Investigating the Semantics of Frame Elements

Sara Tonelli, Volha Bryl, Claudio Giuliano, Luciano Serafini

Fondazione Bruno Kessler, Italy
{satonelli|bryl|giuliano|serafini}@fbk.eu

Abstract. Compared to other existing semantic role repositories, FrameNet is characterized by an extremely high number of roles or *Frame Elements* (FEs), which amount to 8,884 in the last resource release. This represents an interesting issue to investigate both from a theoretical and a practical point of view. In this paper, we analyze the semantics of frame elements by automatically assigning them a set of synsets characterizing the typical FE fillers. We show that the synset repository created for each FE can adequately generalize over the fillers, while providing more informative sense labels than just one generic semantic type. We also evaluate the impact of the enriched FE information on a semantic role labeling task, showing that it can improve classification precision, though at the cost of lower recall.

1 Introduction

FrameNet [29] is one of the most important semantic resources encoding information about situations, the *frames*, and participants, the semantic roles, also called *frame elements* (FEs). Although Filllmore's original formulation of frames included only six generic roles [15], FrameNet is currently characterized by an extremely high number of roles, which amount to 8,884 in the last resource release (version 1.5). This represents an interesting issue to investigate both from a theoretical and a practical point of view, also in the light of its ontological consistency [25].

FrameNet developers assume that events and situations in the world, corresponding to frames, should be described through highly specific roles rather than few, generic ones. Despite this, around 54% of the roles in FrameNet 1.5. have been assigned a semantic type label (40 labels in total), in a partial attempt to identify some common traits among different FEs and provide semantic constraints on FE fillers. For instance, the semantic type *Sentient* was assigned to the *Agent* FE. However, this information does not cover the whole FE set (only 54% of the FEs have a semantic type) and is often very high-level (consider e.g. *Physical_entity* type coupled with *Instrument* FE). Therefore, it could hardly be used for large-scale semantic analysis.

The number of frame elements makes the semantic role labeling (SRL) very challenging, since for many FEs only few training examples are available. In order to improve the performance of SRL systems, it would be important to generalize over the role fillers, while preserving specific FE information.

In this work, we present a generalization strategy over FE lexical fillers that relies on extracting information from a combination of existing Semantic Web resources, namely Wikipedia, WordNet [14] and YAGO [31]. Specifically, we devise an approach that takes advantage of the strengths of each of these resources: Wikipedia is currently the most extensive sense repository and shows an excellent coverage of named entities. Furthermore, robust tools for Wikipedia-based word sense disambiguation (WSD) allow for the disambiguation of the role fillers. WordNet, instead, has a better coverage of nominal entities and its hierarchical structure can be exploited to compute similarity measures between the fillers. Finally, the YAGO ontology combines knowledge about Wikipedia individuals and WordNet taxonomy structure, thus providing a structured description of Wikipedia concepts, which includes the characterization of a concept in terms of WordNet synsets. We show that these three resources are interoperable and can be successfully connected to enrich the frame element repository with additional semantic information. Furthermore, we present an experiment in which we integrate this knowledge into a postprocessor that checks the consistency of FE labels, as annotated by a SRL system, and discards annotations that are unlikely, given our enriched FE repository.

The paper is structured as follows: in Section 2 we present related work on FE analysis and selectional preferences. In Section 3 the workflow for the creation of the sense repository is detailed. In Section 4 we analyze a subpart of the repository, providing a qualitative evaluation of the resource. In Section 5 a task-based evaluation is presented, in which the sense repository is exploited to improve the performance of a SRL system. We draw some conclusions and detail future work in Section 6.

2 Related work

FrameNet structure has undergone several revisions and consistency studies. However, they have been mainly focused on frames and on frame-to-frame relations [25], rather than on its large repository of frame elements. Past attempts to link FrameNet and WordNet concerned mainly finding for each frame a set of corresponding synsets, disregarding FE information [11, 19, 33]. An exception is the work by [2], in which the authors assign a synset to FE fillers based on latent semantic analysis in order to build lexical patterns to be included in a domain ontology. Also [9] present a methodology to enrich frames with argument restrictions provided by a super-sense tagger and domain specialization. The authors employ an off-the-shelf parser, whose tagset comprises 26 labels for nominal FE fillers. Our approach is different from the one presented in [9] in that we devise our own strategy for sense assignment and apply a different generalization method.

