Ontologies and multilinguality

Dra. Guadalupe Aguado de Cea
lupe@fi.upm.es
http://www.oeg-upm.net
Ontological Engineering Group
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo sn,
28660 Boadilla del Monte, Madrid, Spain

Credits

- Some of these transparencies have been prepared by

Elena Montiel-Ponsoda

Outline

- Definition and purpose of multilinguality
- Localization vs. internationalization
- From monolingual to multilingual systems
- NLP systems including multilinguality
- Multilinguality in KB systems
- Multilinguality in ontologies
  - Information
  - Realization
  - Modelling
- A new proposal: Linguistic Information Repository- LIR

Multilinguality. What for?

- Multilinguality is required in different NLP applications
  - Question answering systems
  - Multilingual information retrieval
  - Multilingual speech processing
  - Machine translation
- Knowledge sharing → ontologies
- Reusing Ontologies → Semantic Web

Ontology Localization

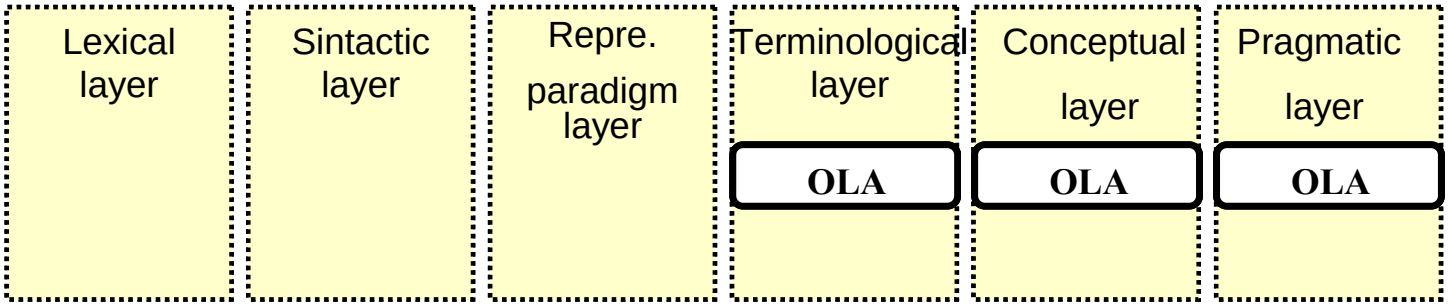
- **Localization** involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold (LISA)
- In **economics**: adapting a product to *a non-native environment*.
- In **software y *web design***: adapting contents, language, and design to the target language and culture
- In **ontologies:** ***Ontology Localization*** *involves the process of adapting an ontology to a particular language and culture*.

Internationalization

- ***Internationalization*** *is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. Internationalization takes place at the level of program design and document development (LISA).*
- Important aspects:
  - Separating text from the source code – > prevents translators from changing the source code
  - Internationalization is also applied to *online help, documentation and web sites*
  - Technical writers have to take internationalization into account: "*writing for a global audience", "web site globalization*"

Ontology Layers implied in the OLA

- **Lexical layer**: characters and symbols that make up the syntax (ASCII encoding, UNICODE, etc.)
- **Syntactic layer**: structure of characters and symbols, i.e., the grammar. It embraces different representation languages (e.g. RDF(S), OWL, etc.)
- **Representation paradigm layer**: paradigm followed in the representation of the ontology (frames, semantic networks, DL, etc.) that allows a certain way of expressing and structuring knowledge
- **Terminological layer**: terms or labels selected to name ontology elements
- **Conceptual layer**: related to conceptualization decisions, such as granularity, expressiveness, perspective, etc.
- **Pragmatic layer**: final layout of the model according to user's needs



J. Barrasa. Modelo para la definición automática de correspondencias semánticas entre ontologías y modelos relacionales. PhD Thesis. Universidad Politécnica de Madrid, Madrid, Spain, 2007

Multilinguality: What for?

- Institutions and organizations worldwide have to **manage** information in different natural languages
- Their **sources of information are multilingual**
- Information results have to be **customized** for different users in different languages

The FAO's main priority is to guarantee effective information access and information management to all levels of society, from farmers and fisher folk to researchers and policy-makers.

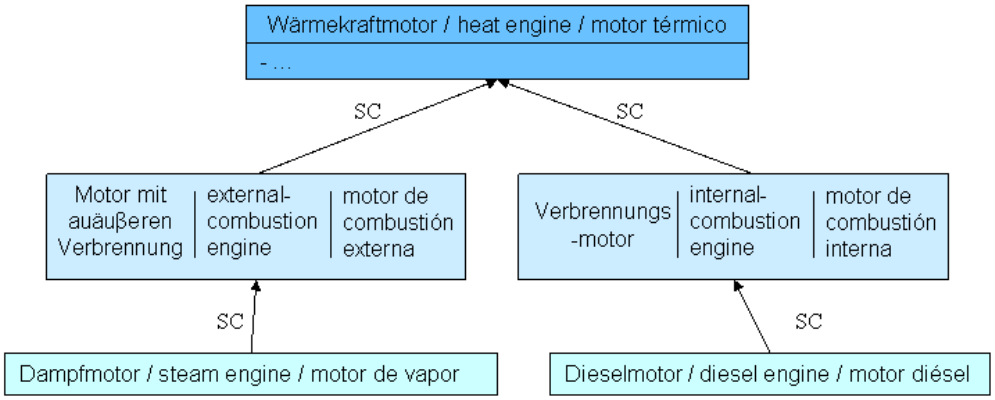
From monolingual to multilingual systems

