

When OntoLex Meets Wikibase: Remodeling Use Cases



David Lindemann

Sina Ahmadi

Anas Fahad Khan

Francesco Mambrini

Federica Iurescia

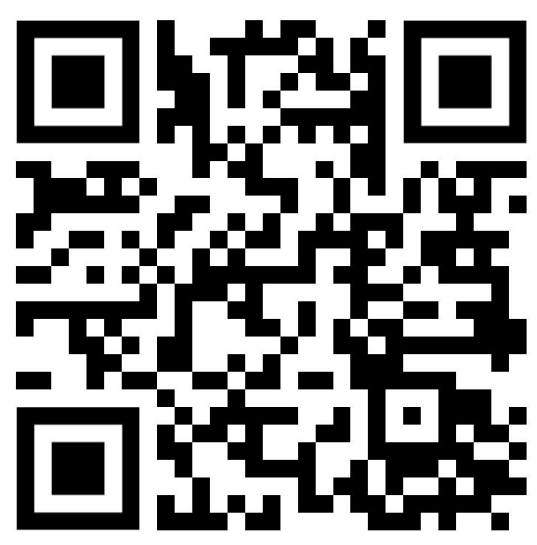
Marco Passarotti

(david.lindemann@ehu.eu)

Highlights

- Task: Remodel OntoLex to Wikidata and convert datasets to create Wikibases
- Two case studies:
 - Kurdî Wikibase (Kurdish)
 - LiLa Wikibase (Latin)
- Contributions:
 - Remodeling OntoLex
 - Conversion to Wikidata
 - Creating Wikibases
 - Enhance multilingualism

Our SPARQL Endpoints:

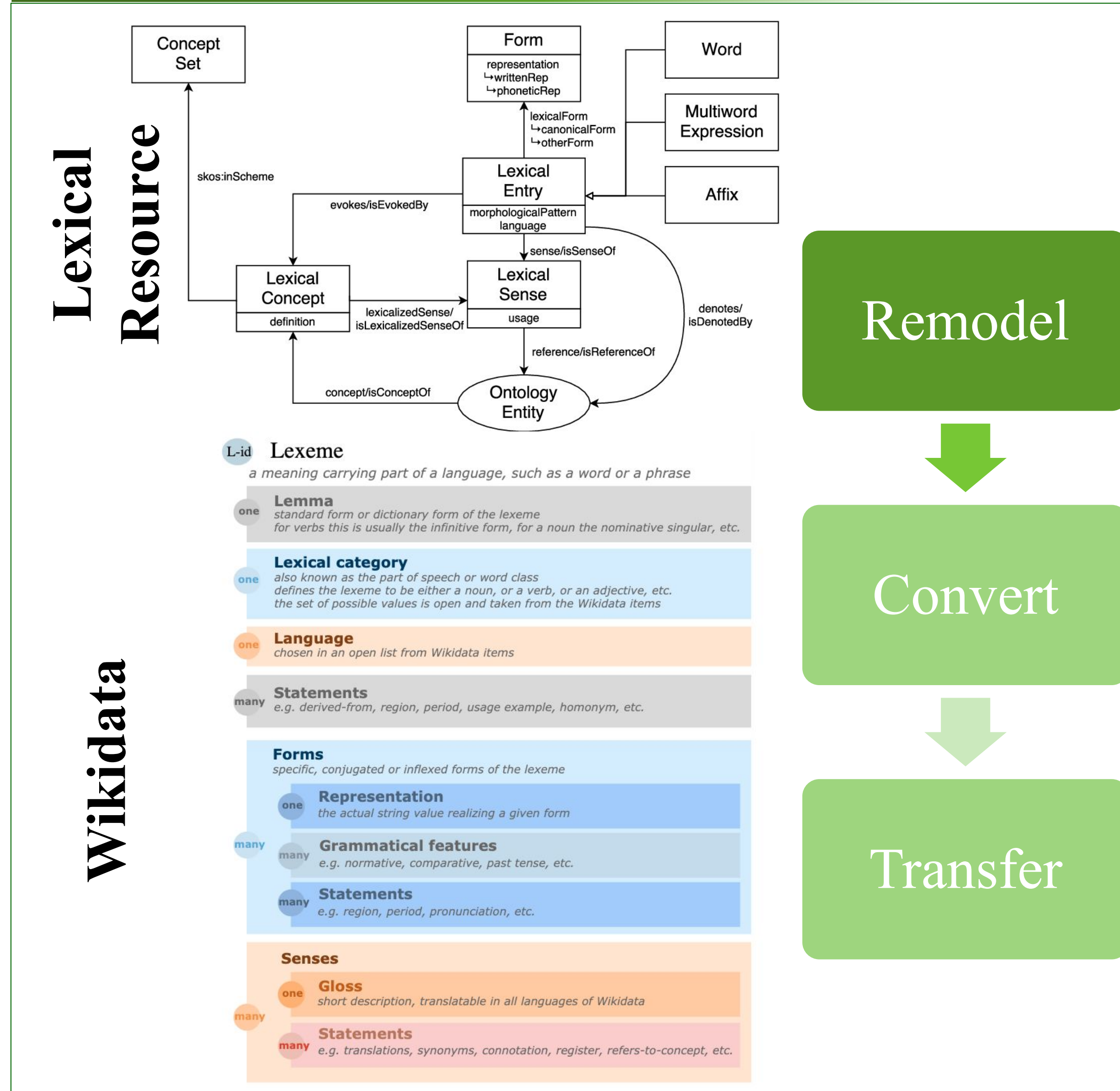


Kurdi Wikibase



LiLa Project

Approach



Remodel

Convert our resources to a Wikibase compatible data model by mapping of source URI to Wikibase URI