Our analysis of the semantics of frame elements leans on past research on *Selectional Preferences* (SPs), which describe typical fillers of a predicate's argument. However, while the original notion of selectional preferences referred to predicates [17, 34], we adopt here a slightly different definition, by considering SPs *on a frame basis*, similar to [12]. In other words, we assume that SPs are shared among all predicates belonging to the same frame.

SPs have been used in different semantically-based NLP tasks such as word sense disambiguation [21], pseudo-disambiguation [13] and semantic role labelling [16, 36]. Starting from [28], most approaches proposed for the acquisition of selectional preferences rely on a corpus-based methodology based on two steps: first, all argument heads are *extracted* from a reference corpus, and then such heads are used to model some selectional preferences over the arguments by *generalizing* to other similar words.

In the generalization step, most relevant models of selectional preferences can be grouped into two classes: *i)* distributional models [27, 12, 4], which rely on word co-occurrence, and *ii)* semantic hierarchy-based models [28, 1, 8], which generalize over seen headwords using an ontology or a semantic hierarchy, usually WordNet. While the former are particularly suited to domain-specific applications, because they can be easily adapted using domain-specific generalization corpora, semantic hierarchy-based models can make predictions also for infrequent words with good accuracy, given that they are included in the hierarchy. In this paper, we focus on the second model for acquiring selectional preferences for FrameNet FEs and building a repository of preferences on a frame basis. We encode these preferences as a list of synsets assigned to each FE.

A first study aimed at including selectional preferences in frame argument classification was presented by [16]. The authors generalize over seen heads of nominal arguments in three ways: with automatic clustering, by bootstrapping training data from unlabeled texts and by ascending the WordNet type hierarchy until reaching a synset for which training data are available. The three approaches show similar precision but a lower coverage for the WordNet-based model. [12] proposes a methodology to compute selectional preferences for semantic roles by using the BNC as a generalization corpus to compute corpus-based similarity metrics between a candidate role headword and the seen headwords. This distributional model achieves smaller error rates than Resnik's WordNet-based model and EM-based clustering models, although it shows lower coverage than EM-based models. The approach is then extended in [13].

[36] show that the integration of SPs in a state-of-the-art SRL system yields statistically significant improvement in classification accuracy. However, our results are not directly comparable because we adopt FrameNet paradigm for semantic role labelling, while [36]'s work is based on PropBank-style roles. Besides, they integrate SPs in the system in the form of features and individual class predictors, which we see as part of our future work.

Most works on selectional preferences that implement WordNet-based models currently follow Resnik's formulation [28], which estimates the argument preference strength without performing word sense disambiguation. Attempts have been made to tackle ambiguity in the acquisition of selectional preferences, for example using Bayesian networks [7], but it still remains an open problem, since existing WordNet-based WSD systems cannot be easily applied to new domains and do not handle named entities. We claim that disambiguation is a necessary step in order to perform an accurate generalization over FEs fillers, but we introduce a new methodology that first disambiguates these fillers by assigning a Wikipedia page and then relies on WordNet and YAGO for generalization.

3 Creation of a FE repository with selectional preferences

The aim of this work is to create a repository in which, for each tuple $(frame, FE_f)$, where FE_f is a frame element of $frame$, a set of senses with a relevance score is listed.

We extract these senses from WordNet because it provides high-quality additional information (e.g. relations, synonyms, etc.) and its hierarchy can be further exploited for computing similarity between FE fillers (see Section 5).

The repository is built based on the following steps:

- Disambiguation of FE fillers from the FrameNet corpus using a Wikipedia-based WSD system
- Linking each disambiguated FE filler to a WordNet synset
- If a synset was not assigned to a FE filler, mapping it to YAGO
- Creation of a synset repository for each $(frame, FE_f)$ by assigning a relevance score to each synset

The four steps are detailed in the following subsections.

3.1 Step 1: Disambiguation of FE fillers

In order to collect a sense repository for each FE filler, we first disambiguate each filler by assigning a Wikipedia page W. We employ *The Wiki Machine*, a kernel-based WSD system (details on the implementation are reported in [32]), which has been trained on the Wikipedia dump of March 2010[1]. Since FE fillers can be both common nouns and named entities, we needed a WSD system that performs satisfactorily on both nominal types. A comparison with state-of-the-art system *Wikipedia Miner* [23] on the ACE05-WIKI dataset [3] showed that *The Wiki Machine* achieves a good performance on both types (0.76 F1 on named entities and 0.63 on common nouns), while *Wikipedia Miner* has a poorer performance on the second noun type (0.76 and 0.40 F1, respectively). These results were confirmed also in a more recent evaluation [22], in which *The Wiki Machine* achieved the highest F1 compared to an ensemble of academic and commercial systems such as DBpedia Spotlight, Zemanta, Open Calais, Alchemy API, and Ontos.

