Ontologies, text and the lexicon

N. Aussenac-Gilles
IRIT- CNRS, équipe MELODI
aussenac@irit.fr

Out-line

- I – Introduction
  - Ontologies, lexicon: some definitions
  - 3 perspectives on the relation between ontology and language
- II – Knowledge representation issues
- III – Text as knowledge source

I – Introduction: ontology

- Ontology – a philosophical perspective
  - *discipline concerned with the study of existence and the formal description of the structure of reality*
- Formal ontology – logical perspective
  - *logical theories describing some aspect of reality, that is of relevance to a set of applications*
  - *a logical theory accounting for the intended meaning of a formal vocabulary i.e., its ontological commitment to a particular conceptualization of the world* (Guarino 1998), *by refering to the nature and structure of the entities.*
- ontologies – knowledge engineering perspective
  - *a formal explicit description of concepts in a domain of discourse (Noy and McGuinness, 2000)*

I – Introduction: structure of an ontology

- Classes (concepts) and their members (instances) : *university*, *Toulouse Capitole*
- subclass axioms: e.g., Science university is_subclass_of University
- non-taxonomic relations: e.g., University hasDean Person
- domain and range restrictions (on relation): restrict which kind of entities can stand at the domain and range positions, respectively
- cardinality constraints: University has (at least) one student
- Part-of relations
- Disjointness: disjoint classes have no common member

I – Introduction: lexicon

- Definition
  - Lexicon / Thesaurus / Lexical Data-base
  - List of words with definitions / and lexical relations / paper or stored in a DB
  - Linguistic grounding: the use of words in language determines their definitions
- *The main difference between an ontology and a lexicon, lexical database or thesaurus is that all of the latter are linguistic objects while ontologies are not (they are logical theories). (Ciminao, Volker, Buitelaar, 2010)*

I – Introduction: ontologies vs lexicons

ontology

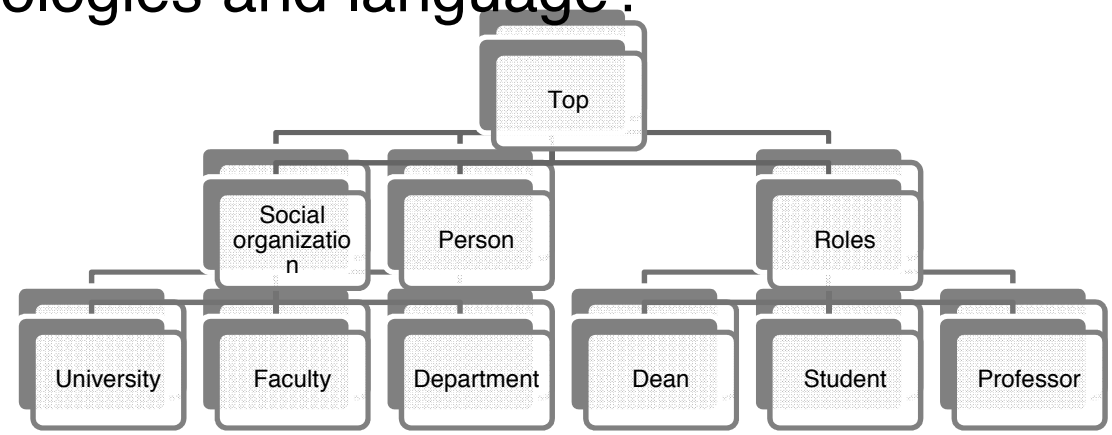
- Concepts, logical primitives
- SubclassOf hierarchy
- Semantic relations (partOf, ingredientOf, causes, hasDean, …)
- Formal definition
- To be created/ modified/ used by a program

lexicon

- Words or terms or entries
- Broader/narrower relations (hypernymy)
- Few lexical relations (seeAlso, meronymy, antinomy, …)
- Linguistic definition
- To be read, used by a human or a program

I - Ontologies versus lexicons

- Ontologies can be used
  - In man-system interaction system
  - For semantic annotation of documents
- Ontologies may be defined from the linguistic expression of knowlege
- Concepts and linguistic expressions are connected
- How to connect ontologies and language?



I- Ontologies, lexicon and text: 3 perspectives on this relation

- Design paradigm for ontologies
  - What theory/domain do we refer to give a meaning to the ontology concepts ?
  - Language can be the reference
- Knowledge representation
  - How to store the lexicon required to make formal concepts understable by a human?
- Knowledge « acquisition »
  - Building Ontologies from text
  - How can linguistic phrases and text be used as knowledge sources ?

I – Ontologies, Lexicon and text: Design paradigms for ontologies

- Issue: which evidences for concept definitions? Which means to access to the “right” concepts to be set in an ontology?
- Ex of paradigms
  - **Philosophical**: Ontology or Formal Ontology, distinction between predicates are based on properties that we are able to perceive or identify in word entities (Husserl)
  - **Linguistic**: language in use, lexical distinctions -> conceptual distinctions
  - **Pragmatic**: only those concepts relevant for the application will be defined
  - **Cognitive**: what is intuitively perceived as different leads to different classes (no need for an explicit and formal specification of differences).

I - Ontologies, text and lexicon: linguistic paradigms

- Possible assumptions
  - Linguistics, terminological distinctions justify conceptual distinctions
  - (or) classes exist independantly of their lexicalization
  - (or) classes are defined in the ontology if required by the application, and are then connected to their lexicalization

Out-line

- I – Introduction
  - Ontologies, lexicon: some definitions
  - 3 perspectives on the relation between ontology and language
- II – Knowledge representation issues
- III – Text as knowledge source

II – Knowledge representation

How to add some linguistic, terminological information to an ontology?

