PreMOn: a Lemon Extension for Exposing Predicate Models as Linked Data

Francesco Corcoglioniti, Marco Rospocher, Alessio Palmero Aprosio, Sara Tonelli
Fondazione Bruno Kessler – IRST
Via Sommarive 18, Trento, I-38123, Italy
{corcoglio,rospocher,aprosio,satonelli}@fbk.eu

Abstract

We introduce *PreMOn* (*pre*dicate *m*odel for *on*tologies), a linguistic resource for exposing predicate models (PropBank, NomBank, VerbNet, and FrameNet) and mappings between them (e.g, SemLink) as Linked Open Data. It consists of two components: (i) the *PreMOn Ontology*, an extension of the *lemon* model by the W3C Ontology-Lexica Community Group, that enables to homogeneously represent data from the various predicate models; and, (ii) the *PreMOn Dataset*, a collection of RDF datasets integrating various versions of the aforementioned predicate models and mapping resources. *PreMOn* is freely available and accessible online in different ways, including through a dedicated SPARQL endpoint.

Keywords: Predicate Models; Predicate Model Mappings; Lemon; Semantic Web; Linguistic Linked Open Data.

1. Introduction

Predicate models such as PropBank (herafter, PB) (Palmer et al., 2005), NomBank (NB) (Meyers et al., 2004), VerbNet (VN) (Schuler, 2005), and FrameNet (FN) (Baker et al., 1998) provide rich descriptions of predicate *semantic classes* — i.e., rolesets in NB and PB, verb classes in VN, and frames in FN (e.g., "Commerce Sell") — and their *semantic roles* (e.g., "Seller" and "Buyer"), abstracting from a number of linguistic phenomena related to their realization in text. Thanks to the mappings of different predicate models, such as SemLink (Palmer, 2009) and the Predicate Matrix (Lacalle et al., 2014), and to their integration in Semantic Role Labeling (SRL) tools, they have become central to a number of tasks such as information extraction, question answering and natural language generation. In particular, due to their laying at the syntactic-semantics interface, predicate models are increasingly used within the Semantic Web (SW) community, for knowledge extraction in tools such as NewsReader (Rospocher et al., 2016) and PIKES (Corcoglioniti et al., 2016), or as the starting point for deriving general-domain ontologies grounded in natural language, such as FrameBase (Rouces et al., 2015) and the ESO ontology (Segers et al., 2015) derived from FN.

Compared to the current situation where each predicate model has its own terminology, structure and proprietary XML format, the availability of a single RDF/OWL ontological model covering the main predicate models with their common aspects, specificities, and mappings would be beneficial to all the applications mentioned above, within and outside the SW area. Indeed, the use of SW technologies and the adoption of the Linked Open Data (LOD) paradigm have already been recognized as particularly beneficial to linguistic resources (see, e.g., Chiarcos et al. (2013)), leading to the creation of the Linguistic Linked Open Data cloud curated by the Open Linguistic subgroup of the Open Knowledge Foundation.[1] In that context, the *lemon* (*le*xical *m*odel for *on*tologies) model by McCrae et al. (2012), recently revised by the W3C Ontology Lexicon Community Group (2015), already sets the basis for modeling the lexical entries referenced in predicate models (e.g., the verb 'sell'), together with their links (lexical senses) to corresponding ontological concepts.

Building on *lemon*, in this paper we present *PreMOn* (*pre*dicate *m*odel for *on*tologies),[2] a linguistic resource for exposing predicate models and mappings between them. It consists of two components: the *PreMOn Ontology*, an OWL 2 ontology that extends *lemon* for modeling semantic classes and semantic roles, the relations among them, their annotations in text, and the mappings between semantic classes and roles in different resources; and (ii), the *PreMOn Dataset*, a freely-available, interlinked RDF dataset for PB, NB, VN, FN, and the associated mapping resources, published online as Linked Open Data according to the conceptual schema defined by the *PreMOn Ontology*. *PreMOn* brings several benefits to users of predicate models:

1. ease of access and reuse of predicate model data, due to the adoption of a common RDF format, stable URIs, and LOD best practices;
2. possibility to abstract and capture the aspects common to different predicate models, while at the same time keeping track of the peculiarities of each model (using RDFS/OWL subclass/subproperty primitives);
3. possibility to apply SW technologies to predicate model data, such as automated reasoning and SPARQL querying, e.g., for retrieving the semantic classes of a lexical entry and the associated mappings;
4. possibility to combine *PreMOn* with other linguistic ontologies, e.g., for providing the SRL annotations of a text according to the NLP Interchange Format (NIF) (Hellmann et al., 2013);
5. possibility for third parties to publish and interlink their datasets with *PreMOn*, extending it in a decentralized way (e.g., with new mappings).

The paper is structured as follows. Section 2 provides an overview of the predicate models considered in *PreMOn* and of related prior works in the SW. Sections 3 and 4 illustrate the *PreMOn Ontology* with its core concepts and specialized modules for each predicate model, while Section 5 describes the *PreMOn Dataset*. Section 6 discusses possible applications, while Section 7 concludes.

[1] http://linguistics.okfn.org/

[2] Available at http://premon.fbk.eu/

Table 1: Comparison of predicate models.