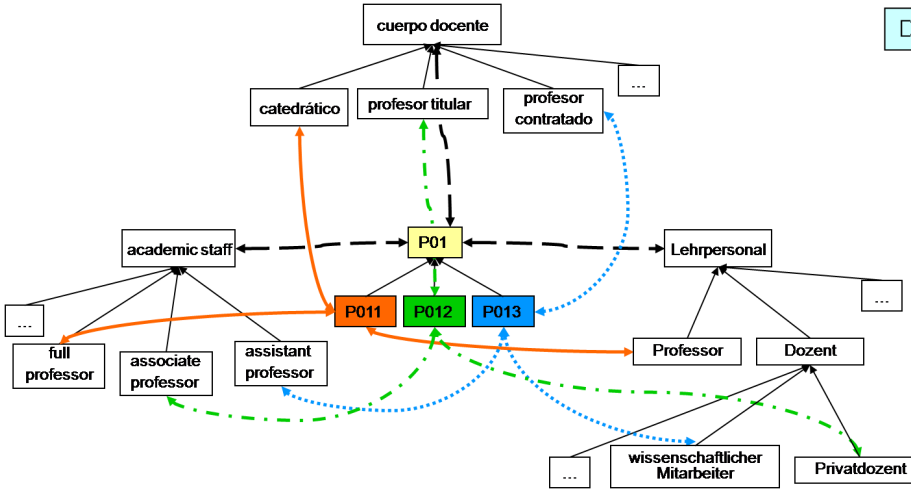
- Few multilingual ontologies
  - http://olp.dfki.de/ontoselect/
  - 1652 ontologies
  - 149 with language information
  - 130 in English, 10 in Spanish
  - 5: en-es, 4: en-es-fr
- Scarce information available about how to represent multilinguality
- Recent interest in international research groups:
- LISA (*Localization Industry Standards Association*)
- OSCAR (*Open Standards for Container/Content Allowing Re-use*)
- OASIS (*Organization for the Advancement of Structured Information Standards*)
- W3C
- ISO *International Standards Organization*

Modelling Modalities

OLA at the Terminological layer

1

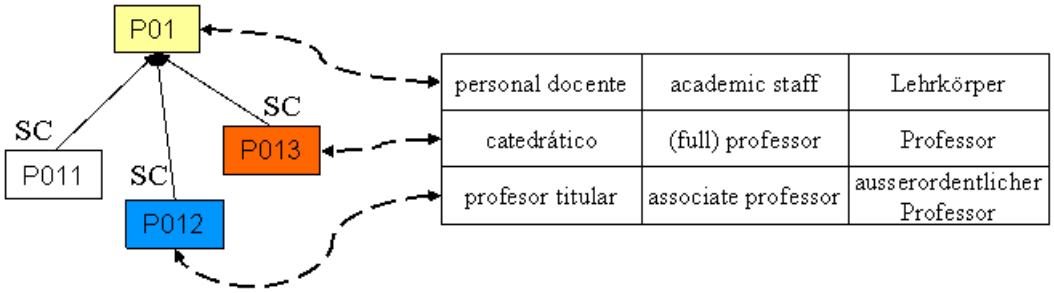




2 **OLA at the Conceptual layer**

OLA at the Terminological and Conceptual layers

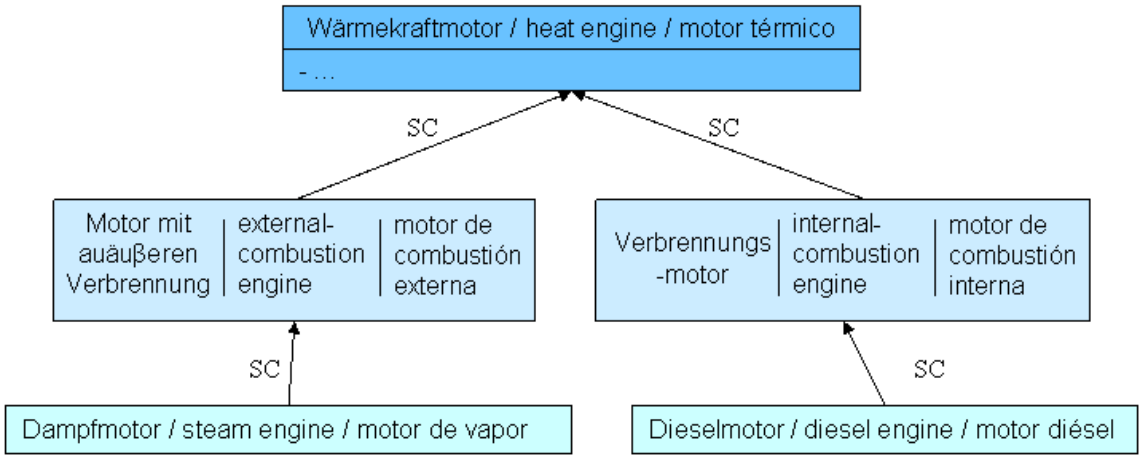
3



personal docente	academic staff	Lehrkörper
catedrático	(full) professor	Professor
profesor titular	associate professor	ausserordentlicher Professor

