LexInfo Model and a comparison with the LIR approach

Tobias Wunner
Unit for Natural Language Processing
DERI - National University of Ireland, Galway

Joint Work with
Paul Buitelaar (DERI), Philipp Cimiano and John McCrae (Univ. of Bielefeld)
Overview

1. Lexicon creation process followed in LexInfo
2. The LexInfo Model
3. Examples
4. LIR and LexInfo
5. Experiments Ontology Alignment (OAEI dataset)
6. Services
7. Summary
Aug 2009 - Master Thesis: “Enhancing the Relation Extraction Process from Natural Language Sources Using Domain Ontologies” at University of Ulm (Dr. Thorsten Liebig)

- Enrich a domain ontology with linguistic information by:
  - adding regular expression patterns to ObjectProperties
  - link Domain/Range arguments to a linguistic domain model
- Semantic Similarity Measures based on Rbox structure to measure the error of the automatically created semantic annotation

Internship at DERI (NUIG) in NLP group (Dr. Paul Buitelaar)

- Development of LexInfo API
- Experiments with ontology alignment with OAEI dataset using LexInfo lexica

PhD in MONNET project
**Lexicon creation process & issues**

What are the requirements for a lexicon?

1. Separation & independence
   - Semantics in ontology
   - Syntax in lexicon
2. Complex linguistic structures
3. Morph. / syntactic decomposition
4. Syntactic behaviours
   - Variants
   - Synonyms
   - Predicate argument structure
5. Support for multilinguality

---

Example:

- freq "fisicoquímicas" vs.
- freq "químico-física" vs.
- freq "física-química"
Lexicon creation process & issues

use of LILAC and corpora statistics in LexInfo
The LexInfo model

- Builds on LMF (Lexical Markup Framework, ISO 24613) model
  - LMF relates ontological concepts i.e. Imf:Sense with a lexicalization via Imf:LexicalEntry
- Additional construct introduced in LexInfo
  - Part-Of-Speech specific LexicalEntry classes: e.g. lexinfo:Verb, lexinfo:Noun
  - Specific Subcategorization Frames: e.g. lexinfo:Transitive, lexinfo:NounPP
  - Phrase structur (e.g. lexinfo:Phrase, lexinfo:NounPhrase) with head
- LexInfo API with LILAC
  - LexInfo Lexicalization, Analysis & Construction (LILAC) rule language to create lexical structure from labels w.r.t. LexInfo model
Publishing the LexInfo vocabulary


LexInfo vocabulary: terms & definitions

---

http://vocab.deri.ie/lexinfo
http://vocab.deri.ie/lmf

---

LexInfo graph

LMF graph

---

LexInfo graph

---

LexInfo: terms & definitions

---

LexInfo: terms & definitions
The LexInfo model

main lexicon classes to describe a label
(1) LexicalEntries
(2) Phrases i.e. a special case of MWEs
(3) Subcategorization Frames
The LexInfo model
The LexInfo model - examples

- Simple Multiword Expression
  - \( \Rightarrow \) great vein of heart

- Prepositional Noun Phrase as Class Label
  - No Semantic Arguments
  - LILAC rules:
    - NP: NP of NP
    - NP: Adjective NP
The LexInfo model - examples

- Simple Multiword Expression
  - \( \Rightarrow \) great cardiac vein

- Adjective Phrase in Noun Phrase as Class Label
  - No Semantic Arguments
  - LILAC rules:
    - AdjectivePhrase: Adjective Adjective
    - NP: AdjectivePhrase NP
The LexInfo model - examples

- Simple MWE
  - => <City> is a **capital of** <Country>

- Prepositional Noun Phrase in Property Label
  - Semantic Arguments map to Domain/Range
  - LILAC rules:
    - NounPP: “is” Noun
      { subject=Range, object=Domain, preposition=”of” }
LIR (Linguistic Information Repository)

- Model which associates linguistic information with an OWL ontology
  - Accessability
  - Interoperability
  - Multilinguality
    - With lexicon semantics to allow for language dependent conceptualizations
- API to generate a lexicon within NeOn toolkit with Editor
  - http://www.neon-toolkit.org/
    - LabelTranslator
      - Localizes and translates domain ontologies
# LexInfo vs. LIR (Linguistic Informa)