Disambiguation is performed on each annotated sentence from the FrameNet database 1.5 [29], including also the documents provided with continuous annotation. The system applies an 'all word' disambiguation strategy, in that it tries to disambiguate each word (or multiword) in a given sentence. Then, we match each disambiguated term with the original frame annotation $(frame, FE_f)$ and, in case the term (partially) matches a string corresponding to a FE, we assume that one possible sense of $(frame, FE_f)$ is represented in Wikipedia through W. The WSD system also assigns a confidence score to each disambiguated term. This confidence is higher in case the words occurring in the same context of the disambiguated term show high similarity, because the system assumes that disambiguation is likely to be more accurate.

We show in Fig. 1 the Wikipedia pages (and confidence score) that the WSD system associates with the sentence 'Sardar Patel was assisting Gandhiji in the Salt Satyagraha with great wisdom', an example sentence for the ASSISTANCE frame originally annotated with four FEs, namely *Helper*, *Benefited_party*, *Goal* and *Manner*.

Since Wikipedia is a repository of concepts, which are usually expressed by nouns, we are able to disambiguate only nominal fillers. This is in line with past research on selectional preferences, that are usually limited to nominal arguments [12, 16].

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

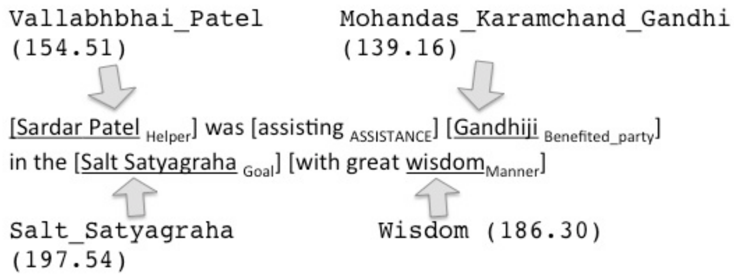

Fig. 1. Disambiguation example with confidence score

3.2 Step 2: Linking with WordNet

Although Wikipedia has an extensive coverage of senses, it is a semi-structured resource, whose concept organization is not rigorous because it is continuously growing and lacks a design a priori. Therefore, after the disambiguation step, we map each Wikipedia page W to a WordNet synset, so that we can better generalize over the lexical fillers using WordNet hierarchy.

For this purpose, we use BABELNET [24], 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 W assigned to $(frame, FE_f)$ with a synset. In this way, we devise a sort of Wikipedia-detour to WordNet-based WSD, because we obtain for each $(frame, FE_f)$ a list of related synsets (henceforth $Syns(FE_f)$). This approach has several advantages: since our WSD system is trained on the whole Wikipedia, it can potentially disambiguate (i.e. assign a Wikipage) to every term or multiword. 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*.

3.3 Step 3: Enrichment through YAGO

Since Wikipedia contains much more concepts (i.e. senses) than WordNet, not all pages W associated with the FE fillers can be mapped to a WordNet synset. Furthermore, BABELNET covers only a subset of all existing Wikipedia pages. As reported in Table 1, 226,520 FE fillers were disambiguated by assigning a Wikipedia page with non-zero confidence score, but 29,924 of them were not linked to a synset because the pages were not found in BABELNET.

In order to cope with this problem, we exploited YAGO [31], an automatically created large-scale ontology, which combines WordNet taxonomy with knowledge about individuals extracted from Wikipedia. Specifically, most of WordNet synsets become classes in YAGO and *SubClassOf* relation between classes is derived from hyponymy relation in WordNet. In turn, Wikipedia categories are mapped to WordNet synsets, that is, they are added to YAGO as subclasses of aforementioned WordNet-derived classes. Thus, given a Wikipedia page URL, one can query the YAGO ontology to obtain a list of WordNet synsets corresponding to its Wikipedia categories. Note that, differently from Step 2 where BABELNET provided us with at most one WordNet synset per Wikipedia page, multiple synsets corresponding to multiple page categories are obtained.

At the end of this step we were able to acquire WordNet synsets for 21,505 FE fillers, thus increasing the linking by 9%. To access the ontology, we used Java API for quering YAGO in SPARQL, which is provided by the YAGO team and works with YAGO dumps in TDB format[2].

