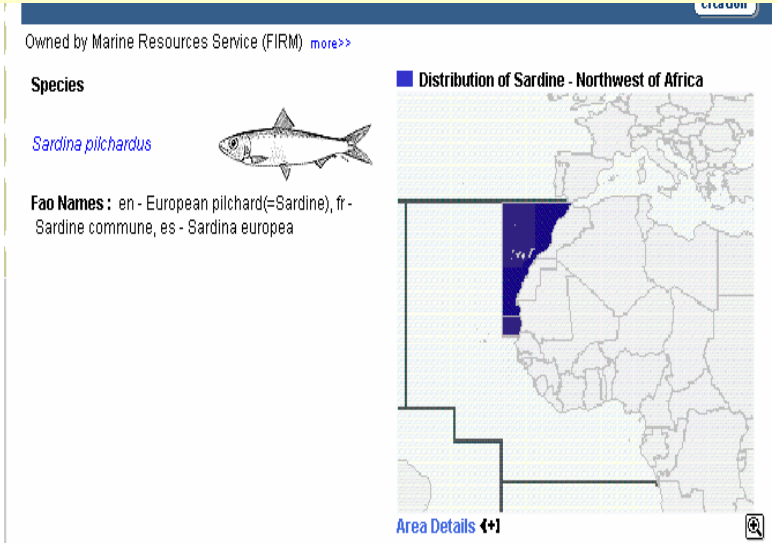
Work on Multilingual Ontologies within the NeOn Project

Mauricio Espinoza, Elena Montiel-Ponsoda,
Oscar Corcho, Guadalupe Aguado, Asunción Gómez-Pérez

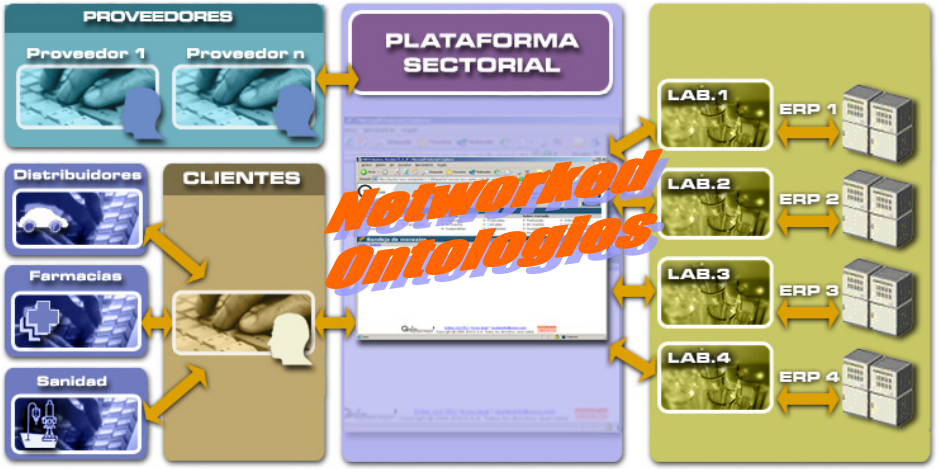
C21 Cost Action Towntology
October 20th, 2008

- Large scale semantic web app.

- Managing fishery knowledge
  to support automatic
  alert mechanisms

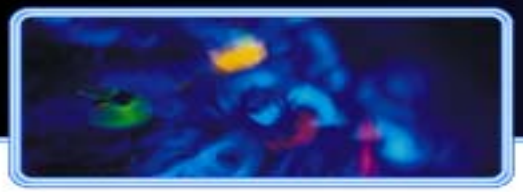


- Pharmaceutical sector
  - E-Invoice management
  - Integration and management of information about pharmaceutical products



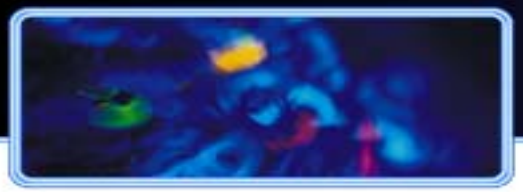
Global: 14.716.896 €

80



Introduction: The issue of Ontology Localization

- What is Ontology Localization? Activity in the Ontology Development process that consist in adapting an ontology to a concrete language or culture community
- Which is the result of the Ontology Localization activity?: (usually) a multilingual ontological system
- Why is Ontology Localization so necessary? Users demanding ontology-based applications do not always speak the same language or have the same cultural background



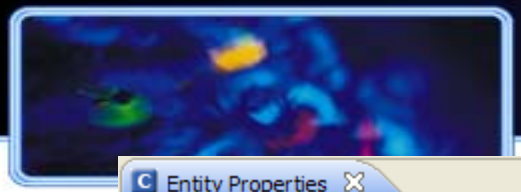
Problem: How to capture and represent the concept-label relation?