- Standard solution
- Early works
- Sophisticated (meta-)models of lexical ontologies
- Recent (meta-)models

II – Standard solution

- in OWL language: “Label” annotation

  Rdf:label “université des sciences”@fr

  Rdf:label “science university”@en

  <rdfs:label xml:lang=« fr »>université des sciences </rdfs:label>
- Annotation
  - All the labels of one concept and in the same langage are synonyms
  - Comment, unstructured data
  - No grammatical information (is it a lemma? An exact form? …)
- Need for a more complex and richer representation

Early works: lexical Ontologies (Maedche and Staab, 2000)

- Ontology:
  - Vocabulary : **predicates** for *categories* ($\mathcal{C}$, *unaires*), *properties* ($\mathcal{P}$, *unaires*) and *relations* ($\mathcal{R}$, *n-aires*)
  - $\mathcal{A}$: ensemble **d'axiomes**, formules logiques
- Lexical component of an ontology
  - **Lexical Entry**: Formes linguistiques exprimant les prédicats de l'ontologie : ens. $\mathcal{L}^{C}$, $\mathcal{L}^{\mathcal{P}}$, $\mathcal{L}^{\mathcal{R}}$ pour les catégories, propriétés et relations
  - **Reference relations** between lexical entries and predicates
- Termino-ontological resource

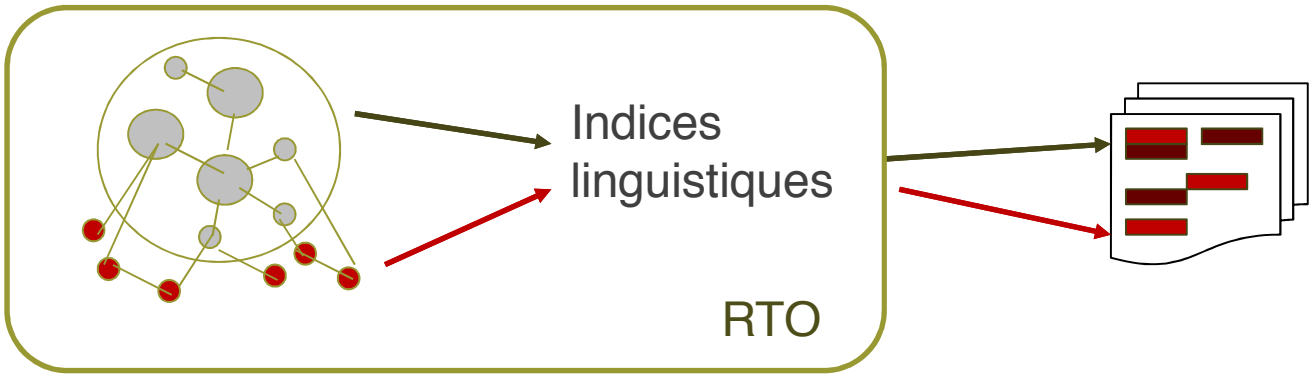
Lexical component of an ontology : example

- concepts atomiques :
  - Personne, Masculin, Féminine, Mère, En
- rôles atomiques :
  - parentDe, mèreDe, pèreDe
- concepts complexes :
  - $\text{Personne} \sqcap \text{Féminine}$
  - $\text{Personne} \sqcap \neg\text{Féminine}$
  - $\text{Personne} \sqcap \exists \text{parentDe}.\top$
  - $\text{Personne} \sqcap \forall \text{parentDe}.\bot$
  - $\text{Personne} \sqcap \exists \text{parentDe}.\top \sqcap \forall \text{parentDe.Fé}$
  - $\text{Personne} \sqcap (\text{Riche} \sqcup \exists \text{parentDe.Riche})$
  - $\text{Personne} \sqcap \exists \text{parentDe}.\exists \text{parentDe}.\top$

- Lexique des catégories
  - Personne, **individu**
  - Masculin
  - Féminin
  - mère, **maman**
  - Enfant
  - riche
- Lexique des relations
  - Parent de, mère de, père de
- Catégories complexes définies par des axiomes
  - Femme, homme
  - Parent
  - Enfant
  - Petit-enfant

Early works: Termino-ontological resources

- 2 meanings in the state of the art
  - Ontologie built from lminguistic resources (text)
  - Enriched ontology with linguistic information



II – Early works

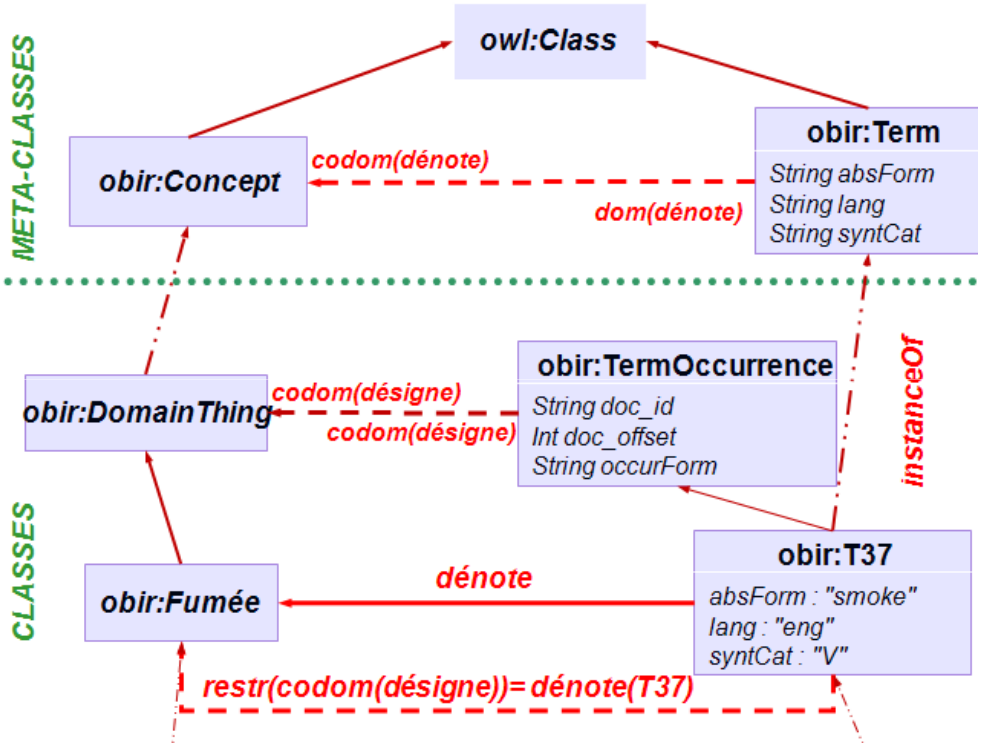
- Terminae (Biébow, Szulman, Aussenac-Gilles, 2005)
  - Distinction between terms and concepts
  - Concepts are associated terminological forms with rich information about the terms, its interpretations in various contexts.
  - Ex of data in a terminological form: POS, acronyms, various meanings, examples of use, standards and recommendations, …

Early works: a TOR meta-model

- Why defining terms as classes?
  - Associer des informations (POS, langue, ..)
- How?

