IJCLA VOL. 2, NO. 1–2, JAN-DEC 2011, PP. 25–40
RECEIVED 01/11/10 ACCEPTED 21/12/10 FINAL 08/02/11

Folding FrameNet – Unfolding Its Potential?

JOSEF RUPPENHOFER AND MANFRED PINKAL

Saarland University, Germany

ABSTRACT

FrameNet holds intuitive appeal as a resource that provides valuable semantic information for NLP systems, but its impact on NLP applications consistently falls short of expectations, due to the poor performance of frame-based semantic parsers. Sparse data and high granularity have been identified as interrelated reasons. In this paper, we study the effect of coarsening FrameNet, using a tool for automatic merging of frames and lexical units. We report results of experiments carried out to assess the effect of granularity reduction on parsing performance and on the discriminative power of frame information for the RTE task. A qualitative examination of annotation changes affected by the merging process leads to interesting conclusions concerning general desiderata in semantic processing and in FN-based lexical-semantic modeling, and on the general usefulness of granularity reduction.

1 INTRODUCTION

Predicate-argument structure has become a central part of information extraction, question answering and other information access tasks [1–3]. The most widely used resources for modeling predicate-argument structure in English are PropBank [4] and FrameNet [5]. PropBank abstracts over variations in the syntactic realization of an individual verb's semantic roles. PropBanks are available for an expanding set of languages, they can be created efficiently and have been integrated into a number of NLP systems [6]. PropBank analysis is, however, rather syntax-dependent, and it does not generalize across lemma boundaries. FrameNet (henceforth also: FN) provides such generalization by grouping relevant senses of different lemmas into the same frame with a shared set of semantic roles, called Frame Elements. Moreover, FN organizes both frames and frame elements into a hierarchy using several types of semantic relations. Via its frame-to-frame relations, FN also models presuppositions of events and generates expectations in the sense that event types (frames) are connected through sequential and other relations. Finally, frame structures are language-independent, and thus suggest themselves for cross-lingual applications.

We use an example from the RTE-2 challenge [7] to illustrate FN's contribution to NLP applications. The RTE task consists in determining whether a text entails a hypothesis in a common-sense way. It uses data from IR, IE, QA and Summarization to form positive and negative text-hypothesis pairs.

(1) T:. : Mr. Fitzgerald revealed he was one of several top officials who told Mr. Libby in June 2003 that Valerie Plame, **wife** of the former ambassador Joseph Wilson, worked for the CIA.
H:. Valerie Plame is **married** to Joseph Wilson.

Frame semantic analysis with a well-performing shallow semantic parser would show *wife* in the text and *married* in the hypothesis to both belong to the `Personal relationship` frame, constituting evidence that entailment holds.

Despite the intuitive appeal of FN, its impact on NLP applications falls short of expectations. An often-mentioned practical drawback is the lack of coverage, due to the incompleteness of the FN database [8]. Even more importantly, it is very difficult to leverage the information which is covered by the FN database for practical tasks [9]. Existing FN-based shallow semantic parsers show consistent low performance, in contrast to the rather impressive accuracy of PropBank role labelers. Of course, the task of assigning frame structures is more ambitious and thus simply harder. But another reason may be that FN is just too fine-grained to allow robust shallow semantic parsing. On the one hand, the high granularity increases the problem of data sparseness: for many frames, frame elements and lexical units, the FN corpus contains only few annotated training instances. On the other hand, it seems that semantic distinctions often are too subtle for parsers – and for humans.

In this paper, we study the effect of systematically coarsening FN. This directly addresses problems of granularity and indirectly issues of coverage and sparsity: the merging of senses provides more annotated instances for the remaining word senses and may eliminate word senses that are defined by FN but not backed by annotations. We use the recently presented FN transformer, which automatically merges frames or word senses and changes annotations of corpus data according to user specifications [10], to systematically vary the granularity of FN, and evaluate the effect of frame folding on parsing accuracy. By themselves, effects on parser performance are of limited interest since frame folding implies a reduction of alternative target annotations, almost trivially entailing an increase in accuracy. The interesting question is this: Does parser improvement come at the cost of losing relevant semantic information? Or could we even gain information through additional cases of semantic relatedness becoming apparent through the merging process? To answer these questions, we build on the experimental setup of Burchardt et al. 2009 [9] for evaluating the impact of FN on the task of Recognizing Textual Entailment (RTE).

