation for SEMbySEM: LIR

We opted for LIR for the following reasons:

- ▶ LIR's lexicographic point of view seemed to fit the SEMbySEM needs better,
- ▶ project seemed more advanced and tested than LexInfo,
- ▶ LIR's alignment with other linguistic and lexicographic standards in addition to LMF: TMX, MLIF and XLIFF.

LIR for SEMbySEM:

a simplified LIR model implemented as a database:



Exact concept ↔ lexicalisation match

Concept in ontology:

```
<smc:Concept
  rdf:about="&sembysesem;#AssetTracking/Wagon">
  <lir:hasLexicalEntry rdf:resource="&lexo;#LE-1-En"
    xml:lang="eng"/>
  <lir:hasLexicalEntry rdf:resource="&lexo;#LE-1-Fr"
    xml:lang="fr"/>
</smc:Concept>
```

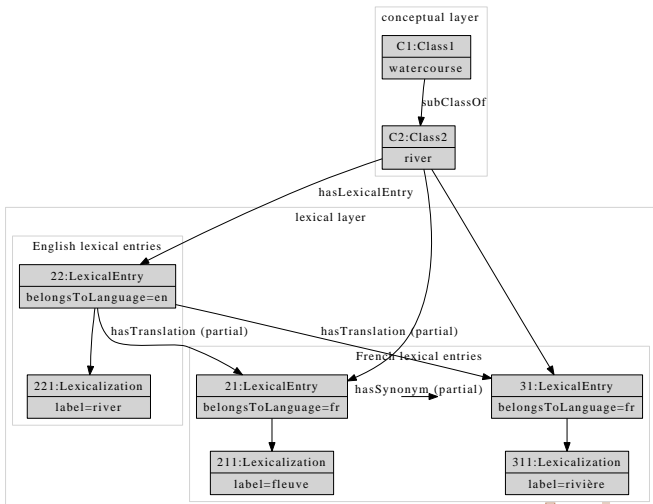
In the LIR lexical ontology:

```
<lir:LexicalEntry rdf:about="&lexo;#LE-1-En">
  <lir:partOfSpeech>noun</lir:partOfSpeech>
  <lir:belongsToLanguage rdf:resource="&lexo;#English"/>
  <lir:hasLexicalization rdf:resource="&lexo;#Lex-1-En"/>
  <lir:hasSense rdf:resource="&lexo;#Sense-1-En"/>
  <lir:hasTranslation rdf:resource="&lexonto;#LE-1-Fr"/>
</lir:LexicalEntry>
```

Exact concept ↔ lexicalisation match ctd.

- ▶ concept ↔ lexical entry link done via *hasLexicalEntry* in domain ontology.
- ▶ lexical entry describes the word *wagon* as an English noun,
- ▶ it's sense is given in a *Sense* instance by one or more definitions.
- ▶ a translation into French is given by the *hasTranslation* relation.

Conceptual mismatches.



Conceptual mismatches ctd.

Localisation choices made explicit:

- ▶ English label *river* lexicalised in English by lexical entry *river* ...
- ▶ in French by the two lexical entries *fleuve* and *rivière*,
- ▶ *fleuve* and *rivière* are partial synonyms,
- ▶ *fleuve* and *rivière* are both (partial) translations of *river*.

Multi-word expressions.

```
<smc:Concept rdf:about=
  "&sembysesem;#AssetTracking/WagonMovement_Notification"/>
```

in **SEMbySEM** one lexical entry corresponding to the entire expression
(*wagon movement notification*), marked *mwe*,

in **LIR** using LMF ListOfComponents construct, which allows to link
each word/component to a lexical entry.

But:

- ▶ syntactic structure \rightsquigarrow *kind of notification about movements of wagons*
- ▶ dependencies between components can be represented in **LexInfo**
through mapping of syntactic and semantic arguments ...
- ▶ ... it would be good if models could be combined.

Implementation issues

- ▶ API proposed by NeOn project too complex for our purposes,
- ▶ \rightsquigarrow lexicon realised as a database:
 - ▶ located at the LORIA,
 - ▶ can be exported to OWL or MLIF,
 - ▶ accessed via web services by the developer/maintainer of the domain ontology,
 - ▶ data transfer in MLIF format.

Summary

- 1 Overview
- 2 The SEMbySEM project.
 - Linguistic needs in SEMbySEM
- 3 Linguistic information on the conceptual level.
 - Structure of the lexical ontology.
 - LIR in action: examples.
 - Implementation
- 4 Linguistic information on the visualisation level.**
- 5 Conclusion

Linguistic information on visualisation level \equiv localisation task

Visual layer: end-user interfaces displaying and giving access to elements of semantic representation. *At design time:*

1. HMIs designed by experts, **language independent** in XUL,
 - ▶ XML User Interface Language developed by the Mozilla project,
 - ▶ **not a standard!**

but:

- ▶ **the most generic and re-usable we could find.**

2. **language dependent** data are provided in MLIF format.

At run time: XUL descriptions and MLIF data are combined to render the user-interface in the end-user's language.

Summary

- 1 Overview
- 2 The SEMbySEM project.
 - Linguistic needs in SEMbySEM
- 3 Linguistic information on the conceptual level.
 - Structure of the lexical ontology.
 - LIR in action: examples.
 - Implementation
- 4 Linguistic information on the visualisation level.
- 5 Conclusion**

Conclusion

We reported about efforts to provide linguistic and lexical information for a sensor supervision and management framework based on a semantic representation.

Linguistic and lexical information at 2 levels:

conceptual lexical ontology based on LMF and other lexical standards aligned with conceptual representations:

- conceptual and lexical representations kept separately,
- standards allow flexible and accurate coupling.

visualisation localisation of end-user interfaces:

- language dependent and independent elements kept separately,
- despite a certain lack of adequate standards.

Thank you!