

Cross-linguistic Projection of Role-Semantic Information

Sebastian Padó
Computational Linguistics
Saarland University
Saarbrücken, Germany
pado@coli.uni-sb.de

Mirella Lapata
School of Informatics
University of Edinburgh
Edinburgh, UK
mlap@inf.ed.ac.uk

Abstract

This paper considers the problem of automatically inducing role-semantic annotations in the FrameNet paradigm for new languages. We introduce a general framework for semantic projection which exploits parallel texts, is relatively inexpensive and can potentially reduce the amount of effort involved in creating semantic resources. We propose projection models that exploit lexical and syntactic information. Experimental results on an English-German parallel corpus demonstrate the advantages of this approach.

1 Introduction

Shallow semantic parsing, the task of automatically identifying the semantic roles conveyed by sentential constituents, has recently attracted much attention, partly because of its increasing importance for potential applications. For instance, information extraction (Surdeanu et al., 2003), question answering (Narayanan and Harabagiu, 2004) and machine translation (Boas, 2002) could stand to benefit from broad coverage semantic processing.

The FrameNet project (Fillmore et al., 2003) has played a central role in this endeavour by providing a large lexical resource based on semantic roles. In FrameNet, meaning is represented by frames, schematic representations of situations. Semantic roles are frame-specific, and are called frame elements. The database associates frames with lemmas (verbs, nouns, adjectives) that can evoke them (called frame-evoking elements or FEEs), lists the possible syntactic realisations of their semantic roles, and provides annotated examples from the British National Corpus (Burnard, 1995). The availability of rich annotations for the surface realisation of semantic roles has triggered interest in semantic parsing and enabled the development of data-driven models (e.g., Gildea and Jurafsky, 2002).

Frame: DEPARTING	
Frame Elements	THEME The officer left the house. The plane leaves at seven. His departure was delayed.
	SOURCE We departed from New York . He retreated from his opponent . The woman left the house .
FEEs	abandon.v, desert.v, depart.v, departure.n, emerge.v, emigrate.v, emigration.n, escape.v, escape.n, leave.v, quit.v, retreat.v, retreat.n, split.v, withdraw.v, withdrawal.n

Table 1: Example of FrameNet frame

Table 1 illustrates an example from the FrameNet database, the DEPARTING frame. It has two roles, a THEME which is the moving object and a SOURCE expressing the initial position of the THEME. The frame elements are realised by different syntactic expressions. For instance, the THEME is typically an NP, whereas the SOURCE is often expressed by a prepositional phrase (see the expressions in boldface in Table 1). The DEPARTING frame can be evoked by *abandon*, *desert*, *depart*, and several other verbs as well as nouns (see the list of FEEs in Table 1).

Although recent advances in semantic parsing¹ have greatly benefited from the availability of the English FrameNet, unfortunately such resources are largely absent for other languages. The English FrameNet (Version 1.1) contains 513 frames covering 7,125 lexical items and has been under development for approximately six years. Although FrameNets are currently under construction for German, Spanish, and Japanese, these resources are still in their infancy and of limited value for modelling purposes. Methods for acquiring FrameNets from corpora automatically would greatly reduce the human effort involved and facilitate their development for new languages.

In this paper, we propose a method which employs parallel corpora for acquiring frame elements

¹Approaches to modelling semantic parsing are too numerous to list; see Carreras and Màrquez (2005) for an overview.

and their syntactic realisations (see the upper half of Table 1) for new languages. Our method leverages the existing English FrameNet to overcome the resource shortage in other languages by exploiting the translational and structural equivalences present in aligned data. The idea underlying our approach can be summarised as follows: (1) given a pair of sentences *E* (English) and *L* (new language) that are translations of each other, annotate *E* with semantic roles; and then (2) *project* these roles onto *L*. In this manner, we induce semantic structure on the *L* side of the parallel text, which can then serve as data for training a statistical semantic parser for *L* that is independent of the parallel corpus.

We first assess if the main assumption of semantic projection is warranted (Section 3), namely whether frames and semantic roles exhibit a high degree of parallelism across languages. Then we propose two broad classes of projection models that utilise lexical and syntactic information (Section 4), and show experimentally that roles can be projected from English onto German with high accuracy (Section 5). We conclude the paper by discussing the implications of our results and future work (Section 6).

