



UNIVERSITÀ
CATTOLICA
del Sacro Cuore



Prefixes Matter

CLARIN and LLD in the light of the LiLa Knowledge Base

Marco Passarotti

Linguistic Linked Data CLARIN Café
29th April, 2021



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.

...starting from the end

Challenges and Perspectives. CLARIN for LLD & LLD for CLARIN



...starting from the end

Challenges and Perspectives. CLARIN for LLD & LLD for CLARIN



- ▶ CLARIN collects **LD** (Linguistic Data) in a centralized repository, but needs a way to efficiently **L** (link) them.
A real **semantic interoperability** between resources is still to come.
Ongoing work/initiatives on integration/interoperability in CLARIN:

...starting from the end

Challenges and Perspectives. CLARIN for LLD & LLD for CLARIN



- ▶ CLARIN collects **LD** (Linguistic Data) in a centralized repository, but needs a way to efficiently **L** (link) them.
A real **semantic interoperability** between resources is still to come.
Ongoing work/initiatives on integration/interoperability in CLARIN:
 - ▶ **Metadata**: integrating, mapping and converting metadata from heterogeneous sources, by linking to external controlled vocabularies and shared concept registries

...starting from the end

Challenges and Perspectives. CLARIN for LLD & LLD for CLARIN



- ▶ CLARIN collects **LD** (Linguistic Data) in a centralized repository, but needs a way to efficiently **L** (link) them.

A real **semantic interoperability** between resources is still to come.

Ongoing work/initiatives on integration/interoperability in CLARIN:

- ▶ **Metadata:** integrating, mapping and converting metadata from heterogeneous sources, by linking to external controlled vocabularies and shared concept registries
- ▶ **Formats:** when launched from a resource in the VLO, the CLARIN Switchboard reads the required format(s) from the CMDI metadata (in a ready-made description format: *profile*) and finds compatible tools

...starting from the end

Challenges and Perspectives. CLARIN for LLD & LLD for CLARIN



- ▶ CLARIN collects **LD** (Linguistic Data) in a centralized repository, but needs a way to efficiently **L** (link) them.

A real **semantic interoperability** between resources is still to come.

Ongoing work/initiatives on integration/interoperability in CLARIN:

- ▶ **Metadata:** integrating, mapping and converting metadata from heterogeneous sources, by linking to external controlled vocabularies and shared concept registries
- ▶ **Formats:** when launched from a resource in the VLO, the CLARIN Switchboard reads the required format(s) from the CMDI metadata (in a ready-made description format: *profile*) and finds compatible tools
- ▶ **Type-based Interoperability:** the *Resource Families* initiative provides an overview per data type of the CLARIN linguistic resources "to facilitate comparative research"

