A Novel FrameNet-based Resource for the Semantic Web

Volha Bryl, Sara Tonelli, Claudio Giuliano, Luciano Serafini
Fondazione Bruno Kessler
via Sommarive 18, 38123 Trento, Italy
{bryl,satonelli,giuliano,serafini}@fbk.eu

ABSTRACT

FrameNet is a large-scale lexical resource encoding information about semantic frames (situations) and semantic roles. The aim of the paper is to enrich FrameNet by mapping the lexical fillers of semantic roles to WordNet using a Wikipedia-based detour. The applied methodology relies on a word sense disambiguation step, in which a Wikipedia page is assigned to a role filler, and then BabelNet and YAGO are used to acquire WordNet synsets for a filler. We show how to represent the acquired resource in OWL, linking it to the existing RDF/OWL representations of FrameNet and WordNet. Part of the resource is evaluated by matching it with the WordNet synsets manually assigned by FrameNet lexicographers to a subset of semantic roles.

Categories and Subject Descriptors

I.2.4 [**Computing Methodologies**]: Artificial Intelligence—*Knowledge Representation Formalisms and Methods*; I.2.7 [**Computing Methodologies**]: Artificial Intelligence—*Natural Language Processing*

Keywords

Semantic web, FrameNet, WordNet, word sense disambiguation, OWL

1. INTRODUCTION

FrameNet [13] has become one of the most important semantic resources encoding information about situations, the *frames*, and participants, the semantic roles, also called *frame elements* (FEs). It has been widely used in natural language processing tasks, from textual entailment [1] to question answering [17]. However, some major issues have emerged regarding coverage and ontological consistency [12]. One crucial aspect that would require improvement is the information about possible lexical fillers for FEs, which may impact on the performance of semantic frame parsers. In the last FrameNet version (1.5), about 40 semantic types were defined to provide semantic constraints on FE fillers, e.g. the semantic type *Sentient* assigned to the *Agent* FE[1]. However, this information does not cover the whole FE set (only 54% of the FEs have a semantic type), is often very high-level (consider e.g. *Physical_entity* type coupled with *Instrument* FE), and only some of the semantic types are mapped to a WordNet synset. Therefore, it could hardly be used during large-scale semantic analysis.

In this work, we aim at enriching FE information by mapping their lexical fillers to WordNet [5] using a Wikipedia-based detour. Our methodology relies first on a word sense disambiguation (WSD) step, in which a Wikipedia page is automatically assigned to nominal FE fillers. The direct mapping to WordNet is not possible as no WordNet-based corpus of considerable dimensions is available to train a WSD system. Then, we apply *BabelNet* [10], an existing resource that maps Wikipedia and WordNet, to create a repository of synsets for each FE. In order to improve *BabelNet* coverage, we further link the fillers having no WordNet information but linked to Wikipedia with the corresponding concepts in YAGO [19] ontology, which is, in turn, linked to WordNet.

The above methodology allows us to create a new complementary FrameNet-based resource: a repository of senses for semantic roles. The repository, on the one hand, is built by making use of the existing semantic web resources and techniques. On the other hand, the acquired resource is beneficial for a number of text processing tasks, primarily, for semantic frame annotation, and thus, it contributes to enriching and linking tasks, which are of high importance for the semantic web. In the paper, we also show how to represent the acquired resource in OWL language linking it to RDF/OWL representations of FrameNet and WordNet, thus making the repository available to the semantic web.

The paper is structured as follows: in Section 2 we describe previous approaches aimed at translating FrameNet into markups for the semantic web and at enriching it with additional semantic information. In Section 3 the methodology for creating the synset repository is detailed and some relevant statistics are given. In Section 4 we present the word sense disambiguation system used for assigning a Wikipedia page to FE fillers. In Sections 5 and 6 we present the OWL representation of the sense repository and evaluate, illustrate and discuss the resource.