2 Related work

A number of recent studies exploit parallel corpora for cross-linguistic knowledge induction. In this paradigm, annotations for resource-rich languages like English are projected onto another language through aligned parallel texts. Yarowsky et al. (2001) propose several projection algorithms for deriving monolingual tools (ranging from part-of-speech taggers, to chunkers and morphological analysers) without additional annotation cost. Hwa et al. (2002) assess the degree of syntactic parallelism in dependency relations between English and Chinese. Their results show that, although assuming direct correspondence is often too restrictive, syntactic projection yields good enough annotations to train a dependency parser. Smith and Smith (2004) explore syntactic projection further by proposing an English-Korean bilingual parser integrated with a word translation model.

Previous work has primarily focused on the projection of morphological and grammatico-syntactic information. Inducing semantic resources from low density languages still poses a significant challenge to data-driven methods. The challenge is recognised by Fung and Chen (2004) who construct a Chinese FrameNet by mapping English FrameNet entries to

concepts listed in HowNet², an on-line ontology for Chinese, however without exploiting parallel texts.

The present work extends previous approaches on annotation projection by inducing FrameNet semantic roles from parallel corpora. Analogously to Hwa et al. (2002), we investigate whether there are indeed semantic correspondences between two languages, since there is little hope for projecting meaningful annotations in nonparallel semantic structures. Similarly to Fung and Chen (2004) we automatically induce semantic role annotations for a target language. In contrast to them, we resort to parallel corpora as a source of semantic equivalence. Thus, we avoid the need for a target concept dictionary in addition to the English FrameNet. We propose a general framework for semantic projection that can incorporate different knowledge sources. To our knowledge, the framework and its application to semantic role projection are novel.

3 Creation of a Gold Standard Corpus

Sample Selection. To evaluate the output of our projection algorithms, we created a gold standard corpus of English-German sentence pairs with manual FrameNet frame and role annotations. The sentences were sampled from Europarl (Koehn, 2002), a corpus of professionally translated proceedings of the European Parliament. Europarl is available in 11 languages with up to 20 million words per language aligned at the document and sentence level.

Recall that frame projection is only meaningful if the same frame is appropriate for both sentences in a projection pair. This constrains sample selection for two reasons: first, FrameNet is as yet incomplete with respect to its coverage. So, a randomly selected sentence pair may evoke novel frames or novel senses of already existing frames (e.g., the “greeting” sense of *hail* which is currently not listed in FrameNet). Second, due to translational variance, there is no a priori guarantee that words which are mutual translations evoke the same frame. For example, the English verb *finish* is often translated in German by the adverb *abschließend*, which arguably cannot have a role set identical to *finish*. Relying solely on the English FrameNet database for sampling would yield many sentence pairs which are either inappropriate for the present study (because they do not evoke the same frames) or simply problematic for annotation since they are outside the

²See http://www.keenage.com/zhiwang/e_zhiwang.html.

present coverage of the database.

For the above reasons, our sample selection procedure was informed by two existing resources, the English FrameNet and SALSA, a FrameNet-compatible database for German currently under development (Erk et al., 2003). We first used the publicly available GIZA++ (Och and Ney, 2003) software to induce English-German word alignments. Next, we gathered all German-English sentences in the corpus that had at least one pair of aligned words (w_e, w_g) , which were listed in FrameNet and SALSA, respectively, and had at least one frame in common. These sentences exemplify 83 frame types, 696 lemma pairs, and 265 unique English and 178 unique German lemmas. Sentence pairs were grouped into three bands according to their frame frequency (High, Medium, Low). We randomly selected 380 pairs from each band. The total sample consisted of 140 sentence pairs.

This procedure produces a realistic corpus sample for the role projection task; similar samples can be drawn for new language pairs using either existing bilingual dictionaries (Fung and Chen, 2004) or automatically constructed semantic lexicons (Padó and Lapata, 2005).

Annotation. Two annotators, with native-level proficiency in German and English, manually labelled the parallel corpus with semantic information. Their task was to identify the frame for a given predicate in a sentence, and assign the corresponding roles. They were provided with detailed guidelines that explained the task using multiple examples. During annotation, they had access to parsed versions of the sentences in question (see Section 5 for details), and to the English FrameNet and SALSA.

The annotation proceeded in three phases: a training phase (40 sentences), a calibration phase (100 sentences), and a production mode phase (1000 sentences). In the calibration phase, sentences were doubly annotated to assess inter-annotator agreement. In production mode, sentences were split into two distinct sets, each of which was annotated by a single coder. We ensured that no annotator saw both parts of any sentence pair to guarantee independent annotation of the bilingual data. Each coder annotated approximately the same amount of data in English and German.