3.4 Step 4: Creation of the repository

In the final FE repository, we do not want to simply list all synsets acquired for each $(frame, FE_f)$. The key idea is rather to exploit WordNet taxonomy to generalize over the extracted synsets. In particular, for each $(frame, FE_f)$ pair a tree of senses was created based on WordNet hyponymy relation, where a node corresponds to a synset s and stores the number of examples linked to s or its hyponym. Then, the distribution of senses for $(frame, FE_f)$ is the lowest (starting from the most specific synsets) tree level, in which the nodes containing at least 10% of examples cover in total at least 60% of examples. In this way, the most representative among the most specific synsets are selected.

For each synset which is finally selected and included in the repository, we also compute the conditional probability $P(s|FE_f)$ as:

$$P(s|FE_f) = \frac{Count(s, FE_f)}{Count(FE_f)} \quad (1)$$

The selected synsets often provide a better characterization of a FE than its original description in FrameNet. For example, the *Undergoer* FE in the BEING_ROTTED frame was described as 'the organic matter that has decayed' by FrameNet lexicographers. However, the *Undergoer* instances in FrameNet annotated sentences include a variety of fillers such as 'house', 'mural', 'nest', 'lung', 'butter', 'reptile', etc. These are all well represented in the synset repository acquired for *Undergoer*:[3]

substance#n#1, 6, 0.27
artifact#n#1, 4, 0.18
body_part#n#1, 3, 0.14
natural_object#n#1, 3, 0.14
food#n#1, 2, 0.09
living_thing#n#1, 2, 0.09

[2] `http://www.mpi-inf.mpg.de/yago-naga/`
[3] Note that the synset entries are in the form *synset, occurrences,* $P(s|FE_f)$.

substance#n#8, 1, 0.04
matter#n#3, 1, 0.04

For more examples the reader is referred to Section 4, to [5] and to the online version of the repository: its text version is available online[4], while the creation of the RDF/OWL version is the work in progress.

We report in Table 1 some statistics about the data processed in the four steps described above. Note that we consider only the fillers for which the WSD system provides a non-zero confidence value. Taking zero confidence links into account would increase the coverage by 12.5%, but it would also increase the amount of noise and, consequently, decrease the quality of the acquired senses.

FE fillers linked to a Wikip. (Step1)	226,520
FE fillers linked to a synset (Step2)	196,596 (87%)
FE fillers linked to a synset through YAGO (Step3)	21,505 (9%)
$(frame, FE_f)$ pairs in the final repository (Step4)	3,847
Avg. synsets for $(frame, FE_f)$	6

Table 1. Statistics about the extracted data.

4 Qualitative evaluation

The acquired repository of senses can be divided into three parts. To a small portion of frame elements a semantic type explicitly mapped to a WordNet synset was assigned by FrameNet lexicographers. There are 242 such $(frame, FE_f)$ pairs, approximately 6% of the repository. For 1573 pairs (41%) a FrameNet semantic type was assigned but no mapping to WordNet was provided. For the remaining 53% of the pairs, no semantic description of a role filler is available in FrameNet. In the following, we present a qualitative evaluation of three repository parts, done manually for those pairs that most frequently appear in the FrameNet corpus. Some more details on the qualitative evaluation of the first version of the sense repository can be found in [5].

FrameNet semantic type is assigned and linked to WordNet: We have considered 88 $(frame, FE_f)$ pairs, each having greater or equal than 50 examples in the FrameNet corpus. For this small portion of the repository, a straightforward evaluation is possible, comparing the acquired synsets with those assigned to the FE semantic types in FrameNet. Specifically, we count as correct the synsets assigned to $(frame, FE_f)$ in

[4] https://dkm.fbk.eu/index.php/FrameNet_extension:_repository_of_senses

the repository if they match or are hyponyms of the synset associated with the semantic type of $(frame, FE_f)$ in FrameNet.