2. RELATED WORK

[1]73 semantic types are listed in FrameNet 1.5. However, some of them refer to lexical units and not to FEs.

SAC'12 March 25-29, 2012, Riva del Garda, Italy.

The first attempt to convert FrameNet into a knowledge representation language is reported by [9], where FrameNet 1.1 is translated into the DAML+OIL markup, a specific kind of RDF markup with a better capability to describe object relationships than simple XML. [15] translate FrameNet to OWL DL, and then show the potential of FrameNet in connection with a DL reasoner to partially solve a question answering task.

[2] present a methodology to enrich frames with argument restrictions provided by a super-sense tagger and different types of constraints, and to encode the results according to a Linguistic MetaModel. Their goal is similar to ours, in that they want to enrich existing frames starting from textual data. However, while we focus on the automatic development of a resource merging frame elements and WordNet synsets, [2] are mainly concerned with the validation of a methodology for domain-specific frame generation that includes a frame detection system, a super-sense tagger and a constraint-based frame selection. Indeed, their manual evaluation is limited to 3 well-defined frames, KILLING, JUDGMENT_COMMUNICATION and COMMERCE.

Finally, [11] describe how FrameNet 1.5 was converted into RDF dataset published on the Linked Open Data (LOD) cloud, and how part of the information related to frames was modeled as full-fledged OWL knowledge patterns. Our work can complement the effort made by [11] in that we contribute to resource linking in the LOD cloud by mapping FEs to WordNet synsets. On the contrary, previous approaches in this direction have focused exclusively on the mapping between FrameNet frames and WordNet synsets, either directly ([8], [20]), or through VerbNet ([18]).

3. CREATION OF A SENSE REPOSITORY FOR ARGUMENT HEADS

Our goal is to automatically acquire from annotated data information about the lexical fillers of frame elements, which allow us to generalize over the fillers and induce constraints for further processing. In particular, we aim at collecting a repository of WordNet synsets for each FE-frame pair. We use the expression 'argument heads' to refer to the semantic heads of arguments, i.e. the word or multiword that contributes the most to the meaning of a phrase corresponding to an argument. For instance, if the prepositional phrase 'in their own backyard' is annotated with the *Place* FE label (role label), we extract its semantic head 'backyard', which is the lexical filler of *Place*.

The workflow for the creation of the repository includes the following steps:

Head extraction: We first extract all argument heads from a reference corpus with semantic role annotation. We use the FrameNet database 1.5 [13], including also all documents in the release provided with continuous annotation, and the frame annotated data developed for the SemEval-10 Evaluation campaign on "Linking Events and their Participants in Discourse" [14]. Since all sentences in the FrameNet database are annotated with gold standard phrase type and part of speech (PoS) tags, we develop a set of rules for the recognition of semantic heads in nominal and prepositional phrases given the PoS sequence. As for the SemEval data, the heads for each role are already given. Similar to previous approaches [4, 6], we limit our extraction process to nominal heads. This is due to the fact that Wikipedia, which we use in the disambiguation step, is a concept-based resource, thus our disambiguation engine achieves better coverage and performance over nouns. Note that for roles expressed by prepositional phrases, we extract the head of the noun phrase child dominated by the prepositional phrase.

Head disambiguation: After the extraction routine, we run the *disambiguation* step and associate each head in a sentence to a Wikipedia concept. In particular, we use the WSD system described in Section 4 that, given a word in context, is able to find the Wikipedia page that best corresponds to the meaning of the word in the given sentence. In this way, we want to associate to each tuple (f, r_f), where f is the frame and r_f is the semantic role, a set of Wikipedia concepts W by linking them to the lexical fillers of r_f.

Creation of a Wikipedia-based repository: We collect all Wikipedia pages (i.e. concepts) associated to each (f, r_f) and the corresponding frequency observed in the reference corpus. For example, the *Forbidden* role in the LAW frame has been associated with the following entry:

Air_pollution(1), Defamation(1), Everything(1), Incest(1), Population_density(1), Tobacco_advertising(1)

containing a list of Wikipedia pages[2] with the number of times they have been linked to the given (f, r_f) pair in the reference corpus.

