Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

A. Gómez-Pérez
Universidad Politécnica de Madrid
asun@fi.upm.es

Acknowledgements: Jorge Gracia (UPM), Victor Rodríguez Doncel (UPM), LIDER Consortium members Nurial Bel (UPF), Marta Villegas (UPF)
Context – Ontology Engineering Group

Directors: A. Gómez-Pérez, O. Corcho
Position: 8th in the UPM ranking (200 groups)
Founded: 1994

- Research Group (30 people)
- Experience on
  1. Ontologies, Semantic Web, Linked Data, Open Data
  2. Semantic E-science
  3. Multilingualism

- ODI Madrid: Madrid Node of the Open Data Institute

- Projects
  - 27 EU projects (7 as coordinator)
  - 54 National Projects
  - 27 contracts with companies

- Standardization activities
  - >25 @ W3C, ISO, OASIS, etc.

- Impact of publications H-index (scholar)
  - Asunción Gómez-Pérez (h: 50, citations 14852)
  - Oscar Corcho García (h: 36, citations 8152)

- Services to the Spanish community
  - esDbpedia
  - linkeddata.es
  - vocab.linkeddata.es

http://www.oeg-upm.net/
https://github.com/oeg-upm
@oeg-upm

170+ Past Collaborators
50+ Past Visitors
• This work is licensed under the Creative Commons Attribution – Non Commercial – Share Alike License

• You are free:
  - to Share — to copy, distribute and transmit the work
  - to Remix — to adapt the work

• Under the following conditions
  - Attribution — You must attribute the work by inserting
    • “[source http://www.oeg-upm.net/]” at the footer of each reused slide

• a credits slide stating: “Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources ” by A. Gómez-Pérez ”
  - Non-commercial
  - Share-Alike
Language Resources
Finding and reusing LR in third party applications is manual and time consuming.
Lack of interoperability of Language resources

- Ecosystem of
  - Open and Closed resources
  - Silos of LRs
  - Complementary resources
    - Lexicon, Corpora, Dictionaries, Grammars, …
  - Heterogeneous formats
    - E.g., for Lexicons: Lexinfo, LMF, LIR, Lemon, …
  - Several repositories with different metadata and schemas
  - Many APIs and services for querying

Discovery and reuse LR in third party applications is hard, manual and time consuming
The Open Data stars

- Make your stuff available on the Web (whatever format) under an open license
- Make it available as structured data (e.g., Excel instead of image scan or a table)
- Use non-proprietary formats (e.g., CSV instead of Excel)
- Use URIs to identify things, so that people can point at your stuff
- Link your data to other data to provide context
Diccionario de la lengua española

El Diccionario de la lengua española (DRAE) es la obra de referencia de la Academia. La edición actual —la 22.ª, publicada en 2001— incluye más de 88 000 entradas.

red

á á é í ó ú ñ

red.

(Del lat. rete).

1. f. Aparejo hecho con hilos, cuerdas o alambres trabados en forma de mallas, y convenientemente dispuesto para pescar, cazar, cercar, sujetar, etc.

2. f. Labor o tejido de mallas.

3. f. redecilla (ll prenda de malla para el pelo).

4. f. Lugar donde se vende pan y otras cosas que se dan por

Artículo enmendado
Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

Asunción Gómez-Pérez

The Luxembourg BabelNet Workshop 2-3rd, March 2016
Language Resources

THE NEED OF CONNECTING LICENSED LANGUAGE RESOURCES
“Red”

Pronunciation: [red]

Grammar category: sustantivo femenino

Singular: “red”

Plural: “redes”

Complementary but not connected

“Red_de_computadores”

Category: redes informáticas

Image

Etimologiy Del latin “rete”

Gender: “f”

Definition.: “Conjunto de ordenadores o de equipos informáticos conectados entre sí....”