<table>
<thead>
<tr>
<th><strong>LexInfo</strong></th>
<th><strong>LIR</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standalone API &amp; WebService</strong></td>
<td>Plugin in GUI with Editor for Knowledge Engineer &amp; Linguist</td>
</tr>
<tr>
<td><strong>Support for Multilinguality</strong></td>
<td>Support for Multilinguality</td>
</tr>
<tr>
<td><strong>Senses: Ontology Concepts i.e., lexicon only reflects syntactic inform.</strong></td>
<td>Senses: Different meanings of a word in the linguistic model</td>
</tr>
<tr>
<td><strong>Rich morphology model and support for subcategorization frames</strong></td>
<td>Support for syntactical variations and some morphology</td>
</tr>
<tr>
<td><strong>Strict separation between semantic and lexical level</strong></td>
<td>Possibility for localization and to have semantics in the lexicon</td>
</tr>
<tr>
<td><strong>LILAC rule language to decompose labels (w.r.t. Language &amp; ontology aspects)</strong></td>
<td>Lexical template database to translate labels</td>
</tr>
</tbody>
</table>
Benchmark Development

- Analyze possible term ('ontology label') variations - in context of ontology alignment task
  - Lexical acronyms
  - Morphological inflectional, derivational
  - Syntactic phrase structure, subcat structure
  - Semantic synonyms, generalizations, specifications
  - Multilingual translations
Syntactic Variation: Phrase Structure

domain ontology

lexicon

voc1

- cell component
  - has component1
  - has component2

- noun phrase
  - has lemma
    - cell

voc2

- component of cell
  - has component1
  - has component2

- noun phrase preposition
  - Prep: of
  - has lemma
    - component

Enabling networked knowledge.
Syntactic Variation: SubCat
## Variations: Mouse-Human Anatomy

<table>
<thead>
<tr>
<th>Mouse OWL ID</th>
<th>Human OWL ID</th>
<th>Description</th>
<th>Mouse</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://mouse.owl#MA_1">http://mouse.owl#MA_1</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 thoracic vertebra 5</td>
<td>t5 vertebra</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_2">http://mouse.owl#MA_2</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 lumbar vertebra 2</td>
<td>l2 vertebra</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_3">http://mouse.owl#MA_3</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 thoracic vertebra 8</td>
<td>t8 vertebra</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_4">http://mouse.owl#MA_4</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 thoracic vertebra 3</td>
<td>t3 vertebra</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_5">http://mouse.owl#MA_5</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 clitoral gland</td>
<td>glans clitoris</td>
<td></td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_6">http://mouse.owl#MA_6</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 left lung hilus</td>
<td>hilar area of the left lung</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_7">http://mouse.owl#MA_7</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 pancreas secretion</td>
<td>pancreatic secretion</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_8">http://mouse.owl#MA_8</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 spleen capsule</td>
<td>splenic capsule</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_9">http://mouse.owl#MA_9</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 larynx mucous gland</td>
<td>laryngeal mucous salivary gland</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_10">http://mouse.owl#MA_10</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 spleen lymphoid follicle</td>
<td>splenic lymphoid follicle</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_11">http://mouse.owl#MA_11</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 pelvis bone</td>
<td>pelvic bone</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_12">http://mouse.owl#MA_12</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 adrenal gland zona reticularis</td>
<td>reticularis zone</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_13">http://mouse.owl#MA_13</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 sensory organ system</td>
<td>special sense organ system</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_14">http://mouse.owl#MA_14</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 spleen venous sinus</td>
<td>splenic sinus</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_15">http://mouse.owl#MA_15</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 adrenal gland zona fasciculata</td>
<td>fasciculata zone</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_16">http://mouse.owl#MA_16</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 kidney collecting duct</td>
<td>collecting tube</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_17">http://mouse.owl#MA_17</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 external naris</td>
<td>nostril</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_18">http://mouse.owl#MA_18</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 efferent arteriole</td>
<td>renal efferent vessel</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_19">http://mouse.owl#MA_19</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 blood vessel endothelium</td>
<td>vascular endothelium</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_20">http://mouse.owl#MA_20</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>4 digestive system fluid/secretion</td>
<td>gastrointestinal fluid or secretion</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_21">http://mouse.owl#MA_21</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 brain grey matter</td>
<td>brain white matter</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_22">http://mouse.owl#MA_22</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 internal nar</td>
<td>internal nar</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_23">http://mouse.owl#MA_23</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 digestive system</td>
<td>gastrointestinal system</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_24">http://mouse.owl#MA_24</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 gastrointestinal system</td>
<td>gastrointestinal tract</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_26">http://mouse.owl#MA_26</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 corpora quadrigemina</td>
<td>quadrigeminal body</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_27">http://mouse.owl#MA_27</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>1 caudal auricular vein</td>
<td>posterior auricular vein</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_28">http://mouse.owl#MA_28</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 cortical layer VI</td>
<td>multiform cell layer</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_29">http://mouse.owl#MA_29</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 ovary mature follicle</td>
<td>graafian follicle</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_30">http://mouse.owl#MA_30</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 great vein of heart</td>
<td>great cardiac vein</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_31">http://mouse.owl#MA_31</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>3 eye sebaceous gland</td>
<td>sebaceous gland of the eyelash</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_33">http://mouse.owl#MA_33</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>3 right lung caudal lobe</td>
<td>lower lobe of the right lung</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_34">http://mouse.owl#MA_34</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>5 lateral cuneiform</td>
<td>external cuneiform bone of the foot</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_36">http://mouse.owl#MA_36</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>peripheral nervous system ganglion</td>
<td>ganglion</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_37">http://mouse.owl#MA_37</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>urinary bladder urothelium</td>
<td>bladder transitional cell epithelium</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://mouse.owl#MA_38">http://mouse.owl#MA_38</a></td>
<td><a href="http://human.owl#">http://human.owl#</a></td>
<td>2 pallidum</td>
<td>globus pallidus</td>
<td>Y</td>
</tr>
</tbody>
</table>

