

Semantic role assignment for event nominalisations by leveraging verbal data

Sebastian Padó
Department of Linguistics
Stanford University
450 Serra Mall
Stanford CA 94305, USA
pado@stanford.edu

Marco Pennacchiotti and Caroline Sporleder
Computational Linguistics
Saarland University
Postfach 15 11 50
66041 Saarbrücken, Germany
{pennacchiotti|csporled}@coli.uni-sb.de

Abstract

This paper presents a novel approach to the task of semantic role labelling for event nominalisations, which make up a considerable fraction of predicates in running text, but are underrepresented in terms of training data and difficult to model. We propose to address this situation by *data expansion*. We construct a model for nominal role labelling solely from verbal training data. The best quality results from salvaging grammatical features where applicable, and generalising over lexical heads otherwise.

1 Introduction

The last years have seen a large body of work on modelling the semantic properties of individual words, both in the form of hand-built resources like WordNet and data-driven methods like semantic space models. It is still much less clear how the combined meaning of phrases can be described. *Semantic roles* describe an important aspect of phrasal meaning by characterising the relationship between predicates and their arguments on a semantic level (e.g., agent, patient). They generalise over surface categories (such as subject, object) and variations (such as diathesis alternations). Two frameworks for semantic roles have found wide use in the community, PropBank (Palmer et al., 2005) and FrameNet (Fillmore et al., 2003). Their corpora are used to train supervised models for semantic role labelling (SRL) of new text (Gildea and Jurafsky, 2002; Carreras and Màrquez, 2005). The resulting analysis can benefit a number of applications, such

as Information Extraction (Moschitti et al., 2003) or Question Answering (Frank et al., 2007).

A commonly encountered criticism of semantic roles, and arguably a major obstacle to their adoption in NLP, is their *limited coverage*. Since manual semantic role tagging is costly, it is hardly conceivable that gold standard annotation will ultimately be available for every predicate of English. In addition, the *lexically specific* nature of the mapping between surface syntax and semantic roles makes it difficult to generalise from seen predicates to unseen predicates for which no training data is available. Techniques for extending the coverage of SRL therefore address an important need.

Unfortunately, pioneering work in unsupervised SRL (Swier and Stevenson, 2004; Grenager and Manning, 2006) currently either relies on a small number of semantic roles, or cannot identify equivalent roles across predicates. A promising alternative direction is *automatic data expansion*, i.e., leveraging existing annotations to classify unseen, but similar, predicates. The feasibility of this approach was demonstrated by Gordon and Swanson (2007) for syntactically similar verbs. However, their approach requires at least one annotated instance of each new predicate, limiting its practicability.

In this paper, we present a pilot study on the application of automatic data expansion to *event nominalisations* of verbs, such as *agreement* for *agree* or *destruction* for *destroy*. While event nominalisations often afford the same semantic roles as verbs, and often replace them in written language (Gurevich et al., 2006), they have played a largely marginal role in annotation. PropBank has only annotated verbs.¹ FrameNet annotates nouns, but covers far fewer nouns than verbs. The same

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

¹A follow-up project, NomBank (Meyers et al., 2004), has since provided annotations for nominal instances, too.

situation holds in other languages (Erk et al., 2003).

Our fundamental intuition is that it is possible to increase the annotation coverage of event nominalisations by data expansion from verbal instances, since the verbal and nominal predicates share a large part of the underlying argument structure. We assume that annotation is available for verbal instances. Then, for a given instance of a nominalisation and its arguments, the aim is to assign semantic role labels to these arguments. We solve this task by constructing *mappings* between the arguments of the noun and the semantic roles realised by the verb’s arguments. Crucially, unlike previous work (Liu and Ng, 2007), we do not employ a classical supervised approach, and thus do not require any nominal annotations.

Structure of the paper. Sec. 2 provides background on nominalisations and SRL. Sec. 3 provides concrete details on our expansion-based approach to SRL for nominalisations. The second part of the paper (Sec. 4–6) provides a first evaluation of different mapping strategies based on syntactic, semantic, and hybrid information. Sec. 8 concludes.