Creation of a WordNet-based repository: We run a further linking step based on BABELNET [10]. This resource is a large semantic network where lexicographic and encyclopedic knowledge from WordNet and Wikipedia have been automatically integrated. The authors report a mapping between 105,797 synsets and 199,735 Wikipedia pages with a precision of 81.9%. In our experiment, we apply such mapping in order to replace each Wikipedia page assigned to (f, r_f) with a synset. In this way, we devise a sort of Wikipedia-detour to WordNet-based WSD, because we obtain for each (f, r_f) a list of related synsets. This approach has several advantages: since the WSD system is trained on the whole Wikipedia, it can potentially disambiguate (i.e. assign a Wikipage) every term or multiword, given that it has been linked at least once in the whole Wikipedia. This includes also many named entities that are not present in WordNet. For example, the expression *American_businessman* is not in WordNet but it was linked to the `Businessperson` page in Wikipedia and then to the synset {*businessperson#n#1, bourgeois#n#1*} through BabelNet. Also *mrs_thatcher* and *mr_gorbachev* cannot be matched with any lemma in WordNet, but they were respectively connected to the `Margaret_Thatcher` and the `Mikhail_Gorbachev` page, and then to *Thatcher#n#1* and *Gorbachev#n#1*.

Extension of the WordNet-based repository: During the previous step, some Wikipedia pages could not be mapped to any WordNet synset due to *BabelNet* coverage. In order to limit this loss of information, we take advantage of existing resources for the semantic web, namely YAGO [19], an automatically created large-scale ontology, with taxonomy structure derived from WordNet and knowledge about individuals extracted from Wikipedia. Each WordNet synset becomes a class in YAGO and *SubClassOf* relation between classes is derived from hyponymy relation in WordNet. Wikipedia categories become subclasses of WordNet-based classes and,

[2]The page can be displayed by preceding it with `http://en.wikipedia.org/`

therefore, given a Wikipedia page one can first find the corresponding YAGO concept (using *hasWikipediaUrl* YAGO property), and then obtain a list of WordNet synsets querying the ontology for the parent classes of the Wikipedia categories of the given concept (using *type* and *hasSynsetId* properties). To access the ontology we used Java API for queering YAGO in SPARQL, which is provided by YAGO team and works with YAGO dump in Jena TDB format[3].

At the end of this step, we have automatically acquired a resource containing 4,395 (f, r_f) tuples, each associated with a list of synsets with the corresponding frequencies. For example, the role *Heating_instrument* in the APPLY_HEAT frame is characterized by the following synsets (number of occurrences between parenthesis):
barbecue#n#1 (2), *fire#n#3 (2)*, *microwave#n#2 (1)*, *oven#n#1 (1)*, *kiln#n#1 (1)*.
We report in Table 1 some statistics about the source dataset, the argument heads linked to Wikipedia and the Wikipedia-WordNet mapping. In the WSD step, we are able to link 77% of all nominal heads to Wikipedia. Then, using BabelNet we are able to map to a synset about 81% of the disambiguated heads, which corresponds to 62% of all nominal heads in the reference corpus. The final linking with YAGO increases the number of heads linked to a synset up to 117,038.

Noun heads in the reference corpus	167,152
Heads linked to Wikipedia	128,573 (77%)
Heads linked to a synset (BabelNet only)	103,712 (62%)
Heads linked to a synset (BabelNet and YAGO)	117,038 (70%)
Frame-FE pairs in FrameNet 1.5	6,369
Frame-FE pairs in reference corpus	4,905
Frame-FE pairs in the final sense repository	4,395
FEs in FrameNet 1.5	1,001
FEs in reference corpus	956
FEs in the final sense repository	911

Table 1: Statistics about the extracted data

Note that the number of frame-FE pairs and of single FEs in FrameNet 1.5 refers only to those that occur at least in one annotated sentence. Indeed, in FrameNet more than 3,000 frame-FEs are reported in the frame description but are not instantiated by any example sentence.