Define a meta-model

  - OBIR (Reymonet et al., 2007) (Reymonet, 2009)
  - LingInfo (Buitelaar et al. 2006)



LingInfo meta-model (Buitelaar et al. 2006)

- Linguistic features added to meta-classes
- Goal= manage multilinguisme

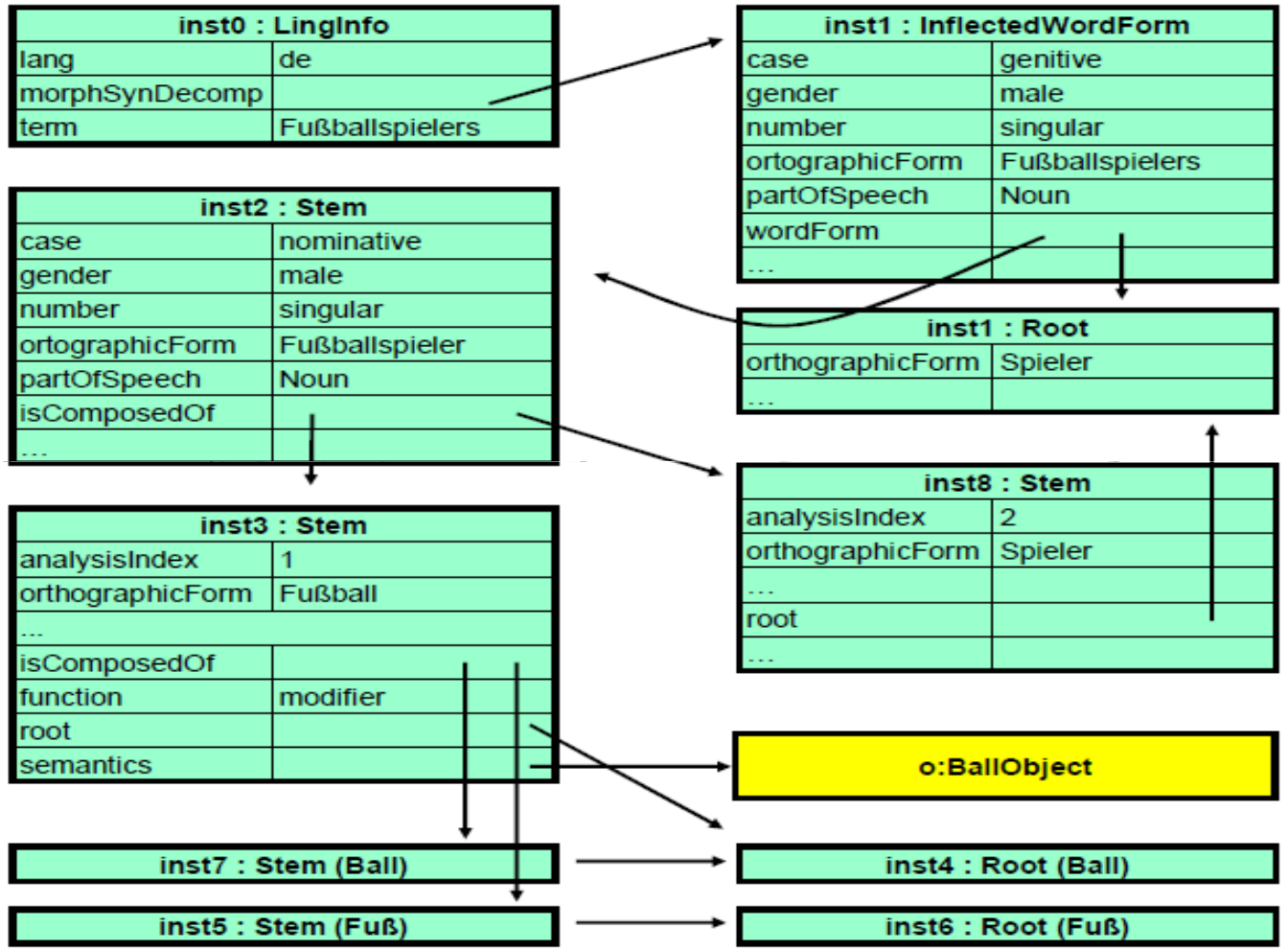


Figure 5: Morphosyntatic Decomposition of "Fußballspielers"

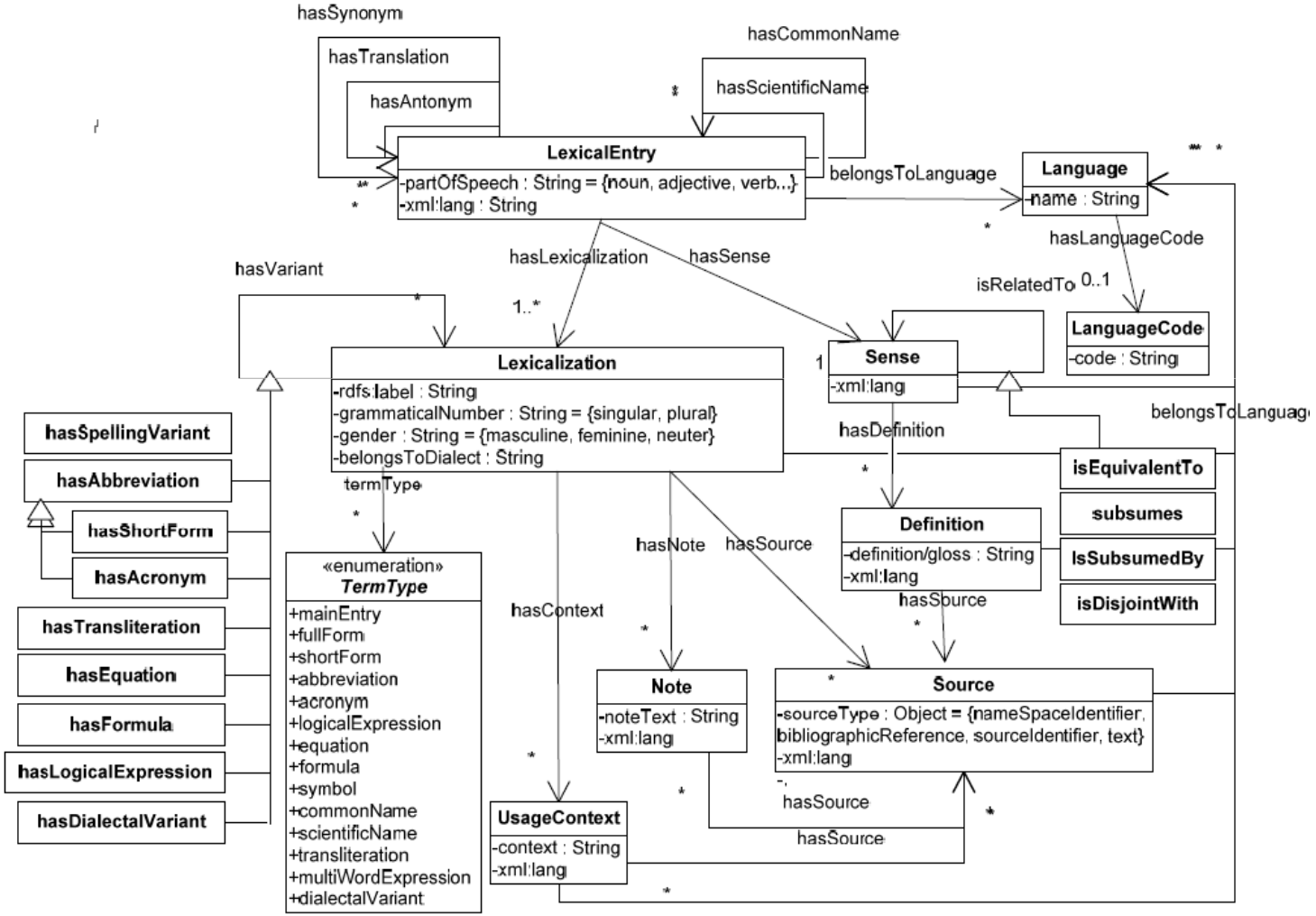
Sophisticated meta-models for lexical ontologies