2 Nominalisations

Nominalisations (or *deverbal nouns*) are commonly defined as nouns morphologically derived from verbs, usually by suffixation (Quirk et al., 1985). They have been classified into at least three categories in the linguistic literature, *event*, *result*, and *agent/patient* nominalisations (Grimshaw, 1990).

Event and *result* nominalisations account for the bulk of deverbal nouns. The first class refers to an event/activity/process, with the nominal expressing this action (e.g. *killing*, *destruction*). Nouns in the second class describe the result or goal of an action (e.g. *agreement*). Many nominals have both an event and a result reading (e.g., *selection* can mean the process of selecting or the selected object). Choosing a single reading for an instance is often difficult; see Nunes (1993); Grimshaw (1990).

A smaller class is *agent/patient* nominalisations. Agent nominals are usually identified by suffixes such as *-er*, *-or*, *-ant* (e.g. *speaker*, *applicant*), while patient nominalisations end with *-ee*, *-ed* (e.g. *employee*). While these nominalisations can be analysed as events (*the baker’s bread* implies that baking has taken place), they more naturally refer to participants. In consequence, agent/patient nominals tend to realise fewer arguments – the average in FrameNet is 1.46 arguments, compared to 1.74

PropBank	
Verbs (Carreras and Márquez, 2005)	80%
Nouns (Liu and Ng, 2007)	73%
FrameNet	
Verbs (Mihalcea and Edmonds, 2005)	72%
Nouns (Pradhan et al., 2004)	64%

Table 1: F-Scores for supervised SRL (end-to-end)

for events/results. As our goal is nominal SRL, we concentrate on the event/results class.

SRL for nominalisations. Compared to the wealth of studies on verbal SRL (e.g., Gildea and Jurafsky (2002); Fleischman and Hovy (2003)), there is relatively little work that specifically addresses nominal SRL. Nouns are generally treated like verbs: the task is split into two classification steps, argument recognition (telling arguments from non-arguments) and argument labelling (labelling recognised arguments with a role). Nominal SRL also typically draws on feature sets that are similar to those for verbs, i.e., comprising mainly syntactic and lexical-semantic information (Liu and Ng, 2007; Jiang and Ng, 2006).

On the other hand, there is converging evidence that nominal SRL is somewhat more difficult than verbal SRL. Table 1 shows some results for both verbal and nominal SRL from the literature. For both PropBank and for FrameNet, we find a difference of 7–8% F-Score. Note, however, that these studies use different datasets and are thus not directly comparable.

In order to confirm the difference between nouns and verbs, we modelled a controlled dataset (described in detail in Sec. 4) of verbs and corresponding event nominalisations. We used Shalmaneser (Erk and Padó, 2006), to our knowledge the only freely available SRL system that handles nouns. SRL models were trained on verbs and nouns separately, using the same settings and features. Table 2 shows the results, averaged over 10 cross-validation (CV) folds. Accuracy was about equal in the recognition step, and 5% higher for verbs in the labelling step. We analysed these results by fitting a *logit mixed model*. These models determine which *fixed factors* are responsible for differences in a response variable (here: SRL performance) while correcting for imbalances introduced by *random factors* (see Jaeger (2008)). We modelled the training and test set sizes and the predicates’ parts of speech as fixed effects, and frames and CV folds as random factors.

For both argument recognition and labelling, the

Step	Verbs	Nouns
Arg recognition (F ₁ , class FE)	0.59	0.60
Arg labelling (Accuracy)	0.70	0.65

Table 2: FrameNet SRL on verbs and nouns

amount of training data turned out to be a significant factor, i.e., more data leads to higher results. While the part of speech was not systematically linked to performance for argument recognition, it was a highly significant predictor of accuracy in the labelling step: Even when training set size was taken into account, verbal arguments were still significantly easier to label ($z=4.5$, $p<0.001$).

In sum, these results lend empirical support to claims that nominal arguments are less tightly coupled to syntactic realisation than verbal ones (Carlson, 1984); their interpretation is harder to capture with shallow cues.