1. OLA at the Terminological layer

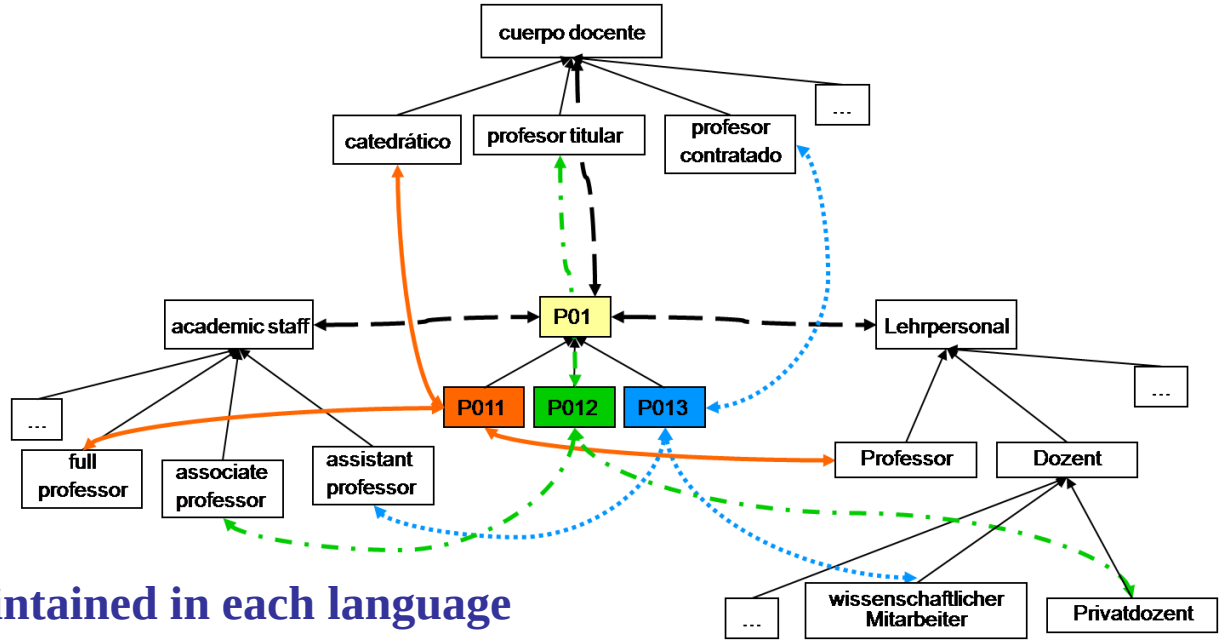
rdfs: label
rdfs: comment



- **Advantages:**
- **Increase of other languages is easily done by including just labels in the ontology.**
- **Suitable for highly specialized domain ontologies: knowledge shareable among different linguistic community experts.**

- **Disadvantages:**
- **Linguistic information included in the ontology is limited.**
- **Exact equivalence is assumed among labels in the different languages (though it may not be true).**