Table 2 shows the results of our inter-annotator agreement study. In addition to the widely used Kappa statistic, we computed a number of different agreement measures: the ratio of frames common

Measure	English	German	All
Frame Match	0.90	0.87	0.88
Role Match	0.95	0.95	0.95
Span Match	0.85	0.83	0.84
Kappa	0.86	0.90	0.87

Table 2: Monolingual inter-annotation agreement on the calibration set

Measure	Precision	Recall	F-score
Frame Match	0.72	0.72	0.72
Role Match	0.91	0.92	0.91

Table 3: Cross-lingual semantic parallelism between English and German

between two sentences (Frame Match), the ratio of common roles (Role Match), and the ratio of roles with identical spans (Span Match). As can be seen, annotators tend to agree in frame assignment; disagreements are mainly due to fuzzy distinctions between frames (e.g., between AWARENESS and CERTAINTY). As can be seen from Table 2, annotators agree in what roles to assign (Role Match is 0.95 for both English and German); agreeing on their exact spans is a harder problem.

Semantic Parallelism. Since we obtained parallel FrameNet annotations for English and German, we were able to investigate the degree of semantic parallelism between the two languages. More specifically, we treated the German annotation as gold standard against which we compared the English annotations. To facilitate comparisons with the output of our automatic projection methods (see Section 4), we measured parallelism using precision and recall. Frames and frame roles were counted as matching if they were annotated in a sentence, regardless of their spans. The results are shown in Table 3.

The cross-lingual data exhibit more than twice the amount of frame differences than monolingual data (compare Tables 2 and 3). This indicates that frame disambiguation methods must be employed in automatic role projection to ensure that two aligned tokens evoke the same frame. However, frame disambiguation is outside the scope of the present paper.

On the positive side, role agreement is relatively high (0.91 F-score). This indicates that in cases where frames match across languages, semantic roles could be accurately transferred (provided that these languages diverge little in their argument structure). This observation offers support for the

projection approach put forward in this paper. Note, however, that a practical projection system could attain this level of performance only if it could employ an oracle to recover annotators’ decisions about the span of roles. We can obtain a more realistic upper bound for an automatic system from the monolingual Role Span agreement figure (F-score 0.84). The latter represents a ceiling for the agreement we can expect from sentences annotated by different annotators.

4 Projection of Semantic Information

In this section, we formalise the semantic projection task and give the details of our modelling approach. All models discussed here project semantic annotations from a source language to a target language. As explained earlier, our present study is only concerned with the projection of roles between matching frames.

4.1 Problem Formulation

We assume that we are provided with source and target sentences represented as sets of entities $e_s \in E_s$ and $e_t \in E_t$. These entities can be words, constituents, phrases, or other groupings. In addition, we are given the semantic annotation of the source sentences from which we can directly read off the source semantic role assignment $a_s : R \rightarrow 2^{E_s}$, where R is the set of semantic roles. The goal of the projection is to specify the target semantic role assignments $a_t : R \rightarrow 2^{E_t}$, which are unknown.³

Clearly, effecting the projection requires establishing some form of match between the source and target entities. We therefore formalise projection as a function which maps the source role assignment and a set of matches $M \subseteq E_s \times E_t$ onto a new target role assignment:

$$proj : (A_s \times M) \rightarrow (R \rightarrow 2^{E_t}) \quad (1)$$

By way of currying, we can state the new target role assignment as a function which directly computes a set of target entities, given the source role assignment, a set of entity matches, and a role:

$$a_t : (A_s \times M \times R) \rightarrow 2^{E_t} \quad (2)$$

According to this formalisation, the crucial part of semantic projection is to identify a correct and exhaustive set of entity matches. Obviously, this raises

³Without loss of generality, we limit ourselves to one frame per sentence, as does FrameNet.

$r \in R$	Semantic role
$t_s \in T_s, t_t \in T_t$	Source, target tokens
$al \in Al : T_s \rightarrow 2^{T_t}$	Word alignment
$a_s \in A_s : R \rightarrow 2^{T_s}$	Source role assignment
$a_t : (A_s \times Al \times R) \rightarrow 2^{T_t}$	Projected target role assignment

Table 4: Notation and signature summary for word-based projection

the question of what linguistic information is appropriate for establishing M . Unfortunately, any attempt to compute a match based on categorical data derived from linguistic analyses (e.g., parts of speech, phrase types or grammatical relations), needs to empirically derive cross-linguistic similarities between categories, a task which must be repeated for every new language pair, and requires additional data.