- Define a linguistic model beside the ontology
  - Dedicated to NLP analyses
  - Makes it possible to manage word derivations, word agglutination, word combination and Multilinguism
  - Be more efficient when associating text and concepts
- examples
  - *LexOnto [Cimiano et al. 2007]*
  - *LexInfo [Buitelaar et al. 2009]*
  - *LIR in NEON (Monteil Ponsoda et al., 2008, 2011)*

Sophisticated meta-models for lexical ontologies: the NEON model

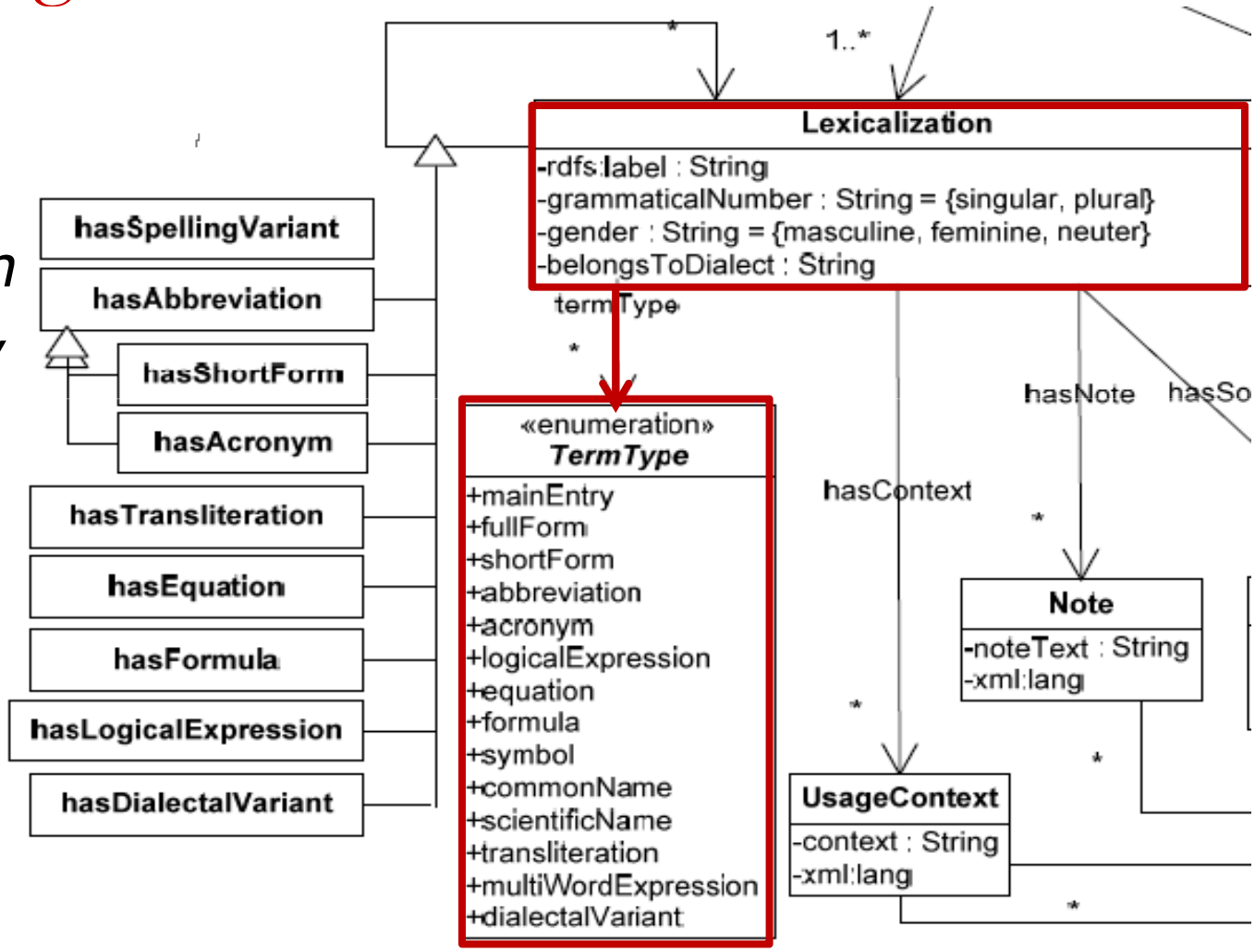
- LIR :