In the reference corpus, we did not consider (i.e. tried to link) about 1,400 frame-FEs pairs, which refer to role fillers that are not nominal. However, we collected at least one synset for 91% of the FEs in FrameNet 1.5 (i.e. 911/1001). This means that most of the FEs having a verbal or adjectival head, can be expressed also by a nominal filler, which we were able to classify.

4. WSD SYSTEM

The problem of assigning to a role filler a Wikipedia page is cast as a WSD exercise, in which each role head has to be disambiguated using Wikipedia to provide the corresponding sense inventory. The idea of using Wikipedia to train a supervised WSD system was first proposed by [3]. Our approach is summarized as follows.

4.1 Training Set

We adopt a word expert approach to create the training set. For each nominal head h being the lexical filler of a role r, we collect from the English Wikipedia dump[4] all contexts where h is linked to a page. In particular, a context corresponds to a line of text in the Wikipedia dump and it is represented as a paragraph in a Wikipedia article. The set of target pages represents the senses of h in Wikipedia and the contexts are used as labeled training examples.

4.2 Learning Algorithm

To disambiguate role heads, we integrate syntactic and semantic knowledge sources, typically used in the WSD literature, by means of a kernel combination [7]. Kernel methods are theoretically well founded in statistical learning theory and have shown good empirical results in many applications [16]. The strategy adopted consists in splitting the learning problem into two parts. First, the input data are embedded in a suitable feature space, and then we use a linear algorithm (e.g. support vector machines) to discover nonlinear patterns in the input space. The kernel function is the only task-specific component of the learning algorithm. For each knowledge source a specific kernel has been defined. By exploiting the property of kernels, basic kernels are then combined to define the WSD kernel. Specifically, we used a linear combination of gap-weighted subsequences, bag-of-words, and latent semantic kernels.

Gap-weighted subsequences kernel This kernel learns syntactic and associative relations between words in a local context. We extend the gap-weighted subsequences kernel to subsequences of word forms, stems, part-of-speech tags, and orthographic features (capitalization, punctuation, numerals, etc.). We define gap-weighted subsequences kernels to work on subsequences of length up to 3. E.g., suppose we have to disambiguate the noun *goal* in the context "Maradona scored Argentina's third goal", given the labeled example "Ronaldo scored two goals in the second half" as training. A traditional approach, that only considers contiguous ngrams, has no clues to return the correct answer because the two contexts have no features in common. The use of gap-weighted subsequences allows us to overcome this problem and extract the feature "score goal", shared by the two examples.

Bag-of-words kernel This kernel exploits domain and topical information. Bag-of-words kernel takes as input a wide context window around the target term. The main drawback of this approach is the need of a large amount of training data to reliably estimate model parameters. E.g., despite the fact that the examples "People affected by AIDS" and "HIV is a virus" express related concepts, their similarity is zero using the bag-of-words model since they have no words in common (they are represented by orthogonal vectors in the vector space model). On the other hand, due to the ambiguity of the word *virus*, the similarity between the contexts "the laptop has been infected by a virus" and "HIV is a virus" is greater than zero, even though they convey very different messages.

Latent semantic kernel To overcome the drawback of bag-of-words, we incorporate semantic information acquired from English Wikipedia in an unsupervised way by means of latent semantic kernel. This kernel extracts semantic infor-

[3]http://www.mpi-inf.mpg.de/yago-naga/yago/

[4]http://download.wikimedia.org/enwiki/20100312

mation through co-occurrence analysis of the corpus. The technique used to extract the co-occurrence statistics relies on a singular value decomposition of the term-by-document matrix. E.g., the similarity in the latent semantic space of the two examples "People affected by AIDS" and "HIV is a virus" is higher than in the bag-of-words representation, because the terms AIDS, HIV and virus very often co-occur in the medicine domain.