Rather than postulating an ad hoc similarity function, we use word alignments to derive information about semantic roles in the target language. Our first model family (Section 4.2) relies exclusively on this knowledge source. Although potentially useful as a proxy for semantic equivalence, automatically induced alignments are often noisy, thus leading to errors in annotation projection (Yarowsky et al., 2001). For example, function words commonly diverge across languages and are systematically misaligned; furthermore, alignments are restricted to single words rather than word combinations. This observation motivates a second model family with a bias towards linguistically meaningful entities (Section 4.3). Such entities can be constituents derived from the output of a parser or non-recursive syntactic structures (i.e., chunks).

In this paper we compare simple word alignment models against more resource intensive models that utilise constituent-based information and examine whether syntactic knowledge significantly contributes to semantic projection.

4.2 Word-based Projection Model

The first model family uses source and target *word tokens* as entities for projection. In this framework, projection models can be defined by deriving the set of matches M directly from word alignments. The resulting signatures are shown in Table 4.

Our first projection model assigns to each role r with source span $s(r)$ the set of all target tokens which are aligned to a token in the source span:

$$a_w(a_s, al, r) = \bigcup_{t_s \in a_s(r)} al(t_s) \quad (3)$$

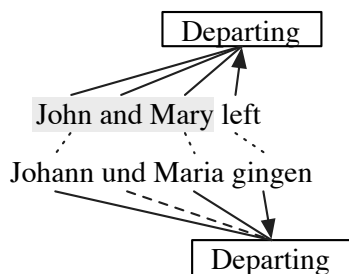


Figure 1: Word alignment-based semantic projection of Role THEME (shadowed), Frame DEPARTING

The main shortcoming of this model is that it cannot capture an important linguistic property of semantic roles, namely that they almost always cover contiguous stretches of text. We can repair non-contiguous projections by applying a “convex complementing” heuristic to the output of (3), which fills all holes in a sequence of tokens, without explicit recourse to syntactic information. We define the convex complementing heuristic as:

$$a_{cw}(a_s, al, r) = \{t_t \mid \min(i(a_{t_1})) \leq i(t_t) \leq \max(i(a_{t_n}))\} \quad (4)$$

where i returns the index of a token t .

The two models just described are illustrated in Figure 1. The frame DEPARTING is introduced by *left* and *gingen* in English and German, respectively. For simplicity, we only show the edges corresponding to the THEME role. In English, the THEME is realised by the words *John and Mary*. The dotted lines show the available word alignments. The projection of the THEME role according to (3) consists only of the tokens $\{Johann, Maria\}$ (shown by the plain black lines); the convex complementing heuristic in model (4) adds the token *und*, resulting in the (correct) convex set $\{Johann, und, Maria\}$.

4.3 Constituent-based Projection Model

Our second model family attempts to make up for errors in the word alignment by projecting from and to *constituents*. In this study, our constituents are obtained from full parse trees (see Section 5 for details). Models which use non-recursive structures are also possible; however, we leave this to future work.

The main difference from word-based projection models is the introduction of constituent information as an intermediate level; we thus construct a *constituent alignment* for which only a subset of word

alignments has to be accurate. The appropriate signatures and notation for constituent-based projection are summarised in Table 5.

In order to keep the model as flexible as possible, and to explore the influence of different design decisions, we model constituent-based projection as two independently parameterisable subtasks: first we compute a real-valued *similarity function* between source and target constituents; then, we employ the similarity function to align relevant constituents and project the role information.

Similarity functions. In principle, any function which matches the signature in Table 5 could be used. In practice, the use of linguistic knowledge runs into the problem of defining similarity between category-based representations discussed above. For this reason, we limit ourselves to two simple similarity functions based on word overlap: Given source and target constituents c_s and c_t , we define the *word overlap* o_w of c_s with c_t as the proportion of tokens within c_t aligned to tokens within c_s . Let $yield(c)$ denote the set of tokens in the yield of a constituent c , then:

$$o_w(c_s, c_t) = \frac{|\bigcup_{t_s \in yield(c_s)} al(t_s) \cap yield(c_t)|}{|yield(c_t)|} \quad (5)$$