These authors use the FATE corpus [11], which contains manual frame semantic annotations for the RTE-2 test set, as a gold corpus. Unlike end-to-end evaluations, this setup allows them to separately measure FN's coverage, the performance of semantic parsers, and the discriminative power of frame-semantic information for entailment recognition. Concerning the latter, they find a discriminative effect of frame information on the entailment task which significantly exceeds a simple word-overlap measure. However, this positive effect can only be measured on the gold corpus; in a realistic setting it is offset by noisy parser output. We replicate their experiments in a somewhat modified version with FNs of different granularity to assess whether coarsening can lead to better automatic parser performance *while avoiding the loss of relevant information*.

Our evaluation shows that parsing results improve significantly by merging, but there is no significant gain in discriminative power. We complement the quantitative evaluation by a qualitative examination of those entailment pairs which are affected by the merging process. This leads to interesting conclusions concerning general desiderata in semantic processing and in FN-based lexical-semantic modeling, and on the general usefulness of frame-folding.

This paper is structured as follows. In section 2, we discuss related work. In section 3, we report on experiments in which we systematically explore the coarsening of FN with the FN transformer tool and its effect on the performance of a statistical semantic parser. In section 4, we use the output of the best transformer setting and replicate parts of Burchardt et al.'s study (2009) on the impact of frame semantic information on the

RTE task. In section 5, we present a qualitative evaluation of the effects of collapsing frame and lexical unit distinctions. In section 6, we offer our conclusions.

2 RELATED WORK

McConville and Dzikovska [12] pursue an approach aimed at reducing FrameNet's granularity. They build a new semantic resource by deriving a verb lexicon for deep syntactic parsing from the annotations in FrameNet using the `Inheritance` relation between frames and the `CoreSet` relation between Frame Elements to reduce the set of semantic roles. Their lexicon comes, however, without an appropriately updated corpus. Therefore, it cannot be used to train a shallow semantic parser. Ruppenhofer et al.'s [10] transformer tool, which we will use in this paper, was specifically created to allow for experimenting with different levels of granularity in FrameNet to suit NLP applications.

Matsubayashi et al. [13] focus on the sparse-data problem for role assignment. They compare various ways of generalizing across semantic roles. Fürstenau and Lapata [14] propose a semi-supervised approach to augment the training material for verbs known to FrameNet based on small seed sets. The performance effects on a semantic role labeler are limited and they are best for smaller seed sets. In another paper, Fürstenau and Lapata [15] propose a semi-supervised approach to assign instances of verbs that are unknown to FrameNet to suitable frames. Further coverage-related work concerns type-based lexical unit induction. Pennacchiotti et al. [16] use vector space models and WordNet relations to assign unknown predicates to the most similar frame.

3 PARSING EXPERIMENTS

In this section, we discuss how some of FrameNet's frame and word sense distinctions can be collapsed in a linguistically motivated way, and measure the effect on the performance of a shallow semantic parser. To this end, we systematically explore different parameter settings for the FrameNet transformer tool to produce modified versions of the FrameNet data. We then train and test a statistical semantic parser on the modified FrameNet data to evaluate its performance.

We use FrameNet's official release 1.3 and modified versions of it that we generate using the transformer tool. FrameNet release 1.3 has 795 frames and 10195 lexical units (word senses), belonging to 8365 different lemmas. 6728 lexical units are exemplified with annotated instances that can be used as training data.[1]

3.1 *Frame Folding*

The FrameNet Transformer tool gives users a possibility to automatically coarsen the FrameNet sense inventory in an iterative procedure. The tool automatically merges either entire frames, if they stand in appropriate frame relations chosen by the user, or word senses of lemmas that belong to frames related by the specified relations. In the latter mode, only word senses standing in a child-ancestor relation are merged. In either mode, the transformer automatically outputs format-compliant FrameNet versions, including modified corpus annotation files that can be used for automatic processing. Users can vary the behavior of the FN transformer by setting paramaters. For our experiment, we selected eight linguistically reasonable parameter combinations. We describe and motivate them in the following.