“Red”

Sinonyms: “sistema”, “malla”, “distribución”
Integration of Linguistic data

**Singular**

[RED]

**Plural**

[REDES]

**Form**

Red

**Feminine**

“red”

**Form**

“malla”

**Red**

**Sense**

“red”

**Translation**

es - en

**Wikienda del español**

“network”

**Written form**

“network”

**Written form**

“red”

**Image**

“red”
Language Resources
The need of connecting licensed LRs
LINKED DATA FOUNDATIONS
RDF(S) models

Unique identifiers: URI
identify or name a resource

Equivalence links to other datasets
Same As

Data navigation

http://iflastandards.info/ns/fr/frbr/frbrer/C1005
http://iflastandards.info/ns/fr/frbr/frbrer/C1001

Person

Is creator of

Is a

Work

El Quijote

http://datos.bne.es/resource/XX1718747
http://datos.bne.es/resource/XX3383563

Cervantes

Same As

http://dbpedia.org/resource/Miguel_de_Cervantes
http://viaf.org/viaf/17220427

Cervantes

http://datos.bne.es/resource/XX1718747
http://datos.bne.es/resource/XX3383563
The model (Ontology) and the data

Idiom

Year

Work

translation

Is creator of

Person

birthPlace

Place

Library

Located at

Has subject

Ontology

Data

1960

Publication date

translation

El Quijote

Is creator of

Cervantes

1960

Located in

18

BNE

Located at

Has subject

Vida de Cervantes

© Asunción Gómez-Pérez

The Luxembourg BabelNet Workshop 2-3rd, march 2016
Linked data is full of URIs

http://iflastandards.info/ns/fr/frbr/frbrer/C1001

http://datos.bne.es/resource/XX3383563

http://datos.bne.es/resource/bimo0002045496
Language Resources
The need of connecting licensed Language Resources
Linked Data foundations

LINGUISTIC LINKED LICENSED DATA
Linked Data interconnects data from resources

- In many domains
- In many languages
- Open and closed License
- Links with other datasets

Linguistic Linked Data Cloud

- Subset of LOD
- Linguistic domain
- Many type of resources
- Interconnected with other Language Resources
- Enables the lexicalization of data on the web, not necessarily data in the LD format
- Enables a new generation of LD-aware NLP and MT Services
What is 3LD?

3LD
Linguistic Linked Licensed Data

Language resources such as:
- Lexica
- Corpora
- Dictionaries
- Grammars ..

Using RDF and standard data models (vocabularies):
- Lexica
- Corpora
- .....

Published along with a machine-readable license.
ODRL
Open Digital Rights Language

www.lider-project.eu
Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

© Asunción Gómez-Pérez

The Luxembourg BabelNet Workshop 2-3rd, march 2016
Linguistic Linked Licensed Data @ Nov 2014

LLOD Cloud in November 2015
- 126 Resources
- 203 Links
- 20 Corpora

Less Centralized: Babelnet, LexVo and LexInfo new hubs

Criteria for inclusion:
- **Resolvable**: URLs that resolve
- **RDF**: resolve to RDF
- **1000 Triples**: self-explaining
- **Links**: to one resource from the cloud or other 50 links
- **Crawlable**: get the whole resource by crawling
- **Linguistic**: data must be a language resources
- **Registered**: at CKAN
The core of the Linguistic Linked Open Data cloud!
How do we represent license information?

Rights expression in ODRL:
Who [can|cannot|must] act what in which resource how

odrl:Rule
   (odrl:Permission)  (odrl:Prohibition)  (odrl:Duty)

odrl:Asset
odrl:Action
odrl:Constraint
odrl:Assignee
odrl:Party
Using Policies to govern conditional access to Linked Data

Example of access to Linked Data for a price (15 EUR for the dataset or 0.01 EUR for a triple thereof)

```xml
@prefix gr: <http://purl.org/goodrelations/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
<http://salonica.dia.fi.upm.es/ldr/policy/cdaddba4-fc2e-4ee0-a784-e62f1db259bf> a odrl:Set ;
  rdfs:label "License Offering Paid Linked Data" ;
  odrl:permission [ a odrl:Permission ;
    odrl:target <http://example.org/dataset/ds01> ;
    odrl:action odrl:reproduce ;
    odrl:duty [ a odrl:Duty ;
      rdfs:label "Pay" ;
      gr:UnitOfMeasurement dcat:Dataset ;
      gr:amountOfThisGood "1" ;
      odrl:action odrl:pay ;
      odrl:target "15,00 EUR"
    ] ,
    [ a odrl:Duty ;
      rdfs:label "Pay" ;
      gr:UnitOfMeasurement rdf:Statement ;
      gr:amountOfThisGood "1" ;
      odrl:action odrl:pay ;
      odrl:target "0,01 EUR"
    ] ]
  ,
  odrl:action odrl:reproduce ;
  odrl:target <http://example.org/dataset/ds01> ;
  [ a odrl:Duty ;
    rdfs:label "Pay" ;
    gr:UnitOfMeasurement rdf:Statement ;
    gr:amountOfThisGood "1" ;
    odrl:action odrl:pay ;
    odrl:target "0,01 EUR"
  ] .
```

http://conditional.linkeddata.es
1. **Agree on vocabularies** for describing
   - Domain datasets
   - LR metadata
   - LR content (Lemon-Ontolex, NIF, …)