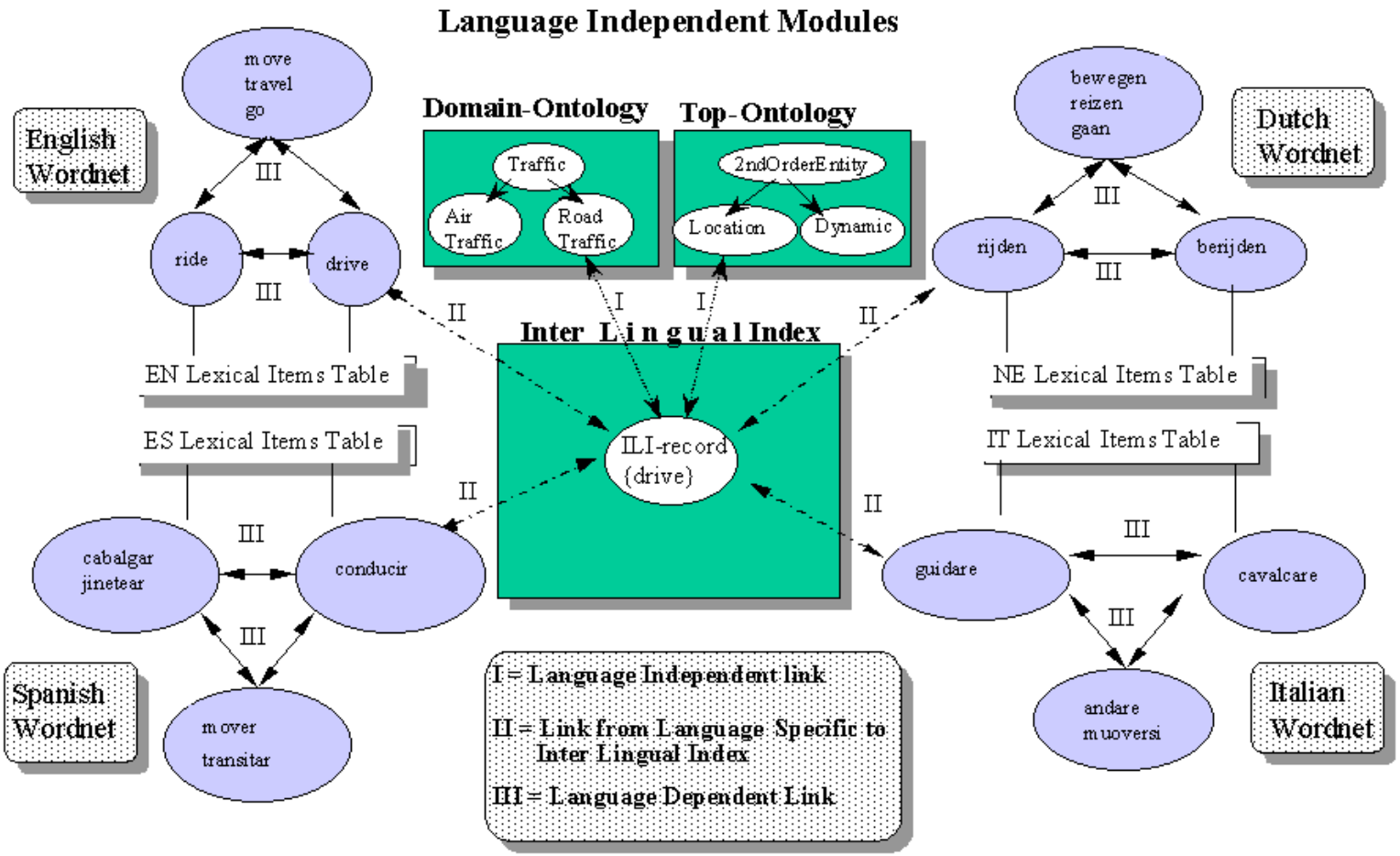
2. OLA at the Conceptual layer



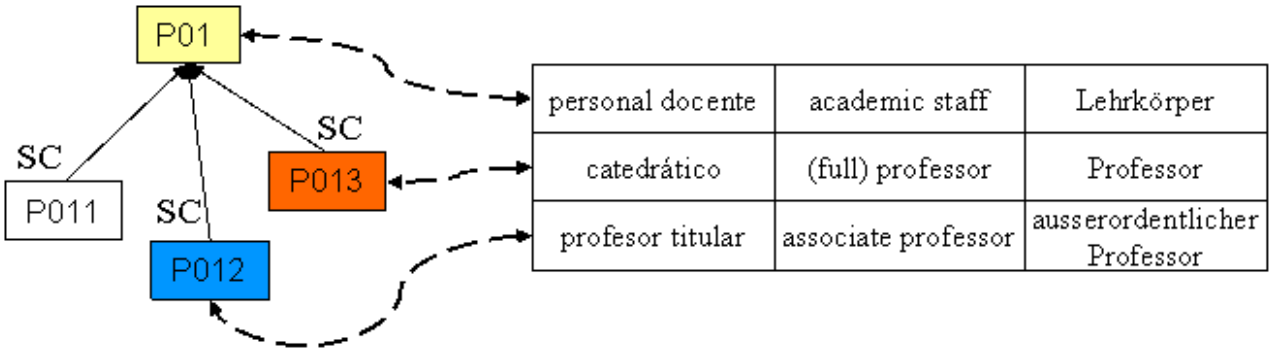
- **Advantages:**
- **Conceptualizations are maintained in each language**
- **Suitable for ontologies highly dependent of a certain culture: the judiciary.**

- **Disadvantages:**
- **Huge effort needed to conceptualize the same domain in different natural languages.**
- **Three types of expertise are required: domain expertise, linguistic expertise, and ontology engineering expertise.**

Example: EuroWordNet



3. OLA at the Terminological and Conceptual layers

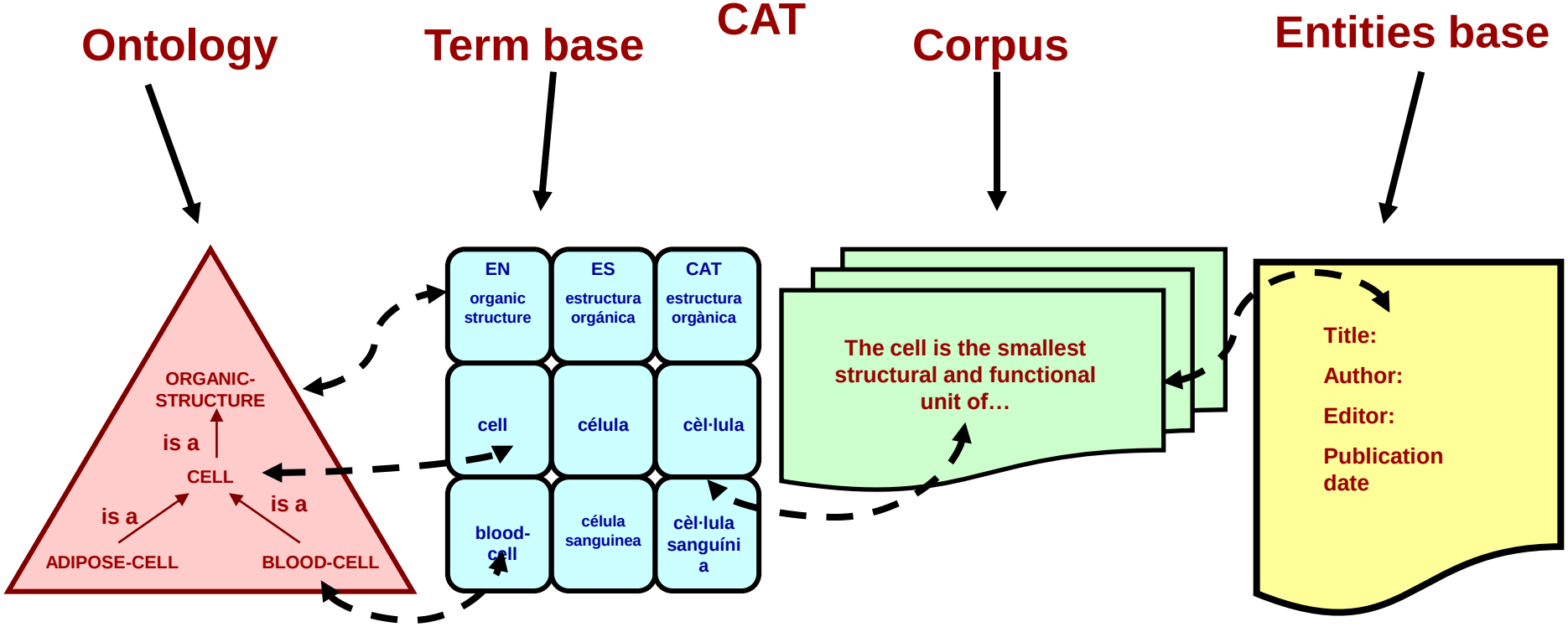


- **Advantages:**
  - **Including as much linguistic information as wished is possible**
  - **Linguistic elements within one language or across languages can be linked.**
  - **Nuances or differences between languages can be reported and formalized at the terminological layer**
  - **Relevant information as, e.g., the provenance of the linguistic elements, can also be included.**

- **Disadvantages:**
  - **Some language specificities could be lost, unless captured in language specific ontology modules, i.e., in the conceptual layer, or in the linguistic model, i.e., at the terminological layer.**