Linguistic Information Repository



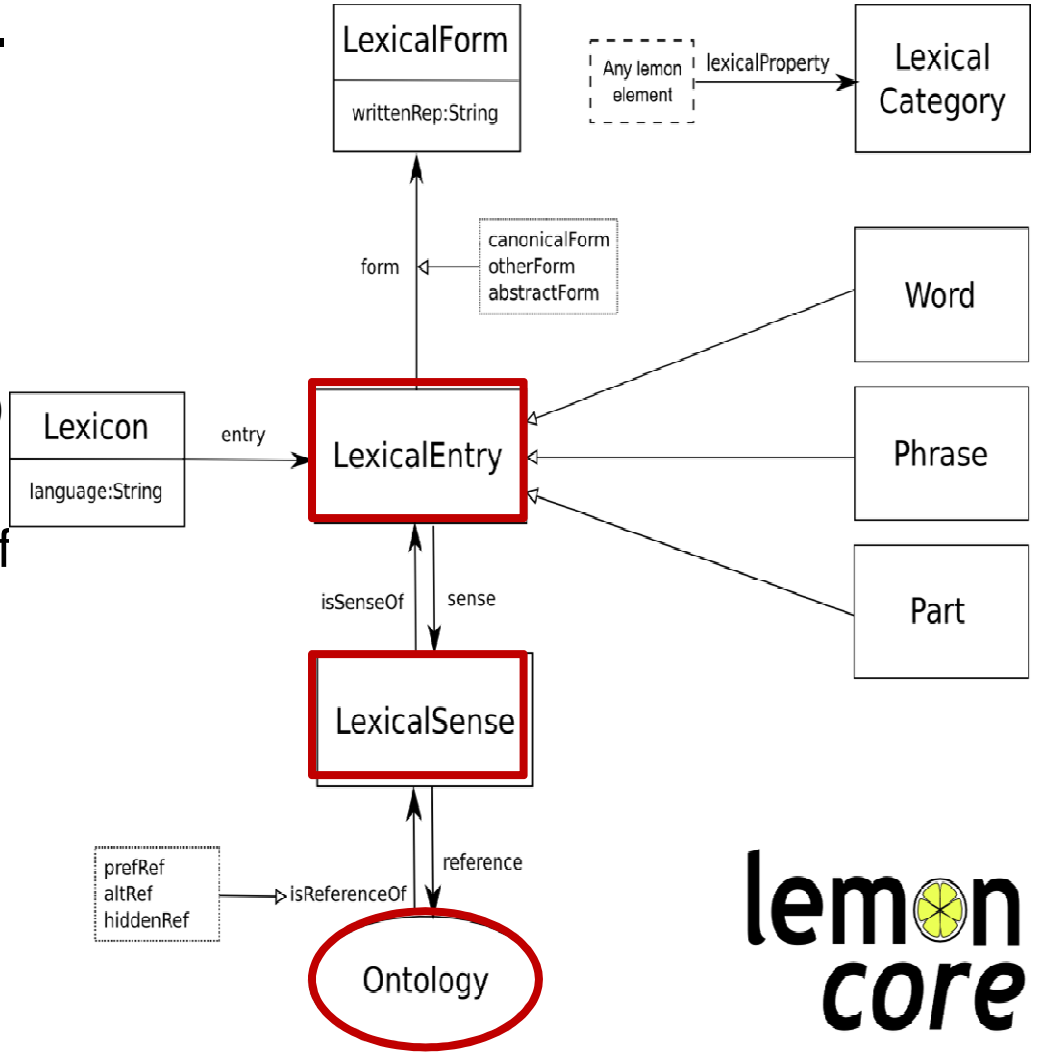
Sophisticated meta-models for lexical ontologies: the NEON model

- LIR : *Linguistic Information Repository*



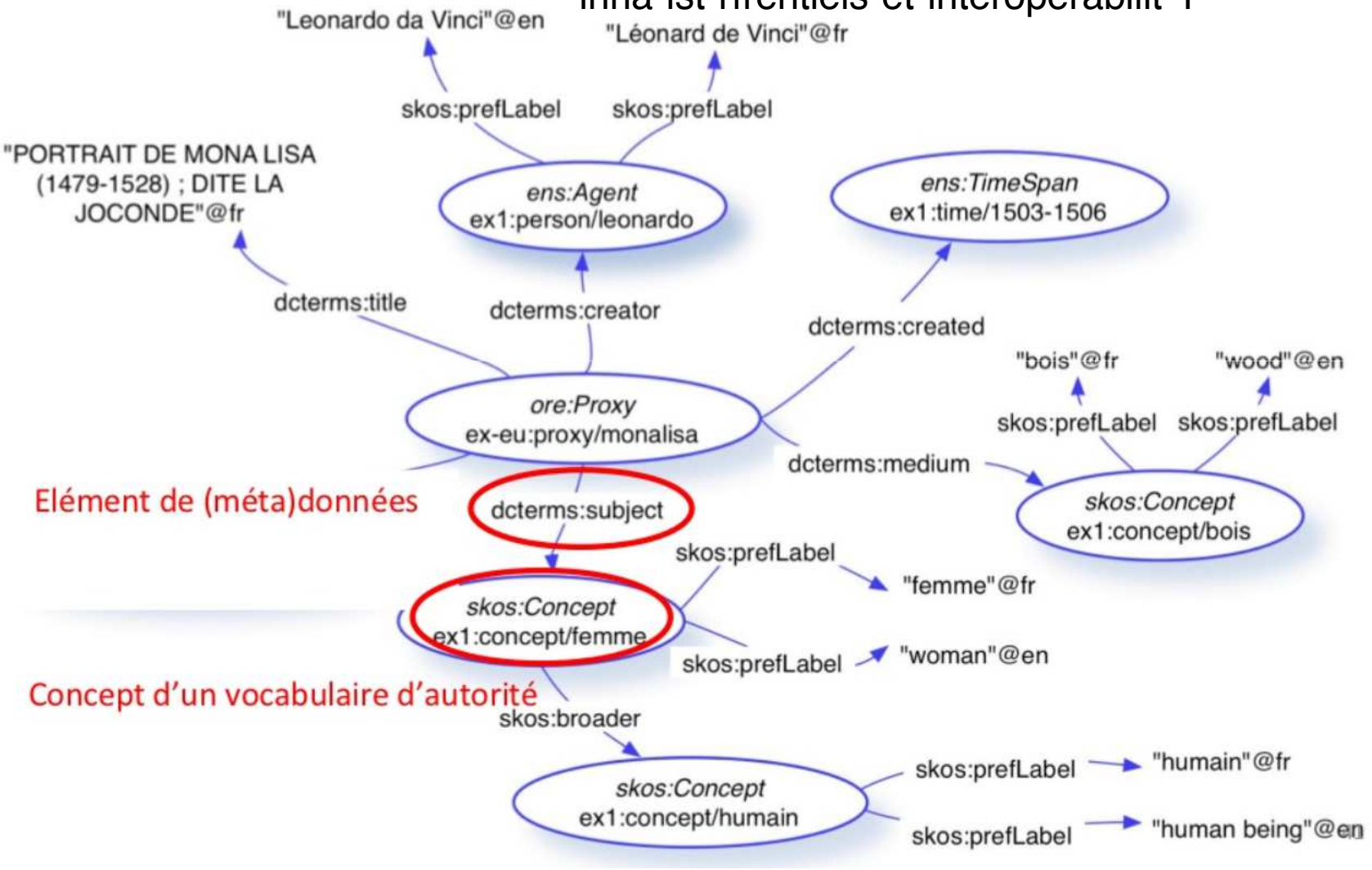
Stabilization : the LEMON meta-model

- **Main Classes of the meta-model:**
  - **Lexical Entry**
  - **Form:** inflectional variant
  - **Representation**: written versions) of the entry
  - **Sense (**one of the meanings of the lexical Entry)
  - **Reference**: actual meaning in the ontology