3 Data Expansion for Nominal SRL

The previous section has established two observations. First, the argument structures of verbs and their event nominalisations correspond largely. Second, nominal SRL is a difficult task, even given nominal training data, which is hard to obtain.

Our proposal in this paper is to take advantage of the first observation to address the second. We do so by modelling SRL for event nominalisations as a *data expansion* task — i.e., using existing *verbal* annotations to carry out SRL for novel *nominal* instances. In this manner, we do away completely with the need for manual annotation of nominal instances that is required for previous supervised approaches (cf. Sec. 2). Consider the following examples, given in format [constituent]**grammatical function**/SEMANTIC ROLE:

- (1) a. [Peter]**Subj**/COGNIZER *laughs*
[about the joke]**PP-about**/STIMULUS
b. [Peter]**Subj**/COGNIZER *laughs*
[at him]**PP-at**/STIMULUS
- (2) [Peter’s]**Prenom-Gen**? [hearty]**Prenom-Mod**?
laughter [about the event]**PP-about**?

The sentences with the verbal predicate *laugh* in (1) are labelled with semantic roles, while the NP containing the event nominalisation *laughter* in (2) is not. The question we face are what information from (1) can be re-used to perform argument recognition and labelling on (2), and how.

In this respect, there is a fundamental difference between lexical-semantic and syntactic information.

Lexical-semantic features, such as the head word, are basically independent of the predicate’s part of speech. Thus, the information from (1) that *Peter* is a COGNIZER can be used directly for the analysis of the occurrence of *Peter* in (2). Unfortunately, pure lexical features tend to be sparse: the head word of the last role, *event*, is unseen in (1), and due to its abstract nature, also difficult to classify through semantic similarity. Therefore, it is necessary to consider *syntactic* features as well. However, these vary substantially between verbs and nouns. When they are applicable to both parts of speech, some mileage can be apparently gained: the phrase in (2) headed by *event* can be classified as STIMULUS because it is an *about*-PP like (1a). In contrast, no direct inferences can be drawn about prenominal genitives or modifiers which do not exist for verbs.

In the remainder of this paper, we will present experiments on different ways of combining syntactic and lexical-semantic information to balance precision and recall in data expansion. We address argument recognition and labelling separately, since the two tasks require different kinds of information. We assume that the frame has been determined beforehand with word sense disambiguation methods.

4 Data

The dataset for our study consists of the annotated FrameNet 1.3 examples. We obtained pairs of verbs and corresponding event/result nominalisations by intersecting the FrameNet predicate list with a list of nominalisations obtained from Celex (Baayen et al., 1995) and Nomlex (Macleod et al., 1998). We found 306 nominalisations with corresponding verbs in the same frame, but excluded some pairs where either the nominalisation was not of the event/result type, or no annotated FrameNet examples were available for either verb or noun. The final dataset, consisting of 265 pairs exemplifying 117 frames, served for both the analysis in Section 2 and the evaluations in subsequent sections. For the evaluations, we used the 26,479 verbal role instances (2,066 distinct role types) as training data and the 6,502 nominal role instances (993 distinct role types) as test data. The specification of the dataset can be downloaded from http://www.coli.uni-sb.de/~pado/nom_data.html.

5 Argument Recognition

Argument recognition is usually modelled as a supervised machine learning task. Unfortunately, ar-

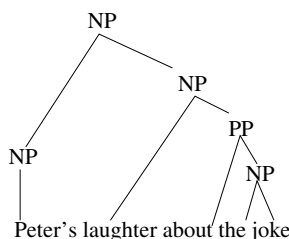


Figure 1: Parse tree for example sentence

argument recognition – at least within predicates – relies heavily on syntactic features, with the grammatical function (or alternatively syntactic path) feature as the single most important predictor (Gildea and Jurafsky, 2002). Since we are bootstrapping from verbal instances to nominal ones, and since there is typically considerable variation between nominal and verbal subcategorisation patterns, we cannot model argument recognition as a supervised task. Instead, we follow up on an idea developed by Xue and Palmer (2004) for verbal SRL, who characterise the set of grammatical functions that could fill a semantic role in the first place. In our application, we simply extract all syntactic arguments of the nominalisation, including any premodifiers. We make no attempt to distinguish between adjuncts and compulsory arguments. Fig. 1 shows an example: in the NP *Peter's laughter about the joke*, the noun *laughter* has two syntactic arguments: the PP *about the joke* and the premodifying NP *Peter's*. Both are extracted as (potential) arguments.