Ontolex-Lemon	Wikibase
rdfs:label	
ontolex:writtenRep	wikibase:lemma
ontolex:sense	ontolex:sense
(skos:definition)	skos:definition
ontolex:lexicalForm	ontolex:lexicalForm
ontolex:writtenRep (lexinfo properties)	ontolex:representation
	wikibase:grammaticalFeatures

LiLa		Wikidata	
URI	Label	URI	Label
lila:n2	2nd declension (m/f) nouns	wd:Q3953983	2nd decl.
lila:n2e	2nd decl irregular nouns	wd:Q3953983	2nd decl.
lila:n6	1st class adj.	wd:Q3606519	Adj. of 1st class
lila:v3r	3rd conj. verb	wd:Q54295441	Latin 3rd conj.
lila:v3e	3rd conj. impersonal verb	wd:Q54295441	Latin 3rd conj.
lila:n	uninflected noun/adj.	—	—

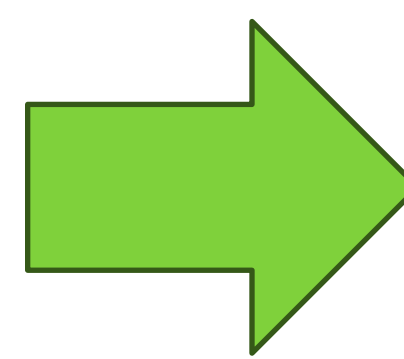
Why Wikibase?

Customized knowledge base, community support, collaborative editing & interoperability and integration

Use Case 1: Kurdî Wikibase

```

1 :lex_kmr_7802180323 a ontolex:LexicalEntry, ontolex:Word ;
2 ontolex:canonicalForm :form_kmr_7802180323 ;
3 dct:language <www.lexvo.org/page/iso639-3/kmr> ;
4 rdfs:label "parti"@kmr-latn ;
5 lexinfo:partOfSpeech lexinfo:noun ;
6 ontolex:sense :kmr_8301494711_sense ;
7 lexinfo:gender lexinfo:feminine .
8
9 :form_kmr_7802180323 a ontolex:Form ;
10 ontolex:writtenRep "parti"@kmr-latn ;
11 lexinfo:number lexinfo:singular .
12
13 :kmr_8301494711_sense a ontolex:LexicalSense ;
14 skos:definition "political party"@en.
15
16 :kmr_8301494711_sense ontolex:usage [
17 rdf:value "partiyên kurd yên siyasi"@kmr-latn ;
18 rdf:value "Kurdish political parties"@en ] .
    
```



```

1 wd:L1083983 a ontolex:LexicalEntry ;
2 wikibase:lemma "parti"@kmr-latn ;
3 rdfs:label "parti"@kmr-latn ;
4 wikibase:lexicalCategory Q1084 ; # noun
5 dct:language wd:Q36163 ; # Kurmanji
6 ontolex:sense wd:L1083983-S1 ;
7 ontolex:lexicalForm wd:L1083983-F1 ;
8 wdt:P5185 wd:Q1775415 ; # gender fem.
9 p:P5831 [ps:P5831 "partiyên kurd yên siyasi"@kmr-latn ;
10 pq:P2441 "Kurdish political parties"@en ;
11 pq:P6072 wd:L1083983-S1] . # usage example
12
13 wd:L1083983-S1 a ontolex:LexicalSense ;
14 skos:definition "political party"@en ;
15 wdt:P5137 wd:Q7278 . # item-for-this-sense political party
16
17 wd:L1083983-F1 a ontolex:Form ;
18 ontolex:representation "parti"@kmr-latn ;
19 wikibase:grammaticalFeature wd:Q110786 . # singular
    
```

> 20,000 lexemes in four under-represented dialects of Kurdish with individual codes and in two scripts

Use Case 2: Latin and Wikidata

- Based on the LiLa project (<https://lila-erc.eu>)
- Richly described lexemes with senses and inflections
- 56,202 lexemes transferred to Wikidata
- Link to LiLa knowledge base using `wdt:P11033`
- Linked at the lemma-level to LiLa corpora
- Aligned with other lexical resources such as WordNet

Takeaways

- Wikibase is ideal as a customizable knowledge base
- Beneficial to linguistic communities and multilingualism
- Future developments for direct & indirect mappings, e.g. lexinfo to Wikidata
- Community support before transferring data to Wikidata
- Provide guidelines for future projects



University of the Basque Country



Istituto di Linguistica Computazionale "Antonio Zampolli"

Consiglio Nazionale delle Ricerche



UNIVERSITÀ CATTOLICA del Sacro Cuore