

ISO working meeting in Berlin

Gil FRANCOPOULO (INRIA-Loria)

gil.francopoulo@wanadoo.fr

Opinion regarding to 4 different points about the work in progress:

- 1) **Distinction of MRD vs NLP lexicon**
- 2) **Core model : suggestion for improvement**
- 3) **LMF/NLP extension presentation**
- 4) **What's next ?**

1) MRD vs NLP lexicons

- What is common and what is different
=> avoid confusion



- **What is common ?**
 - describe words in 1 or # languages
 - data organized in a network of small granules of information, alphabetical list being one of the possible views.
 - maintained by human beings with sometimes the help of programs

● What is different ?

User

MRD human being
NLP program

Precision

MRD can be loosely defined and filled (readable)
NLP need to be systematic (processable)

Structural organisation of information

MRD general rules + exceptions to be interpreted
NLP generic schema is preferred

Linguistic material

MRD light morphology, etymology, mostly only single words
NLP full morphology, no etymology, single words + MWE
and often complex syntax and semantics

Multilinguality

MRD actors specialised in one language + agreement for other L
NLP all companies process # languages (in Europe)

Lexical creation and maintenance

2 different (small) communities
the employers are not the same

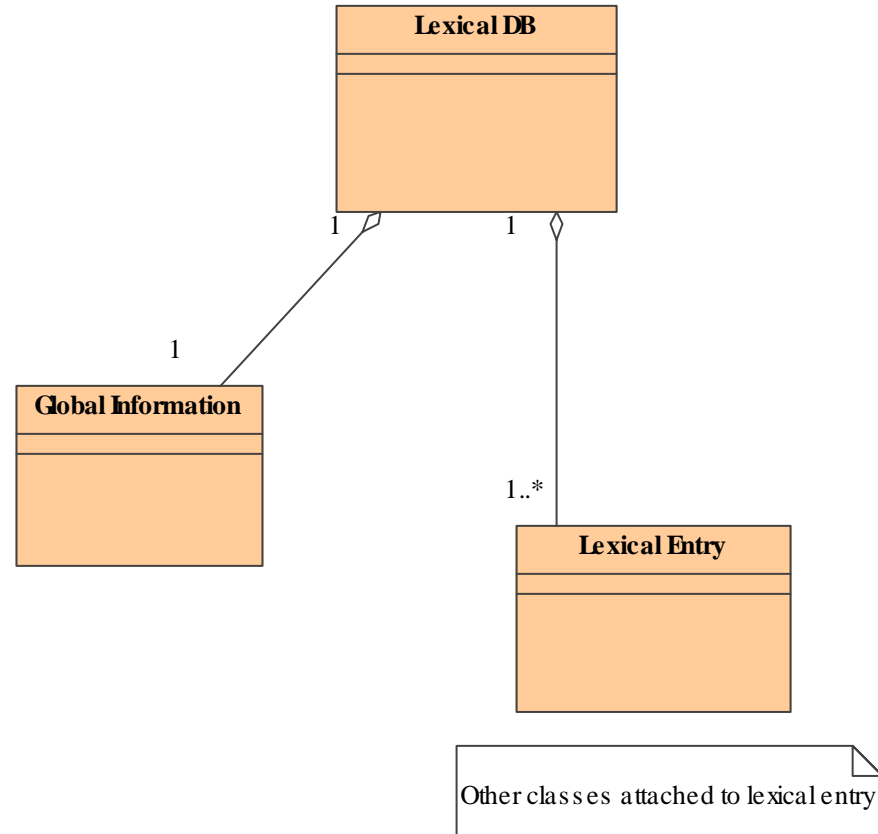
● May be the confusion comes from links between MRD & NLP ?

MRD->NLP A lot of NLP lexicons include data taken from MRD lexicons
(e.g. textual definition used for maintenance)

NLP->MRD Modern MRD lexicographers use NLP technics

2) Core model (suggestion for improvement)