Ontologies are conceptual constructs **without** linguistics

Concepts are abstract notions whose labels are **arbitrary**

Lexicalizations that function as labels for these concepts are only considered to be **evocative** of the ontological meaning of the concepts

Intensional senses from a lexical model are mapped to **extensional** interpretations on ontology elements



F-Logic - NeOn Toolkit

Window Help

Entity Properties

Ontology element

Name: Plant

Namespace:

Lexical Entries

Entries

Identifier	Part Of Speech	Language
LexicalEntry-1	noun	English

Lexical Entry Relationships

- Synonyms
- Translations
- Antonyms
- Scientific names

Lexicalizations

Entries

Label	G. Number	Gender	Dialect	Language
Plant	Singular			English

Lexicalization Term type

- Acronym / Abbreviation
- Full form / Short form
- Common Name / Scientific Name
- Main Entry / Formula
- Dialectal Variant / M.Word Expression
- Transliteration / Logical Expression
- Symbol / Equation

Lexical Entry Senses

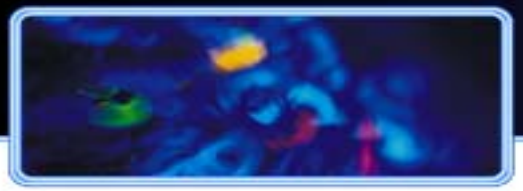
Usage Context

Class Restrictions | Taxonomy | Annotations | Linguistic Information

IATE

EWN

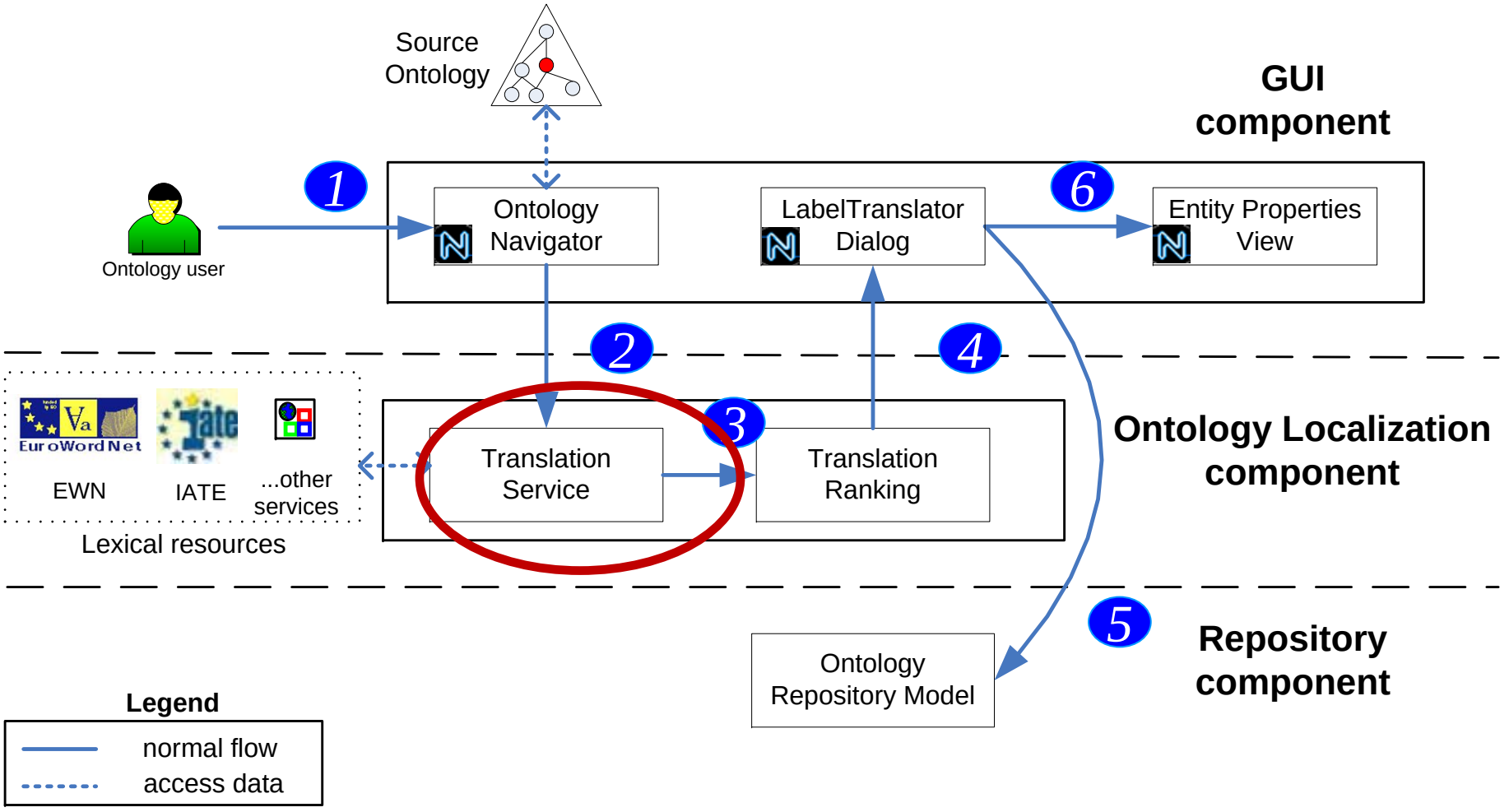
Close

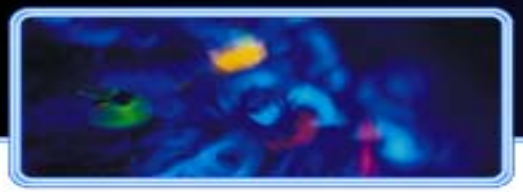


Introduction

- Used resources
  - Online Web Translation → translation equivalents
  - EWN databases
  - Online Web Ontologies → translation senses
- Linguistic Meta-model
  - captures all the relevant linguistic/terminological information associated with concepts
- Supported languages
  - English, German and Spanish

