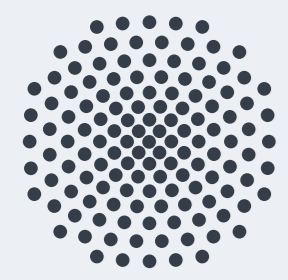


LVF-lemon – Towards a Linked Data Representation of “Les Verbes français”



About

How to convert traditional **lexical syntactico-semantic resource** – “Les Verbes français” (LVF, [?, ?]) into standardised and normalised **linked open data model**?

Objective

- ▶ explicit and make accessible **linguistic knowledge**
- ▶ **interoperability** with other linguistic resources (e.g. corpus data and subcat lexicons).

To do

RDF conversion \approx straightforward

Data modeling: model content in terms of established vocabularies:

- general: RDFS, OWL, SKOS, ...
- linguistic: lemon, LexInfo, OLIA, LMF, ...
- LVF specific

Linking the data

Focus: **Syntax**, **semantics** and **syntax-semantics** interface.

The Lexicon – LVF, Les Verbes français [?, ?]

▶ \approx 12 300 verbs, \approx 25 610 readings, \approx 15 readings/verb

▶ elaborate morpho-syntactic and semantic description.

Sample uses, semantic descriptions

Syntactic descriptions

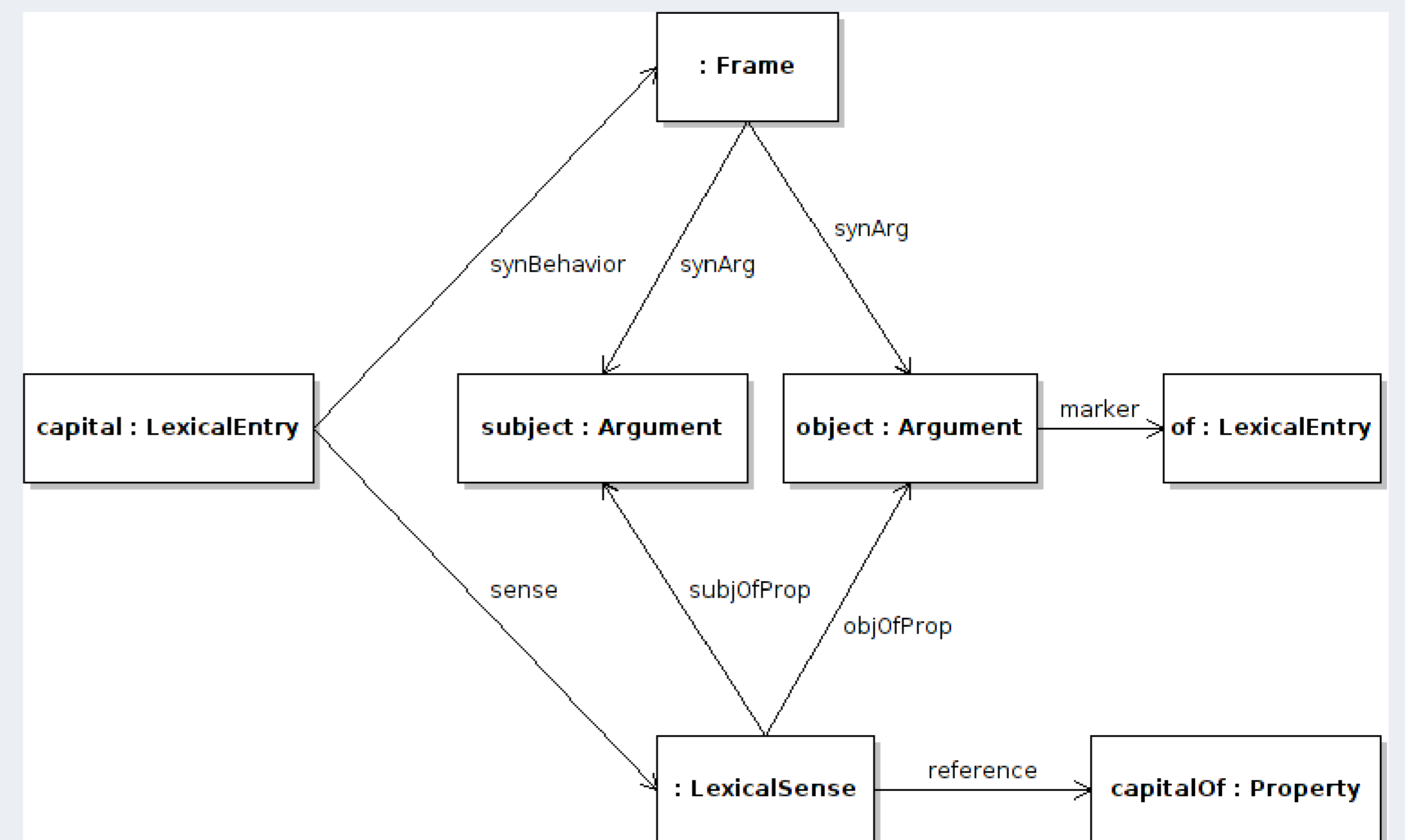
id	example	semantic decomposition	sem. primitive	sem. class	id	schema encoded information
01	On élargit une route/ La route s'élargit .	r/d+qt	large	(make) become	01 A30	intransitive with adjunct, inanimate subject
02	Cette veste élargit Paul aux épaules.	d	large	become	T1308	transitive, human subject, inanimate direct object, instrumental adjunct
03	On élargit ses connaissances.	r/d	large abs	(make) become, figurative	P3008	reflexive, inanimate subject, instrumental adjunct
04	On élargit le débat à la politique étrangère.	f.ire	abs VRS	directed move, figurative	02 N1i A90 T3900	intransitive, animate subject, prep. phrase headed by <i>de</i> (of) intransitive with adjunct, subject human or thing transitive, inanimate subject, object human or thing