This method cannot identify roles that are syntactically *non-local*, i.e., those that are not in the maximal projection of the frame evoking noun. Such roles are more common for nouns than for verbs. Example 3 shows that an “external” NP like *Bill* can be analysed as filling the HELPER role of the noun *help*. However, the overall proportion of non-local roles is still fairly small in our data (around 10%).

- (3) [Bill]_{HELPER} offered *help* in case of need.

Table 3 gives the argument recognition results for our rule-based system on all roles in the gold standard and on the local roles alone. This simple approach is surprisingly effective, achieving an overall F-Measure of 76.89% on all roles, while on local roles the F-Measure increases to 82.83% due to the higher recall. Precision is 82.01%, as not all syntactic arguments do fill a role. For example, modal modifiers such as *hearty* in (2) rarely fill a (core) role in FrameNet. False-negative errors, which affect recall, are partly due to parser errors and partly

<i>Roles</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
all roles	82.01	72.37	76.89
local roles	82.01	83.66	82.83

Table 3: Argument recognition (local / all roles).

to role fillers that do not correspond to constituents or that are embedded in syntactic arguments. For instance, in (4) the PP *in this country*, which fills the PLACE role of *cause*, is embedded in the PP *of suffering in this country*, which fills the role EFFECT. We extract only the larger PP.

- (4) the *causes* [of suffering
[in this country]PP-in]PP-of

6 Argument Labelling

Argument labelling presents a different picture from argument recognition. Here, both syntactic and lexical-semantic information contribute to success in the task. We present three model families for nominal argument labelling that take different stances with respect to this observation.

The first (*naive-semantic*) and the second (*naive-syntactic*) model families represent extreme positions that attempt to re-use verbal information as directly as possible. Models from the third family, *distributional* models infer the role of a noun’s arguments by computing the semantic similarity between nominal arguments and semantic representations of the verb roles given by the role fillers’ semantic heads.² In the *lexical-level* instantiation, the mapping is established between individual noun arguments and roles. In the *function-level* instantiation, complete nominal grammatical functions are mapped onto roles.³

6.1 Naive semantic model

The naive semantic model (*naive sem*) implements the assumption that lexical-semantic features provide the same predictive evidence for verbal and nominal arguments (cf. Sec. 3). It can be thought of as constructing the trivial *identity mapping* between the values of nominal and verbal semantic features. To test the usefulness of this model, we train the Shalmaneser SRL system (Erk and Padó, 2006) on the verbal instances of the dataset described

²Usually, the semantic head of a phrase is its syntactic head. Exceptions occur e.g. for PPs, where the semantic head is the syntactic head of the embedded NP.

³We compute grammatical functions as phrase types plus relative position; for PPs, we add the preposition.

in Sec. 4, using only the lexical-semantic features (head word, first word, last word). We then apply the resulting models directly to the corresponding nominal instances.

6.2 Naive syntactic model

The intuition of this model (*naive syn*) is that grammatical functions shared between verbs and nouns are likely to express the same semantic roles. It maps all grammatical functions of a verb g_v onto the identical functions of the corresponding noun g_n and then assigns the most frequent role realised by g_v to all arguments with grammatical function g_n . For example, if PPs headed by *about* for the verb *laugh* typically realise the STIMULUS role, all arguments of the noun *laughter* which are realised as *PP-about* are also assigned the STIMULUS role.

We predict that this strategy has a “high precision–low recall” profile: It produces reliable mappings for those grammatical functions that are preserved across verb and noun, in particular prepositional phrases; however, it fails for grammatical functions that only occur for one part of speech.