Aspect	NomBank	PropBank	VerbNet	FrameNet
Considered parts-of-speech	noun	verb	verb	any (9 total)
Term for semantic classes	roleset	roleset	verb (sub-)class	frame
Lexical entries per class	exactly one	exactly one	zero or more	zero or more
Scope of semantic roles	local to semantic class	local to semantic class	global	local to semantic class (core roles)
Types of semantic roles	numbered, modifier	numbered, modifier, secondary agent	thematic role (hierarchy)	core + other 3 types
Semantic class relations	–	–	subclass	inheritance + 8 relations
Semantic role relations	–	–	–	inheritance + 9 relations
Additional features	mappings to PB, VN	mappings to VN	selectional restrictions, syntactic frames	semantic types on classes, roles, lexical units

2. Background

PropBank PropBank (PB)[3] by Palmer et al. (2005) is a predicate model for verbs, later extended to other parts-of-speech in OntoNotes (Hovy et al., 2006). PB associates a lexical entry to one or more semantic classes called *rolesets* that are not shared in general with other entries. Semantic roles are defined locally to each semantic class and categorized as *numbered arguments* (e.g., Arg0 and Arg1, usually the proto-agent and proto-patient), *modifiers* (ArgM plus a function tag, such as LOC for location), and *secondary agent* (ArgA in OntoNotes). Annotated examples are provided for each semantic class. A summary of PB features, compared to other predicate models, is reported in Table 1.

NomBank NomBank (NB)[4] by Meyers et al. (2004) is a model for noun predicates that closely mirrors PB, in that it associates nouns to (noun-specific) semantic classes also called *rolesets*, and defines semantic roles locally to semantic classes, categorizing them as *numbered arguments* and *modifiers*. Examples are provided for each semantic class.

VerbNet VerbNet (VN)[5] by Schuler (2005) is a predicate model for verbs inspired by Levin classes. Semantic classes in VN are organized in a hierarchy of *verb classes*. A verb class is associated to multiple verb lexical entries and to multiple globally defined *thematic roles* (e.g., agent, patient), which describe the semantic roles for that class (with possible selectional restrictions) and form themselves a role hierarchy. Verb classes are also associated to *syntactic frames*. They define how semantic roles can be realized syntactically (e.g., "Agent V Theme") and are associated to a specification of the conveyed semantics based on logical-like predicates applied to thematic role and event variables (e.g., "cause(Agent, E)") . Both thematic roles and syntactic frames are inherited by subclasses. An example sentence is provided for each frame of a verb class.

FrameNet FrameNet (FN)[6] by Baker et al. (1998) builds on the theory of frame semantics to provide semantic classes called *frames* that define prototypical situations *evoked* by lexical entries of different parts-of-speech. The pair ⟨frame, lexical entry⟩ is called *lexical unit*. The semantic roles of a frame are called *frame element*s (FEs): *core* and *core unexpressed* FEs classify mandatory arguments and characterize a frame, differently from *peripheral* FEs, while *extra-thematic* FEs (e.g., "Iteration") situate the frame in a larger context. A set of typically co-expressed FEs form a *CoreSet*. Several Frames and FE relations are defined, starting from frame inheritance, along which Core FEs (but not unexpressed ones) are propagated. A small subset of frames and FEs is annotated with *semantic types*, that for FEs are used to express selectional constraints. Annotated examples are provided for each lexical unit.

SemLink SemLink[7] by Palmer (2009) is a resource providing a mapping between the ⟨lexical entry, semantic class⟩ pairs of (i) VN and PB and (ii) VN and FN, as well as between the semantic roles of these resources.

Predicate models in RDF/OWL Despite the advantages mentioned in Section 1 and the availability of ontologies such as *lemon*, few attempts have been made for representing predicate models in RDF. To the best of our knowledge, no RDF version of PB and NB exists. An ontological model of VN is presented in Gangemi (2010) but no RDF data is available. An RDF version of FN 1.5 is available,[8] but its schema is not aligned to *lemon* and instead closely mirrors the structure and naming used in FN frame files. Finally, VN and FN have been exposed in RDF/*lemon* as part of lemonUby (Eckle-Kohler et al., 2015), a *lemon* version of Uby, but several modeling decisions[9] are no more aligned with the latest developments of *lemon* by the W3C Ontology-Lexica Community Group (besides, only RDF data for an old FN version is available).

3. The PreMOn Ontology – Core Module

Namespace: http://premon.fbk.eu/ontology/core#
Prefix: pmo

The *PreMOn Ontology* is an extension of *lemon* (W3C Ontology Lexicon Community Group, 2015) for representing predicate models and their mappings. An overview of the *PreMOn Ontology* core module, and its relation with *lemon*, is shown in Figure 1 using a UML-like notation. To guide the exposition, we will also refer to Figure 2, showing an example of instantiation of semantic classes and roles for

[3] http://verbs.colorado.edu/~mpalmer/projects/ace.html
[4] http://nlp.cs.nyu.edu/meyers/NomBank.html
[5] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html
[6] http://framenet.icsi.berkeley.edu/fndrupal/

[7] http://verbs.colorado.edu/semlink/
[8] http://ontologydesignpatterns.org/ont/framenet/html/
[9] E.g., modeling of semantic classes as ontolex:LexicalSenses and owl:sameAs links between syntactic and semantic arguments.

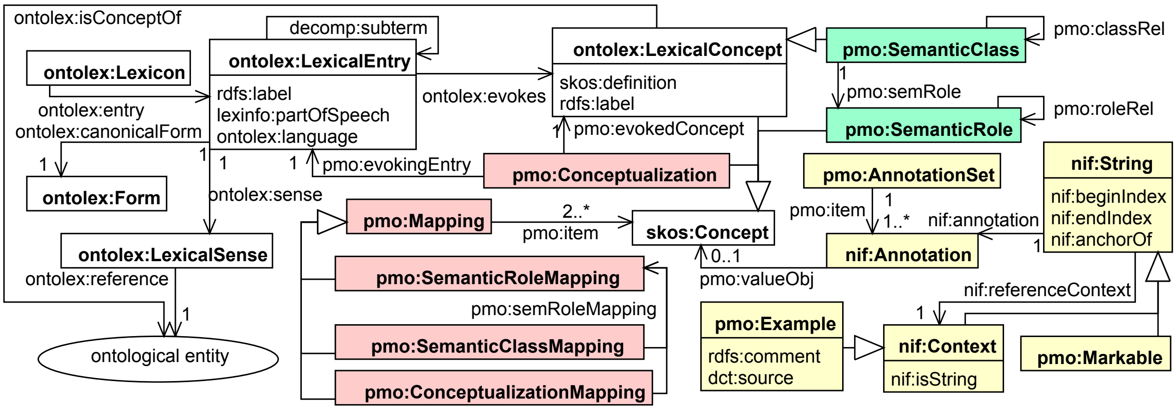

Figure 1: The *PreMOn Ontology* extending lemon.