**Variations Types:**
- ‘acronymic’ variations
- morphologic variations
- multilingual/semantic variations
- semantic variations
LexInfo Webservice (First Version)

LexInfo Lexicalization Service

This service creates a lexicon for a given Domain Ontology based on the Lexinfo model [1] and API [2]. The service analyzes the String values stored in the labels. The labels can be encoded in various formats (e.g., RDFs label, SKOS). In case the label is language dependent, the service can be set up for a specific language (e.g., extract all German labels). If the Domain Ontology does not use label format the URI fragment is extracted. The format of the output lexicon is described by the Lexinfo ontology (light version on vcosb.de/lexinfo or OWL-Full version on www.lexinfo.net/lexinfo).

Enter ontology file: [Browse...]
Labels stored by: [Automatic]
Language: [English]
Lexicon URI: [http://server/path]
Show statistics: [☐]
Corpora: [None]

LILAC Rules [show/hide]

Generate [this may take a long time.]
Alternatively we can e-mail you the result when we are done [☐]

References:

http://greententacle.techfak.uni-bielefeld.de:8009/LexiconService/
## LexInfo Webservice (First Version)

**OWL Lexicon:**
- Download OWL Ontology

### Results:

<table>
<thead>
<tr>
<th>URI</th>
<th>OWL Type</th>
<th>Aggregate type</th>
<th>POS Structure</th>
<th>Lexicalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://nlp.deri.ie/2010/2/madrid.owl#City">http://nlp.deri.ie/2010/2/madrid.owl#City</a></td>
<td>Class</td>
<td>Noun</td>
<td>[Noun]</td>
<td>city</td>
</tr>
<tr>
<td><a href="http://www.w3.org/2002/07/owl#Thing">http://www.w3.org/2002/07/owl#Thing</a></td>
<td>Class</td>
<td>Noun</td>
<td>[Noun]</td>
<td>thing</td>
</tr>
<tr>
<td><a href="http://nlp.deri.ie/2010/2/madrid.owl#chemical_physical_property">http://nlp.deri.ie/2010/2/madrid.owl#chemical_physical_property</a></td>
<td>DataProperty</td>
<td>NounPhrase_NounPP</td>
<td>[[Noun, [Adjective, [Noun]]]]</td>
<td>The chemical physical property of Y is X</td>
</tr>
</tbody>
</table>

Total aggregates generated = 13/13 (100%)

By Type:
- Noun: 7 (53.85%)
- NounPhrase: 3 (23.08%)
- NounPP: 2 (15.38%)
- NounPhrase_NounPP: 1 (7.69%)
Summary

- LexInfo model based on ISO standard LMF
  - Automatic lexicon generation for a domain ontology
  - Linguistic grounding of a domain ontology
  - Published in vocab.deri.ie (http://vocab.deri.ie/lexinfo)
- LexInfo API (http://code.google.com/p/lexinfo/)
  - LILAC rules for en+es+de
  - POS Taggers (LingPipe, TreeTagger, Stanford) for en+es+de+fr+it+nl
  - Corpora methods under development
- LexInfo Webservice (http://greententacle.techfak.uni-bielefeld.de:8009/LexiconService/)
  - Webservice to online generate LexInfo lexica
LexInfo & Previous Models

- **LingInfo**: modeling morphosyntactic decomposition of (complex) terms [Buitelaar et al. 2006]

- **LexOnto**: capturing syntactic behaviour and syntax-semantics links [Cimiano et al. 2007]

- **Lexical Markup Framework (LMF)**: ISO standardized model for representing machine readable lexica (agnostic about connection with ontology) [Francopoulo et al. 2007]

- **LexInfo**: building on LMF as a core, develop a model which “subsumes” LingInfo and LexOnto for flexibly associating linguistic information to ontologies [Buitelaar, Cimiano, Haase, Sintek 2009]

- **Localizing Ontologies in OWL**, [E. Montiel-Ponsado, G. Aguado de Cea, A. Gómez-Pérez 2007]