4.3 Implementation Details

The latent semantic model is derived from the 200,000 most visited Wikipedia articles. After removing terms that occur less than 5 times, the resulting dictionary contains about 300,000 terms. We use the SVDLIBC package to compute the SVD, truncated to 100 dimensions.[5] To classify each role head, we employ the LIBSVM package.[6] No task-specific parameter optimization was performed.

A recent evaluation of the disambiguation performance of the system on the English Automatic Content Extraction (ACE 2005) dataset[7] enriched with ground-truth links to Wikipedia showed that the linking quality over common nouns and named entities achieves 0.72 precision, 0.71 recall and 0.71 F1.

5. RESOURCE STRUCTURE

One of the objectives of our work is to make the acquired sense repository available to the semantic web. Therefore, in this section we present the OWL version of the resource and explain the modelling choices made.

To produce a compact representation of the repository, we have generated *distributions of senses* for each frame-FE pair. In particular, for each frame-FE a tree of senses is created based on WordNet hyponymy relation, where each node (synset) stores the number of examples of a given frame-FE pair classified with this synset or its hyponym. Then, the distribution of senses for a frame-FE pair is the lowest (starting from the most specific synsets) tree level, in which each node contains at least 10% of the examples of the respective frame-FE pair and all nodes in total contain at least 60% of the examples. For instanse, for *Observable_body_parts-Possessor* frame-FE pair the resulting set is

$$animal\#n\#1,\ 23\%;\quad person\#n\#1,\ 51\%$$

In this way, the most representative among the most specific senses are selected. The selection is done automatically, therefore the two above thresholds can be varied to compare the outcomes.

The OWL version of the sense repository is based on the existing RDF/OWL representations of both WordNet[8] and Framenet [15]. We have used Protégé ontology editor[9], which allowed us to create an intermediate level between these two ontologies. As illustrated in Figure 1, the main structural element of our ontology is *learntSemType* class, each instance of which is classified as a specific frame-FE (which is a class in FrameNet OWL representation), and has as properties a WordNet synset (which is an individual in RDF/OWL WordNet representation), the total number

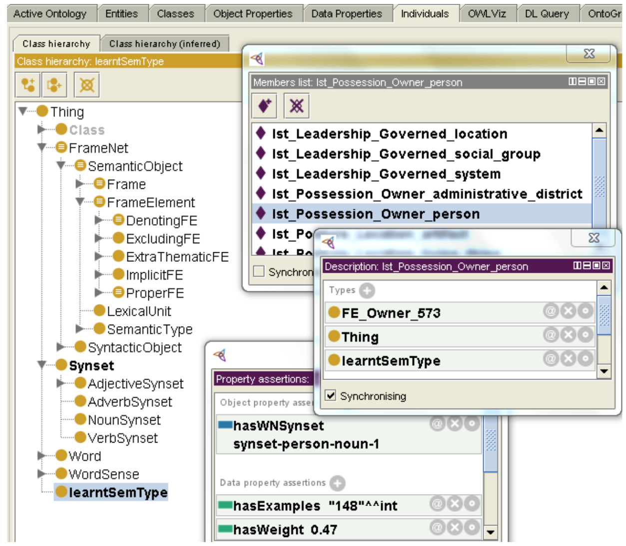

Figure 1: Modelling the sense repository with Protégé ontology editor

of examples for a given frame-FE and a number (percentage) of the examples classified with a given synset or its hyponym. The creation and evaluation of the OWL version of the resource is in progress, and its online availability is our first priority among the future work directions.

6. RESOURCE EVALUATION

In the first evaluation step, we take advantage of the WordNet synsets that were manually associated by FrameNet lexicographers to specific semantic types. For example, the *Shape* semantic type assigned to several FEs is explicitly linked to the synset WN *shape#n#2*. In FrameNet, 19 semantic types are associated with a synset.