18 out of the 88 pairs considered were originally assigned the *Human* semantic type (and *person#n#1* synset) in FrameNet. For 16 of them, the average matching score is around 96%. The other 2 examples are (COLLABORATION, *Partners*) and (COLLABORATION, *Partner_2*) (100 examples per pair). Let us consider the latter case, *Partner_2*: for this frame element, the acquired distribution of senses includes not only *person#n#1* synset, but also *social_group#n#1, organization#n#1, company#n#1* and some others, which not only justifies the low matching score but allows for a more complete description of possible FE fillers. In some other cases, one sense has been acquired for a FE, being more specific than the corresponding FrameNet semantic type. This would make a replacement possible between the original FE semantic type and the newly acquired synset. For instance, the pair (WEAPON, *Weapon*) (165 examples) was assigned the semantic type *artifact#n#1* in FrameNet, while we have acquired the *weapon#n#1* synset in 98% of the examples. High matching scores are obtained for the *artifact#n#1* and the *body_part#n#1* semantic types. For a number of types (e.g. *location#n#1*) the matching score is low, while the suggested distribution includes, in addition to *location#n#1*, the *structure#n#1* (hyponym of *room#n#1* and *house#n#1*), *organization#n#1* and *event#n#1* synsets.

FrameNet semantic type is assigned but not linked to WordNet: For this part of the repository, evaluation is more complex because we do not rely on gold synsets. Therefore, we perform a post-hoc analysis and count as correct the synsets assigned to $(frame, FE_f)$ in the repository if they comply with the semantic type associated with $(frame, FE_f)$ in FrameNet, according to human judgment. We have considered 69 $(frame, FE_f)$ pairs of this group, each having more than 400 examples in the FrameNet corpus. 64 of these pairs were assigned the *Sentient* semantic type in FrameNet, and for them the average matching score is 93%: more specifically, for 63 pairs the score is greater than 80%, and for 30 pairs greater or equal than 95%. One exception with 76% match is the (KILLING, *Victim*) pair (520 examples), where another 12% of the occurrences are learnt to be *animal#n#1* and *group#n#1* (e.g. ethnical groups). Other semantic types for these 69 pairs are *Goal* (3 pairs), *Path* and *State_of_affairs* (1 pair each). To give an example, for (SELF_MOTION, *Goal*) (FrameNet semantic type is *Goal*) the distribution of senses includes *structure#n#1* (hyponym of *room#n#1, area#n#1, building#n#1*), *location#n#1* (hyponym of e.g. *area#n#4*), *vehicle#n#1, event#n#1, social_group#n#1, instrumentality#n#1* (hyponym of e.g. *furniture#n#1*).

No semantic type associated with FE in FrameNet: This is the largest part of the repository and the most problematic from the evaluation point of view. Also in this case, we perform a post-hoc analysis, and accept as correct the synsets assigned to $(frame, FE_f)$ if they comply with the FE_f definition in FrameNet. We have considered 31 $(frame, FE_f)$ pairs, each having more than 400 examples in the FrameNet corpus. For 16 pairs, the *person#n#1* synset was suggested as the most frequent role filler type, with 13 pairs having the average matching score of 91%. For other pairs, like, for instance, (MAKE_NOISE, *Sound_source*), the distribution includes, in addition to *person#n#1*, *animal#n#1, instrument#n#1, atmospheric_phenomenon#n#1*, etc. Another example of a semantic role description is that acquired for *(*LEADERSHIP, *Governed*), whose distribution of senses includes *person#n#1, location#n#1, socail_group#n#1* and *structure#n#1* synsets.

Although the presented analysis is limited to a small set of $(frame, FE_f)$ pairs, we still can conclude that in many cases the acquired senses are not only correct, but also provide a much more accurate semantic description of possible FEs fillers than FrameNet semantic types.

5 Task-based evaluation

An issue we want to address with the present study is the possibility to exploit the information on FE fillers collected so far in a real NLP task. Specifically, we want to assess if, despite the possible mistakes introduced in our pipeline through the linking steps (from Wikipedia to WordNet through BABELNET and YAGO), the sense repository can be used as it is to improve the performance of existing NLP systems.

We test this hypothesis in a semantic role labelling (SRL) task. For each $(frame, FE_f)$ pair, we employ the assigned synset list as selectional preferences and apply them to the output of a SRL system, in order to accept or discard the FE labels based on such preferences.

The evaluation process comprises the following steps:

1. Annotate unseen text with a frame-based SRL system
2. For each $(frame, FE_f)$ assigned by the system, disambiguate the lexical filler by assigning a Wikipedia page
3. Map the Wikipedia page to a WordNet synset s_0 using BABELNET and YAGO (see Sections 3.2 and 3.3.)
4. Compute a similarity score between s_0 and the synsets $Syns(FE_f)$ previously associated with $(frame, FE_f)$
5. If the similarity score is above a certain threshold, the FE_f label assigned by the SRL system is accepted, otherwise it is discarded.