Type of merger. Frame-based merging aims at reduction of frame granularity and providing larger sets of training data per frame. Lexical-unit based merging aims at reduction of word-sense ambiguity, and leaves the frame structure unchanged. We tested frame-based and LU-based merging, as well as combinations of the two in either order. In Table 1 below, we use F to designate a frame based merger, L to designate lemma-based one, FL for a sequence of frame-based merging followed by LU-based merging, and LF for a sequence of mergers in the reverse order.

Frame relations licensing merge. The FrameNet database employs different frame-to-frame relations, a subset of which can be specified by the user of the transformer tool. In the frame-based mode, the transformer merges those pairs of frames which are connected by licensing relations. In the LU-based mode, the transformer traverses licensing relations downward from the frame of the target LU to find lemma-identical LUs to merge into the target LU. E.g., with the lemma *depart*, the sense in the frame `Departing` can serve as a target for the sense in the child frame `Quitting a place`, which is related by INHERITANCE. The PERSPECTIVE_ON, SUBFRAME, INCHOATIVE_OF, and CAUSATIVE_OF relations are reliable indicators of close semantic relatedness. We use

[1] Very recently, FrameNet made available a new release, 1.5. Its format is different from release 1.3 and therefore not compatible with the FrameNet transformer tool we use.

them throughout in our experiments. We exclude the USING relation because it often connects frames that are only vaguely related. INHERITANCE is not a clear positive candidate because the semantic step length between frames connected by this relation varies greatly. We carry out every experiment in two versions, one with and one without INHERITANCE as a licensing relation. In the Setting column of table 1, we refer to these versions as F+I, F-I, etc.

Stop frames. Stop frames are frames which may not serve as target frames for frame-based mergers. Stop frames are highly abstract frames like Event, Process, Transitive_action, which are connected to large numbers of very different frames that we do not want to collapse. We use the same set of 12 stop frames in all experiments. Stop frames are used only in frame-based merging.

Number of iterations performed. The FN transformer proceeds iteratively, always merging (lexical units in) adjacent frames. We set the number of iterations to 3 for frame-based merging, and 1 for LU-based merging because only few additional mergers take place if one performs additional iterations.

We use eight different parameter combinations, each starting from the official FrameNet release and producing a different modified FrameNet version. These parameterizations share the appropriate fixed settings but vary with regard to merger type and the use of the INHERITANCE relation in identifying source frames and LUs.

3.2 *Parsing*

To automatically annotate the nine (original and derived) versions of FN, we used the Shalmaneser shallow semantic parser [17], which was the only freely available parser for frame semantic role labeling at the time.[2] Shalmaneser breaks down the task of frame semantic annotation into three ordered sub-problems (frame assignment, argument recognition, and argument labeling) which are modeled as modular, supervised learning tasks. We used the default Naive Bayes classifiers that come with Shalmaneser.

We trained and tested the Shalmaneser semantic parser in a 10-fold cross validation setting. We trained each system only on those lemmas that were affected in at least one of our eight experiments, because only in

[2] The system Johansson and Nugues built for the Semeval-2007 task was no longer available [18]. CMU's SEMAFOR-system [19] became available too late for use in this work.

Table 1. Performance of SRL system trained on different versions of FN

ID	Setting	LU red.	avg. # senses	frame assign.	arg. recognition			arglab
				acc.	prec.	rec.	f-score	acc.
-	ORIG	0	1.66	0.938	0.852	0.697	0.766	0.762
1	F, +I	153	1.55	0.944	0.846	0.682	0.755	0.746
2	**LF, +I**	**397**	**1.36**	**0.963**	**0.842**	**0.674**	**0.749**	**0.741**
3	L, +I	353	1.39	0.958	0.845	0.691	0.760	0.746
4	FL,+I	369	1.38	0.956	0.843	0.676	0.750	0.738
5	F, -I	112	1.58	0.943	0.848	0.680	0.755	0.750
6	LF, -I	339	1.40	0.958	0.846	0.677	0.752	0.742
7	L, -I	352	1.39	0.959	0.848	0.692	0.762	0.748
8	FL, -I	360	1.39	0.957	0.849	0.683	0.757	0.745

these cases the parser's behavior may change. The data set contains 1300 lemmas and 36681 annotated frame instances (annotation sets). The 1300 lemmas involved have a total of 2163 word senses associated with them in the official FrameNet release. 755 of the lemmas are unambiguous. They are included in the data set because their single sense was affected by a frame-based merger. 354 lemmas have 2 senses and 191 lemmas have more than 2 senses. The reduction of senses in the sub-corpus under consideration can be read off the third column of Table 1.