This mapping is seen as a gold standard against which we compare our resource. More precisely, we consider the subset of FEs in our resource having a semantic type manually associated with a synset, and we match the corresponding list of automatically acquired synsets with the gold standard ones. Since the gold synsets are usually the topmost synsets that can be possibly linked with a semantic type, we consider as a positive match a synset in the repository that is either identical to the gold one or is a hyponym of it. Note that the scale of this evaluation step is somehow limited as only 266 (6%) out of 4395 frame-FE pairs in the final sense repository have a semantic type which was mapped to a WordNet synset in FrameNet.

In Table 2 we report the results of this evaluation. The number of matches per synset_i, which is assigned to a semantic type in FrameNet, refers to the number of frame-FE examples for which we were able to classify the example head with synset_i or with a hyponym of it (e.g., *head#n#1* or *palm#n#1* were counted as matches in case synset_i=*body_part#n#1*).

We do not report the results for the semantic type *Physical_entity* mapped to *entity#n#1* WordNet synset, for which we have 852 frame-FE examples. This is a top synset in WordNet taxonomy, so such a semantic type does not provide any information. Also we do not report results for *Event*, *Running-water*, *Organization*, *Human_act* and *State*

[5] http://tedlab.mit.edu/~dr/svdlibc/
[6] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[7] http://www.itl.nist.gov/iad/mig//tests/ace/ace05/
[8] http://www.w3.org/TR/wordnet-rdf/
[9] http://protege.stanford.edu/

FN semtype	WN synset	#ex	%mtch
Living_thing	organism#n#1	49	73%
Artifact	artifact#n#1	855	73%
Human	person#n#1	1301	71%
Group	group#n#1	40	65%
Physical_object	object#n#1	5703	62%
Quantity	measure#n#2	79	59%
Location	location#n#1	2197	26%
Message	message#n#2	1004	4%
Content	content#n#5	1042	0.5%

Table 2: Comparing FrameNet semantic types linked to WordNet and the acquired senses

FN semtype	Learned synset	#ex	%mtch
Message (message#n#2)	communication#n#2	1004	24%
	object#n#1		22%
	psychological_feature#n#1		20%
	relation#n#1		15%
Artifact (artifact#n#1)	commodity#n#1	855	26%
	instrumentality#n#3		25%
	structure#n#1		15%
Sentient	person#n#1	42669	67%
Body_part	body_part#n#1	1502	85%
Time	fundamental_quantity#n#1	1272	42%
	event#n#1		15%
	whole#n#2		12%

Table 3: Examples of acquired senses for existing FrameNet semantic types

semantic types linked to *event#n#1*, *watercourse#2*, *organization#n#1*, *act#n#2* and *state#n#2* respectively, as the number of frame-FE examples for these types is limited.

For the first 6 semantic types of Table 2 the number of matches is high, which allows for the positive evaluation of this part of the sense repository. But what happens in case of *Location*, *Message* and *Content* semantic types? Let us explain this based on the example of the *Location* semantic type. The point is that WordNet synsets are organized into a strict taxonomy (of hyponymy relations) which in some cases makes it difficult to group the related concepts. In particular, the learned distribution of senses we get for the frame-FEs assigned *Location* semantic type in FrameNet, includes, among others, *house#n#1*, *room#n#1* and *area#n#4* synsets which in WordNet are classified as *structure#n#1* (a hyponym of *artifact#n#1*) but not as *location#n#1*. Another synset in the acquired distribution is *organization#n#1* (a hyponym of *social_group#n#1*), which was extracted from the examples in which the name of a university, company, etc. was used to denote a location (e.g. "at Harvard University", "inside the school"). As a result, in the repository of senses for the *Location* semantic type, *location#n#1* has a weight of 26%, whereas there are also *artifact#n#1* with a weight of 40%, *event#n#1* and *social_group#n#1* with weights of 5% and 6%, respectively.

