An Overview of the LiLa Knowledge Base

Motivation, Architecture, Services, and Resources

Marco Passarotti

LiLa Closing Workshop
Università Cattolica del Sacro Cuore, Milan
25th-26th May 2023

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.

Table of Contents

Table of Contents

Research question
State of affairs

We have built and collected (for Latin and other languages):

Research question
State of affairs

We have built and collected (for Latin and other languages):

- Textual Resources

Research question
State of affairs

We have built and collected (for Latin and other languages):

- Textual Resources
- Lexical Resources

Research question
State of affairs

We have built and collected (for Latin and other languages):

- Textual Resources
- Lexical Resources
- NLP Tools

Research question
State of affairs

We have built and collected (for Latin and other languages):

- Textual Resources
- Lexical Resources
- NLP Tools

Scattered and unconnected

ERC Consolidator Grant
2018-2023

A collection of multifarious, interoperable linguistic resources represented with the same vocabulary for knowledge description (by using common data categories and ontologies)

Interlinking as a Form of Interaction

The Linked Data Principles

…just to be FAIR

The Linked Data Principles
...just to be FAIR

- Use URIs for things (e.g. an entry in a lexicon, a token in a corpus)

The Linked Data Principles
…just to be FAIR

- Use URIs for things (e.g. an entry in a lexicon, a token in a corpus)
- Use HTTP URIs to allow people (and machines) to look up things

The Linked Data Principles

…just to be FAIR

- Use URIs for things (e.g. an entry in a lexicon, a token in a corpus)
- Use HTTP URIs to allow people (and machines) to look up things
- Use web standards to represent/query (meta)data, such as RDF and SPARQL

The Linked Data Principles

…just to be FAIR

- Use URIs for things (e.g. an entry in a lexicon, a token in a corpus)
- Use HTTP URIs to allow people (and machines) to look up things
- Use web standards to represent/query (meta)data, such as RDF and SPARQL
- Include links to other URIs

Benefits of Applying LD to Linguistic Resources

Chiarcos et al. (2013)

Benefits of Applying LD to Linguistic Resources

Chiarcos et al. (2013)

- Representation and Modelling: RDF is a very versatile data model to represent stand-off annotations, dependency parses etc.

Benefits of Applying LD to Linguistic Resources

Chiarcos et al. (2013)

- Representation and Modelling: RDF is a very versatile data model to represent stand-off annotations, dependency parses etc.
- Structural/Syntactic Interoperability: data formats and communication protocols (HTTP, URIs, RDF) to ensure data exchange. The systems involved can process the exchanged information, but there is no guarantee that the interpretation is the same

Benefits of Applying LD to Linguistic Resources

Chiarcos et al. (2013)

- Representation and Modelling: RDF is a very versatile data model to represent stand-off annotations, dependency parses etc.
- Structural/Syntactic Interoperability: data formats and communication protocols (HTTP, URIs, RDF) to ensure data exchange. The systems involved can process the exchanged information, but there is no guarantee that the interpretation is the same
- Conceptual Interoperability: when two systems have the ability to automatically interpret exchanged information meaningfully, through common ontologies and re-usable vocabularies: resources are unambiguously linked. What is sent is the same as what is understood

Benefits of Applying LD to Linguistic Resources

Chiarcos et al. (2013)

- Representation and Modelling: RDF is a very versatile data model to represent stand-off annotations, dependency parses etc.
- Structural/Syntactic Interoperability: data formats and communication protocols (HTTP, URIs, RDF) to ensure data exchange. The systems involved can process the exchanged information, but there is no guarantee that the interpretation is the same
- Conceptual Interoperability: when two systems have the ability to automatically interpret exchanged information meaningfully, through common ontologies and re-usable vocabularies: resources are unambiguously linked. What is sent is the same as what is understood
- Federation: to combine information from physically separated repositories

Benefits of Applying LD to Linguistic Resources

Chiarcos et al. (2013)

- Representation and Modelling: RDF is a very versatile data model to represent stand-off annotations, dependency parses etc.
- Structural/Syntactic Interoperability: data formats and communication protocols (HTTP, URIs, RDF) to ensure data exchange. The systems involved can process the exchanged information, but there is no guarantee that the interpretation is the same
- Conceptual Interoperability: when two systems have the ability to automatically interpret exchanged information meaningfully, through common ontologies and re-usable vocabularies: resources are unambiguously linked. What is sent is the same as what is understood
- Federation: to combine information from physically separated repositories
- Dynamicity: to provide access to the most recent version of a resource