Since the asymmetry of this overlap measure leads to high overlap scores for small target constituents, we define *word overlap similarity*, as the product of two constituents’ mutual overlap:

$$sim(c_s, c_t) = o(c_s, c_t) \cdot o(c_t, c_s) \quad (6)$$

Simple word-based overlap has one undesired characteristic: larger constituents tend to be less similar because of missing alignments (e.g., between function words). Since content words are arguably more important for the role projection task, we define a second overlap measure, *content word overlap* o_{wc} , which takes only nouns, verbs and adjectives into account. Let $yield_c(c)$ denote the set of tokens in the yield of c that are content words, then:

$$o_{wc}(c_s, c_t) = \frac{|\bigcup_{t_s \in yield_c(c_s)} al(t_s) \cap yield_c(c_t)|}{|yield_c(c_t)|} \quad (7)$$

Constituent alignment. Considerable latitude is available in interpreting a similarity function to derive a constituent alignment. Due to space limitations, we demonstrate two basic models.

Our first *forward constituent alignment* model (a_{fc}), aligns source constituents that form the span

$r \in R$	Semantic role
$c_s \in C_s, c_t \in C_t$	Source and target constituents
$yield : C \rightarrow T$	Yield of a constituent
$yield_c : C \rightarrow T$	Content word yield of a constituent
$al \in Al : T_s \rightarrow 2^{T_t}$	Word alignment
$a_s \in A_s : R \rightarrow 2^{C_s}$	Source role assignment
$sim : C_s \times C_t \rightarrow \mathbb{R}^+$	Constituent similarity
$a_t : A_s \times Sim \times R \rightarrow 2^{C_t}$	Projected target role assignment

Table 5: Notation and signature summary for constituent-based projection

of a role to a single target constituent. We compute the similarity of a target constituent c_t to a set of source constituents $c_s \in a_s(r)$ by taking the product similarity for each source and target constituent pair:

$$a_{fc}(a_s, sim, r) = \operatorname{argmax}_{c_t \in C_t} \prod_{c_s \in a_s(r)} sim(c_s, c_t) \quad (8)$$

This projection model forces the target role assignment to be a function, i.e., it makes the somewhat simplifying assumption that each role corresponds to a single target constituent.

Our second *backward constituent alignment* model (a_{bc}) proceeds in the opposite direction: it iterates over target constituents and attempts to determine their most similar source constituent for each c_t . If the aligned source constituent is labelled with a role, it is projected onto c_t :

$$a_{bc}(a_s, sim, r) = \{c_t | (\operatorname{argmax}_{c_s \in C_s} sim(c_s, c_t)) \in a_s(r)\} \quad (9)$$

In general, a_{bc} allows for more flexible role projection: it will sometimes decide not to project a role at all (if the source constituents are dissimilar to any target constituents), or it can assign a role to more than one target constituent; however, this means that there is less control over what is projected, and wrong alignments can lead to wrong results more easily.

Finally, if no word alignments are found for complete source or target constituents, the maximal similarity rating in a_{bc} or a_{bf} will be zero. This is often the case for semantically weak single-word constituents such as demonstrative pronouns (e.g., *[That] is right./ [Das] ist richtig.*). When we observe this phenomenon, we heuristically skip unaligned constituents (*zero skipping*).

Figure 2 contrasts the two constituent-based projection models using the frame QUESTIONING as

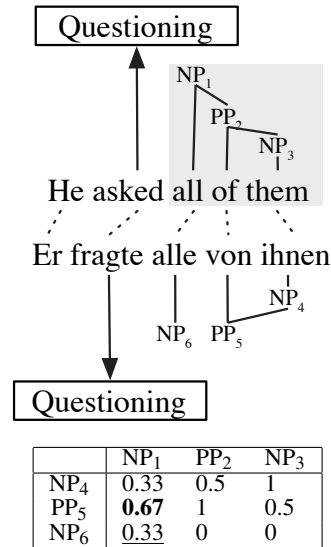


Figure 2: Constituent-based semantic projection of role ADDRESSEE (shaded), frame QUESTIONING. Below: Constituent similarity matrix.

an example. Again, we only show one role, ADDRESSEE, indicated by the shaded box in Figure 2. Note that the object NP in German was mis-parsed as an NP and a PP, a relatively frequent error. The difference between the two decision procedures can be explained straightforwardly by looking at the table below the graph, which shows the similarity matrix for the constituents according to equation (6). In this table, the source constituents (indices 1–3) correspond to columns, and the target constituents (indices 4–6) to rows. The alignment model in (8) iterates over labelled source constituents (here only NP₁) and chooses the row with the highest value as the target constituent for a candidate role. In our case, this is the PP₅ (cell in bold-face). In contrast, model (9) iterates over all target constituents (i.e., rows) and checks if the most similar source constituent bears a role label. Since NP₁ is the most similar constituent for NP₆ (underlined cell), (9) assigns the QUESTIONING role to NP₆.