3.3 *Results*

The results of our experiments, shown in Table 1, confirm that we can improve parser performance by collapsing certain distinctions made by FrameNet. Not too surprisingly, frame assignment accuracy correlates with the degree of frame ambiguity. The best performance on frame assignment results from experiment 2, where polysemy was reduced the most, by combining LU-based merging with frame-based merging in that order, and allowing INHERITANCE. Table 1 further shows that, unlike with frame assignment, the modified FN versions cannot outperform the original release when it comes to argument recognition and labeling.

4 RTE EXPERIMENTS

In our second experiment, we turn to the question how frame folding affects the relevant semantic information contained in frame annotations.

We draw on the work of Burchardt et al.(2009), who made a study of the impact of frame semantic information on the RTE task. Like them, we base our study on the FATE corpus.

4.1 *The FATE Corpus*

The FATE corpus [11] contains manual annotations of the RTE-2 test set, consisting of 800 entailment pairs, 400 positive, 400 negative. In positive entailment pairs, the sentences are not fully annotated with semantic frames but only on relevant spans which annotators deemed to have a bearing on making the entailment decision. The FATE corpus contains 1686 sentences with 4490 annotated frame instances and 9518 role instances.

Lemmas in FATE were annotated with frames in a rather flexible way. For instance, annotators were allowed to apply frames to occurrences of lemmas if the frames seemed to match the occurring word senses, even if the lemmas were not listed in the relevant frames in FrameNet. Due to this generous annotation policy in FATE, Burchardt et al. report a surprisingly high coverage of 92% on frame instances. If we only consider annotated instances of frames for which the frame-evoking lemma is listed in FN 1.3 and has training data, we are left with a total of 1519 frame instances (34%) that the Shalmaneser parser can possibly handle.[3] We call this subset of FATE annotations FATE-strict.

To fairly assess the performance of the Shalmaneser semantic parser, we use FATE-strict as a gold standard. We train the parser on FrameNet release 1.3 and test it on FATE-strict. On the 1519 annotatable instances in FATE-strict, the system achieves a precision of 86% and a recall of close to 100% for frame assignment. Accordingly, accuracy also stands at close to 86%, which is somewhat lower than the 93% accuracy value we obtained when training and testing with Shalmaneser on FrameNet release 1.3. The precision and recall figures we obtained in our experiment contrast markedly with the precision and recall values of 0.35 and 0.40 for Shalmaneser given by Burchardt et al., who evaluated Shalmaneser against the full set of FATE annotations, including the cases that Shalmaneser was not equipped to handle.

---

[3] The Shalmaneser semantic parser can only assign a frame to a lemma instance if it has seen training data for that lemma-frame pair. It cannot transfer what it has learned on a lemma with training data for a given frame, to another lemma without training data.

4.2 *Experimental Setup*

We follow Burchardt et al 2009 in using the FATE corpus but we will not attempt to replicate their experiments in full. The focus of our analysis is on comparing versions of the FATE corpus which are annotated according to frame schemas reflecting different levels of granularity.

The basic method is to extract frame-based statistical information from the positive and negative entailment pairs in the annotated corpus, respectively, and to measure the overlap of frame structures between the text and the hypothesis sentences. The key assumption behind this method is that the more of the semantic material in the hypothesis can be 'embedded' into the text, the more likely it is that an entailment relation exists between text and hypothesis. For the level of frames or word senses, Burchardt et al. define a **frame label overlap** measure between hypothesis sentences and the text sentences paired with them. Frame label overlap is defined as the percentage of frame labels in the hypothesis which have a counterpart in the text. Frame label overlap is calculated separately for the frames in all positive and negative pairs. The difference between these two scores is taken as a measure of the discriminative power contributed by using frame semantic information.

Our major purpose is to compare the effect of frame folding on frame label overlap information and difference values. We decided to only use the maximally reduced FN version, produced by parsing experiment 2, and compare it to the original FrameNet annotation because it leads to the highest parser quality and provides us with the greatest amount of possible frame assignment differences. We compute the overlap and difference scores for three versions of the corpus,

- the original generously annotated version of FATE, which we use as a gold corpus; the scores give us information about how optimally available FN information (at the original and reduced level of granularity) will be able to discriminate positive and negative entailment cases
- FATE-strict, i.e., the corpus constrained to the parsable frame instances; this is not very interesting in itself, but is needed for comparison with the third version:
- the constrained version of the corpus annotated by the Shalmaneser system; the scores for this version are an indicator of how much frame information can be accessed in a realistic setting.