2. **Unified and standardized language** for describing resources (RDF(S))

3. **Unified and standardized query language** (SPARQL)

4. **Standardized non-proprietary APIs**

5. **Links** to other resources

www.lider-project.eu
Language Resources

The need of connecting licensed Language Resources
Linked Data foundations
Linguistic Linked Licensed Data

OUTCOMES OF THE LIDER PROJECT
Discovery LR metadata with LingHub

Language resource metadata transformation and aggregation

- Metadata OWL ontology @ LD4LT W3C group
- Exposure of Metashare, Clarin, LREMap, datahub metadata as LD

LingHub

- Collect and Homogenise metadata
- Enhance metadata
- Helps you to search LRs

www.lider-project.eu

http://linghub.lider-project.eu/
Conversion of 86 LRs data into LD

86 Resources transformed:
- Princeton WordNet,
- BabelNet
- IATE
- Wikipedia Bitaxonomy
- Terminesp (spanish)
- Termcat
- 22 Apertium bilingual dictionaries
- the Brown Corpus
- Wikilinks Corpus
- 27 wordnets from the Open Multilingual Wordnet
- 22 lexical resources for Swedish, including SALDO

Use of models and guidelines from WP4 standardization activities
- Use of Ontolex-lemon Model from the Ontolex group at W3C
- BPMLOD@W3C guidelines

NIF Refinement by the community

Resources transformed by the community
- LDL 2014
- Summer datathon on LLD

www.lider-project.eu
8 Guidelines for Linguistic Linked License Data
- Generation of
  - Multilingual Dictionaries (BabelNet)
  - Bilingual Dictionaries
  - Multilingual Terminologies (TBX)
  - WordNets
- Corpus creation using NIF
- NIF-based NLP Web Services
- LLD Exploitation
- LLOD aware services

Best practices for Multilingual Linked Data Publication
- Practices for Naming (URIs)
- Practices for Dereferencing
- Practices for Textual Information
- Practices for Linking
- Practices for Language Identification

http://www.lider-project.eu/guidelines

www.lider-project.eu
Reference Architecture

Certification

Benchmarking & Validation

Discovery

LLD Linking

Service Composition

LLD Publishing

LLD-aware Services

Metadata

Licensing

Provenance

Multilingual Data

Guidelines and Standardization

Vocabularies

Hosting

Scalability

Streaming

Interoperability

www.lider-project.eu
Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

© Asunción Gómez-Pérez

The Luxembourg BabelNet Workshop - 3rd, March 2016

Reference Architecture and Reference Cards

- Certifications
- Benchmarking & Validation
- Discovery
- LLD Linking
- LLD Publishing
- Metadata
- Service Composition
- LLD-aware Services
- Licensing
- Provenance

Vocabularies
Hostings
Scalability
Streaming
Interoperability

Guidelines and Standardization
Multilingual Data

www.lider-project.eu

Dataset description with DataID

Contributing to the Linguistic Linked Open Data Cloud

This card describes how you can contribute to the Linguistic Linked Open Data Cloud.

Documenting metadata of language resources in Datahub

This card explains how to add some basic metadata to resources documented in Datahub.

Target audience and Scope

The card is intended for people who have a dataset in linked data and wish to document it in Datahub. This card extends the card "Contributing to the Linguistic Linked Open Data Cloud" by adding more details on how to document some basic metadata. See also, as an alternative, the "Dataset description with DataID" card.