Example: GENOMA-KB

Biomedical knowledge base for the human genome in EN, ES, and CAT



A new proposal
Linguistic Information Repository - LIR

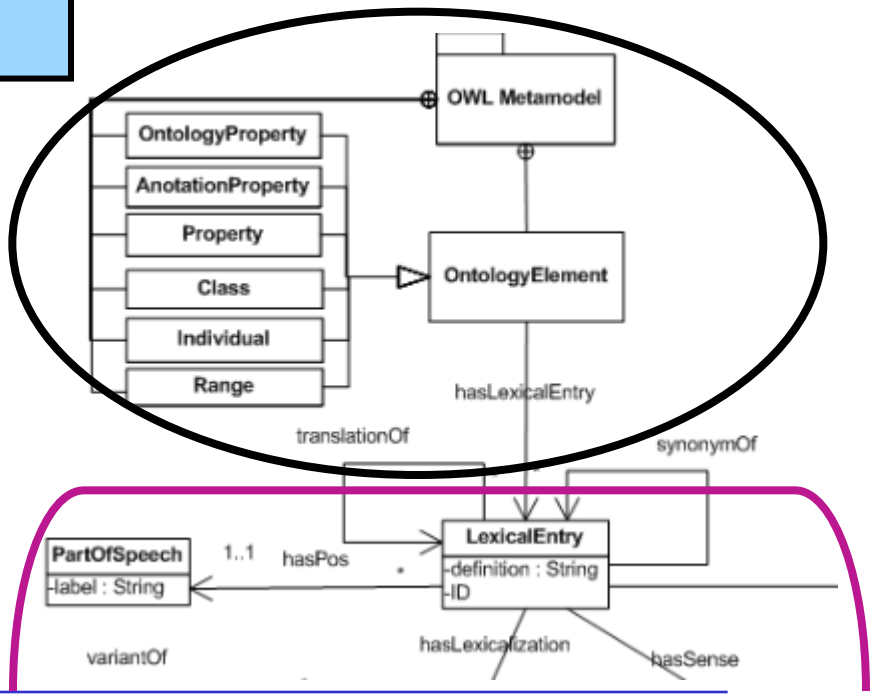
- It is modelled as an ontology.
- The linguistic information captured in the LIR is organized around the `LexicalEntry` class.
- A lexical entry is a ternary relation: `Lexicalization`, `Sense` and `LanguageContext`.
- `Note` is linked to the `Lexicalization`, but it could be linked to any other class in the model to include supplemental information.
- By linking `Note` to the `Sense` or `Definition` classes, possible differences or nuances among senses in different languages can be made explicit.

LIR – Linguistic Information Repository

- The linguistic model should allow for:
  - Addition of multilingual data to any element of the ontology
  - Selection of the ontology element (class, property, instance, etc.) the user wants to be multilingual

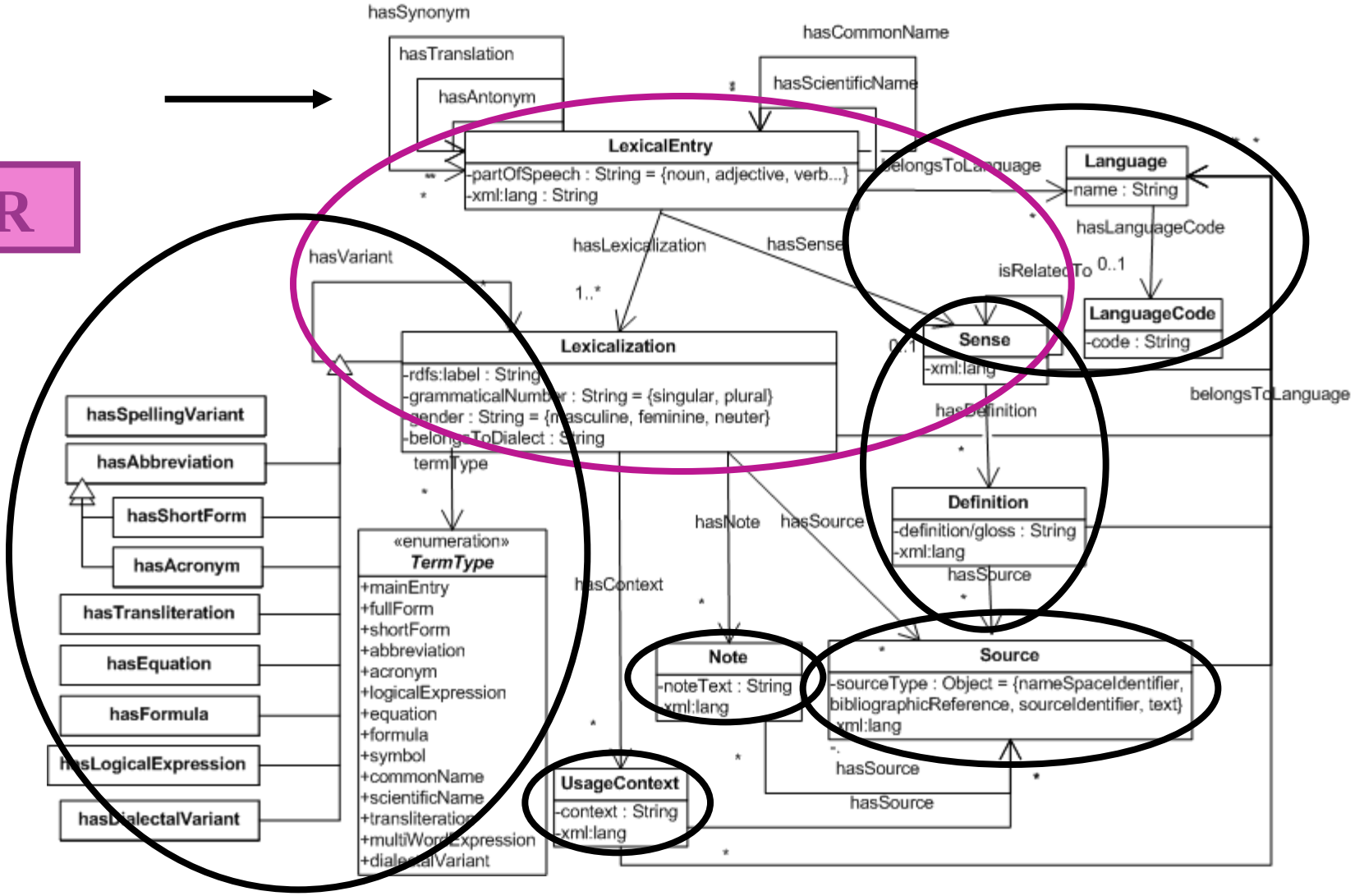
HOW?