Terminae 2013 model and Europeana Model (Isaac, 2012)

http://fr.slideshare.net/antoineisaac/sminaire-inria-ist-rfrentiels-et-interoperabilit-1



III – Texts as knowledge sources

- What are the issues ?
- A layered approach
- Some tools

Quels sont les problèmes ?

- Given a linguistic formulation, how can I represent it?
- How can I identify all the required concepts and their linguistic formulations?
- Building ontologies versus populating ontologies with instances

How to represent a linguistic phrase

hémopéritoine : « *épanchement hématique localisé au niveau du péritoine* »

Hiérarchie de concepts

ETAT_PATHOLOGIQUE
- ETAT_PATHOLOGIQUE_LOCAL
  - **LESION**
    - adénopathie
    - …
    - épanchement
      - épanchement gazeux
      - épanchement liquidien
        - épanchement de pus
        - **épanchement hématique**

ANATOMIE
- …
- ANA_TISSU_ENVEL
  - capsule
  - duremère
  - mésentère
  - peau
  - …
  - **péritoine**

Hiérarchie de relation

LOCALISATION
- à_côté_de
- à_l'extérieur_de
- au_dessus_de
- …
- **au_niveau_de**

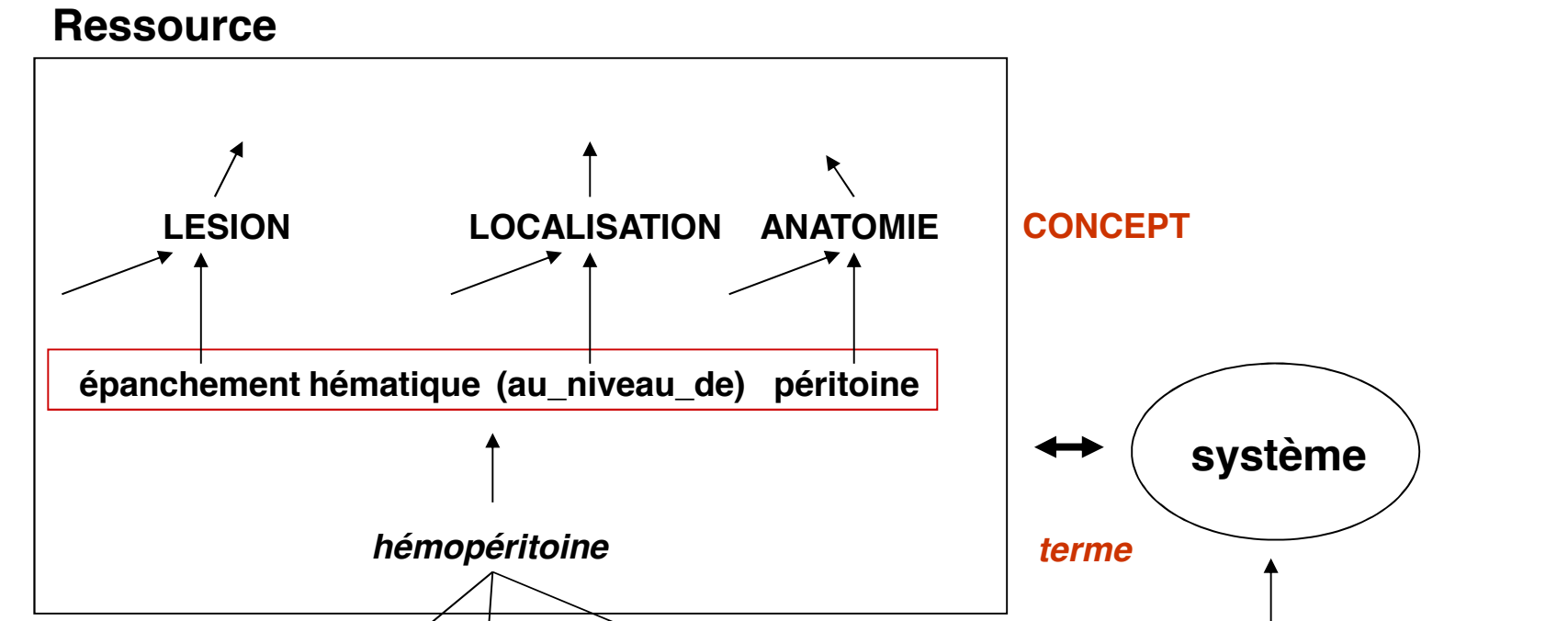
Liens

LESION (**LOCALISATION**) **ANATOMIE**

Concept défini

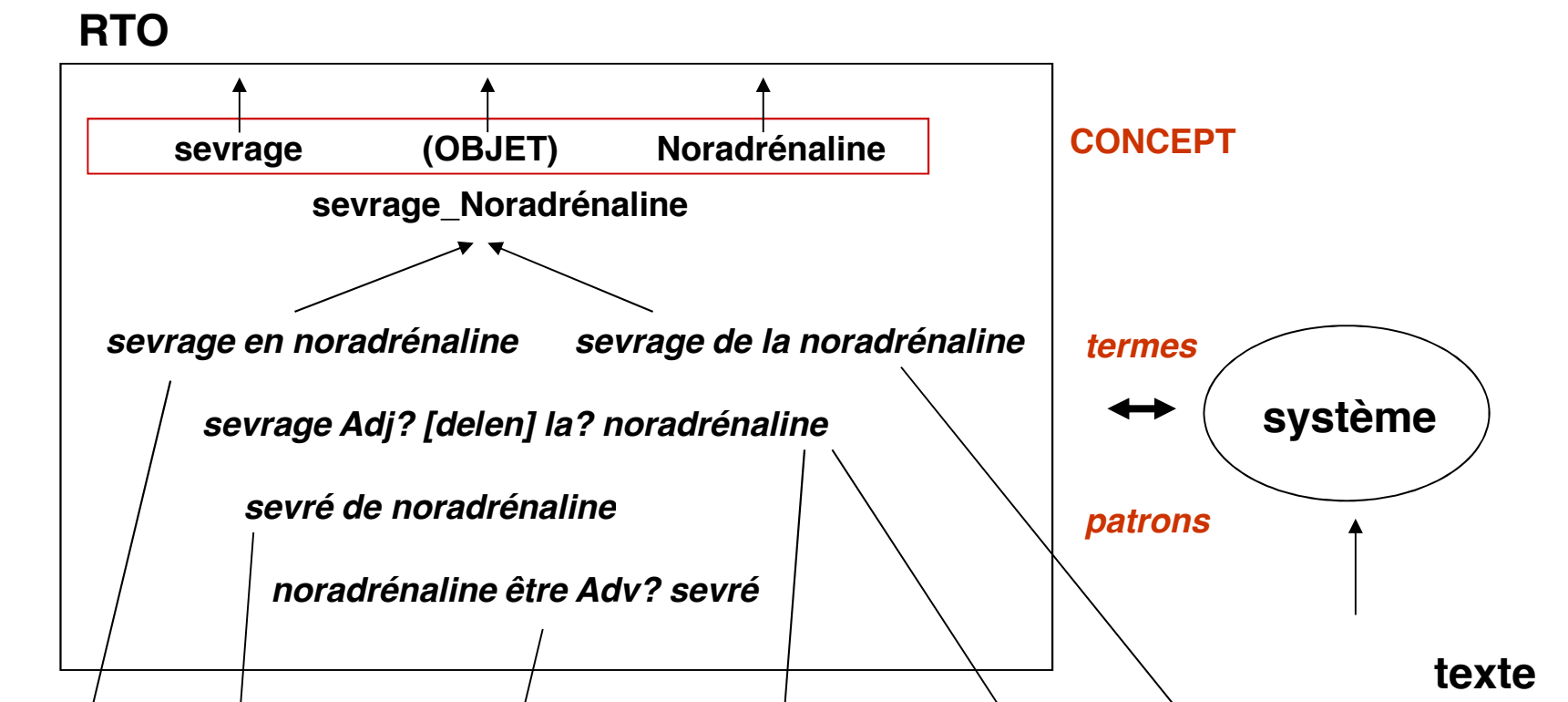
épanchement hématique **(au_niveau_de)** **péritoine**

Many lexical forms for one concept



L'échographie abdominale retrouve une contusion splénique et un **hémopéritoine.** Traumatisme abdominal avec une fracture de rate et un **hémopéritoine** abondant, ayant nécessité une splénectomie en urgence. Un nouvel état de choc apparaît associé à syndrome compartimentaire abdominal (**hémopéritoine** + hématome rétropéritonéal)) justifiant la laparotomie exploratrice. L'évacuation de **l'hémopéritoine** ne permet pas de mettre en évidence une cause nette au saignement ;Instabilité hémodynamique