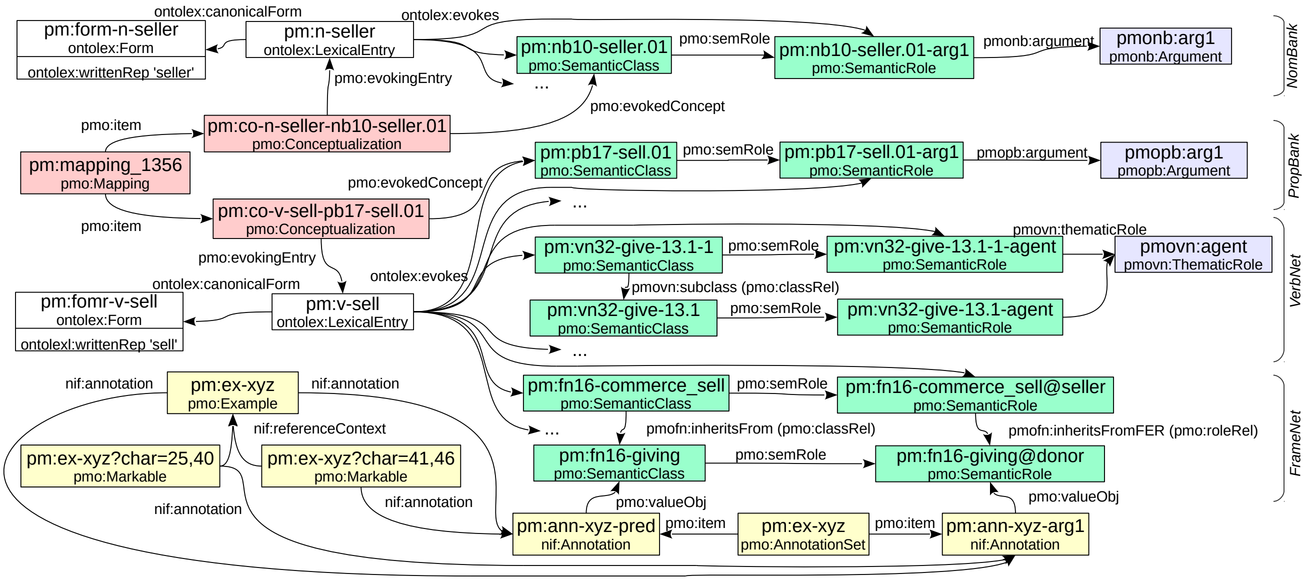

Figure 2: Example of predicate model data represented according to the *PreMOn Ontology*.

the predicate models here considered, as well as an example of mapping between resources from different models.

3.1. Semantic Classes and Roles

lemon represents lexical entries (class ontolex:LexicalEntry) with their associated lexical forms, and allows relating entries to the ontological entities they denote (classes, properties, individuals) using the ontolex:LexicalSense reified relation. Besides mapping to an ontology, which provides the extensional (formal) interpretation of lexical entries, *lemon* supports mapping entries to ontolex:LexicalConcepts (subclass skos:Concept), each denoting an intensional (∼informal) meaning evoked by a set of lexical entries. Example of lexical concepts are WordNet synsets, whose semantics is not formally encoded in an ontology.

We extend *lemon* by introducing classes pmo:SemanticClass and pmo:SemanticRole (filled in light green in Figure 1). pmo:SemanticClass homogeneously represents the semantic classes from the various predicate models. That is, individuals of this class correspond to rolesets in PB and NB (e.g., pm:nb10-seller.01 and pm:pb17-sell.01[10] in Figure 2), verb classes in VN (e.g., pm:vn32-give-13.1-1), and frames in FN (e.g., pm:fn15-commerce_sell). An instance of pmo:SemanticClass typically has (via property pmo:semRole) a number of pmo:SemanticRoles, representing, from a semantic point of view, the roles the arguments of that pmo:SemanticClass can play. For instance, the triple

pm:pb17-sell.01 pmo:semRole pm:pb17-sell.01-arg1

states that pm:nb10-seller.01 has the semantic role pm:nb10-seller.01-arg1. Importantly, semantic roles are defined locally to semantic classes, so VN 'agent' is represented as multiple semantic roles, one for each verb class it occurs in, and with each semantic role linked to its specific selectional restrictions (if any). Note that pmo:SemanticClass is defined as subclass of ontolex:LexicalConcept, as we see pmo:SemanticClasses as essentially informal concepts rather than well defined concepts of a formal ontology (although an ontology can be derived from them, cf., FrameBase and ESO). Being ontolex:LexicalConcepts, pmo:SemanticClasses inherit the link to lexical entries as well as the link (via ontolex:isConceptOf) to the ontological entities formalizing them, typically event classes.