4.3 *Results*

Table 2 gives the detailed results of the experiment. In addition to the frame overlap scores for positive and negative entailment pairs, and the difference between the two, it shows the corresponding values for word overlap measured on the sets of annotated instances in the generous and the constrained corpus versions.

The results show that for all versions of FATE, frame label overlap scores are higher than lemma overlap scores, underlining the fact that frame semantic normalization brings out latent semantic overlap that is not captured by lemma overlap alone. Also, we could replicate the result of Burchardt et al. that the difference in frame overlap outperforms word overlap by 3.5%. Moreover, we achieve consistently higher frame label overlap scores for the merged versions than on the corresponding versions annotated according to the original FrameNet scheme. This is as expected since the modified scheme reduces the number of frames. However, unlike what we had hoped for, we find that frame label overlap not only increases on positive entailment pairs but also on negative ones. In fact, for all three of the merged versions, there is a small decrease compared to their full FrameNet counterpart versions.

Overall, the quantitative results seem to paint a disappointing picture for the possible contribution that the collapsing of frame and lexical unit distinctions could make to the task of entailment recognition. However, before we accept such a general assessment we will take a closer look at the corpus data themselves.

Table 2. Frame label and lemma overlap on entailment pairs

	Pos	Neg	Diff
Gold Corpus Generous			
Original FN	0.571	0.459	0.112
Merged	0.575	0.490	0.085
Lexical Overlap Baseline	0.450	0.373	0.076
Gold Corpus Constrained			
Original FN	0.549	0.453	0.096
Merged	0.570	0.480	0.090
Lexical Overlap Baseline	0.525	0.410	0.115
Shalmaneser			
Original FN	0.524	0.433	0.091
Merged	0.530	0.450	0.080

5 QUALITATIVE STUDY

In order to get a better sense of the effect of collapsing frame and lexical unit distinctions, we will examine the differences between original and merged FrameNet annotation in detail. We will focus on the generous version of the gold corpus, to have a maximum amount of clean data available. Actually, the subset of text-hypothesis pairs seeing changes in their overlap score is rather small: 49 pairs out of 774 (6.3%). However, note that 541 of the 774 pairs have at least one changed frame instance in the transposed FATE (69.9%). Based on manual inspection of 50 randomly chosen pairs with changing annotation and unchanged score, we find that in the majority of cases (about 80%), a frame that is being shifted occurs only on the T(ext) or only on the H(ypothesis) side, its shift thus having no impact on the overlap score. In the remaining 20% of cases, the same frame is changed in identical ways on both T and H.

We now focus on the 49 entailment pairs where relevant changes occur. In 17 cases, the change that occurs works in our favor. 11 cases are textbook examples of what we hoped to achieve by folding frames: different lemmas and frames on positive pairs in the original version of the corpus come to be aligned in the transformed corpus. An example of this is (2), where *accuse* shifts from `Notification_of_charges` to `Crime_scenario` on the H side, while *arrest* and *try* on the T side shift to the same frame from `Arrest` and `Trial`, respectively.

(2) T: Wyniemko , now 54 and living in Rochester Hills , was **arrested** and **tried** in 1994 for a **rape** in Clinton Township.
H: Wyniemko was **accused** of rape .

In the other 6 cases, frame instances in negative pairs that are aligned in the original FATE come apart, yielding lower overlap scores. However, under ideal conditions none of these changes would have happened, and the lack of entailment would have been recognized by another module of an RTE system. 4 of these cases are due preprocessing errors in the FATE corpus (incorrect lemmatization). 1 case is the result of a coverage gap in FrameNet and the two lemmas in question both should have moved to the same frame. The final case involves two different lemmas on the text and hypothesis side, which do not share the same kind of polysemy and thus cannot move to the same frame during the lexical-unit based second merger phase of parsing experiment 2. The relevant case is (3), where *hang* and *execute* both have senses in the `Killing` frame but only *execute* has a sense in the `Intentionally_act` frame, with which its `Killing` sense can merge in the LU-based merger step.