Main Components





Translation Service

- Input
  - ontology label (*l*) → source language
- Output
  - set of translations $T = (t_1, t_2, ...t_n)$ → target language
- Resources
  - multilingual dictionaries: Babelfish, Wiktionary, etc.
  - remote lexical databases: EuroWordNet
  - other lexical resources: IATE
- *A buffer* stores previously translations

Translation Service (II)

- $t_i = \langle trs; senses \rangle \quad t_i \in T$

  where,

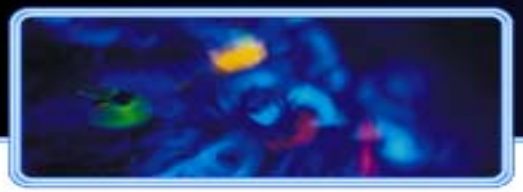
  trs, translated label in the specific target language

  senses, semantic senses[1] extracted from different knowledge pools.

list of synonym names *ontology context* *description in natural langua*

$$\langle plant; \left[\begin{array}{l} s^1_{class} = \left\{ (flora, \ldots), \begin{array}{c} organism \\ plant \\ \underbrace{aquatic \quad endemic}_{} \quad \ldots \end{array}, "a\ living\ \ldots" \right\} \\ s^2_{class} = \left\{ (works, \ldots), \begin{array}{c} building\ complex \\ plant \\ \underbrace{factory \quad bottiling}_{} \quad \ldots \end{array}, "buildings\ \ldots" \right\} \\ \ldots \end{array} \right] \rangle$$

[1] *R. Trillo, J. Gracia, M. Espinoza, and E. Mena. Discovering the semantics of user keywords. Journal on Universal Computer Science. Special Issue: Ontologies and their Applications, 2007.*



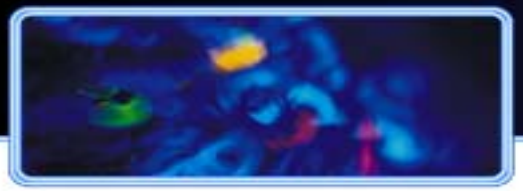
Discovering Semantic Senses of each ti

$$s^{1}_{class} = \{(EWN1\#\,flora, EWN1\#\,plant\ life),\ \overset{natural-object}{\overset{plant}{aquatic}}\ \overset{organism}{\overset{}{endemic}}\ \overset{}{apomict}\ \ldots,\ \text{"a living organism and typically immobile living ..."}\}$$

$$s^{2}_{class} = \{(EWN1\#\,works, EWN1\#\,industrial\ plant),\ \overset{building\,complex}{\overset{plant}{factory}}\ \overset{}{bottiling}\ \overset{}{distillery}\ \ldots,\ \text{"buildings for carrying..."}\}$$

$$s^{3}_{class} = \{(ONT1\#\,plant),\ \overset{natural-object}{\overset{plant}{aquatic}}\ \overset{organism}{\overset{}{endemic}}\ \overset{}{apomict}\ \ldots,\ \text{"typically immobile living things"}\}$$

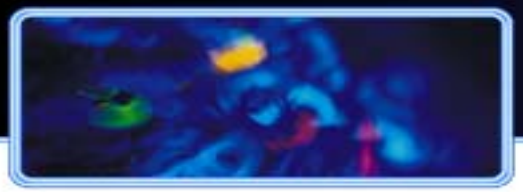
s^{3}_{class} Sense Alignment



Method to Translate Compound Labels

- Lexical templates → high quality translations
  - guarantees grammatical output
  - source language meaning is preserved

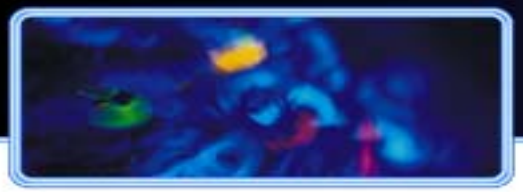
- A template:
  - source pattern → label to be localized (source language)
  - target pattern → translation (target language)



Learning Lexical Templates

Templates	Sample of source and target patterns	
	Spanish	English
$[N_1\ J_2]_{es} \rightarrow [J_2\ N_1]_{en}$	región espacial → producto industrial → planta ornamental →	spatial region industrial product ornamental plant
$[N_1\ [pre]\ N_2]_{es} \rightarrow [N_2\ N_1]_{en}$	vehículo de transporte → dominio del conocimiento →	transport vehicle knowledge domain
$[VB_1\ [pre]\ J_2]_{es} \rightarrow [J_2\ VB_{1ger}]_{en}$	detección remota; detección a distancia →	remote sensing
$[N_1\ [pre]\ N_2\ VB_3]_{es} \rightarrow [VB_{3part}\ N_2\ N_1]_{en}$	dominio de conocimiento asociado →	associated knowledge domain
…	…	

J: adjective N: noun VB: verb pre: preposition [ ]: optional

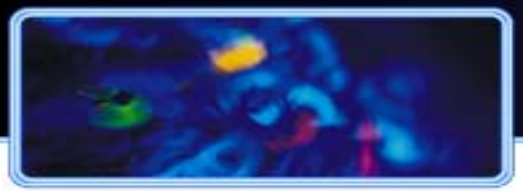


Algorithm to Translate Compound Labels

- Input: PlantaOrnamental (Spanish → English)

- compound label is normalized
  - planta
  - ornamental

- translate each token
  - planta: plant(noun), sow(verb), to plant(verb), …(8)
  - ornamental: ornamental(adj), decorative(adj)