Properties pmo:classRel and pmo:roleRel, and their resource-specific subproperties, are introduced to express the

[10]pm: http://premon.fbk.eu/resource/ is the namespace of *PreMOn Dataset* (Section 5); the fragments nb10, pb17, and so on identify the resource and its version (e.g., NB 1.0, PB 1.7).

relations between elements at each level, such as subtyping, and predicate and role inheritance (e.g., pmofn:inheritsFrom and pmofn:inheritsFromFER for FN). Additional resource-specific classes (e.g., pmovn:ThematicRole, filled in light blue in Figure 1) and properties (e.g., pmovn:thematicRole) further characterize important aspects of each predicate model, like commonalities between semantic roles.

3.2. Mappings

Mappings between different predicate models are practically relevant but cannot be expressed using only the classes above, as they are often defined (e.g., in SemLink and PredicateMatrix) in terms of ⟨pmo:SemanticClass, ontolex:LexicalEntry⟩ pairs. To model these pairs, one could reuse the notion of ontolex:LexicalSense. However, its formalization in *lemon* as reified relation depends on the existence of (exactly) one ontological entity for each ⟨ontolex:LexicalConcept, ontolex:LexicalEntry⟩ pair, a strong constraint that we do not necessarily need for our purposes. Therefore, we introduce the pmo:Conceptualization class (filled in light red in Figure 1, together with other mapping related classes). Structurally, a pmo:Conceptualization can be seen as the reification of the ontolex:evokes relation between ontolex:LexicalEntry and ontolex:LexicalConcept. Semantically, it can be seen as a very specific intensional concept (among many, in case of polysemy) evoked by a single ontolex:LexicalEntry, which can be generalized to a ontolex:LexicalConcept when multiple entries are considered but with a possible loss of information that prevents precise alignments to be represented.

Mappings are explicitly represented as individuals of class pmo:Mapping, and can be seen as sets of (or n-ary relations between) either (i) pmo:Conceptualizations, (ii) pmo:SemanticClasses, and (iii) pmo:SemanticRoles, with role mappings anchored to conceptualization or class mappings via property pmo:semRoleMapping. Figure 2 shows an example of mapping (pm:mapping_1356) between two pmo:Conceptualizations, one from NB (pm:co-n-seller-nb10-seller.01) and one from PB (pm:co-v-sell-pb17-sell.01). We rely on this set-like modeling, since mappings are not necessarily represented as binary relations in predicate mapping resources: e.g., in the PredicateMatrix, each row represents the mapping of a semantic role / lexical entry pair over the different resources (e.g., ⟨13.1-1-agent, deal⟩ in VN, ⟨sell.01-arg0, sell⟩ in PB, ⟨Commerce_Sell-seller, sell⟩ in FN) as well as the corresponding WordNet verb sense. Reifying the n-ary mapping relation also allows us, if needed, to further characterize each single mapping, asserting additional information such as confidence and reliability. Moreover, it is possible to further specialize mappings (e.g., to model mappings holding only in one direction, from a resource to another one, or to represent different types of relationships among the members of the mapping) by subtyping the pmo:Mapping class or the property (pmo:item) relating a pmo:Mapping to its members.

3.3. Annotations

Predicate models are typically complemented by examples showing concrete occurrences of semantic classes and roles in text. More generally, a text can be annotated with semantic classes and roles as a result of manual or automatic SRL. The *PreMOn Ontology* provides some common primitives (filled in light yellow in Figure 1), based on the NLP Interchange Format (NIF) by Hellmann et al. (2013), which aim at properly modeling the heterogeneous annotations of a text for different predicate models. NIF introduces the general notion of nif:String to represent arbitrary text strings. nif:Context is a particular subclass of nif:String, representing a whole string of text. Any substring (itself a nif:String) has a nif:referenceContext relation to the nif:Context individual representing the whole text containing it.

To specifically model the aforementioned examples complementing predicate models, we introduce pmo:Example, subclass of nif:Context, to represent the string associated with the example. The occurrence of a ontolex:LexicalEntry, pmo:SemanticClass, or pmo:SemanticRole in a nif:Context is denoted by an instance of nif:Annotation, related to the given ontolex:LexicalEntry, pmo:SemanticClass, or pmo:SemanticRole via property pmo:valueObj (the value attached to the annotation), and to the nif:Context instance via property nif:annotation. If detailed information on the specific span of text (i.e., substring) denoting the ontolex:LexicalEntry, pmo:SemanticClass, or pmo:SemanticRole is available (e.g., FN provides the specific offsets of lexical units, frames, and frame elements, in the example text) an additional instance of pmo:Markable, subclass of nif:String, is created and linked to the specific nif:Annotation and nif:Context via properties nif:annotation and nif:referenceContext, respectively. As the same nif:Context may contain multiple nif:Annotations referring to one or more semantic classes and their corresponding roles, an additional pmo:AnnotationSet instance is created to cluster annotations from the same predicate structure.

4. Specializing the PreMOn Ontology

While the *PreMOn Ontology* Core Module provides an homogeneous abstraction over heterogeneous predicate models, additional ontology modules, specializing or extending the *PreMOn Ontology* core elements, can be defined to properly capture and model resource-specific aspects (including terminology) in a way compatible with the underlying *PreMOn Ontology* assumptions. We developed four ontology modules, one for each predicate model: PropBank, NomBank, VerbNet, and FrameNet. An overview of the main additional classes (filled in light blue) and properties defined is shown in Figures 3, 4, 5, and 6.

4.1. PreMOn Ontology – PropBank

Namespace: http://premon.fbk.eu/ontology/pb#
Prefix: pmopb

We define classes pmopb:Roleset and pmopb:SemanticRole[11] as subclasses of pmo:SemanticClass and pmo:SemanticRole, respectively. Each pmopb:SemanticRole instance is related (via property pmopb:argument) to exactly one pmopb:Argument, which is defined as the disjoint union

[11] While this class does not necessarily specialize pmo:SemanticRole with additional properties or restrictions, we add it to ease the retrieval of PB-specific semantic roles, something handy when the same repository contains roles from several predicate models.