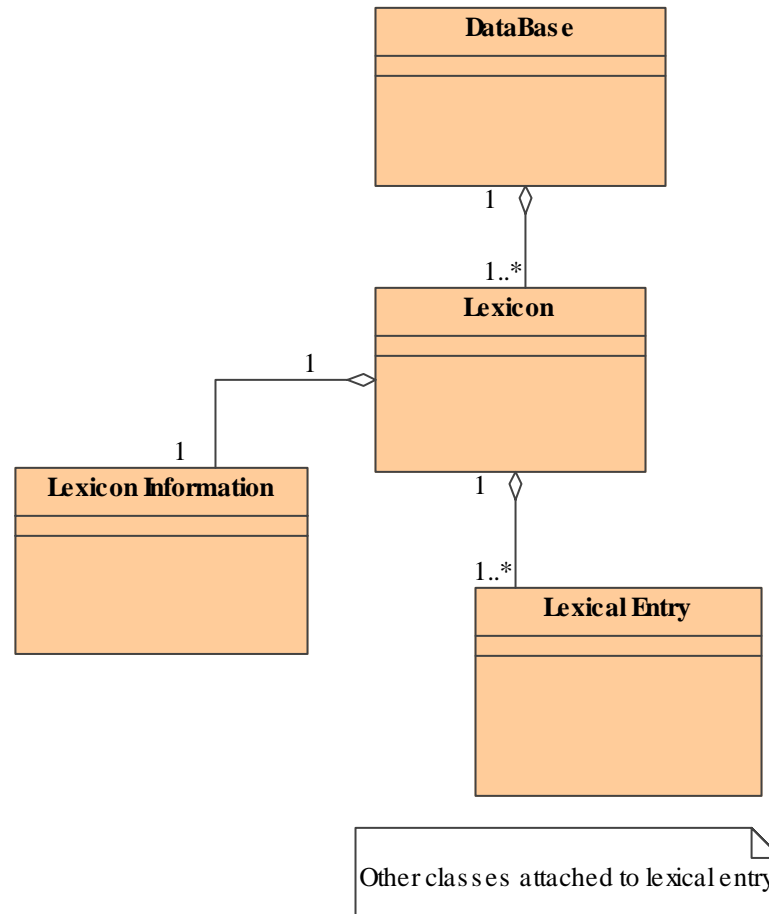
- Today (recall)