This problem becomes particular pertinent for two prominent role types, namely AGENT-style roles (deep subjects) and PATIENT-style roles (deep objects). These roles are usually expressed via different and ambiguous noun and verb functions (Gurevich et al., 2006). For verbs, the AGENT is typically expressed by the *Subject*, while for nouns it is expressed by a *Pre-Modifier*. The PATIENT is commonly realised as the *Object* for verbs, and either as a *Pre-Modifier* or as a *PP-of* for nouns. As the noun’s *Pre-Modifier* is highly ambiguous, it is also ineffective to apply a *non-identity* mapping such as ($subject_v, Pre-Modifier_n$) or ($object_v, Pre-Modifier_n$).⁴

A final variation of this model is the *generalised naive syntactic model (naive sem-gen)*, where we assign the role most frequently realised by a given function across *all* verbs in the frame. This method alleviates data sparseness stemming from functions never seen with particular verbs and is fairly safe, since mapping within frames tends to be uniform.

6.3 Distributional models

The distributional models construct mappings between verbal and nominal semantic heads. In con-

⁴Lapata (2002) has shown that the mapping can be disambiguated for individual nominalisations. Her model, using lexical-semantic, contextual and pragmatic information, is outside the scope of the present paper.

trast to the naive semantic model, they make use of some measure of semantic similarity to find mappings, and optionally use syntactic constraints to guide generalisation. In this manner, distributional models can deal with unseen feature values more effectively. In sentences (1) and (2), for example, an ideal distributional model would find the head word *event* in (2) to be more semantically similar to the head *joke* in (1a) than to head *him* in (1b). The resulting mapping (*joke, event*) leads to the assignment of the role STIMULUS to *event*.

Semantic Similarity. *Semantic similarity measures* are commonly used to compute similarity between two lexemes. There are two main types of similarity: *Ontology-based*, computed through the closeness of two lexemes in a lexical database (e.g., WordNet); and *distributional*, given by some measure of the distance between the lexemes’ vector representations in a semantic co-occurrence space. We chose the latter approach because it tends to have a higher coverage and it is knowledge-lean, requiring just an unannotated corpus.

We compute distributional-similarity with a semantic space model based on lexical co-occurrences backed by syntactic relations (Padó and Lapata, 2007).⁵ The model is constructed from the British National Corpus (BNC), using the 2.000 most pairs of words and grammatical functions as dimensions. As similarity measure, we use cosine distance on log-likelihood transformed counts.

Lexical level model. The lexical level model (*dist-lex*) assigns to each nominal argument the verb role that it is semantically most similar to. Each role is represented by the semantic heads of its fillers. For example, suppose that the role STIMULUS of the verb *laugh* has been realised by the heads *story, scene, joke, and tale*. Then, in “*Peter’s laughter about the event*”, we analyse *event* as STIMULUS, since *event* is similar to these heads.

Formally, each argument head l is represented by a co-occurrence vector \vec{l} . A verb role $r_v \in R_v$ is modelled by the centroid \vec{r}_v of its instances’ heads:

$$\vec{r}_v = \frac{1}{|L_{r_v}|} \sum_{l \in L_{r_v}} \vec{l}$$

Roles are assigned to nominal argument heads $l_n \in L_n$ by finding the semantically most similar role r

⁵We also experimented with bag-of-words based vector spaces, which showed worse performance throughout.

while the grammatical function g_n is ignored:

$$r(l_n, g_n) = \operatorname{argmax}_{r_v \in R_v} \operatorname{sim}_{\cos}(\vec{l}_n, \vec{r}_v)$$