While the widely used WordNet-based model proposed by [28] estimates selectional preferences for an argument filler by splitting the filler frequency equally among all synsets containing it, we can now take advantage of the outcome of the disambiguation step, and apply a model for selectional preferences directly to synsets. Another advantage is that we generalize over fillers based on the synsets, therefore we can admit also unseen lexical fillers (i.e. terms that were not included in the FrameNet data used to create the synset repository).

5.1 Similarity function

In order to compute the similarity score mentioned in Section 5 at (4), we implement a function that computes the similarity measure between s_0 and each $s \in Syns(FE_f)$ and multiplies it by the probability of s $P(s|FE_f)$ as described in Eq. 1. The highest product obtained during this comparison corresponds to the similarity score $S_{FE_f}(s_0)$

between s_0 and the most similar s in the sense repository. The probability value should boost the similarity between s_0 and s, if s was frequently observed for FE_f in the FrameNet corpus:

$$S_{FE_f}(s_0) = arg \max_{s \in Syns(FE_f)} sim(s_0, s) \times P(s|FE_f) \tag{2}$$

5.2 Similarity measures

The similarity metrics instantiating the sim function of Equation 2 are based on existing implementations of WordNet-based metrics of semantic *similarity* implemented in the *WordNet::Similarity* library [26][5]. In particular, we test 4 different measures that are either based on information content or on path length.

The information-based measures of *similarity* rely on the *information content* of the least common subsumer (LCS) of two synsets, given the sense-tagged corpus SemCor[6] as reference source for information. The measures we use are *jcn* [18] and *lin* [20]. *jcn* and *lin* are based on the sum of the information content of the two synsets being considered. The *jcn* measure corresponds to the difference between the sum and the information content of LCS, while the *lin* measure is equal to twice the information content of LCS divided by the sum of the information content of each input synset.

As for path-length based measures of semantic similarity, they rely on the assumption that the length of the path between a pair of synsets in the WordNet taxonomy can be used to compute their similarity. In particular, *path* corresponds to the inverse of the shortest path between two synsets and *wup* [35] divides the depth of the synsets' LCS by the sum of the depth of the single synsets.

5.3 Evaluation results

Our task-based evaluation relies on a postprocessor that, given the output of a FrameNet-based SRL system, assesses for each annotated FE if it is correct or not. Since our post-processing algorithm requires a boolean true/false judgment, whereas each of the similarity measures we applied returns a numerical value of similarity, we try different *acceptance thresholds* for each measure in order to estimate the best cutoff value for accuracy optimization.

We use the test set released for SemEval 2010, Task 10 [30], which we first annotate using the SEMAFOR frame semantic parser [10] in the standard setting. Then, we disambiguate each nominal filler of an annotated FE by assigning a Wikipedia link and then a WordNet synset. Finally, we compute a similarity score between the disambiguated filler and the sense repository previously associated with the given FE by applying the sim function. We test the similarity function described in Section 5.1 combined with the four WordNet metrics. Further, we experiment with different acceptance thresholds.

[5] For an overview see [6]

[6] `http://www.cse.unt.edu/~rada/downloads.html`

We report in Table 2 the results of this evaluation. SEMAFOR performance[7] is compared with the *Exact match* score, which is obtained by retaining only the FE labels assigned by SEMAFOR whose associated synset appears in the sense repository for the given FE. For instance, if the token *house* was annotated by SEMAFOR as belonging to the LOCATING frame and having the *Location* FE, we first disambiguate it by assigning the *house#n#1* synset and consider it correct only if we find *house#n#1* in the sense repository for *Location*, LOCATING. The *Exact match* implements a basic version of our strategy for FE selection, in that no similarity function is applied. In Table 2 we report also the performance achieved by applying to SEMAFOR output our strategy for FE validation based on selectional preferences. The acceptance thresholds of the four WordNet similarity metrics are set in order to maximize precision.

	Precision	**Recall**	$\mathbf{F_1}$
SEMAFOR	0.446	0.492	0.468
Exact match	0.467	0.371	0.414
jcn	0.469	0.412	**0.439**
lin	0.474	0.375	0.419
wup	0.476	0.399	0.434
path	**0.478**	0.399	0.435

Table 2. Comparative evaluation of SEMAFOR performance and FE selection strategy

5.4 Results

Table 2 shows that the information collected on the semantics of FEs can be used to improve SRL precision. The best strategy for FE selection is based on path distance between the current synset and previously observed synsets. In general, however, our approach does not seem to be effective in improving SRL performance, because none of the proposed strategies yields an improvement over SEMAFOR F_1. The information collected on the semantics of FEs should probably be integrated into the system in the form of features in order to be fully exploited and interact with the existing syntactic and lexical features. Although the system is available as an open source project,[8] however, it is extremely difficult to manipulate its complex architecture and extend it in order to include new features. This extension is left to future work.