...starting from the end

Challenges and Perspectives. CLARIN for LLD & LLD for CLARIN

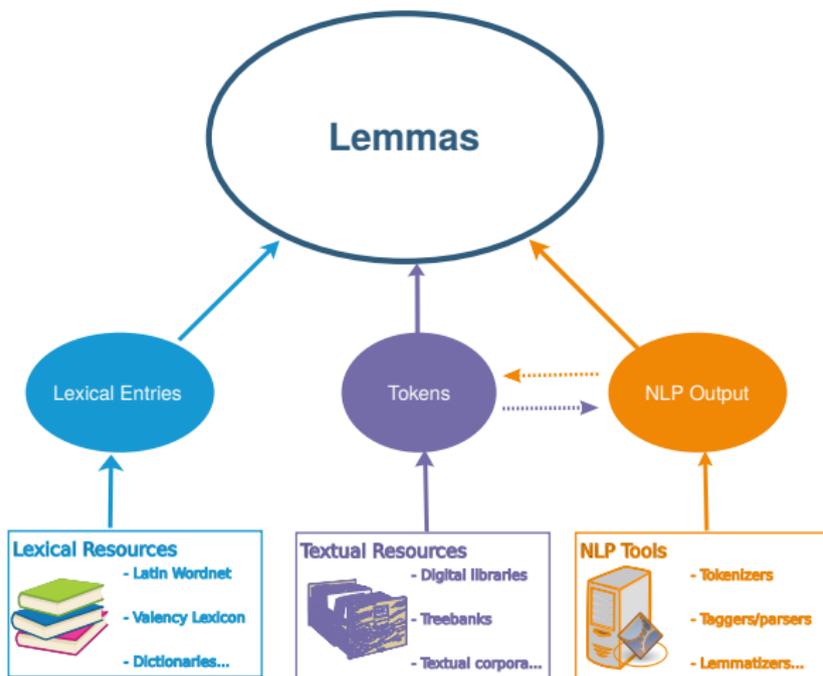


- ▶ CLARIN collects **LD** (Linguistic Data) in a centralized repository, but needs a way to efficiently **L** (link) them.
A real **semantic interoperability** between resources is still to come.
Ongoing work/initiatives on integration/interoperability in CLARIN:
 - ▶ **Metadata:** integrating, mapping and converting metadata from heterogeneous sources, by linking to external controlled vocabularies and shared concept registries
 - ▶ **Formats:** when launched from a resource in the VLO, the CLARIN Switchboard reads the required format(s) from the CMDI metadata (in a ready-made description format: *profile*) and finds compatible tools
 - ▶ **Type-based Interoperability:** the *Resource Families* initiative provides an overview per data type of the CLARIN linguistic resources "to facilitate comparative research"
- ▶ LLD needs **LD**, but provides ways to semantically **L** them: common vocabularies, ontologies, data categories, etc. to represent and interlink different kinds of linguistic resources

ERC Consolidator Grant 2018-2023

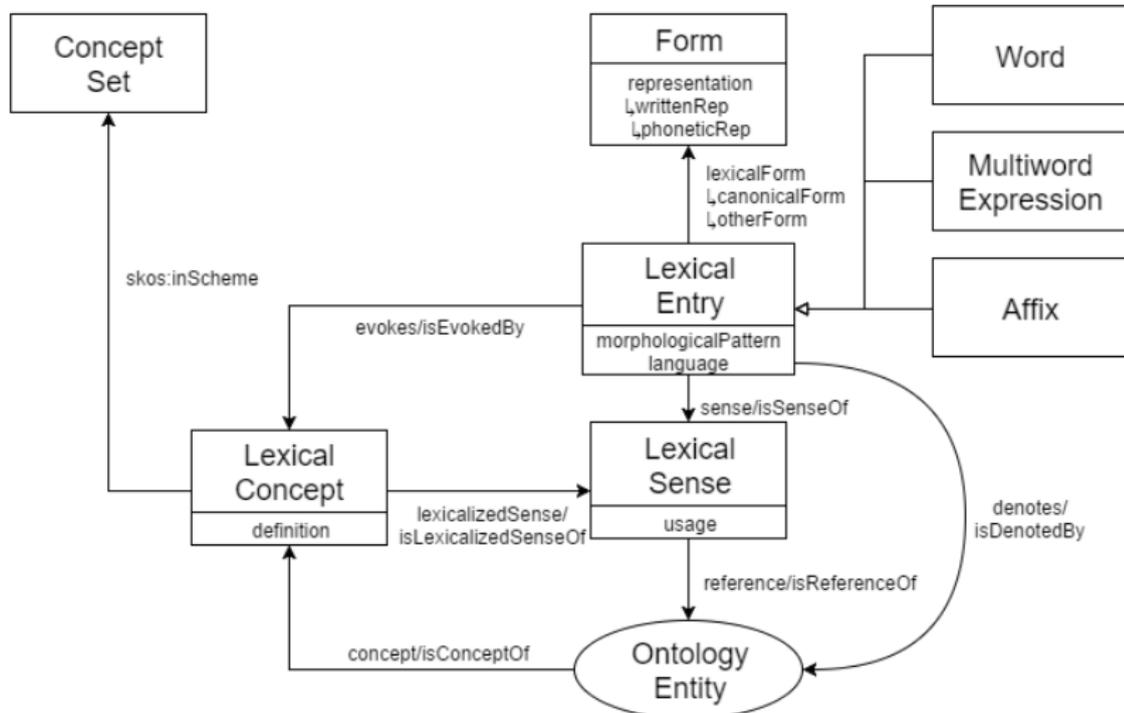
A collection of multifarious, interoperable linguistic resources described with the same vocabulary for knowledge description (by using common data categories and ontologies)





LiLa and Ontolex Lemon

A *de facto* W3C standard for publishing lexical data as LLD



Lemma *admiror* 'to admire, to respect'

<https://lila-erc.eu/data/id/lemma/87541>

Thanks!

Get in touch



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.