initiale, avec TA 80 / 60 et fréquence cardiaque à 120 / min en relation avec des pertes sanguines sur les foyers de fractures (échographie abdominale normale, absence d' **hémopéritoine** ou de lésion viscérale intrapéritonéale évidente)).deux échographies abdominales successives un **hémopéritoine** évolutif, sans pneumopéritoine, ainsi qu'un décollement péricardique postérieur.un traumatisme thoraco-abdominal avec fracture des arcs postérieurs des 7e, 8e et 9e côtes droites, contusion hépatique au niveau des segments VI-VII, **hémopéritoine** de moyenne abondance, contusion splénique,Instabilité hémodynamique nécessitant remplissage, transfusion de culots globulaires, PFC et plaquettes dans le cadre d'un **hémopéritoine** sur fracture du bassin.

Variants and their charactrization



sevrage en noradrénaline le 23 / 10. Stabilité cardiorespiratoire, le **sevrage de la Noradrénaline** est fait à J5 de la réintervention et l'extubation est réalisée à J6. Evolution favorable avec **sevrage** rapide **en Noradrénaline** le 20 / 06 / 99. Stabilité hémodynamique initiale correcte avec **sevrage** progressif **en Noradrénaline** obtenu à J3. Les suites du choc hypovolémique hémorragique sont simples, avec **sevrage de noradrénaline** à J3 et extubation à J2. L'évolution est alors favorable avec Instauration d'une corticothérapie ayant permis un **sevrage** rapide **de la Noradrénaline**, La patiente est **sevrée de noradrénaline** le 16 / 06 / 00. Le patient est **sevré de Noradrénaline** le 13 / 05 / 00. La **Noradrénaline est sevrée** dans la nuit du 20 au 21 / 01. Après remplissage vasculaire par 1000 cc de Plasmion et 1000 cc d'Elohes, la **Noradrénaline est** rapidement **sevrée**

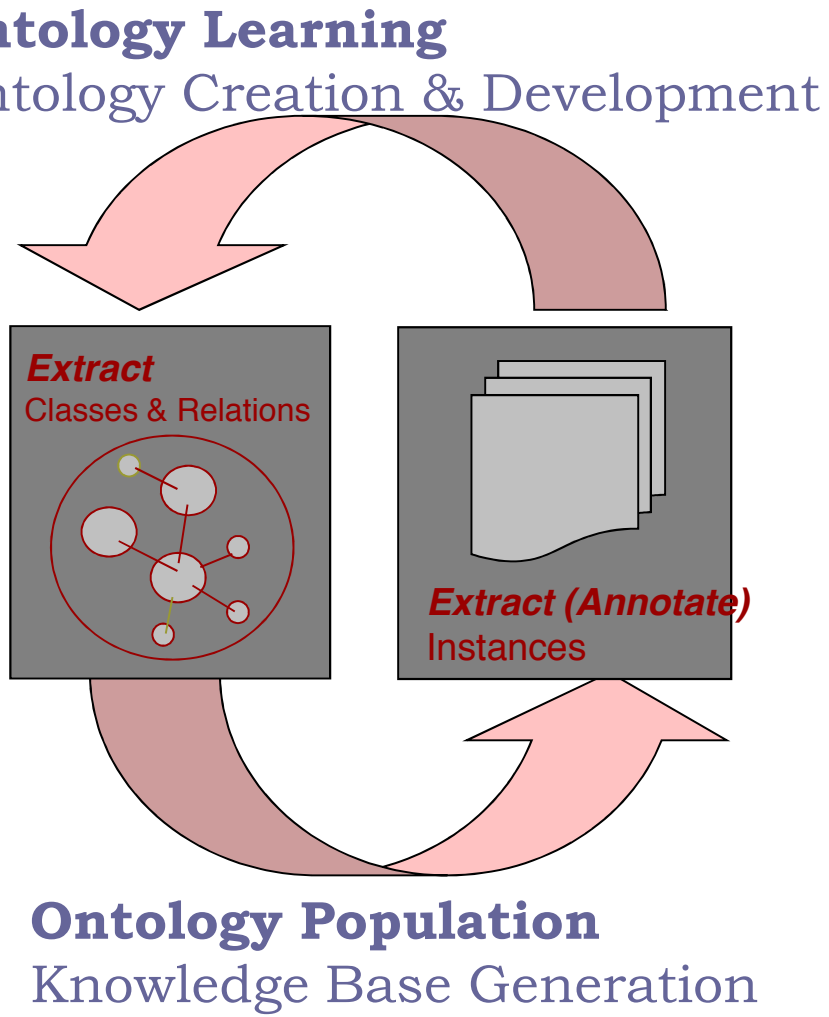
Ontology learning vs ontology population