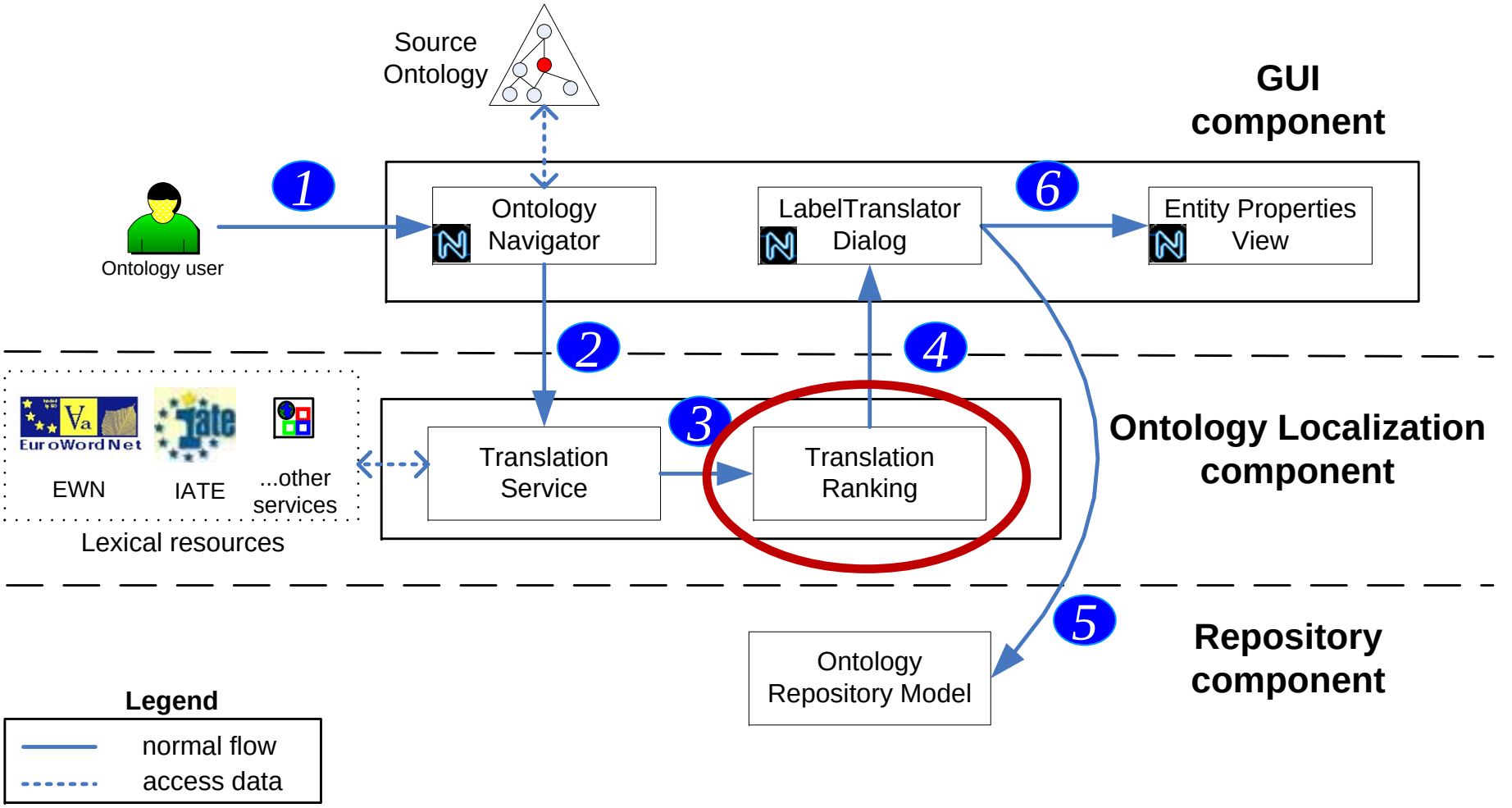
Algorithm to Translate Compound Labels

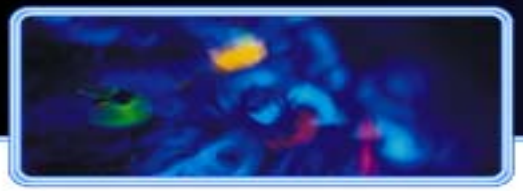
- create candidate translations → lexical templates

Lexical Template for "PlantaOrnamental" [N_1 J_2]es → [J_2 N_1]en		
Source Pattern	Candidate translations	Target pattern
[N_1 J_2]es	ornamental plant ornamental sow decorative plant decorative sow …	[J_2 N_1]en [J_2 VB_1]en [J_2 N_1]en [J_2 VB_1]en
J: adjective N: noun	VB: verb	

- select translations that fulfill the target pattern
  - ornamental plant
  - decorative plant