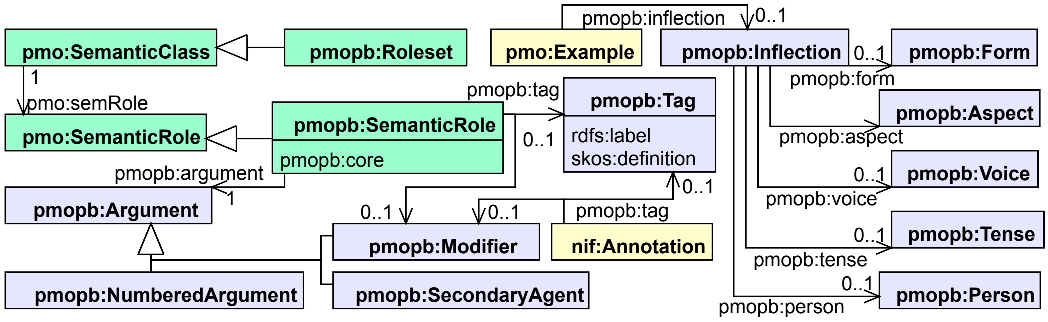

Figure 3: *PreMOn Ontology*-PropBank module

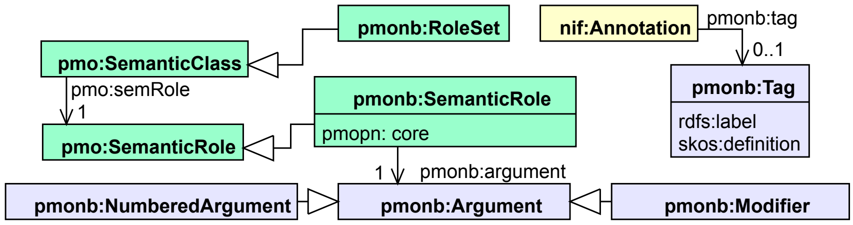

Figure 4: *PreMOn Ontology*-NomBank module

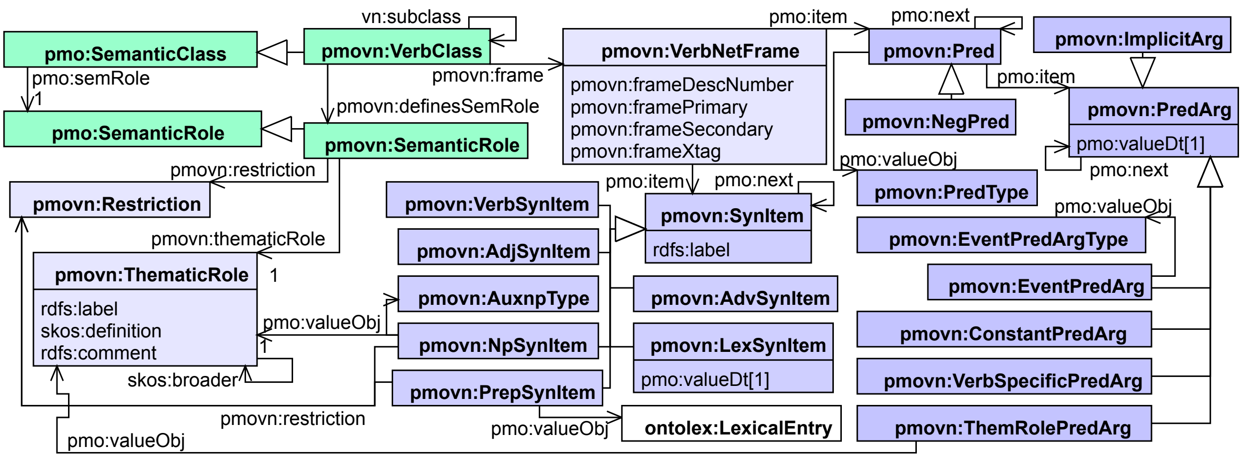

Figure 5: *PreMOn Ontology*-VerbNet module

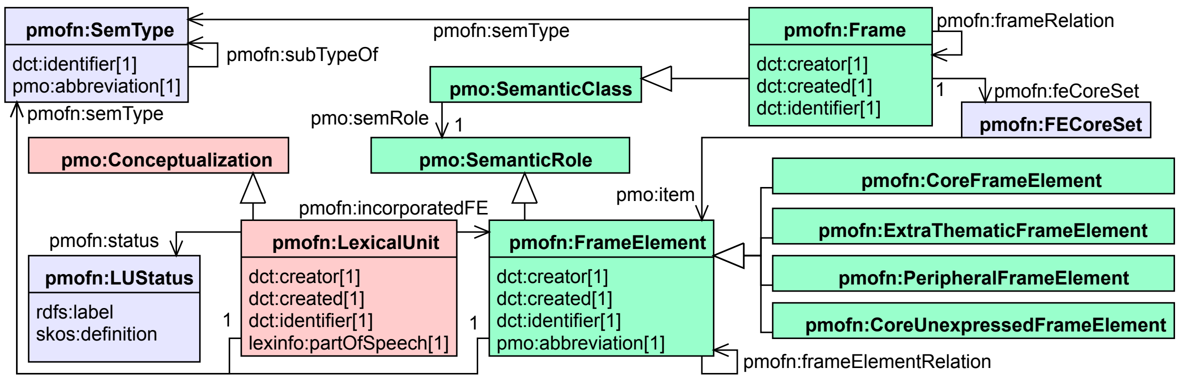

Figure 6: *PreMOn Ontology*-FrameNet module

of three subclasses: pmopb:NumberedArgument, containing the individuals corresponding to numbered arguments (e.g., Arg0, Arg1); pmopb:Modifier, containing the individuals corresponding to modifiers (e.g., ArgM-LOC, ArgM-TMP); and, pmopb:SecondaryAgent, containing the single individual annotating secondary agents (ArgA). While PB annotation guidelines define a single modifier (ArgM) with multiple function tags (e.g., LOC, TMP), we opt to specialize the modifier for each function tag, similarly to the way these arguments are actually annotated by state-of-the-art SRL tools. Property pmopb:tag enables associating possible tags, either a pmopb:Modifier or some additional tag defined in class pmopb:Tag, to pmopb:SemanticRoles, or nif:Annotations of semantic roles in examples. Additional classes (and related properties) are defined to represent inflectional information about examples: pmopb:Inflection, pmopb:Person, pmopb:Tense, pmopb:Aspect, pmopb:Voice, and pmopb:Form.

4.2. PreMOn Ontology – NomBank

Namespace: http://premon.fbk.eu/ontology/nb#
Prefix: pmonb

Similarly to PB, we define pmonb:Roleset and pmonb:SemanticRole as subclasses of pmo:SemanticClass and pmo:SemanticRole, respectively. Each pmonb:SemanticRole instance is related (via property pmonb:argument) to exactly one pmonb:Argument, which is defined as the disjoint union of two subclasses: pmonb:NumberedArgument, containing the individuals corresponding to numbered argument (e.g., Arg0, Arg1), and pmonb:Modifier, containing the individuals corresponding to modifiers (e.g., ArgM-LOC, ArgM-TMP). We also define class pmonb:Tag to capture (via property pmonb:tag) some specific annotations of markables (e.g., PRD, REF, SUPPORT) in the examples.

4.3. PreMOn Ontology – VerbNet

Namespace: http://premon.fbk.eu/ontology/vn#
Prefix: pmovn

We define classes pmovn:VerbClass and pmovn:SemanticRole as subclasses of pmo:SemanticClass and pmo:SemanticRole, respectively. Class members are modeled as instances of ontolex:LexicalEntry, connected to their class via property ontolex:evokes. The VN class hierarchy is modeled via the pmovn:subclassOf property (subproperty of skos:broader), that relates a verb class (e.g., 13.1-1) with its parent class (e.g., 13.1). Given the propagation of semantic roles along the class hierarchy, we introduce property pmovn:definesSemRole to differentiate the pmovn:SemanticRole instances defined on a class from the ones inherited from its ancestor classes. Each pmovn:SemanticRole instance is related (via property pmovn:thematicRole) to exactly one pmovn:ThematicRole, which contains all the thematic roles defined in VB. These thematic roles are organized in a hierarchy, which is formalized via the skos:broader property. For instance,

pmovn:agent skos:broader pmovn:actor