Registering with Datahub

Go to http://datahub.io and follow the steps in the "Contributing to the Linguistic Linked Open Data Cloud" reference card. As result, Datahub automatically creates an RDF metadata file for the dataset, based on the Data Catalogue Vocabulary (DCAT). This metadata file is accessible at

http://datahub.io/dataset/DATASET_ID

where DATASET_ID is the identifier of your dataset.
Reference Architecture and Reference Cards

GUIDELINES
- Multilingual Dictionaries (BabelNet)
- Bilingual Dictionaries
- Multilingual terminologies (TBX)
- WordNets

Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

© Asunción Gómez-Pérez
The Luxembourg BabelNet Workshop 2-3rd, March 2016

How to publish Linguistic Linked Data
Publishing data on the Web in a way that it can be easily consumed is a daunting task. The Web community has converged on the idea that Linked Data provides the best practices for publishing highly interoperable data. This card explains in detail why it is important to use Linked Data and how to do it.

Linked Data Corpus Creation with NIF
This document describes best practices to follow for the generation of Linked Data text corpora, using the NLP Interchange Format (NIF).

Target audience
Corpus creators and users seeking to make corpora interoperable and to publish them as linked data. Basic knowledge of RDF is mandatory for conversion. Basic knowledge of linked data and web server access is needed for publication.

Scope
Conversion of existing corpora into RDF using NIF, as well as creation of linked data corpora from textual data.

Core concepts
We understand a corpus as a collection of documents. Documents contain text, represented as strings...
Reference Architecture and Reference Cards

Certification & Benchmarking

Discovery

LLD Linking

LLD Publishing

Metadata

Licensing

Multilingual Data

How to represent cross-lingual LD links

This card explains the basic mechanisms available to represent links between linked data in different natural languages.

Target audience

Data publishers and developers with basic acquaintance of RDF and LD who need to interlink data in different languages in the linked data cloud.

Scope

The card review the different types of cross-lingual LD links. It does not discuss the methods/techniques needed to infer the links, but deals with representation mechanisms only. This cards extends the card "How to publish Linguistic Linked Data".

Core concepts

Cross-lingual mappings can be established at two different levels: conceptual level and linguistic level. In the first case, links are

www.lider-project.eu
Reference Architecture and Reference Cards

Certification

Benchmarking & Validation

Discovery

LLD Linking

LLD Publishing

Metadata

Licensing

Multilingual

www.lider-project.eu
GUIDELINES

- NIF-based NLP Web Services
- LLOD aware services
- LLD exploitation

LLD Linking

Service Composition

LLD Publishing

LLD-aware Services

Vocabularies
Hosting
Scalability
Streaming
Interoperability

Metadata
Licensing
Provenance

Multilingual Data

www.lider-project.eu
Relation between W3C groups

Linked Data for Language Technologies (LD4LT)

Use Cases

Best practices (BPMLOD)

Ontology lexica (Ontolex)

BP for LD in LT

lemon specification

BP for using lemon

BP for Data on the Web

BP for Multilingual Data on the Web

Data on the Web Best Practices

www.lider-project.eu
Language Resources
The need of connecting licensed Language Resources
Linked Data foundations
Linguistic Linked Licensed Data
Outcomes of the LIDER project

THE APERTIUM RDF CASE AND THEIR LINKS
Motivation

Apertium Platform: Machine Translation.

> 40 bilingual dictionaries available in XML.

Afrikaans <-> Dutch
Breton --> French
Catalan <-> Italian
Welsh <-> English
Danish <-> Norwegian
English <-> Catalan
English <-> Spanish
English <-> Galician
Esperanto <-> Catalan
Esperanto <-> English
Esperanto <-> Spanish
Esperanto <-> French
Spanish <-> Aragonese
Spanish <-> Asturian
Spanish <-> Catalan
Spanish <-> Galician
Spanish <-> Italian
Spanish <-> Portuguese
Spanish <-> Romanian
Basque --> English
Basque --> Spanish
French --> Catalan
French --> Spanish
Serbo-Croatian --> English
Serbo-Croatian --> Macedonian
Serbo-Croatian --> Slovenian
Indonesian --> Malaysian
Icelandic --> Swedish
Icelandic --> English
Kazakh --> Tatar
Macedonian --> Bulgarian
Macedonian --> English
Norwegian Nynorsk <-> Norwegian Bokmål
Occitan <-> Catalan
Occitan <-> Spanish
Portuguese <-> Catalan
Portuguese <-> Galician
Northern Sami --> Norwegian Bokmål
Swedish --> Danish
......
GOAL: to expose translations contained in bilingual dictionaries as Linked Data on the Web

Joint effort by
Open data formats and publication

+40 Apertium Open bilingual Dictionaries proprietary format

22 Apertium bilingual Dictionaries in LMF

22 Apertium RDF

690K links with LexInfo
277K links with BabelNet
Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

The Luxembourg BabelNet Workshop 2-3rd, march 2016

# Apertium English lexicon:
http://linguistic.linkeddata.es/id/apertium/lexiconEN

# Apertium Spanish lexicon:
http://linguistic.linkeddata.es/id/apertium/lexiconES

# Apertium English-Spanish translation set:
http://linguistic.linkeddata.es/id/apertium/tranSetEN-ES
Modelling: for each bilingual dictionary....