Benefits of Applying LD to Linguistic Resources

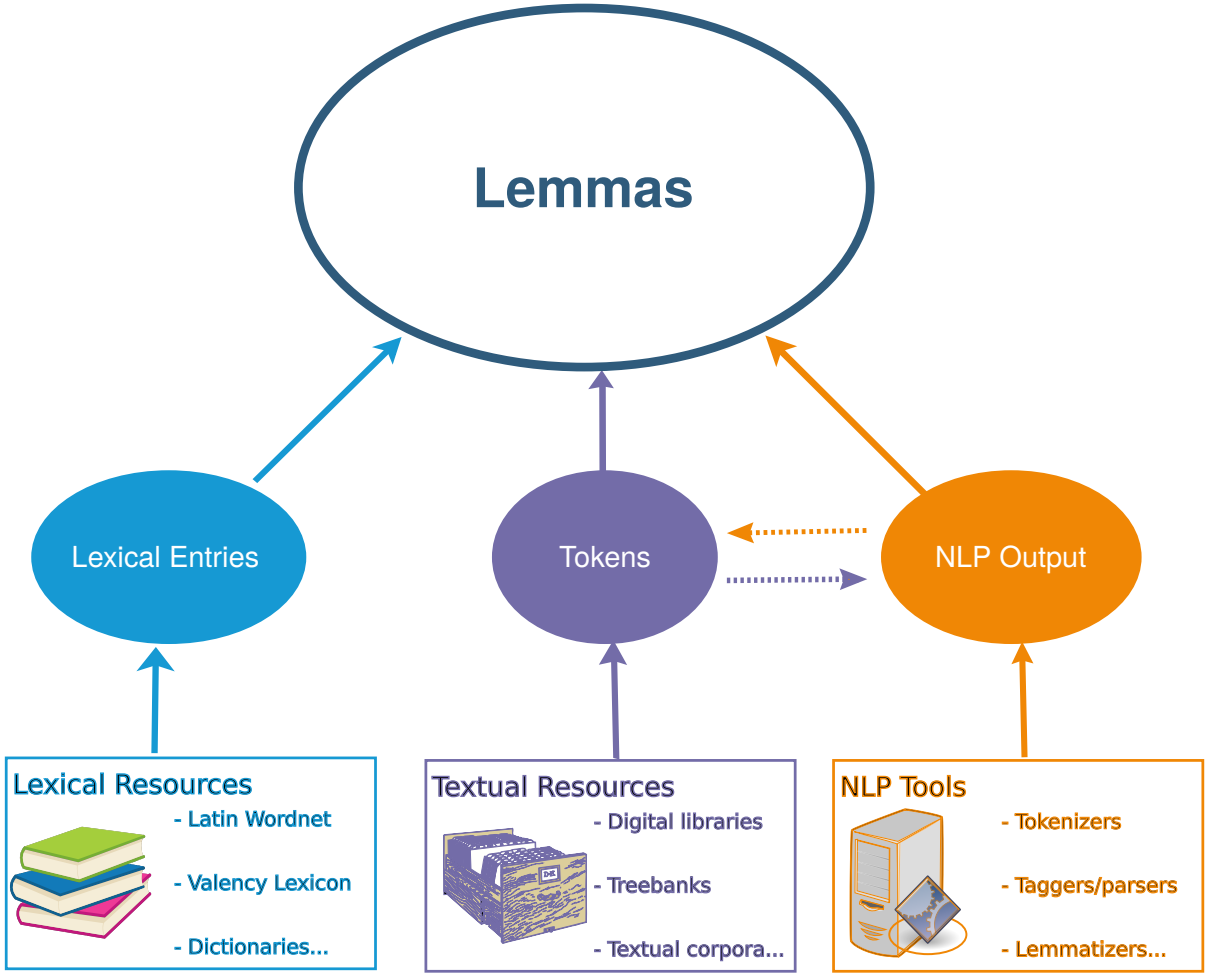
Chiarcos et al. (2013)

- Representation and Modelling: RDF is a very versatile data model to represent stand-off annotations, dependency parses etc.
- Structural/Syntactic Interoperability: data formats and communication protocols (HTTP, URIs, RDF) to ensure data exchange. The systems involved can process the exchanged information, but there is no guarantee that the interpretation is the same
- Conceptual Interoperability: when two systems have the ability to automatically interpret exchanged information meaningfully, through common ontologies and re-usable vocabularies: resources are unambiguously linked. What is sent is the same as what is understood
- Federation: to combine information from physically separated repositories
- Dynamicity: to provide access to the most recent version of a resource
- Ecosystem: maintained by a large and active community with common tools and practices