Table 3 presents the distribution of senses for some of the existing FrameNet semantic types, among which are *Message* and *Artifact*, which are mapped to WordNet synsets. The distribution of senses for the *Message* semantic type explains the low number of matches in Table 2, whereas the senses acquired for the *Artifact* semantic type provide more detailed constraints for the FE filler compared to a synset assigned to it in FrameNet. Table 4 in its first 3 lines reports the examples of sense distributions acquired for the frame-FE pairs whose semantic types are linked to WordNet. These either provide more details on possible senses of a role filler (*Buildings-Building* and *Artifact-Artifact* in lines 1 and 3), or correct the existing FrameNet semantic type (*Collaboration-Partner_1* in line 2).

The second evaluation step concerns analyzing the rest of the sense repository, namely, 94% of the frame-FE pairs for which no mapping to WordNet sysnset is provided by FrameNet lexicographers. The last 3 lines of Table 3 are the examples of the acquired senses for the semantic types which are assigned a type but are not mapped to any WordNet synset in FrameNet, whereas in Table 4 one finds 4 examples of senses acquired per frame-FE pair (assigned a FrameNet semantic type but not a WordNet synset). Such frame-FE pairs constitute around 40% (1740 pairs) of the sense repository. As the examples in Tables 3 and 4 illustrate, the acquired senses can be either compared with the existing FrameNet semantic types (e.g. there is a clear match between *Body_part* semantic type and *body_part#n#1* WordNet synset), or specialize them (compare e.g. the distribution of senses for the *Sentient* semantic type for *Wearing-Wearer* and *Telling-Addressee* FEs in Table 4).

Finally, for 54% (2389) of the frame-FE pairs of our sense repository FrameNet lexicographers do not provide any semantic type, and thus the senses we have acquired (see the last 4 rows of Table 4 for the examples) constitute a completely new part of the resource.

To summarize, the presented repository of senses allows us to specialize/complement and in some cases correct/replace the semantic constraints on FEs fillers available in FrameNet 1.5. The acquired resource is beneficial for a number of NLP tasks, and in particular, for frame annotation tools, where one can potentially increase the annotation precision by comparing the WordNet synsets of candidate FE fillers, which can be acquired following the same methodology as in Section 3, with the senses available in the repository.

7. CONCLUSIONS

In this paper we have presented an extension of FrameNet with a sense repository in which we map the fillers of the semantic roles to WordNet synsets. The methodology we present uses interlinked web-based resources such as Wikipedia, YAGO and BabelNet. We have shown how the repository can be modelled as an intermediate level between FrameNet and WordNet RDF/OWL representations, and have directly evaluated the part of the resource for which such an evaluation is feasible.

The main ongoing developments of this work are the indirect evaluation of the resource in which the precision of semantic frame parsers is supposed to be improved and the finalization of the OWL version of the repository. Among the future work directions, we plan to acquire binary constraints for FE fillers, as well as to exploit the existing relations between FEs in FrameNet in order to improve the quality of the acquired constraints.

8. ACKNOWLEDGMENTS

The research leading to these results has received funding from the Copilosk project (`http://copilosk.fbk.eu`), a Joint Research Project under Future Internet – Internet of Content program of Fondazione Bruno Kessler.