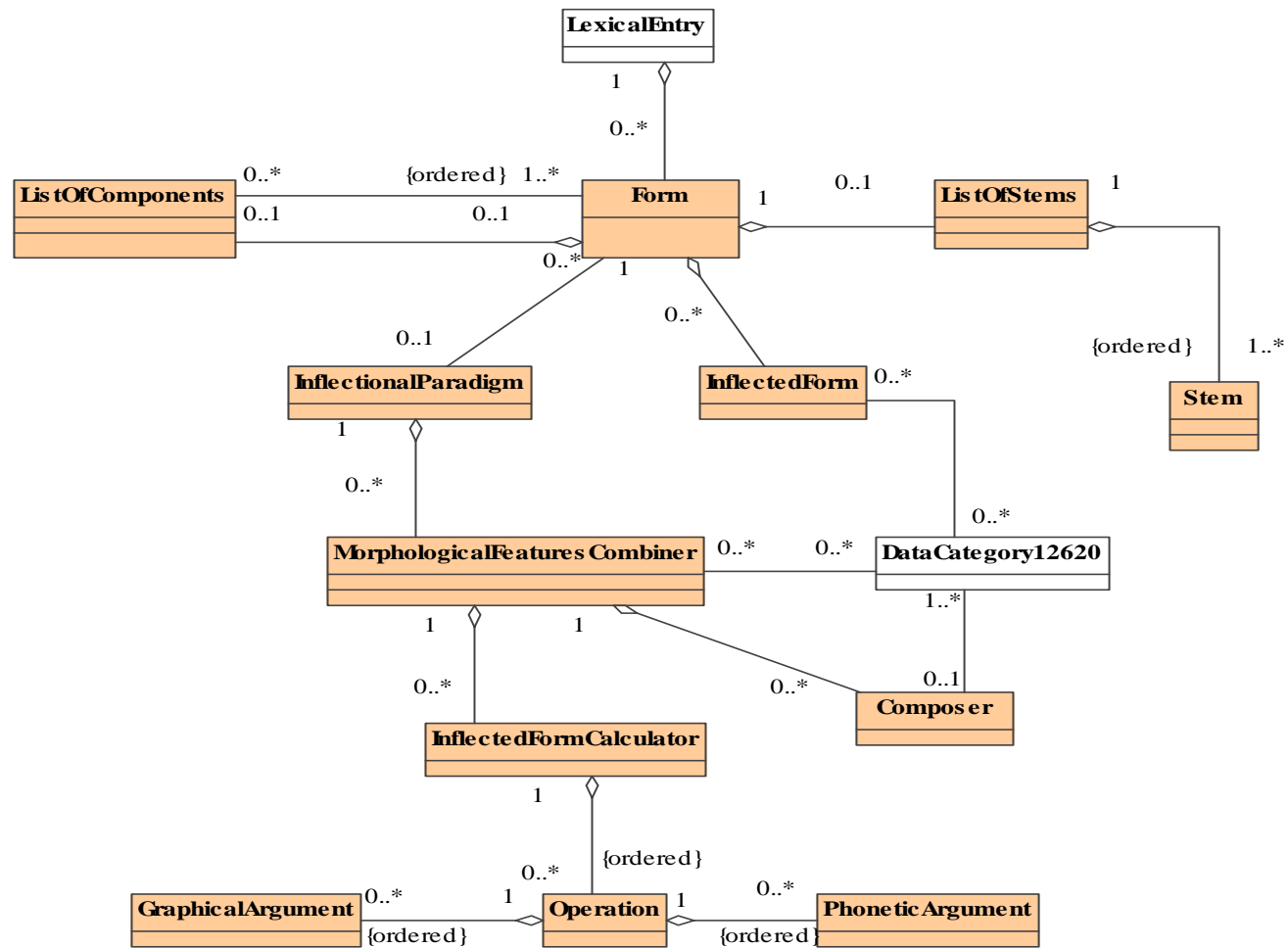
The bug

- Imagine that you manage a database containing two lexicons: for instance one in English and one in Spanish
- Use case on metadata:
 - Part-1:** you need a copyright mention « John Smith » to all English words
 - Part-2:** you need a copyright mention « Pedro Dovar » to all Spanish words
 - Part-3:** you need a global version number « version-1 » to the whole database before delivery
- **You cannot do that**
- This is not an exotic situation
- Conceptually, current core model is not suited for multilingual lexicon: nothing distinguish a monolingual lexicon from the whole database