EN-ES dictionary

EN-ES RDF dictionary

Lexicon ES

Lexicon EN

Translation Set EN-ES

Specification

Modelling

EN-ES

Trans. module
Modeling lexicons with Lemon

- Specification
- Modelling
- RDF Generation
- Links Generation
- Publication
- Exploitation

**LexicalEntry**

- **LexicalForm**
  - writtenRep: String

- **LexicalSense**
  - isSenseOf: sense
  - prefRef, altRef, hiddenRef
  - isReferenceOf: reference

- **LexicalCategory**
  - lexicalProperty
  - canonicalForm, otherForm, abstractForm

- **Ontology**
- **Lexicon**
  - language: String

- **Word**
- **Phrase**
- **Part**
The translation module

Translation Module
http://purl.org/net/translation.owl

LexicalSense

Translation

Translation Set

Resource

Translation Categories
http://purl.org/net/translation-categories

directEquivalent

culturalEquivalent

lexicalEquivalent
Modeling lexicons with Lemon

Specification

Modelling

lemon:LexicalEntry

lemon:LexicalEntry

lemon:LexicalForm

lemon:LexicalSense

tr:Translation

tr:TranslationSet

tr:translationSource

tr:translationTarget

tr:trans

translationSetEN-ES
Translation example

“bench”@en

lemon:Lexicon
lexiconEN

lemon:LexicalEntry

lemon:entry

lemon:Lexicon
lexiconES

lemon:LexicalEntry

lemon:entry

lemon:Entry

lemon:writtenRep

“banco”@es

lemon:LexicalSense

lemon:isSenseOf

tr:translationSource

tr:TranslationSet
translationSetEN-ES

tr:Translation

tr:TranslationSet

tr:trans

tr:translationTarget

lemon:LexicalSense

lemon:isSenseOf

tr:translationTarget

tr:translationSource

lemon:LexicalSense

lemon:isSenseOf

tr:trans

tr:TranslationSet

tr:Translation
Who will be the mapping generator?

RDF generation using Open Refine

For each bilingual dictionary:
1. Lexicon of the source language
2. Lexicon of the target language
3. Translations

```
apertium:lexiconEN a lemon:Lexicon ;
dc:source <http://hdl.handle.net/10230/17110> .
...
apertium:lexiconEN lemon:entry apertium:lexiconEN/bench-n-en .
apertium:lexiconEN/bench-n-en a lemon:LexicalEntry ;
lemon:lexicalForm apertium:lexiconEN/bench-n-en-form ;
lexinfo:partOfSpeech lexinfo:noun .
apertium:lexiconEN/bench-n-en-form a lemon:Form ;
lemon:writtenRep "bench"@en .
```
## Apertium RDF: 22 generated datasets