Frame-FE	FN semtype	#ex	Learned synsets	%mtch
Buildings-Building	Artifact (artifact#n#1)	117	building#n#1	29%
			housing#n#1	42%
Collaboration-Partner_1	Human (person#n#1)	61	living_thing#n#1	40%
			organization#n#1	19%
			region#n#3	15%
Artifact-Artifact	Physical_entity (entity#n#1)	48	engineering#n#2	92%
Wearing-Wearer	Sentient	372	person#n#1	77%
Telling-Addressee	Sentient	313	person#n#1	51%
			political_unit#n#1	12%
Experience_bodily_harm-Body_part	Body_part	148	external_body_part#n#1	39%
			joint#n#1	34%
Travel-Goal	Goal	114	location#n#1	45%
			system#n#2	11%
			whole#n#2	17%
Make_noise-Sound_source	—	777	instrumentality#n#3	14%
			organism#n#1	51%
Manipulation-Entity	—	343	artifact#n#1	18%
			body_part#n#1	31%
			living_thing#n#1	22%
Leadership-Governed	—	297	location#n#1	34%
			social_group#n#1	25%
			system#n#2	13%

Table 4: Examples of acquired senses for frame-FE pairs

9. REFERENCES

[1] A. Burchardt, N. Reiter, S. Thater, and A. Frank. A semantic approach to textual entailment: System evaluation and task analysis. In *Proceedings of Pascal RTE-3 Challenge*, 2007.

[2] B. Coppola, A. Gangemi, A. M. Gliozzo, D. Picca, and V. Presutti. Frame Detection over the Semantic Web. In *Proceedings of ESWC-2009*, pages 126–142, 2009.

[3] S. Cucerzan. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of EMNLP-CoNLL*, pages 708–716, 2007.

[4] K. Erk. A simple, similarity-based model for selectional preferences. In *Proceedings of the 45^{th} Annual Meeting of the Association for Computational Linguistics*, pages 216–223, 2007.

[5] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[6] D. Gildea and D. Jurafsky. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288, 2002.

[7] C. Giuliano, A. M. Gliozzo, and C. Strapparava. Kernel methods for minimally supervised wsd. *Computational Linguistics*, 35(4):513–528, 2009.

[8] E. Laparra and G. Rigau. Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm. In *Proceedings of RANLP-2009*, pages 208–213, 2009.

[9] S. Narayanan, C. J. Fillmore, C. F. Baker, and M. R. L. Petruck. FrameNet Meets the Semantic Web: A DAML+OIL Frame Representation. In *Proceedings of the The 18^{th} National Conference on Artificial Intelligence*, 2002.

[10] R. Navigli and S. P. Ponzetto. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48^{th} Annual Meeting of the Association for Computational Linguistics*, pages 216–225, 2010.

[11] A. G. Nuzzolese, A. Gangemi, and V. Presutti. Gathering lexical linked data and knowledge patterns from framenet. In *Proceedings of K-CAP 2011*, pages 41–48, 2011.

[12] E. Ovchinnikova, L. Vieu, A. Oltramari, S. Borgo, and T. Alexandrov. Data-Driven and Ontological Analysis of FrameNet for Natural Language Reasoning. In *Proceedings of LREC-2010.*.

[13] J. Ruppenhofer, M. Ellsworth, M. R. Petruck, C. R. Johnson, and J. Scheffczyk. *FrameNet II: Extended Theory and Practice*. 2010.

[14] J. Ruppenhofer, C. Sporleder, R. Morante, C. F. Baker, and M. Palmer. SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proceedings of SemEval-2010*.

[15] J. Scheffczyk, C. F. Baker, and S. Narayanan. Ontology-based reasoning about lexical resources. In *Proceedings of OntoLex-2006 Workshop*, 2006.

[16] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[17] D. Shen and M. Lapata. Using semantic roles to improve question answering. In *Proceedings of EMNLP-CoNLL*, pages 12–21, 2007.

[18] L. Shi and R. Mihalcea. Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic parsing. In *Computational Linguistics and Intelligent Text Processing*, pages 100–111, 2005.

[19] F. M. Suchanek, G. Kasneci, and G. Weikum. Yago: a core of semantic knowledge. In *Proceedings WWW-2007*, pages 697–706, 2007.

[20] S. Tonelli and D. Pighin. New Features for FrameNet - WordNet Mapping. In *Proceedings of CoNLL-2009*, pages 219–227, 2009.