Building ontologies

From linguistic clues to concepts, relations and properties

Ontology population

Instance extraction and semantic annotation



Layers of linguistic analyses (Buitelaar, EACL 2006)

Each person has one natural mother

Disjoint(person, city)

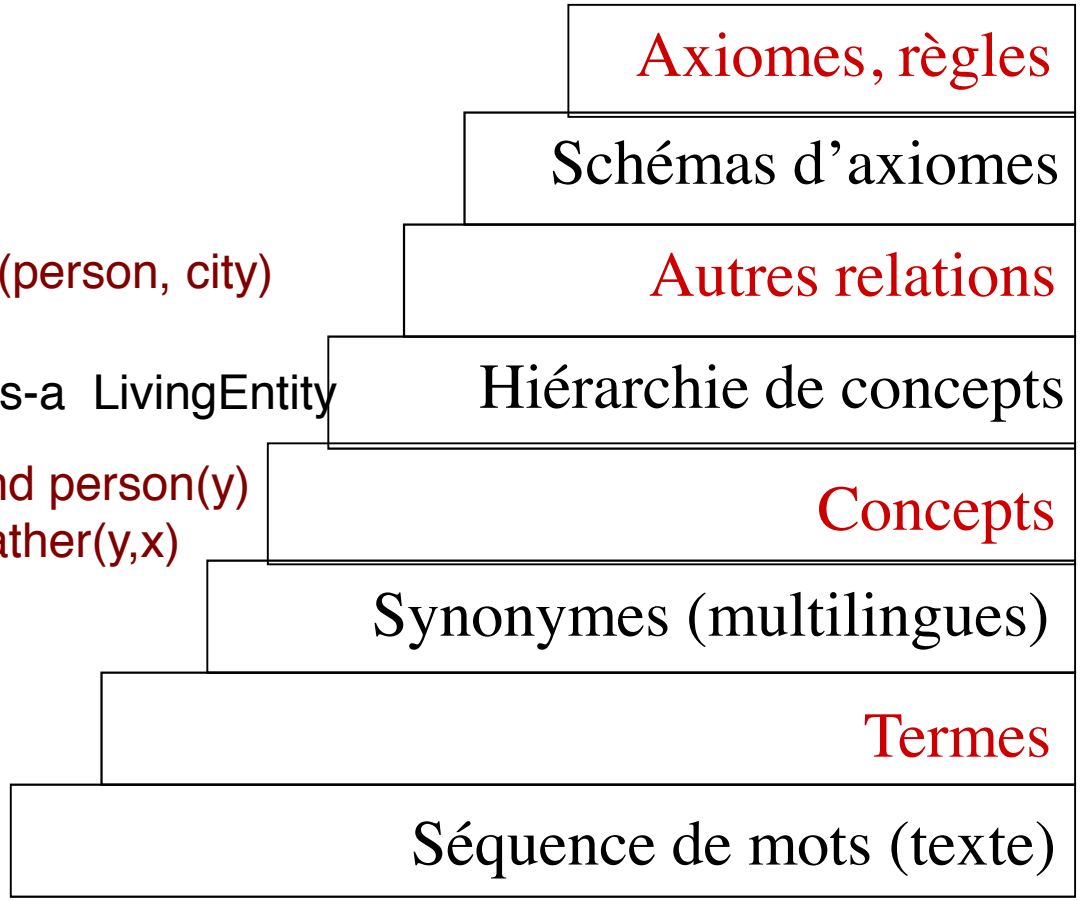
HasParent(Person, person) ; LivesIn(person, city)

AdoptiveFather is-a Father ; person is-a LivingEntity

C:father = {René,…} or (person(x) and person(y)
And hasParent(x,y) and male(y) -> father(y,x)

{father, dad, géniteur}
{husband, spouse, partner}

father, adoptive father, child, male …



Tools to build ontologies from text

- Dafoe http://dafoe4app.fr
- Text-to-Onto > Text2Onto
- OntoLearn (Navigli, 2004), TaxoLearn (2011)
- Protégé + OntoLing
- NEON + Gate *www.**neon**-project.org/*

Références

- MAECHE A., 2002, Ontology Learning for the semantic web, Kluwer
- Roberto NAVIGLI, and Paola VELARDI. (2004). Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites. *Computational Linguistics*, 50
- Paul BUITELAAR, Philipp CIMIANO, Bernardo MAGNINI. *Ontology learning from text: Methods, evaluation and applications*. Amsterdam : IOS Press, 2005
- Philipp CIMIANO. *Ontology learning and population from text: Algorithms, evaluation and applications*. Berlin : Heidelberg : Springer-Verlag, 2006
- Alexander MAEDCHE, Steffen STAAB. « Ontology learning ». *In* : S. Staab, R. Studer (eds.). *Handbook on ontologies*. 2nd ed. P. 245-268. Berlin : Heidelberg : Springer-Verlag, 2009
- Roberto NAVIGLI, Paola VELARDI, Stefano FARALLI. « A graph-based algorithm for inducing lexical taxonomies from scratch ». *In* : *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI 2011*), Barcelona, Catalonia, Spain, July 16-22, 2011