Table: Sample readings for *élargir/enlarge*

Lemon: the linguistic LLOD model [?]

- ▶ links lexical entries to knowledge repositories (ontology)
- ▶ lexicon and knowledge (ontology) layers separated
- ▶ linguistically sound structure based on Linguistic Markup Framework – LMF [?]
- ▶ linking to data categories: allows for arbitrarily complex linguistic description, integration with annotated corpora
- ▶ RDF-native form: enables leverage of existing Semantic Web technologies
- ▶ **syntax-semantics interface** based on LexInfo ontology
 - ▷ lexical entries \leftrightarrow ontology sense
 - ▷ link between subcat frame \leftrightarrow sense not represented

Problem in LVF: syntax & semantics intricately interleaved!



lemonUBY– [?]

- ▶ representation of UBY, network of multilingual, interlinked lexical semantic resources
- ▶ based on LMF, ISOcat data categories
- ▶ conversion by mapping LMF representations
- ▶ **synsem mapping** implicit:
 - ▷ via subcats, syntactic arguments \leftrightarrow thematic roles

PDEV-lemon – [?]

- ▶ representation of PDEV, the Pattern Dictionary of English verbs
 - ▷ verbs associated with attested patterns of use.
- ▶ **synsem mapping**: patterns \leftrightarrow PDEV extension **frame sense**
 - ▷ to model phraseological or idiomatic expressions
- ▶ ontology of semantic types: *human, animate, etc.*

LVF-lemon: road-map

▶ **morphology** can be appropriately represented by **lemon** core model (morphology module)

syntax: schemes can be mapped to *lemon* or LMF syntactic arguments

semantics: need LVF specific representation

▶ syntactic/semantic types assigned to syntactic arguments:

- ▷ LVF specific, related to PDEV semantic type ontology
- ▷ 8 LVF specific semantic roles, related to VerbNet/FrameNet role inventories

semantic classes:

based on linguistic theories

semantic primitives in *opérateur*: [?, ?] among others

\implies relate to VerbNet/FrameNet representations

References