6 Conclusions

In this paper, we have investigated how to cast the semantics of frame elements without relying on generic semantic type labels. Specifically, we have created a repository in

[7] SEMAFOR performance is lower than the one reported in the SemEval task [30], because the system release used for this evaluation was not trained on SemEval training data. Also our sense repository does not include these training data.

[8] `http://code.google.com/p/semafor-semantic-parser/`

which each $(frame, FE_f)$ pair has been associated with a list of synsets that are representative of the fillers' senses. The repository has been obtained by mapping the FE fillers to WordNet synsets through Wikipedia, BABELNET and YAGO.

The repository has been manually inspected, showing that, in some cases, the acquired synsets provide a more accurate and multifaceted semantic description of possible FEs fillers than FrameNet semantic types. When using the sense repository to improve the performance of a SRL system, however, recall is a main issue, while precision achieves some improvement.

In the future, we plan to further improve the consistency of the sense repository by performing a semi-automatic check of the acquired synsets. In addition, we will investigate the possibility to merge the synsets of different $(frame, FE_f)$ pairs, if the FEs share the same label, such as (ABUSING, *Victim*), (ATTACK, *Victim*), (OFFENSES, *Victim*).

As for SRL, we will model the acquired knowledge on typical FE fillers as features, so that it can be integrated directly into a supervised SRL system. This will allow for a more thorough task-based evaluation.

References

1. Abe, N., Li, H.: Learning word association norms using tree cut pair models. In: Proceedings of the 10^{th} International Conference on Machine Learning. pp. 3–11. Bari, Italy (1996)
2. Basili, R., Croce, D., De Cao, D., Giannone, C.: Learning Semantic Roles for Ontology Patterns. In: Proceedings of the Workshop on Natural Language Processing and Ontology Engineering (NLPOE 2009), held in conjunction with the Conference on Web Intelligence (WI/IAT-09). Milan, Italy (2009)
3. Bentivogli, L., Forner, P., Giuliano, C., Marchetti, A., Pianta, E., Tymoshenko, K.: Extending English ACE 2005 Corpus Annotation with Ground-truth Links to Wikipedia. In: 23rd International Conference on Computational Linguistics. pp. 19–26 (2010)
4. Bergsma, S., Lin, D., Goebel, R.: Discriminative learning of selectional preferences from unlabeled text. In: Proceedings of the 13^{th} Conference on Empirical Methods in Natural Language Processing. pp. 59–68. Honolulu, HI, US (2008)
5. Bryl, V., Tonelli, S., Giuliano, C., Serafini, L.: A Novel FrameNet-based Resource for the Semantic Web. In: In Proceedings of ACM Symposium on Appliced Computing, Technical Track on The Semantic Web and Applications (2012)
6. Budanitsky, A., Hirst, G.: Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics 31(1), 13–48 (2006)
7. Ciaramita, M., Johnson, M.: Explaining away ambiguity: Learning verb selectional preference with Bayesian networks. In: Proceedings of the 18^{th} International Conference on Computational Linguistics (COLING). pp. 187–193. Saarbrücken, Germany (2000)
8. Clark, S., Weir, D.: Class-based probability estimation using a semantic hierarchy. In: Proceedings of the 2^{nd} Annual Meeting of the North American Chapter of the Association for Computational Linguistics. pp. 95–102. Pittsburgh, PA, US (2001)
9. Coppola, B., Gangemi, A., Gliozzo, A.M., Picca, D., Presutti, V.: Frame Detection over the Semantic Web. In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009). pp. 126–142. Heraklion, Crete, Greece (2009)
10. Das, D., Schneider, N., Chen, D., Smith, N.A.: Probabilistic frame-semantic parsing. In: Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference. Los Angeles, CA, USA (2010)