Main Components





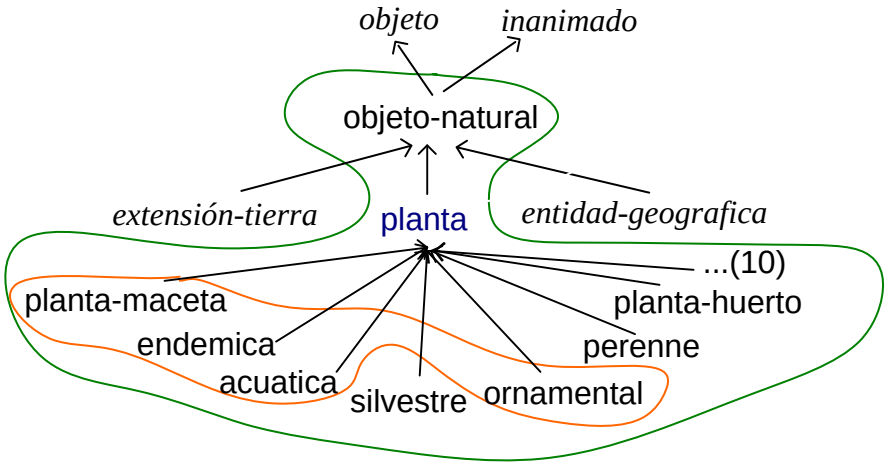
Translation Ranking Method

- For each ontology label
- Input
  - set of translations T obtained in the previous step
- Output
  - sorted set of translations

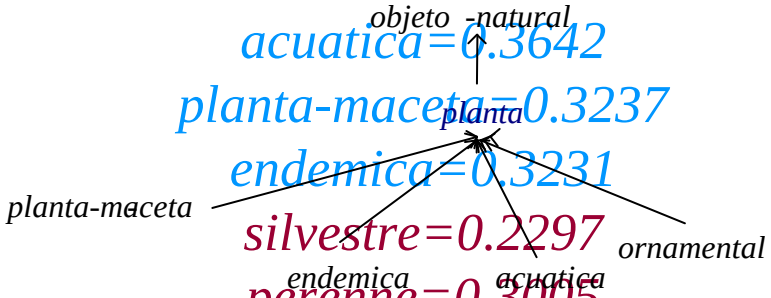
- To disambiguate translation senses
  - Context of each ontology label

Context of an Ontology Term

Context	Ontology Term		
	Class	Property	Relation
Hypernyms	X	X	
Hyponyms	X	X	
Attributes	X		
Domain		X	X
Range		X	X



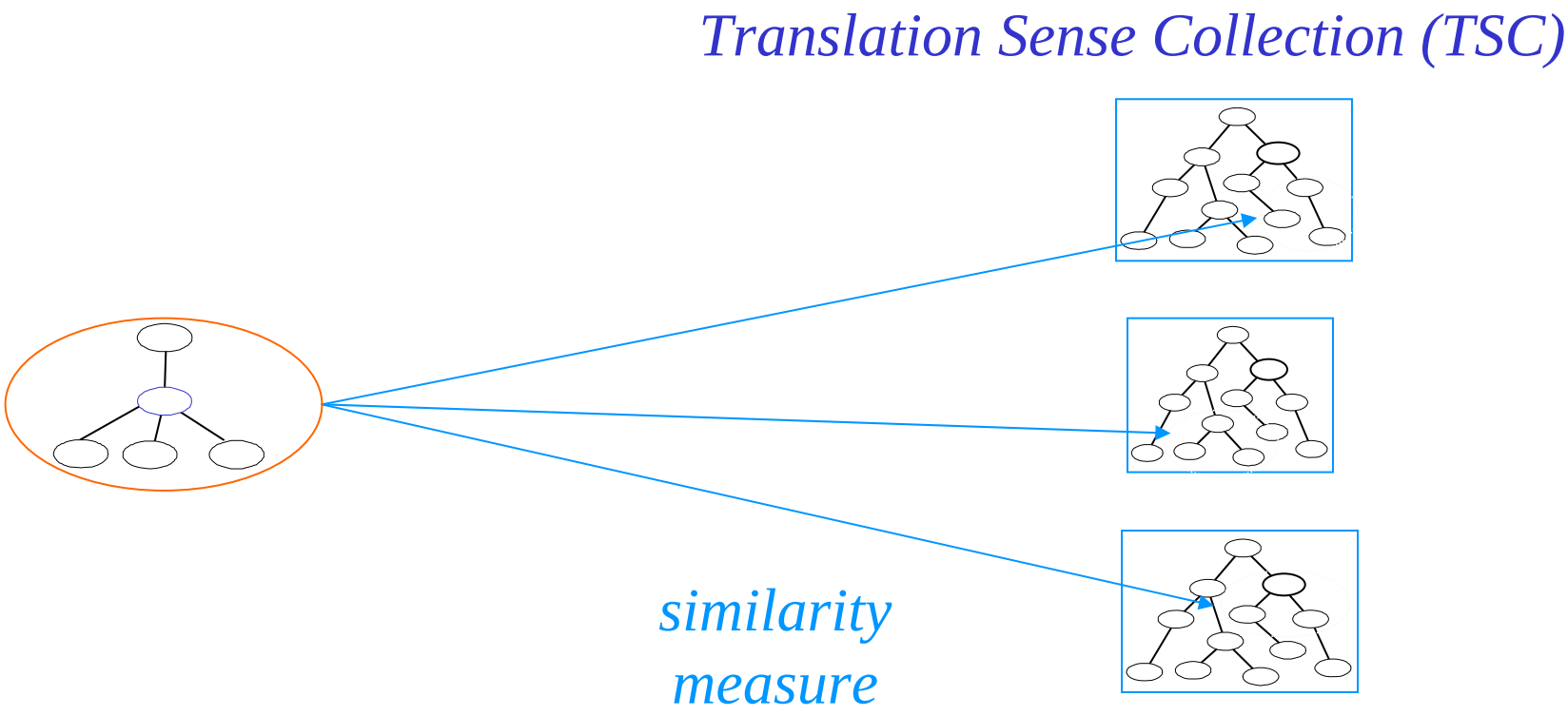
NGD between planta and:
ornamental=0.3790
acuatica=0.3642
planta-maceta=0.3237
endemica=0.3231
silvestre=0.2297
perenne=0.3005
planta-huerto=0.2197