<table>
<thead>
<tr>
<th>Lang. pair</th>
<th># triples</th>
<th># trans.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-IT</td>
<td>180,851</td>
<td>7,869</td>
</tr>
<tr>
<td>EN-CA</td>
<td>759,601</td>
<td>33,029</td>
</tr>
<tr>
<td>EN-ES</td>
<td>576,316</td>
<td>25,830</td>
</tr>
<tr>
<td>EN-GL</td>
<td>425,117</td>
<td>20,034</td>
</tr>
<tr>
<td>EO-CA</td>
<td>426,301</td>
<td>19,964</td>
</tr>
<tr>
<td>EO-EN</td>
<td>617,772</td>
<td>31,474</td>
</tr>
<tr>
<td>EO-ES</td>
<td>380,198</td>
<td>17,212</td>
</tr>
<tr>
<td>EO-FR</td>
<td>726,281</td>
<td>35,791</td>
</tr>
<tr>
<td>ES-AN</td>
<td>71,997</td>
<td>3,110</td>
</tr>
<tr>
<td>ES-AST</td>
<td>825,54</td>
<td>36,096</td>
</tr>
<tr>
<td>ES-CA</td>
<td>730,501</td>
<td>31,291</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lang. pair</th>
<th># triples</th>
<th># trans.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES-GL</td>
<td>206,284</td>
<td>8,985</td>
</tr>
<tr>
<td>ES-PT</td>
<td>279,245</td>
<td>12,054</td>
</tr>
<tr>
<td>ES-RO</td>
<td>400,366</td>
<td>17,318</td>
</tr>
<tr>
<td>EU-ES</td>
<td>262,336</td>
<td>11,838</td>
</tr>
<tr>
<td>EU-EN</td>
<td>265,466</td>
<td>13,089</td>
</tr>
<tr>
<td>FR-CA</td>
<td>152,002</td>
<td>6,550</td>
</tr>
<tr>
<td>FR-ES</td>
<td>495,614</td>
<td>21,475</td>
</tr>
<tr>
<td>OC-CA</td>
<td>346,346</td>
<td>15,983</td>
</tr>
<tr>
<td>OC-ES</td>
<td>317,162</td>
<td>14,561</td>
</tr>
<tr>
<td>PT-CA</td>
<td>163,149</td>
<td>7,111</td>
</tr>
<tr>
<td>PT-GL</td>
<td>234,065</td>
<td>10,144</td>
</tr>
</tbody>
</table>
1. Links between Apertium Bilingual dictionaries
2. Links with other resources in the Linguistic LOD
   1. Babelnet
   2. LexInfo
   3. Zhishi.me
Linking with other Apertium RDF

Specification → Modelling → RDF Generation → Links Generation

Apertium LMF

Monolingual lexicons

Lexicon EN

Lexicon ES

Lexicon CA

Translation sets

Translation Set EN-ES

Translation Set EN-CA

EN-ES

EN-CA

Granted by URI Design

# Apertium English lexicon:
http://linguistic.linkeddata.es/id/apertium/lexiconEN
Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

© Asunción Gómez-Pérez

The Luxembourg BabelNet Workshop 2-3rd, March 2016

LexiconES

TranslationSetEN-ES

LexiconEN

TranslationSetES-PT

LexiconPT
Added value of linking Apertiums

- Apertium RDF Graph
  - Dijkstra algorithm to choose shortest path
  - Indirect translations for “bank” EN-> ES -> PT

<table>
<thead>
<tr>
<th>Pivot translation written repres.</th>
<th>Indirect translation written repres.</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;banco&quot;@es</td>
<td>&quot;banco&quot;@pt</td>
</tr>
<tr>
<td>&quot;orilla&quot;@es</td>
<td>&quot;orla&quot;@pt</td>
</tr>
</tbody>
</table>
# Direct translations for “bank”@en

<table>
<thead>
<tr>
<th>Translated written repr.</th>
<th>Part of Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;banc&quot;@ca</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;riba&quot;@ca</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;banco&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;orilla&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;ribera&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;beira&quot;@gl</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;banco&quot;@gl</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;ourela&quot;@gl</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;orela&quot;@gl</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;banku&quot;@eu</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;erribera&quot;@eu</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;ertz&quot;@eu</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#noun">http://www.lexinfo.net/ontology/2.0/lexinfo#noun</a></td>
</tr>
<tr>
<td>&quot;amuntegar&quot;@ca</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#verb">http://www.lexinfo.net/ontology/2.0/lexinfo#verb</a></td>
</tr>
<tr>
<td>&quot;agolpar&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#verb">http://www.lexinfo.net/ontology/2.0/lexinfo#verb</a></td>
</tr>
<tr>
<td>&quot;amontonar&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#verb">http://www.lexinfo.net/ontology/2.0/lexinfo#verb</a></td>
</tr>
<tr>
<td>&quot;apelotonar&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#verb">http://www.lexinfo.net/ontology/2.0/lexinfo#verb</a></td>
</tr>
<tr>
<td>&quot;hacinar&quot;@es</td>
<td><a href="http://www.lexinfo.net/ontology/2.0/lexinfo#verb">http://www.lexinfo.net/ontology/2.0/lexinfo#verb</a></td>
</tr>
<tr>
<td>....</td>
<td>...</td>
</tr>
</tbody>
</table>
Linking Apertium with BabelNet using senses