Function level model. The syntactic level model (*dist-fun*) generalises the lexical level model by exploiting the intuition that, within nouns, most semantic roles tend to be consistently realised by one specific grammatical function. This function can be identified as the one most semantically similar to the role’s representation. Following the example above, suppose that the grammatical function *PP-about* of *laughter* has as semantic heads the lexemes: *event*, *story*, *news*. Then, it is likely to express the role STIMULUS, as its heads are semantically similar to those of the verbal fillers of this role: *story*, *scene*, *sentence*, *tale*. For each nominalisation, this model constructs mappings (r_v, g_n) between a verbal semantic role r_v and a nominal grammatical function g_n . The representations for roles are computed as described above. We compute the semantic representations for grammatical functions, in parallel to the roles’ definition above, as the centroid of their fillers’ representations L_{g_n} :

$$\vec{g}_n = \frac{1}{|L_{g_n}|} \sum_{l \in L_{g_n}} \vec{l}$$

The assignment of a role to a nominal arguments is now determined by the argument’s grammatical function g_n ; its lemma l_n only enters indirectly, via the similarity computation:

$$r(l_n, g_n) = \operatorname{argmax}_{r_v \in R_v} \operatorname{sim}_{\cos}(\vec{r}_v, \vec{g}_n)$$

This strategy guarantees that each nominal grammatical function is mapped to exactly one role. In the inverse direction, roles can be left unmapped or mapped to more than one function.⁶

6.4 Hybrid models

Our last class combines the naive and distributional models with a back-off approach. We first attempt to harness the reliable naive syntactic approach whenever a mapping for the argument’s grammatical function is available. If this fails, it backs off to a distributional model. This strategy helps to recover the frequent AGENT- and PATIENT-style roles that cannot be recovered on syntactic grounds.

⁶We also experimented with a global optimisation strategy where we maximised the overall similarity between roles and functions subject to different constraints (e.g., perfect matching). Unfortunately, this strategy did not improve results.

		System		Accuracy
BL	baseline	random		17.09
	baseline	most common		42.97
Naive	naive syn			15.29
	naive syn-gen			21.56
	naive sem			24.00
Dist	dist-lex			44.57
	dist-fun			52.00
Hybrid	naive syn	+ dist-lex		48.22
	naive syn-gen	+ dist-lex		50.54
	naive syn	+ dist-fun		54.39
	naive syn-gen	+ dist-fun		56.42

Table 4: Results for nominal argument labelling

In (2), a hybrid model would assign the role STIMULUS to the argument headed by *event*, using the naive syntactic mapping (*PP-about_v*, *PP-about_n*) derived from (1a). For the prenominal modifier, no syntactic mapping is available; thus, it backs off to lexical-semantic evidence from (1a-b) to analyse *Peter* as COGNIZER.

We experiment with two hybrid models: naive syntactic plus lexical level distributional (*naive syn + dist-lex*), and naive syntactic plus functional level distributional (*naive syn + dist-fun*).

7 Experimental results

The results of our experiments are reported in Table 4. The models are compared against two baselines: A *random baseline* which randomly chooses one of the verb roles for each of the arguments of the corresponding noun; a *most common baseline* which assigns to each nominal argument the most frequent role of the corresponding verb – i.e. the role which has most fillers. All models with the exception of *naive syn* significantly outperform the random baseline, but only *dist-fun* and all hybrid models outperform the *most common* baseline.

In general, the best performing methods are the hybrid ones, with best accuracy achieved by *naive syn-gen + dist-fun*. Non-hybrid approaches always have lower accuracy. This validates our main hypothesis in this paper, namely that the combination of syntactic information with distributional semantics is a promising strategy.

Matching our predictions, the low accuracy of the naive syntactic model is mainly due to a lack of *coverage*. In fact, the model leaves 5,010 of the 6,502 gold standard noun fillers unassigned since they realise syntactic roles that are unseen for the verbs in question. A large part of these are *Pre-Modifier* and *PP-of* functions, which are central for nouns, but mostly ungrammatical for verbs. On

the 1,492 fillers for which a role was assigned, the model obtains an accuracy of 67%, indicating a reasonably high, but not perfect, accuracy for shared grammatical functions. The remaining errors stem from two sources. First, many grammatical functions are ambiguous, causing wrong assignments by a “syntax-only” model. For example, *PP-in* can indicate both TIME and PLACE for many nominalisations. Second, a certain number of grammatical functions do not preserve their role between verb to noun (Hull and Gomez, 1996). For example, *PP-to* realises the MESSAGE role of the verb *require* but the ADDRESSEE role of the noun *request*.