states that pmovn:agent is more specific than pmovn:actor. VN selectional restrictions on pmovn:SemanticRoles (e.g., restricting "theme" to something not animate) are formalized using property pmovn:restriction and class pmovn:Restriction.[12] A verb class may have one or more pmovn:VerbNetFrames (via property pmovn:frame, or its subproperty pmovn:definesFrame, to distinguish frames defined on the class or inherited from ancestors), which have one or more ordered[13] pmo:SynItems, modeling a syntactic construction (e.g., "Agent V Theme [-sentential]") shared by all members of the class, and one or more ordered semantic pmo:Preds, modeling the meaning of the event, and its participants, expressed by the verb class for that syntactic construction (e.g., "approve(during(E), Agent, Theme)"). pmo:SynItems are specialized according to their syntactical function (e.g., pmovn:NpSynItem for noun phrases). A pmovn:NpSynItem can point (via pmo:valueObj) to a pmovn:SemanticRole, and define, via pmovn:restriction, (i) a selectional restriction holding for the pmovn:SemanticRole in that frame (e.g., "animate"), or (ii) some other syntactic restriction (e.g., "np_to_inf"). Similarly, selectional restrictions can be modelled on pmovn:PrepSynItem (e.g., "spatial"). Predicates in pmovn:Pred have a type (pmovn:PredType, e.g., "approve") and can be further decomposed in pmovn:PredArg (e.g., "during(E)") of various types (e.g., pmovn:EventPredArg). Negation of a predicate is expressed by typing the corresponding instance as pmovn:NegPred, while implicit pmovn:PredArgs are typed as pmovn:ImplicitArg.

[12]A detailed presentation of the formalization of selectional / syntactic restrictions is omitted in the paper due to lack of space. More details are provided on *PreMOn* website.

[13]We relied on the standard first/next/item pattern for lists.

4.4. PreMOn Ontology – FrameNet

Namespace: http://premon.fbk.eu/ontology/fn#
Prefix: pmofn

We define classes pmofn:Frame and pmofn:FrameElement as subclasses of pmo:SemanticClass and pmo:SemanticRole, respectively. pmofn:FrameElement is further specialized in four subclasses, denoting the four typologies of FN frame elements (e.g., pmofn:CoreFrameElement). Being pmo:SemanticRoles, in *PreMOn Ontology* frame elements are always specific to the frame where they are defined, even for extra thematic frame elements that are typically shared across frames in FN (e.g., the "Circumstances" extra thematic frame element corresponds to multiple individuals of type pmofn:ExtraThematicFrameElement, one for each frame where it is defined). Frame element core sets of a pmofn:Frame are represented as reified objects of type pmofn:FECoreSet, having as members some pmofn:FrameElements. Relations between pmofn:Frames are modeled using the subproperties of pmofn:frameRelation (e.g., pmofn:inheritsFrom). Similarly, mappings between pmofn:FrameElements of pmofn:Frames related via some pmofn:frameRelation are represented using frame relation-specific subproperties of pmofn:frameElementRelation (e.g., pmofn:inheritsFromFER). Within a frame, a frame element may exclude/require the presence of another frame element (pmofn:excludesFrameElement/pmofn:requiresFrameElement). pmofn:LexicalUnit, associating a lexical entry with a frame, is defined as subclass of pmo:Conceptualization. A pmofn:LexicalUnit may have a development status (pmofn:LUStatus) and can incorporate a pmofn:FrameElement (e.g., "microvawe.v', besides evoking frame "Apply heat", also incorporates the frame element "Heating instrument"). Finally, pmofn:Frames, pmofn:FrameElements and pmofn:LexicalUnits can be constrained according to some semantic types, defined in pmofn:SemType, and organized in a hierarchy according to pmofn:subTypeOf relations between them.