- Specification
- Modelling
- RDF Generation
- Links Generation

**LexiconES**
- `ribera`
- Sense-1
  - bank-ribera
  - Sense-2
- TranslationSetEN-ES
- `bank`
- LexiconEN
  - Sense-1
  - bank
  - Sense-2
  - `bank`@en

“ribera”@es

s00008363n

“bank”@en

“ribera”@es

“Sloping land (especially the slope beside a body of water)”@en
<table>
<thead>
<tr>
<th>Translated Written Repr.</th>
<th>BabelSynset</th>
<th>BabelNet gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;banco&quot; @es</td>
<td><a href="http://babelnet.org/rdf/s00008371n">http://babelnet.org/rdf/s00008371n</a></td>
<td>“A building in which the business of banking transacted”</td>
</tr>
<tr>
<td>&quot;banco&quot; @es</td>
<td><a href="http://babelnet.org/rdf/s00008366n">http://babelnet.org/rdf/s00008366n</a></td>
<td>“An arrangement of similar objects in a row or in tiers”</td>
</tr>
<tr>
<td>&quot;banco&quot; @es</td>
<td><a href="http://babelnet.org/rdf/s15346085n">http://babelnet.org/rdf/s15346085n</a></td>
<td>“An ocean bank, sometimes referred to as a fishing bank or simply bank, …”</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>&quot;orilla&quot; @es</td>
<td><a href="http://babelnet.org/rdf/s00008363n">http://babelnet.org/rdf/s00008363n</a></td>
<td>“Sloping land (especially the slope beside a body of water)”</td>
</tr>
<tr>
<td>&quot;ribera&quot; @es</td>
<td><a href="http://babelnet.org/rdf/s00008363n">http://babelnet.org/rdf/s00008363n</a></td>
<td>“Sloping land (especially the slope beside a body of water)”</td>
</tr>
</tbody>
</table>
Zhishi.me Lexicalization and linking

- Zhishi.me (http://zhishi.me)
- First effort to publish large scale Chinese semantic data as Linked Open Data

It has over 8 million distinct instances and 200 million RDF triples.

Acknowledgements: Zhijia Fang and Haofen Wang
East China University of Science and Technology (ECUST)
Lexicalisation of Zhishi.me

- ECUST (China) and UPM (Spain) collaboration
- New dataset created Zhishi.lemon with:
  - The lexical layer of Zhishi.me in pinyin, traditional, and simplified Chinese
  - Translations into English and Spanish
  - Links to DBpedia, esDBpedia, and BabelNet
- For the links to BabelNet, a disambiguation algorithm based on the categories was used
- Initially, 16,000 links to BabelNet were created. More to come in the future
Advantages of Apertium RDF

- Apertium data on the Web following SW standards
- Common entry point for all the Apertium dictionaries
- Unique graph that contains direct and indirect translations
- Linked with BabelNet
Linguistic Linked License Data in Use

Language Resources
The need of connecting licensed Language Resources
Linked Data foundations
Linguistic Linked Licensed Data
Outcomes of the LIDER project
The apertium RDF case and their links

LINGUISTIC LINKED LICENSED DATA IN USE
Multilingual search: retrieve an author

- Servantes Saavedra, Migel
- Therbantes, Minkel nte
- Cervantes di Saavedra, Michele
- سرفنتس سافدرا، ميجيل دى
- Zerbantes eta Saabedra, Mikel
- Θερβάντες, Μιγκέλ ντε
- Cervantes
- Sirfantis Saafedrā, Mīgīl dī
- Сервантес Сааведра, Мигель де
- Sewantisi Saweidela, Migai'er de
- 塞万提斯·萨维德拉, 米盖尔德
- Cervantes, Miguel de
Cervantes Sevedra, Miguel de (1547-1616)

Miguel de Cervantes Sevedra fue un soldado, novelista, poeta y dramaturgo español. Suponemos que nació el 29 de septiembre de 1547 en Alcalá de Henares y murió el 22 de abril de 1616 en Madrid.
Search a book in Spanish about the “Athénée de Madrid”
Linguistic Linked Data allows Multilingual data integration.
1. Programmers built applications making queries in SPARQL and get RDF

2. Citizens/Users access LD through a user interface (they do not see RDF)


Culture  Geographical  Smart Cities
Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources

© Asunción Gómez-Pérez

The Luxembourg BabelNet Workshop 2-3rd, march 2016
Thanks for your attention!