Distributional models show in general better performance than the naive syntactic approach (approx. +25% accuracy). They do not suffer from the coverage problem, since they assign a role to each filler. Yet, the accuracy over *assigned* roles is lower than for the syntactic approach (52% for dist-fun).

We conclude that in the limited cases where a pure syntactic mapping is applicable, it is far more reliable than methods which are mainly based on lexical-semantic information. The major limitation of the latter is that lexical-semantics tend to fail when roles are semantically very similar. For example, for the noun *announcement*, the syntactic-level distributional model wrongly builds the mapping (ADDRESSEE, *PP-by*) instead of (SPEAKER, *PP-by*), because the two roles are very similar semantically (the computed similarities of the *PP-by* arguments to ADDRESSEE and SPEAKER in the semantic space are 0.94 and 0.92, respectively).

The syntactic-level distributional model outperforms the lexical-level, suggesting that generalising the mapping at the argument level offers more stable statistical evidence to find the correct role, i.e. a *set* of noun arguments better defines the semantics of the mapping than a *single* argument. This is mostly the case when the context vector of the argument is not a good representation because the semantic head is ambiguous, infrequent or atypical. Consider, for example, the following sentence for the noun *violation*:

- (5) Sterne’s Tristram Shandy consists of a series of *violations* [of literary conventions]**PP-OF/NORM**

The syntactic-level model builds the correct mapping (NORM, *PP-of*), as the role fillers of the verb *violate* (e.g. *principle*, *right*, *treaty*, *law*) are very similar to the noun’s category fillers (e.g. *conven-*

tion, *rule*, *agreement*, *treaty*, *norm*), causing the centroids of NORM and *PP-of* to be close in the space. The lexical-level model, however, builds the incorrect mapping (PROTAGONIST, *convention*). This happens because *convention* is ambiguous, and one of its senses (“*a large formal assembly*”) is compatible with the PROTAGONIST role, and happens to have a large influence on the position of the vector for *convention*. Unfortunately, this is not the sense in which the word is used in this sentence.

8 Conclusions

We have presented a data expansion approach to SRL for event nominalisations. Instead of relying on manually annotated nominal training data, we harness annotated data for verbs to bootstrap a semantic role labeller for nouns. For argument recognition, we use a simple rule-based approach. For argument labelling, we profit from the fact that the argument structures of event nominalisations and the corresponding verbs are typically similar. This allows us to learn a mapping between verbal roles and nominal arguments, using syntactic features, lexical-semantic similarity, or both.

We found that our rule-based approach for argument recognition works fairly well. For argument labelling, our approach does not yet attain the performance of supervised models, but has the crucial advantage of not requiring any labelled data for nominal predicates.

We achieved the highest accuracy with a hybrid syntactic-semantic model, which indicates that both types of information need to be combined for optimal results. A purely syntactic approach results in a high precision, but low coverage because frequent grammatical functions in particular cannot be trivially mapped. Backing off to semantic similarity provides additional coverage. However, semantic similarity has to be considered on the level of complete functions rather than individual instances to promote “uniformity” in the mappings.

In this paper, we have only considered nominal SRL by *data expansion*, i.e. we only applied our approach to those nominalisations for which we have annotated data for the corresponding verbs. However, even if no data is available for the corresponding *verb*, it might still be possible to bootstrap from other verbs in the same *frame* (assuming that the frame is known for the nominalisation) and we plan to pursue this idea in future research. We also intend to investigate whether a joint optimisation of

the mapping constrained by additional syntactic information such as subcategorisation frames leads to better results. Finally, we will verify that our methods, which we have evaluated on English FrameNet data, carry over to other corpora and languages.

Acknowledgments. Our work was partly funded by the German Research Foundation DFG (grant PI 154/9-3).