- **Linking the `OntologyElement` class (from the OWL Ontology Meta-model) to the Linguistic Information Repository**

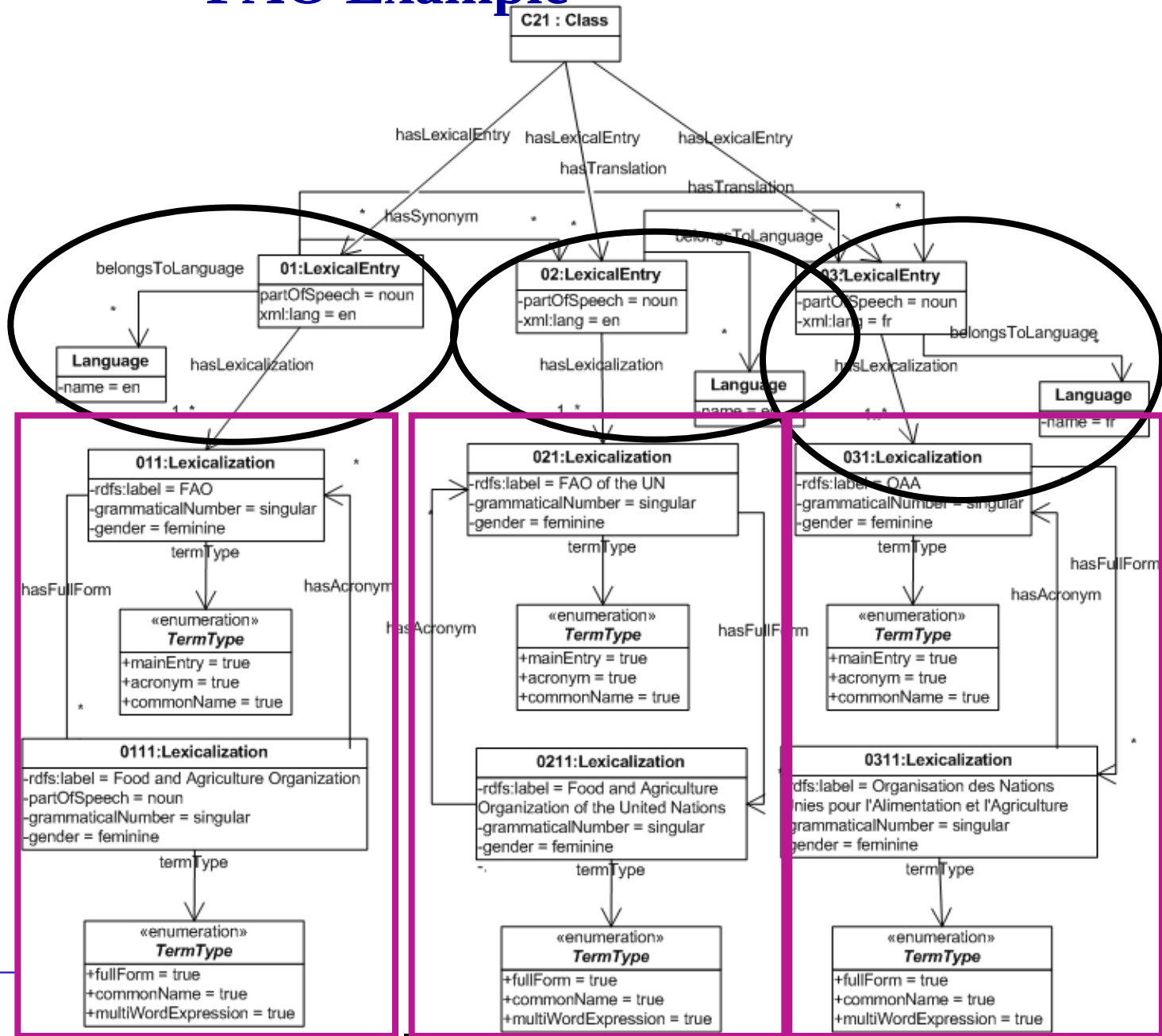


LIR – Linguistic Information Repository

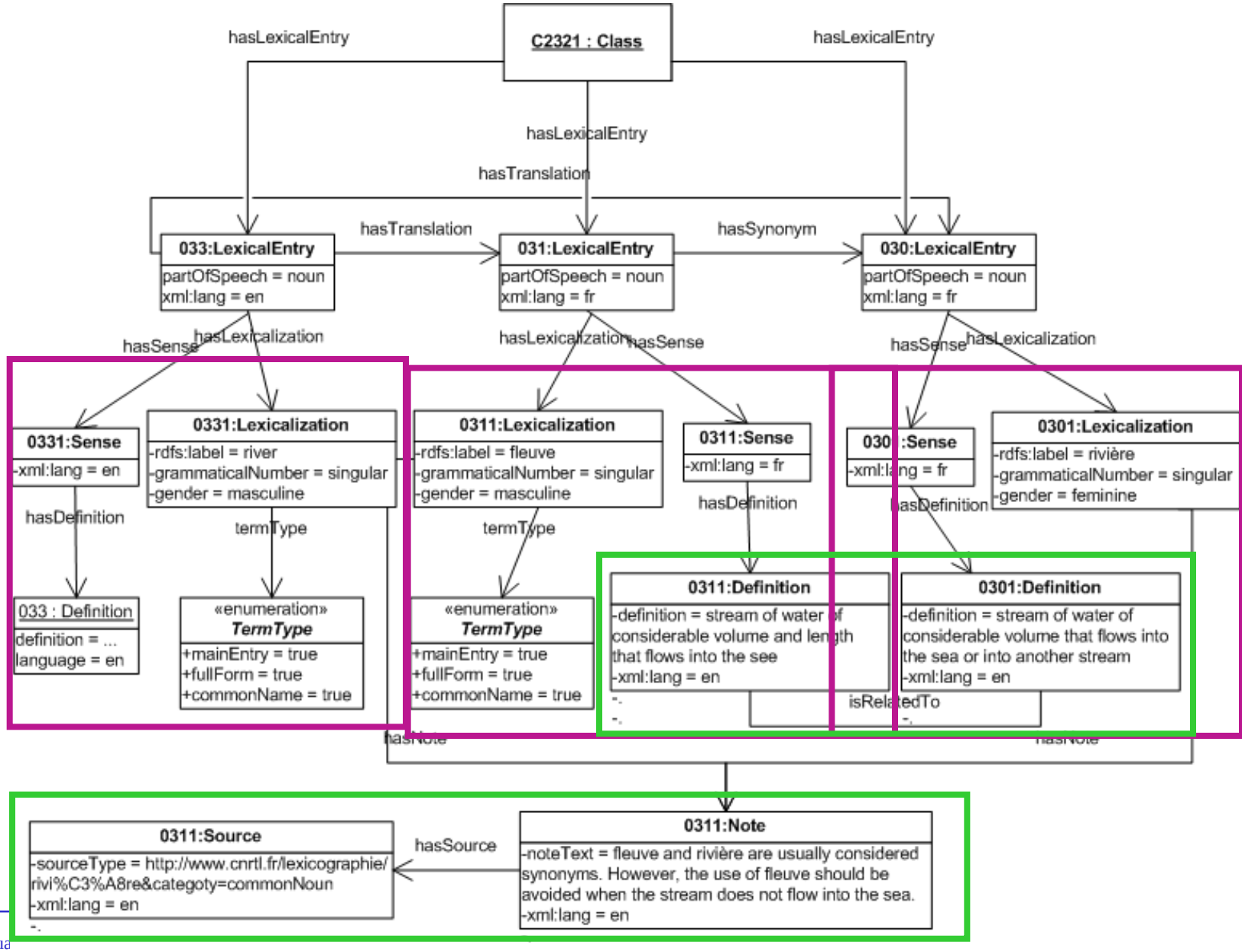
LIR



FAO Example

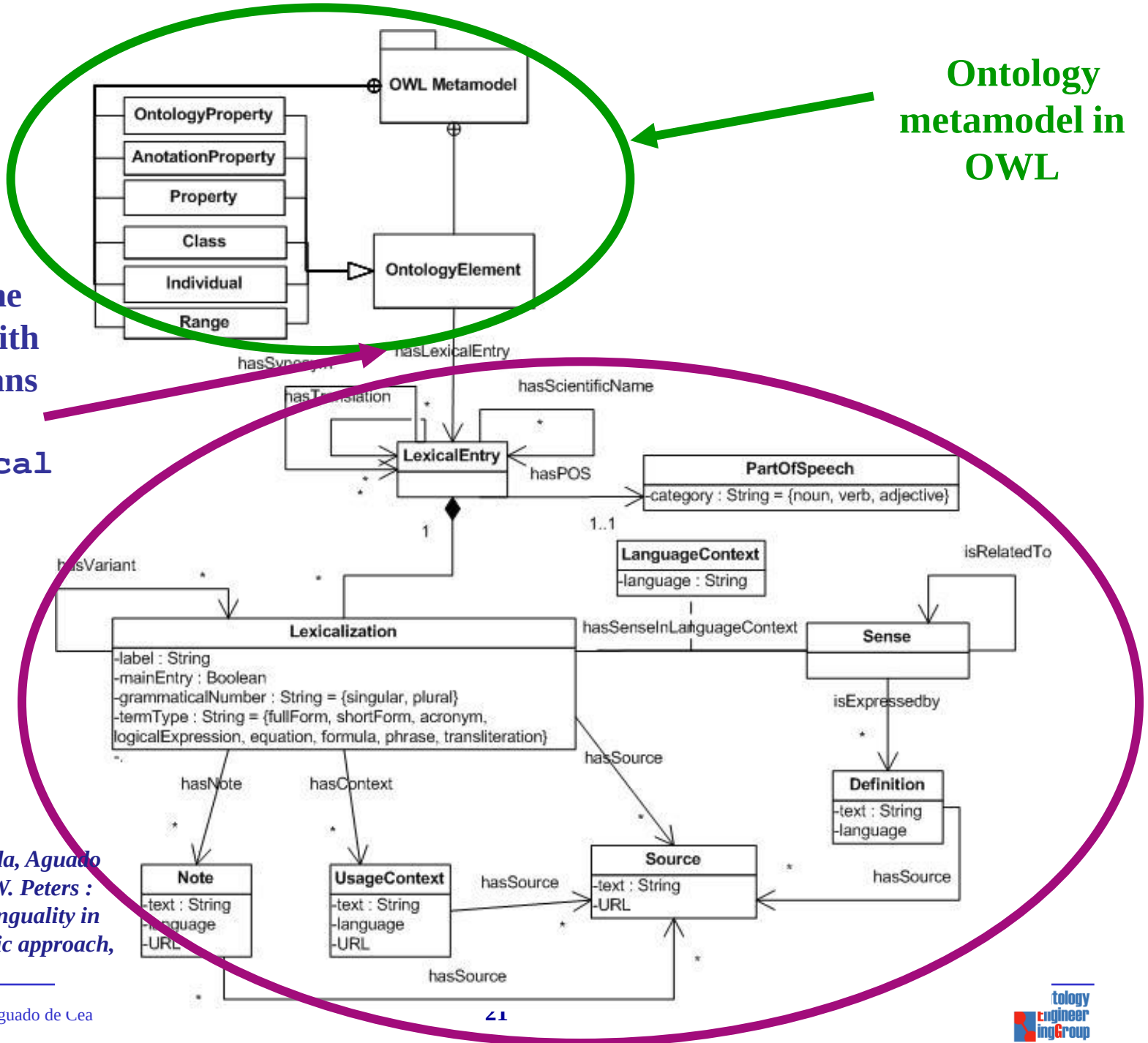


River Example



Ontology metamodel in OWL

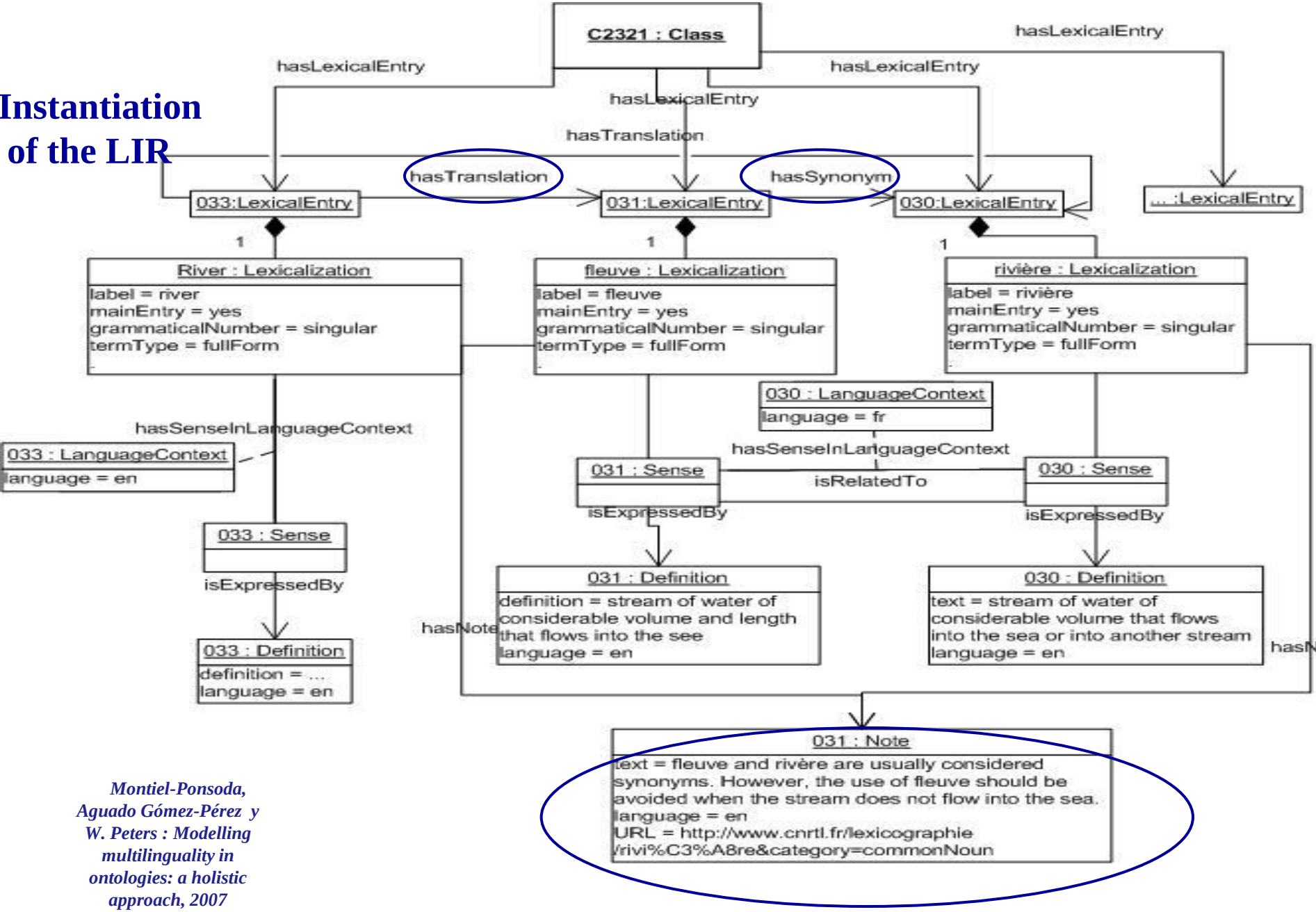
Linking the Ontology with LIR by means of the "`hasLexicalEntry`" relation



LIR

Montiel-Ponsoda, Aguado Gómez-Pérez y W. Peters : Modelling multilinguality in ontologies: a holistic approach, 2007

Instantiation of the LIR



Montiel-Ponsoda, Aguado Gómez-Pérez y W. Peters : Modelling multilinguality in ontologies: a holistic approach, 2007

Advantages of the LIR proposal

- Preserves the independence between the ontology and the linguistic layer.
- Links multilingual information with all ontology elements.
- By adopting linguistic standards for describing linguistic features helps to maintain language specificities
- Allows localization at the terminological and conceptual level.
- Facilitates interoperability and extensibility if more information is needed.
- Solves conceptualization mismatches
- Access to multilingual resources is possible thanks to certain tools: LabelTranslator.

Ontologies and multilinguality

Dra. Guadalupe Aguado de Cea
lupe@fi.upm.es
http://www.oeg-upm.net
Ontological Engineering Group
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo sn,
28660 Boadilla del Monte, Madrid, Spain