11. De Cao, D., Croce, D., Pennacchiotti, M., Basili, R.: Combining Word Sense and Usage for modeling Frame Semantics. In: Proceedings of STEP 2008. Venice, Italy (2008)
12. Erk, K.: A simple, similarity-based model for selectional preferences. In: Proceedings of the 45^{th} Annual Meeting of the Association for Computational Linguistics (ACL-07). pp. 216–223. Prague, Czech Republic (2007)
13. Erk, K., Padó, S., Padó, U.: A flexible, corpus-driven model of regular and inverse selectional preferences. Computational Linguistics 36(4), 723–763 (2010)
14. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press (1998)
15. Fillmore, C.J.: The Case for Case. In: Bach, E., Harms, R.T. (eds.) Universals in Linguistic Theory, pp. 1–88. Holt, Reinhart and Winston, New York (1968)
16. Gildea, D., Jurafsky, D.: Automatic labeling of semantic roles. Computational Linguistics 28(3), 245–288 (2002)
17. Hirst, G.: Semantic Interpretation and the Resolution of Ambiguity. Cambridge University Press (1987)
18. Jiang, J.J., Conrath, D.W.: Semantic similarity based on corpus statistics and lexical taxonomy. In: Proceedings of the International Conference on Research in Computational Linguistics (ROCLING X). pp. 19–33. Taiwan (1997)
19. Laparra, E., Rigau, G.: Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm. In: Proceedings of RANLP-2009. pp. 208–213 (2009), `http://www.aclweb.org/anthology/R09-1039`
20. Lin, D.: Automatic retrieval and clustering of similar words. In: Proceedings of the 36^{th} Annual Meeting of the Association for Computational Linguistics. pp. 768–774. Montreal, Quebec, Canada (1998)
21. McCarthy, D., Carroll, J.: Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. Computational Linguistics 29(4), 639–654 (2003)
22. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: DBpedia Spotlight: Shedding Light on the Web of Documents. In: In Proceedings of I-SEMANTICS, the 7th Conference on Semantic Systems (2011)
23. Milne, D., Witten, I.H.: Learning to link with Wikipedia. In: CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge management. pp. 509–518. ACM, NY, USA (2008)
24. Navigli, R., Ponzetto, S.P.: Babelnet: Building a very large multilingual semantic network. In: Proceedings of the 48^{th} Annual Meeting of the Association for Computational Linguistics. pp. 216–225. Uppsala, Sweden (2010)
25. Ovchinnikova, E., Vieu, L., Oltramari, A., Borgo, S., Alexandrov, T.: Data-Driven and Ontological Analysis of FrameNet for Natural Language Reasoning. In: Proceedings of LREC 2010. Valletta, Malta (2010)
26. Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet::Similarity - Measuring the Relatedness of Concepts. In: Proceedings of the 5^{th} Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-04). pp. 38–41. Boston, MA, USA (2004)
27. Pereira, F., Tishby, N., Lee, L.: Distributional clustering of English words. In: Proceedings of the 31^{st} Annual Meeting of the Association for Computational Linguistics. pp. 183–190. Columbus, OH, US (1993)
28. Resnik, P.: Semantic classes and syntactic ambiguity. In: Proceedings of the Workshop on Human Language Technology. pp. 278–283. Morristown, NJ, USA (1996)
29. Ruppenhofer, J., Ellsworth, M., Petruck, M.R., Johnson, C.R., Scheffczyk, J.: FrameNet II: Extended Theory and Practice. Available at http://framenet.icsi.berkeley.edu/book/book.html (2010)

30. Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C.F., Palmer, M.: SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In: Proceedings of SemEval-2010: 5th International Workshop on Semantic Evaluations (2010)
31. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: WWW '07: Proceedings of the 16th international conference on World Wide Web. pp. 697–706. ACM Press (2007)
32. Tonelli, S., Giuliano, C., Tymoshenko, K.: Wikipedia-based WSD for multilingual frame annotation. Artificial Intelligence (2012)
33. Tonelli, S., Pighin, D.: New Features for FrameNet - WordNet Mapping. In: Proceedings of CoNLL-2009. pp. 219–227. Boulder, Colorado (2009), `http://www.aclweb.org/anthology/W09-1127`
34. Wilks, Y.: An intelligent analyzer and understander of English. Communications of the ACM 18(5), 264–274 (1975)
35. Wu, Z., Palmer, M.: Verb Semantics and Lexical Selection. In: Proceedings of the 32^{nd} Annual Meeting of the Association for Computational Linguistics. Las Cruces, New Mexico (1994)
36. Zapirain, B., Agirre, E., Màrquez, L., Surdeanu, M.: Improving semantic role classification with selectional preferences. In: Proceedings of the 2010 Annual Conference of the North American Chapter of the ACL. pp. 373–376. Los Angeles, CA, USA (2010)