References

- Baayen, R., R. Piepenbrock, and L. Gulikers, 1995. *The CELEX Lexical Database (Release 2)*. LDC.
- Carlson, G. 1984. Thematic roles and their role in semantic interpretation. *Linguistics*, 22:259–279.
- Carreras, X. and L. Màrquez, editors. 2005. *Proceedings of the CoNLL shared task: Semantic role labelling*, Ann Arbor, MI.
- Erk, K. and S. Padó. 2006. Shalmaneser – a flexible toolbox for semantic role assignment. In *Proceedings of LREC*, Genoa, Italy.
- Erk, K., A. Kowalski, S. Padó, and M. Pinkal. 2003. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In *Proceedings of ACL*, pages 537–544, Sapporo, Japan.
- Fillmore, C., C. Johnson, and M. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16:235–250.
- Fleischman, M. and E. Hovy. 2003. Maximum entropy models for FrameNet classification. In *Proceedings of EMNLP*, pages 49–56, Sapporo, Japan.
- Frank, A., H.-U. Krieger, F. Xu, H. Uszkoreit, B. Crysmann, B. Jörg, and U. Schäfer. 2007. Question answering from structured knowledge sources. *Journal of Applied Logic*, 5(1):20–48.
- Gildea, D. and D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- Gordon, A. and R. Swanson. 2007. Generalizing semantic role annotations across syntactically similar verbs. In *Proceedings of ACL*, pages 192–199, Prague, Czechia.
- Grenager, T. and C. Manning. 2006. Unsupervised discovery of a statistical verb lexicon. In *Proceedings of EMNLP*, pages 1–8, Sydney, Australia.
- Grimshaw, J. 1990. *Argument Structure*. MIT Press.
- Gurevich, O., R. Crouch, T. King, and V. de Paiva. 2006. Deverbal nouns in knowledge representation. In *Proceedings of FLAIRS*, pages 670–675, Melbourne Beach, FL.
- Hull, R. and F. Gomez. 1996. Semantic interpretation of nominalizations. In *Proceedings of AAAI*, pages 1062–1068, Portland, OR.
- Jaeger, T. 2008. Categorical data analysis: Away from ANOVAs and toward Logit Mixed Models. *Journal of Memory and Language*. To appear.
- Jiang, Zheng Ping and Hwee Tou Ng. 2006. Semantic role labeling of NomBank: A maximum entropy approach. In *Proceedings of EMNLP*, pages 138–145, Sydney, Australia.
- Lapata, M. 2002. The disambiguation of nominalizations. *Computational Linguistics*, 28(3):357–388.
- Liu, C. and H. Ng. 2007. Learning predictive structures for semantic role labeling of NomBank. In *Proceedings of ACL*, pages 208–215, Prague, Czechia.
- Macleod, C., R. Grishman, A. Meyers, L. Barrett, and R. Reeves. 1998. Nomlex: A lexicon of nominalizations. In *Proceedings of EURALEX*, Liège, Belgium.
- Meyers, A., R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. Annotating Noun Argument Structure for NomBank. In *Proceedings of LREC*, Lisbon, Portugal.
- Mihalcea, Rada and Phil Edmonds, editors. 2005. *Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, Barcelona, Spain.
- Moschitti, A., P. Morarescu, and S. Harabagiu. 2003. Open-domain information extraction via automatic semantic labeling. In *Proceedings of FLAIRS*, pages 397–401, St. Augustine, FL.
- Nunes, M. 1993. Argument linking in English derived nominals. In Valin, Robert D. Van, editor, *Advances in Role and Reference Grammar*, pages 372–432. John Benjamins.
- Padó, S. and M. Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.
- Palmer, M., D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.
- Pradhan, S., H. Sun, W. Ward, J. Martin, and D. Jurafsky. 2004. Parsing arguments of nominalizations in English and Chinese. In *Proceedings of HLT/NAACL*, pages 141–144, Boston, MA.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.
- Swier, R. and S. Stevenson. 2004. Unsupervised semantic role labelling. In *Proceedings of EMNLP*, pages 95–102.
- Xue, N. and M. Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of EMNLP*, pages 88–94, Barcelona, Spain.