5. The PreMOn Dataset

Namespace: http://premon.fbk.eu/resource/
Prefix: pm

To populate *PreMOn* with content from the various resources (predicate models, mappings), we developed an open-source Java command-line tool available on *PreMOn* website. The tool applies pluggable, resource-specific converters to the original distribution files of each resource, instantiating the proper individuals and assertions according

to the *PreMOn Ontology*. If available, mappings to additional resources (e.g., WordNet synsets, OntoNotes groupings) are also extracted. OWL 2 RL inference, statistics extraction and some cross-resource cleanup (e.g., for dropping inconsistent mappings) are applied to extracted triples, leveraging RDFpro (Corcoglioniti et al., 2015) for RDF processing. The resulting triples are placed in distinct *named graphs* identified by the resource name and version (e.g., pm:fn15 for FN v1.5), so to track provenance at a coarse-grained level and allow querying only data of specific predicate models using SPARQL clauses FROM and FROM NAMED. Examples, and related triples, are placed in further separated named graphs (e.g., pm:fn15-ex), and their extraction can be enabled/disabled at runtime.
Specific conversion strategies had to be implemented for each predicate model. E.g., in VN, semantic roles (with selectional constraints) and frames have to be propagated from a class to its subclasses, unless redefined in the latter. In PB (and NB), the instantiation of pmopb:SemanticRoles requires creating an individual for each ⟨pmopb:Roleset, pmopb:Argument⟩ pair, as no information is provided on which arguments a predicate may have (besides explicit occurrence in frame files, in which case semantic role attributes pmopb:core/pmonb:core are set to "true").
We applied the conversion suite on a large collection of resources, producing a comprehensive dataset, namely the *PreMOn Dataset*, containing: PB v1.7 (pb17), PB v2.1.5 released with OntoNotes v5 (pb215), NB v1.0 (nb10), VN v3.2 (vn32), FN v1.5 (fn15), FN v1.6 (16), and SemLink 1.2.2c (sl122c). The *PreMOn Dataset* contains the mappings between semantic classes and roles provided by each predicate model and SemLink, as well as the mappings between VN classes and lexical senses in WordNet 3.1 (wn31) and OntoNotes 5 groupings. Table 2 reports the number of individuals (e.g., of pmo:SemanticClasses), relations (e.g., pmo:classRel), and triples for the considered predicate models and their versions. Table 3 reports the numbers of mappings, which may be lower than the ones in the original resources as we drop mappings involving non-existing semantic classes, roles, or conceptualizations (e.g., because removed/reorganized in the newer versions of the resources, released after the mapping was defined).
All the datasets generated (as well as the *PreMOn Ontology*) are documented and published online as Linked Open Data on *PreMOn* website. Data can be accessed in three ways: URI dereferencing with content negotiation (RDF and HTML formats supported); SPARQL querying on our endpoint (see Figure 7); and bulk dataset download.

6. Leveraging PreMOn

By adopting an homogeneous schema for heterogeneous predicate models, *PreMOn* facilitates the joint querying of content from different resources. Figure 7 shows an example of query, run through the SPARQL endpoint on *PreMOn* website, that looks for lexical entries (?lexEnt) evoking semantic classes in different resources (?resource), for which no mappings are defined. Results are ordered by decreasing number of resources defining the lexical entry. This query hints a way to exploit *PreMOn* to investigate, and possibly extend, mappings between predicate models.

Figure 7: SPARQL querying *PreMOn* data.

Other interesting queries can be answered given *PreMOn* data (see website) as, e.g., to map a semantic class or role (for a lexical entry) from a resource to another, possibly navigating chains of mappings via SPARQL property paths, or, together with WordNet RDF data, to find synonymous lexical entries that can extend a resource lexicon.
The use of a single RDF model for multiple resources, amenable to the use of reasoning, querying and other SW data processing techniques, opens up new opportunities for analyzing, validating, and possibly cleaning up predicate model data. We already implemented a simple technique for identifying inconsistent mappings, which we currently remove, and we plan to further pursue this direction.
By exploiting *PreMOn* data and ontology, NLP and knowledge extraction tools can now express SRL annotations in RDF in a way compatible with the increasingly used NIF. Indeed, the *PreMOn* formalization of the usage examples released with predicate models shows a concrete case of SRL annotation data exposed according to NIF principles. We are currently working on adapting the PIKES system to produce annotation of SRL according to this formalization.
By extending *lemon*, *PreMOn* endorses the latter's goal to provide proper linguistic grounding for ontologies. In particular, *PreMOn* enables to ground event and frame ontologies, such as FrameBase and ESO, on a proper representation of the linguistic information for predicates, thus supporting the development of a comprehensive catalogue, from the linguistic to the knowledge level, of event and event participant types.