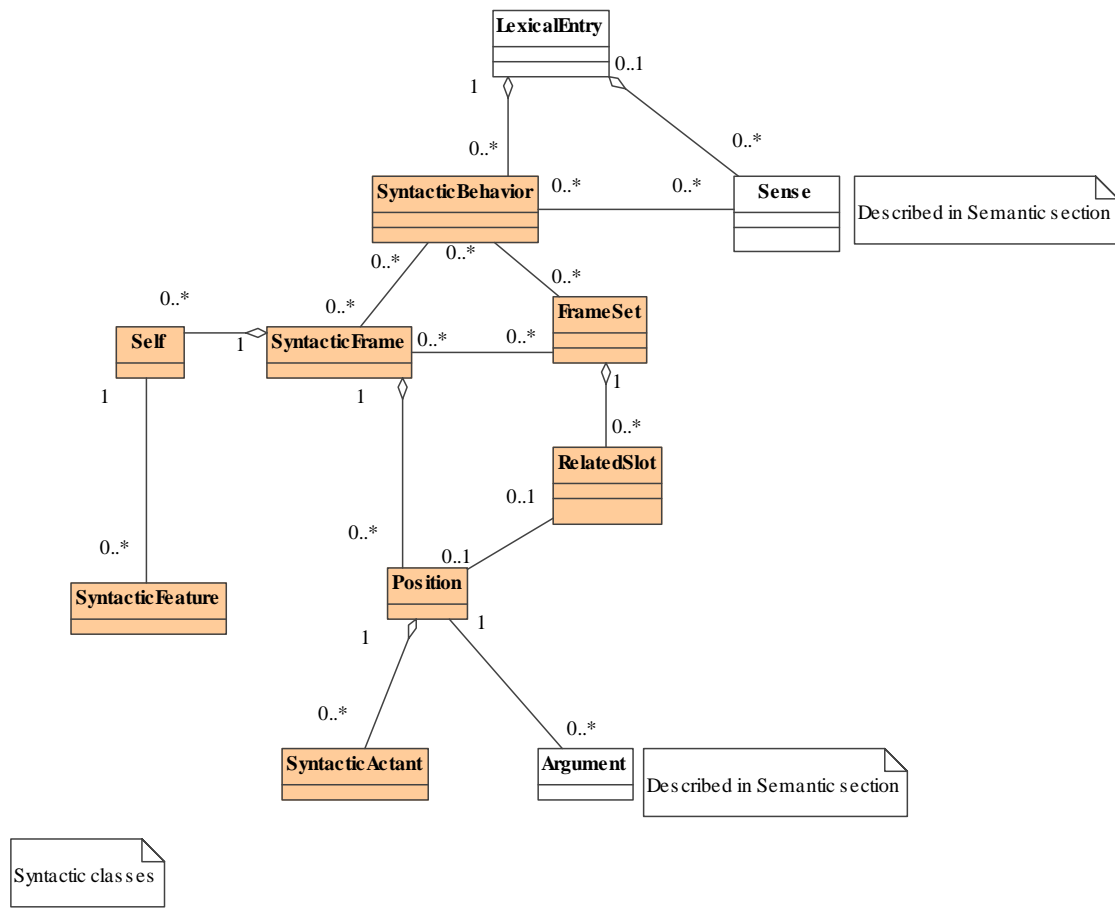
My suggestion for LMF-rev-6



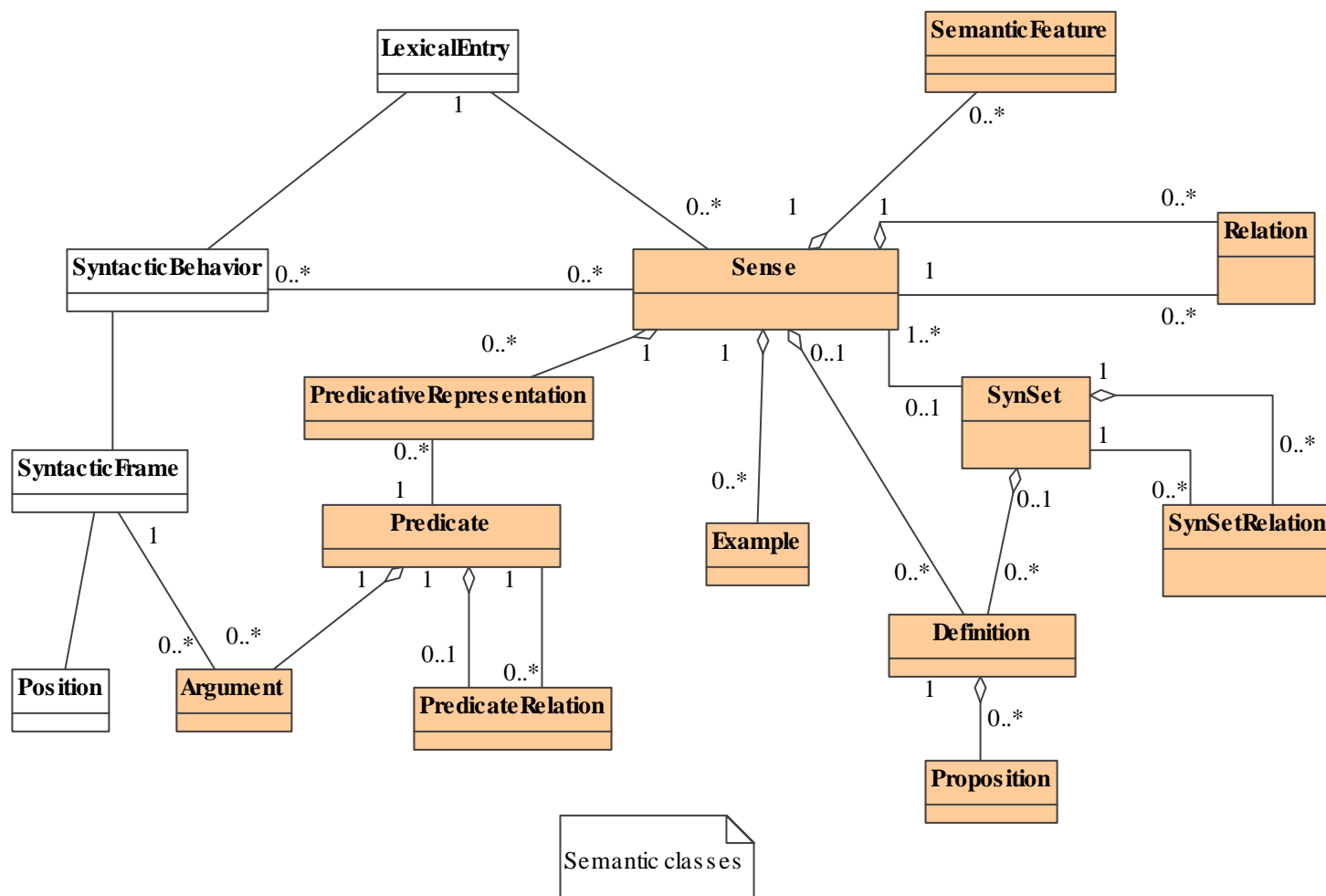
3) LMF/NLP extension presentation



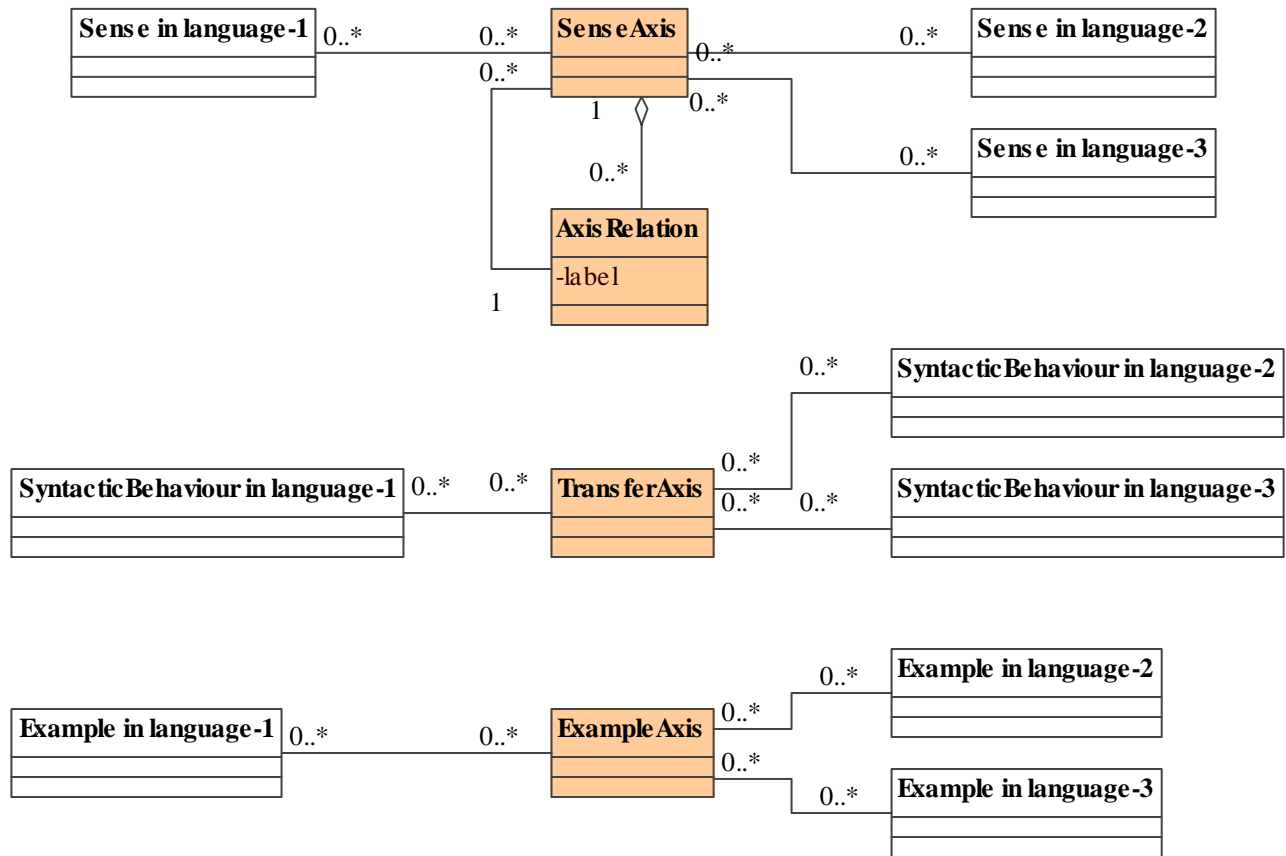
LMF/NLP extension, syntax



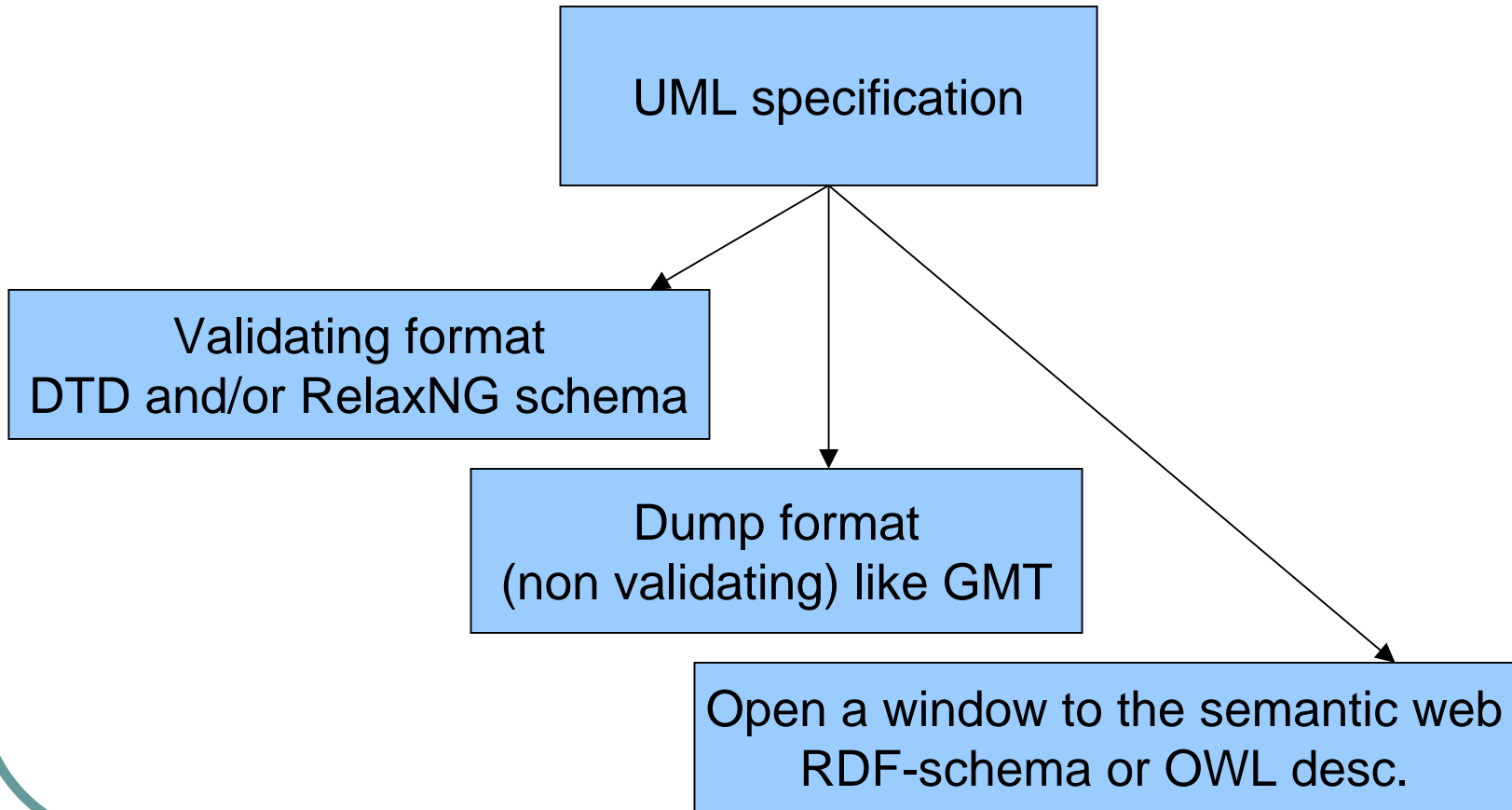
LMF/NLP extension, semantics



LMF/NLP ext, multilingual notations



4) The future, what's next?



Semantic web ?

I have the feeling the semantic web will need NLP lexicons ?

- a) As direct ontologies like WordNet
(see www.semanticweb.org/library/)
- b) Incorporated in tools for building & maintaining ontologies
- c) Incorporated in tools for mapping the great number of foreseen ontologies (known as the « ontology mapping problem »)
 - for jumping from one ontology to another in the same language
 - behavior of the driver /
 - driver's behavior /
 - driving behavior /
 - pilot behavior etc.
 - for jumping between ontologies expressed in different languages
 - driver's behavior /
 - comportement du conducteur