Table of Contents

LiLa Knowledge Base

Lexically-based architecture and (meta)data sources



LiLa and Ontolex Lemon

The **Lemma Bank**: the core of LiLa

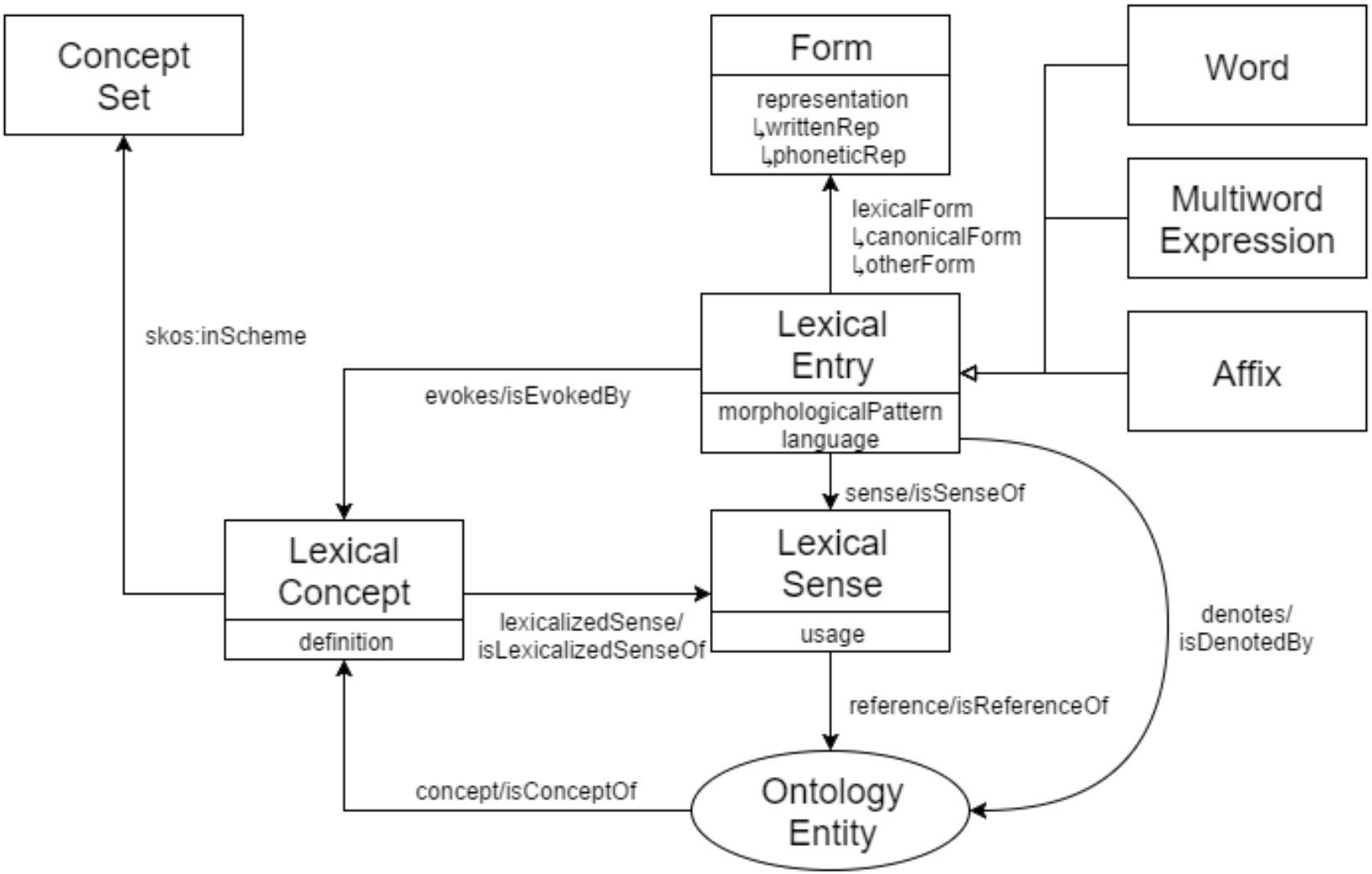


Table of Contents

Lemma Bank Query Interface

https://lila-erc.eu/query/

LiLa Search Platform

http://lila-erc.eu:8080/lila-lisp/

SPARQL Access Point

`https://lila-erc.eu/sparql/`

TextLinker

http://lila-erc.eu:8080/LiLaTextLinker/

Table of Contents

Linguistic Resources Interlinked in LiLa

...and some numbers

Resources connected so far

https://lila-erc.eu/data-page/

- **Canonical Forms** in the Lemma Bank: approx. 215K
- **Lexical Entries** in Lexical Resources: approx. 145K
- **Tokens** in Textual Resources (158 works): approx. 3,5M
- **Triples**: approx. 70M

Enjoy the LiLa Workshop!

Thank you

LiLa: Linking Latin
Università Cattolica del Sacro Cuore
CIRCSE Research Centre

info@lila-erc.eu

https://github.com/CIRCSE

https://lila-erc.eu

@ERC_LiLa

Largo Gemelli 1, 20123 Milan, Italy

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.