(3) T: Some 420 people have been **hanged** in Singapore since 1991 , mostly for drug trafficking , an Amnesty International 2004 report said . That gives the country of 4.4 million people the highest execution rate in the world relative to population .
H: 4.4 million people were **executed** in Singapore .

In 32 of the 49 pairs with changes in the overlap score, the change runs against our interest. In 9 positive entailment pairs, two correctly aligned frame instances come apart during the lexical unit-based second merger phase because they have two different lemmas that do not share the same polysemy, similarly to what happens on the negative pair exemplified in (3). In another case, two correctly aligned frame instances come apart because of a lemmatization error. In the remaining 22 cases, all in negative entailment pairs, two frame instances are correctly brought into alignment in the course of frame-based merging. The lack of entailment in most of these cases depends on aspects which are out of the scope of lexical semantics. In some cases, the compositional process brings targets bearing identical frame information into completely different contexts. For instance, in example (4), the *Entity* frame elements of the two frame instances do not refer to the same entities. In other cases, modal operators or negation influence the veridicality of a predicate-argument structure and thereby prevent entailment. In (7), the embedding of the `Forming_relationships` frame evoked by *married* under the modal operator introduced by *possibility* prevents entailment. Example (5) is a different case. Here, the opposed scalar values of *cut* and *rise* are not differentiated in FrameNet, although the opposite polarity is a matter of lexical semantics.

(4) T: **Changes** [in the cell-cycle and apoptotic machineries , or in the signaling pathways that control them ^{Entity}] allow cancer cells to escape the normal control of cell proliferation and cell death .
H: The **altered** [cellular networks of molecular pathways ^{Entity}] sustain cancer cell growth and make them resistant to certain therapies .

(5) T: The National Institute for Health and Clinical Excellence estimates the move would **cut** the number of unplanned pregnancies by 70,000 each year .
H: Unplanned pregnancies **rise** by 70,000 each year .

(6) Still , violence continued : Insurgents killed five U . S . soldiers , set off a suicide car bomb that **killed** [four Iraqi policemen ^{Victim}] in Baghdad and targeted more polling sites across the country .
H: Five U S soldiers were killed , and at least [10 Iraqis ^{Protagonist}] **died** in Baghdad

(7) T: The possibility for Yevgenia Timoshenko , daughter of the Ukrainian ex-prime minister , to be **married** to the English rock singer Sean Carr , in church , is still questionable
H: Yevgenia Timoshenko is the **wife** of Sean Carr .

Overall, our examination of the entailment pairs with changed overlap scores suggests that collapsing frame distinctions is useful, if we look past the simple metric of frame overlap scores. First, the predominantly negative effects on merging by pre-processing errors can be set aside as orthogonal issues. More importantly, while the easy cases of frame instances coming into correct alignment to increase overlap on positive pairs are an argument for the merging, so are the cases of frame instances coming into correct alignment on negative pairs (5-7). For these latter cases, too, it is crucial that they come into alignment since other RTE modules checking on correct semantic composition, quantification etc presuppose that the relations they are operating on are correctly identified as equivalent.

The handling of antonyms could readily be improved by augmenting FrameNet's semantic representation. Antonyms should either be handled by separating them into distinct frames or recording their scalar properties on each lexical unit. Finally, based on this data set, lexical-unit based merging seems to mostly have ill effects, often breaking up correct alignments on positive pairs. With this type of merger, the benefit of improved parsing accuracy in the abstract seems to be outweighed by the loss of relevant information in the RTE setting.

6 Conclusion

We investigated the effect of automatic coarsening of FrameNet on the performance of a semantic parser, and on the usefulness of frame-semantic information for NLP tasks, using the example of RTE. We have shown that coarsening FrameNet improves performance on the frame assignment task but incurs a slight drop on the role recognition and labeling tasks. The quantitative evaluation of the impact of frame-folding on the RTE task did not show a positive result. A qualitative study of the induced changes in the annotation gave a differentiated, but basically positive picture. Frame-merging is responsible for most changes, they almost always lead to locally correct alignments, but must be complemented with mechanisms to handle non-local operators and compositional structure. LU-based merging seems to be less advisable because it often has negative effects. The FrameNet database should not only be completed, but extended to include further layers of lexical-semantic information such as polarity.

ACKNOWLEDGMENTS This work was funded by the German Research Foundation DFG (PI 154/9-3).