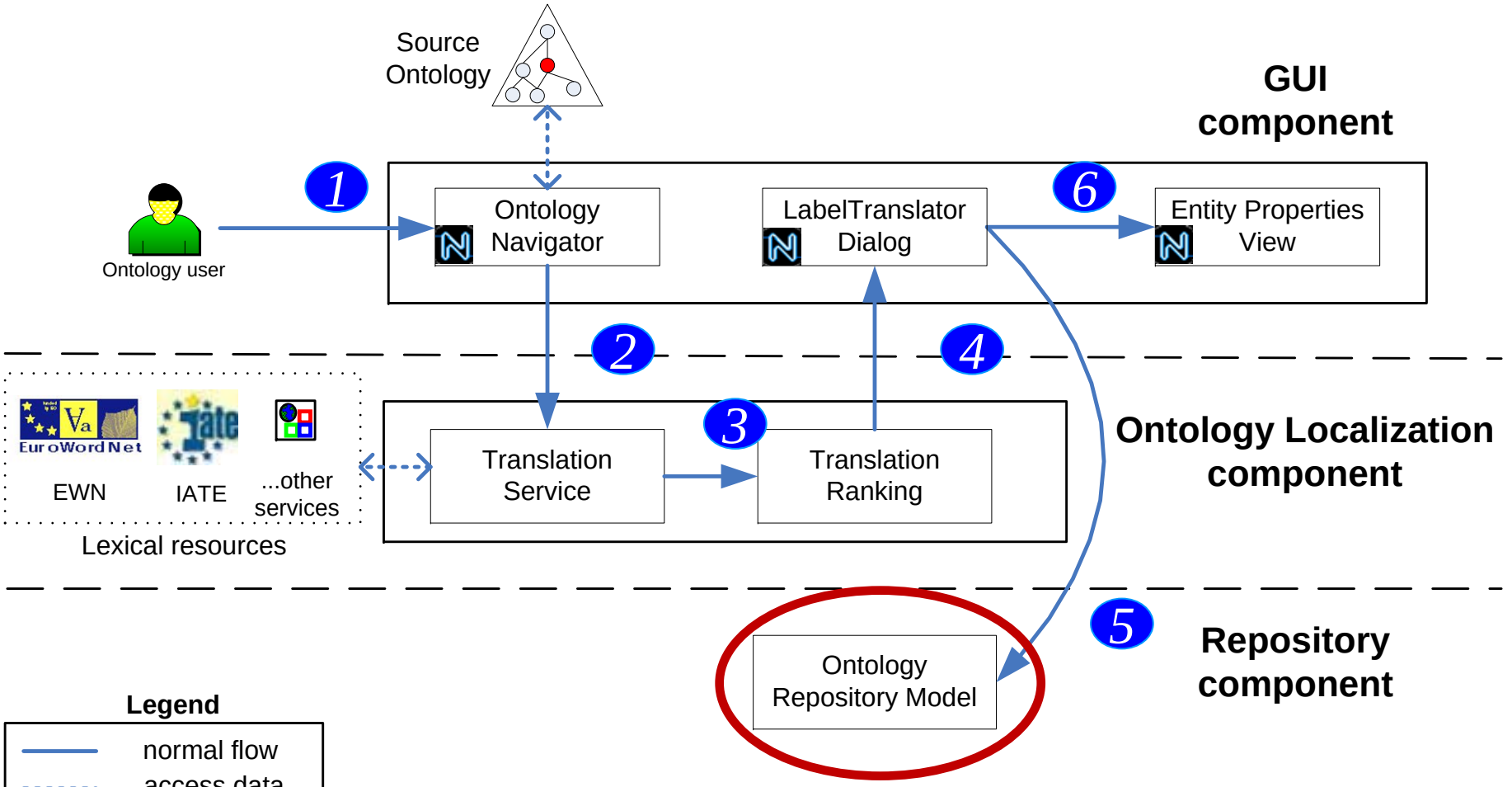
hyponyms sampled context

Translation Ranking (II)

- select one of the senses from the set *TSC* as the most appropriate sense of term.



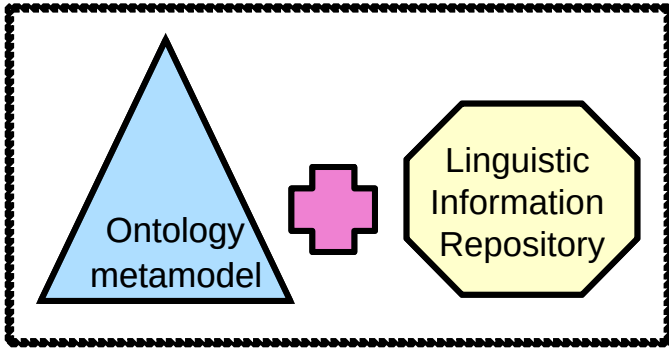
Main Components

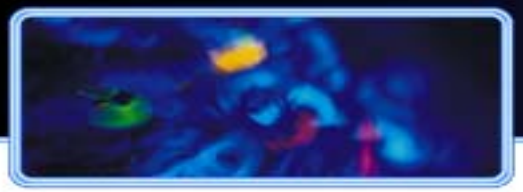


Requirements

Define a model for linking ontological and linguistic/terminological meanings, taking into account the following requirements:

- Independence between the ontology structure & the multilingual layer
- Multilinguality to any ontology element
- Localization of a concept in a certain language
- Interoperability with linguistic representation standards
- Accessibility to external linguistic resources



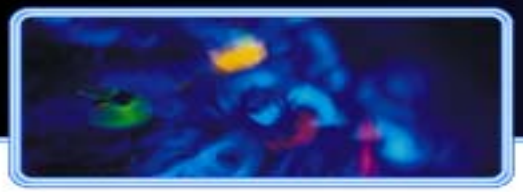


Linguistic Model
Localization

- The linguistic model should allow for:
  - Enrichment of the conceptual knowledge with multilingual natural language data to localize the ontology, i.e. to make the ontology suitable for a specific language and culture community

HOW?

- Selecting a set of data categories (ISO 12620) or classes that allow for an unambiguous identification of a concept in a certain language:
  - Lexicalization, Sense, Definition, Source, Usage Context, and Note (cf. LIR Figure)



Linguistic Model Interoperability

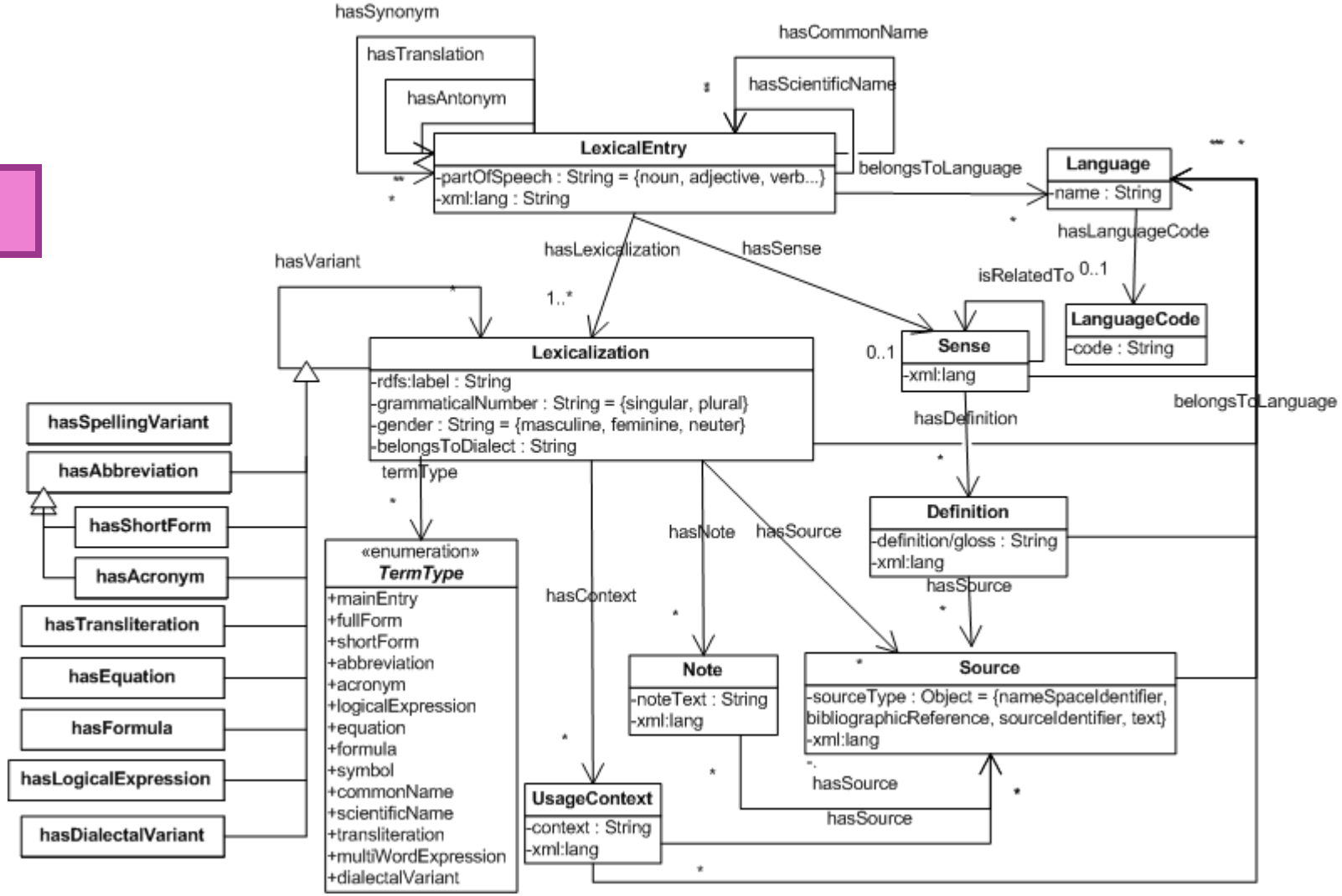
- The linguistic model should allow for:
  - Interoperability with existing and proposed standards for representing and integrating linguistic knowledge in order to guarantee reusability and interoperability

Which standards ?