7. Conclusions

PreMOn is both an ontology extending *lemon* for representing predicate models and their mappings, and a LOD dataset based on that ontology that contains interlinked predicate data for several predicate models.
Future works will leverage the possibility to query and reason on the *PreMOn Dataset* and will focus on: (i) extending the mappings between predicate models, integrating data from the Predicate Matrix; (ii) mapping semantic classes and roles to concepts in the FrameBase and ESO ontologies; and (iii) mapping selectional constraints from VN and FN, as well as from works such as (Bryl et al., 2012), to a

Table 2: *PreMOn Dataset* General Statistics.

	nb10	pb17	pb215	vn32	fn15	fn16	Total
pmo:SemanticClass	5,576	6,181	8,751	484	1,018	1,204	23,214
pmo:SemanticRole	76,241	145,586	206,701	1,396	9,633	11,251	450,808
pmo:Conceptualization	5,576	6,196	9,208	6,338	11,887	13,247	52,452
ontolex:LexicalEntry	4,704	5,187	7,671	4,402	9,413	10,107	17,195
pmo:Example	9,583	10,221	15,832	1,396	153,589	174,425	365,046
nif:AnnotationSet	9,583	10,221	15,832	1,396	154,573	196,610	388,215
pmo:classRel	–	–	–	212	1,643	1,968	3,823
pmo:roleRel	–	–	–	–	8,807	10,599	19,406
Core triples	517,495	629,932	926,466	236,526	332,386	411,554	3,035,670
Example triples	339,496	429,953	653,266	22,404	5,752,132	7,122,995	14,320,246
Inferred triples	252,738	371,586	539,087	69,202	548,291	632,890	2,414,220
Total	1,109,729	1,431,471	2,118,819	328,132	6,632,809	8,167,439	19,770,136

Table 3: *PreMOn Dataset* Mapping Statistics.

source-target (resource)	# mappings		
	conceptualization	class	role
pb17-to-vn32 (pb17)	2,754	3,899	7,950
pb215-to-vn32 (pb215)	4,765	6,186	12,917
pb215-to-fn15 (pb215)	1,706	1,792	–
nb10-to-vn32 (nb10)	1,081	1,664	3,410
nb10-to-pb17 (nb10)	3,474	3,474	9,841
nb10-to-pb215 (nb10)	3,473	3,473	9,905
vn32-to-wn31 (vn32)	5,248	–	–
vn32-to-on5 (vn32)	3,357	–	–
vn32-to-fn15 (sl122c)	3,617	1,713	1,904
vn32-to-pb17 (sl122c)	4,506	4,617	10,056
vn32-to-pb215 (sl122c)	5,416	5,563	12,430
fn15-to-fn16 (fn16)	–	891	9,132
Total	36,253	33,272	77,545

common ontology (e.g., Yago), possibly propagating these constraints to PB and NB. All these lines of work are part of a larger effort to create a coherent resource for knowledge extraction that maps predicate models used by available SRL tools to FrameBase and other SW ontologies.

Acknowledgements Partially funded by the European Union's FP7 via the NewsReader Project (ICT-316404).

8. References

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proc. COLING-ACL '98*.

Bryl, V., Tonelli, S., Giuliano, C., and Serafini, L. (2012). A novel FrameNet-based resource for the Semantic Web. In *Proc. of ACM SAC 2012*, pages 360–365.

Chiarcos, C., McCrae, J., Cimiano, P., and Fellbaum, C. (2013). Towards open data for linguistics: Linguistic Linked Data. In *New Trends of Research in Ontologies and Lexical Resources*.

Corcoglioniti, F., Rospocher, M., Mostarda, M., and Amadori, M. (2015). Processing billions of RDF triples on a single machine using streaming and sorting. In *Proc. of ACM SAC 2015*, pages 368–375.

Corcoglioniti, F., Rospocher, M., and Palmero Aprosio, A. (2016). A 2-phase frame-based knowledge extraction framework. In *Proc. of ACM SAC 2016*. to appear.

Eckle-Kohler, J., McCrae, J. P., and Chiarcos, C. (2015). lemonUby - A large, interlinked, syntactically-rich lexical resource for ontologies. *Sem. Web. J.*, 6(4):371–378.

Gangemi, A. (2010). What's in a schema? In *Ontology and the Lexicon*, pages 144–182. Cambridge Uni. Press.

Hellmann, S., Lehmann, J., Auer, S., and Brümmer, M. (2013). Integrating NLP using Linked Data. In *Proc. of ISWC 2013*, pages 98–113.

Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., and Weischedel, R. (2006). OntoNotes: The 90% solution. In *Proc. of HLT-NAACL (Short Papers)*, pages 57–60.

Lacalle, M. L. D., Laparra, E., and Rigau, G. (2014). Predicate Matrix: extending SemLink through WordNet mappings. In *Proc. of LREC 2014*.

McCrae, J., Aguado-De-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., and Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Lang. Resour. Eval.*, 46(4):701–719.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004). The NomBank project: An interim report. In *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106.

Palmer, M. (2009). SemLink: Linking PropBank, VerbNet and FrameNet. In *Proc. of GenLex 2009*.

Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., Soroa, A., Ploeger, T., and Bogaard, T. (2016). Building event-centric knowledge graphs from news. *Journal of Web Semantics*. to appear.

Rouces, J., de Melo, G., and Hose, K. (2015). FrameBase: Representing n-ary relations using semantic frames. In *Proc. of ESWC 2015*.

Schuler, K. K. (2005). *Verbnet: A Broad-coverage, Comprehensive Verb Lexicon*. Ph.D. thesis.

Segers, R., Vossen, P., Rospocher, M., Serafini, L., Laparra, E., and Rigau, G. (2015). ESO: a frame based ontology for events and implied situations. In *MAPLEX 2015*.

W3C Ontology Lexicon Community Group. (2015). Final model specification. Technical report. http://w3.org/community/ontolex/wiki/Final_Model_Specification.