REFERENCES

1. Kulick, S., Bies, A., Liberman, M., Mandel, M., McDonald, R., Palmer, M., Schein, A., Ungar, L.: Integrated annotation for biomedical information extraction. In: HLT-NAACL 2004 Workshop: BioLINK 2004, Linking Biological Literature, Ontologies and Databases, Boston, ACL (2004) 61–68
2. Kaisser, M., Webber, B.: Question answering based on semantic roles. In: Proceedings of the Workshop on Deep Linguistic Processing, Prague, ACL (2007) 41–48
3. Fliedner, G.: Linguistically Informed Question Answering. Ph.d., Saarland University, Saarbrücken (2007)
4. Palmer, M., Gildea, D., Kingsbury, P.: The Proposition Bank: A Corpus Annotated with Semantic Roles. Computational Linguistics **31**(1) (2005) 71–106
5. Baker, C., Fillmore, C., Lowe, J.: The Berkeley FrameNet Project. In: Proceedings of the 17th International Conference on Computational Linguistics, Montreal, ACL (1998) 86–90
6. Carreras, X., Màrquez, L.: Introduction to the conll-2005 shared task: semantic role labeling. In: Proceedings of the Ninth Conference on Computational Natural Language Learning, Ann Arbor, ACL (2005) 152–164
7. Bar-Haim, R., Dagan, I., Dolan, B., Ferro, L., Giampiccolo, D., Magnini, B., Szpektor, I.: The Second PASCAL Recognising Textual Entailment Challenge. In: Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment, Venice (2006)
8. Palmer, A., Sporleder, C.: Evaluating framenet-style semantic parsing: the role of coverage gaps in framenet. In: Coling 2010: Posters, Beijing, Coling 2010 Organizing Committee (2010) 928–936
9. Burchardt, A., Pennacchiotti, M., Thater, S., Pinkal, M.: Assessing the Impact of Frame Semantics on Textual Entailment. Journal of Natural Language Engineering **15**(4) (2009) 527–550

10. Ruppenhofer, J., Sunde, J., Pinkal, M.: Generating FrameNets of various granularities: The FrameNet Transformer. In: Proceedings of the Seventh conference on International Language Resources and Evaluation, La Valletta, ELRA (2010)
11. Burchardt, A., Pennacchiotti, M.: Fate: A framenet-annotated corpus for textual entailment. In: Proceedings of the Sixth conference on International Language Resources and Evaluation, Marrakesh, ELRA (2008)
12. McConville, M., Dzikovska, M.: Using inheritance and coreness sets to improve a verb lexicon harvested from FrameNet. In: Proceedings of the Second Linguistic Annotation Workshop, ELRA (2008) 33–40
13. Matsubayashi, Y., Okazaki, N., Tsujii, J.: A Comparative Study on Generalization of Semantic Roles in FrameNet. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Singapore, ACL (2009) 19–27
14. Fürstenau, H., Lapata, M.: Semi-Supervised Semantic Role Labeling. In: Proceedings of the 12th Conference of the European Chapter of the ACL, Athens, ACL (2009) 220–228
15. Fürstenau, H., Lapata, M.: Graph Alignment for Semi-Supervised Semantic Role Labeling. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, ACL (2009) 11–20
16. Pennacchiotti, M., De Cao, D., Basili, R., Croce, D., Roth, M.: Automatic induction of framenet lexical units. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Honolulu, ACL (2008) 457–465
17. Erk, K., Pado, S.: Shalmaneser - a flexible toolbox for semantic role assignment. In: Proceedings of the Fifth conference on International Language Resources and Evaluation, Genoa, ELRA (2006)
18. Johansson, R., Nugues, P.: Syntactic representations considered for frame-semantic analysis. In: Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories, Bergen (2007)
19. Das, D., Schneider, N., Chen, D., Smith, N.: Probabilistic frame-semantic parsing. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, Los Angeles, ACL (2010) 948–956

Josef Ruppenhofer
Saarland University
Computational Linguistics and Phonetics
Saarbrücken,
Germany
E-mail: <josefr@coli.uni-saarland.de>

MANFRED PINKAL
SAARLAND UNIVERSITY
COMPUTATIONAL LINGUISTICS AND PHONETICS
SAARBRÜCKEN,
GERMANY
E-MAIL: <PINKAL@COLI.UNI-SAARLAND.DE>