- TMF (Terminological Markup Framework - ISO 16642:2003), and the associated TBX (TermBase eXchange format) of greatest impact in the terminological field
- LMF (Linguistic Markup Framework – ISO 24613) for encoding linguistic information
- MLIF (MultiLingual Information Framework) for integrating different representation models according to linguistic needs.

LIR – Linguistic Information Repository

LIR

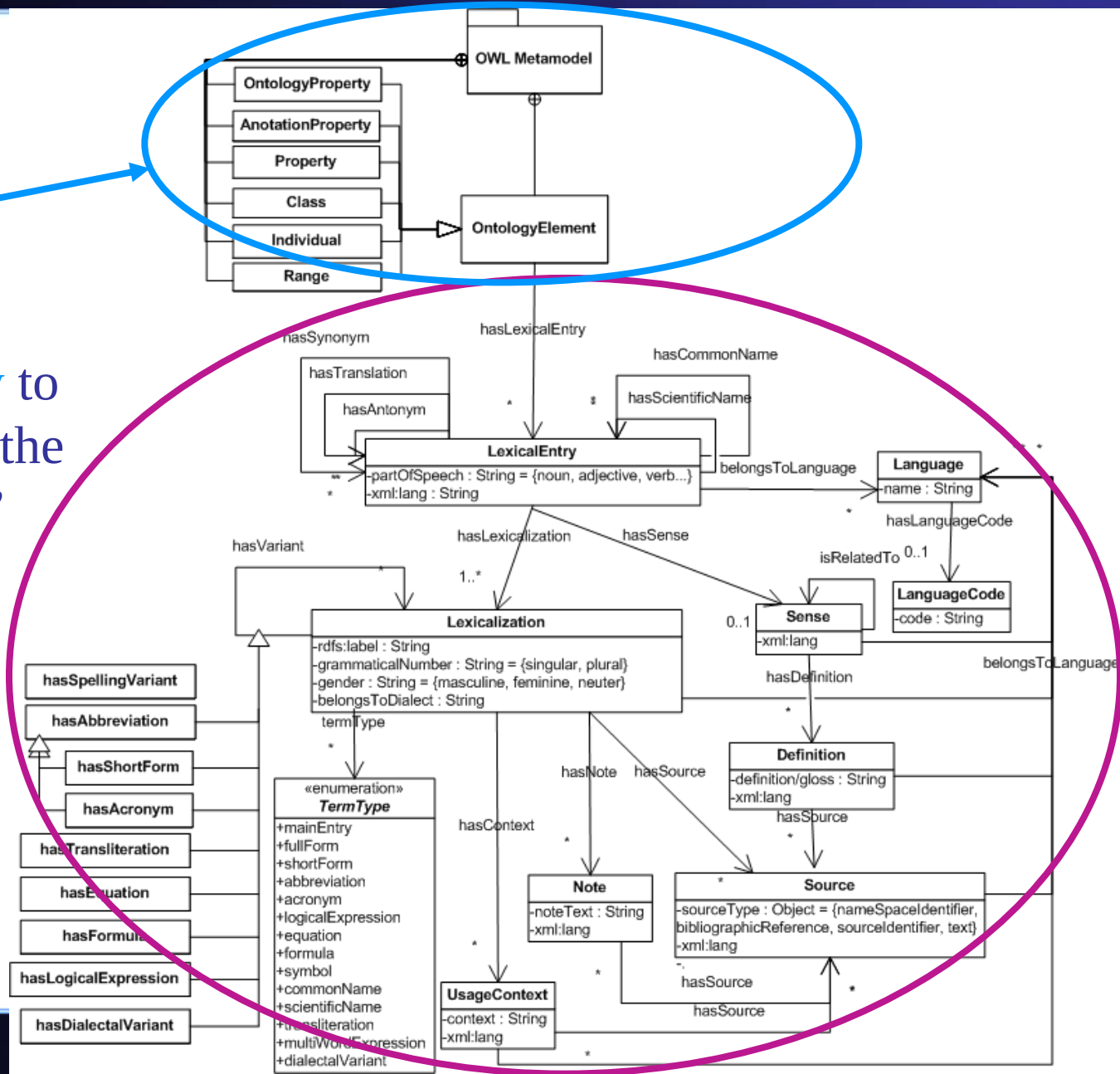


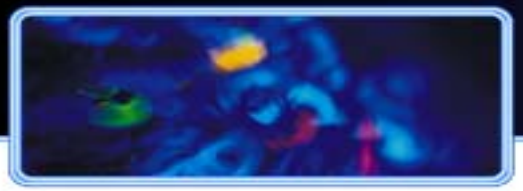
Linking the LIR and the Ontology

OWL Ontology Metamodel

Link of the Ontology to the LIR by means of the "*hasLexicalEntry*" relation

LIR





Conclusions

- LabelTranslator
  - A NeOn plugin that localizes ontologies
  - Lexical templates for compound labels
  - Disambiguation method to rank the translations
- LIR (multilingual meta-model)
  - Maintains the independence between the ontology structure and the multilingual layer
  - Allows for the linking of multilingual data to any ontology element
  - Localization of the ontological meaning as well as the capture of translation specificities
  - Interoperability and interchange with linguistic representation standards, as well as extensibility

Work on Multilingual Ontologies within the NeOn Project

Mauricio Espinoza, Elena Montiel-Ponsoda,
Oscar Corcho, Guadalupe Aguado, Asunción Gómez-Pérez

C21 Cost Action Towntology
October 20th, 2008