5 Experiments

Evaluation Framework. We implemented the models described in the previous section and used them to project semantic information from English onto German. For the constituent-based models, constituent information was obtained from the output of Collins’ parser (1997) for English and Dubey’s parser (2004) for German. Words were

Model	Precision	Recall	F-score
<i>w</i>	0.41	0.40	0.41
<i>cw</i>	0.46	0.45	0.46
Upper bound	0.85	0.84	0.84

Table 6: Results for word-based projection models

aligned using the default setting⁴ of GIZA++ (Och and Ney, 2003), a publicly available implementation of the IBM models and HMM word alignment models. We evaluated the projected roles against the “gold standard” roles obtained from the manual annotation (see Section 3). We also compared our results to the upper bound given by the inter-annotator agreement on the calibration data set.

Results. Table 6 shows our results for the word-based projection models. The simplest word-based model (*a_w*), obtains an F-score of 0.41. This is a good result considering that the model does not exploit any linguistic information (e.g., parts of speech or syntactic structure). It also supports our hypothesis that word alignments are useful for the role projection task. The convex complementing heuristic (*a_{cw}*) delivers an F-score increase of five points over the “words only” model, simply by making up for holes in the word alignment.

We evaluated eight instantiations of the constituent-based projection models; the results are shown in Table 7. The best model (in boldface) uses forward constituent alignment, content word-based overlap similarity, and zero skipping. We observe that backward constituent alignment-based models (1–4) perform similarly to word-based projection models (the F-score ranges between 0.40 and 0.45). However, they obtain considerably higher precision (albeit lower recall) than the word-based models. This may be an advantage if the projected data is destined for training target-language semantic parsers. This precision/recall pattern appears to be a direct result of *a_{bc}*, which only projects a role from *c_s* to *c_t* if *c_s* “wins” against all other source constituents, thus resulting in reliable, but overly cautious projections, which cannot not be further improved by zero skipping.

The forward constituent alignment models (5–8) show consistently higher performance than word-based models and models 1–4, indicating that the stronger assumptions made by forward alignment

⁴The training scheme involved five iterations of Model 1, five iterations of the HMM model, five iterations of Model 3, and five iterations of Model 4.

Model	<i>al</i>	<i>o</i>	0-skip	Precision	Recall	F-score
1	<i>bc w</i>	no		0.70	0.33	0.45
2	<i>bc w</i>	yes		0.70	0.33	0.45
3	<i>bc wc</i>	no		0.65	0.32	0.42
4	<i>bc wc</i>	yes		0.65	0.32	0.42
5	<i>fc w</i>	no		0.61	0.60	0.60
6	<i>fc w</i>	yes		0.66	0.60	0.63
7	<i>fc wc</i>	no		0.62	0.60	0.61
8	<i>fc wc</i>	yes		0.70	0.60	0.65
Upper bound				0.85	0.84	0.84

Table 7: Results for constituent-based projection models (*al*: constituent alignment model; *o*: overlap measure; 0-skip: zero skipping)

are justified in the data. In addition, we also find that we can increase precision by concentrating on reliable alignments. This is achieved by using the zero skipping heuristic (compare the odd vs. even-numbered models in Table 7) and by computing overlap on content words (compare Models 6 vs. 8, and 5 vs. 7).

We used the χ^2 test to examine whether the differences observed between the two classes of models are statistically significant. The best constituent-based model significantly outperforms the best word-based model both in terms of precision ($\chi^2 = 114.47$, $p < 0.001$) and recall ($\chi^2 = 400.40$, $p < 0.001$). Both projection models perform significantly worse than humans ($p < 0.001$).

Discussion. Our results confirm that constituent information is important for the semantic projection task. Our best model adopts a conservative strategy which enforces a one-to-one correspondence between roles and target constituents. This strategy leads to high precision, however recall lags behind (see Model 8 in Table 7). Manual inspection of the projection output revealed that an important source of missing roles are word alignments gaps. Such gaps are not only due to noisy alignments, but also reflect genuine structural differences between translated sentences. Consider the following (simplified) example for the STATEMENT frame (introduced by say) and its semantic role STATEMENT (introduced by we):

- (10) We claim X and we say Y
Wir behaupten X und — sagen Y

The word alignment correctly aligns the German pronoun *wir* with the first English *we* and leaves

the second occurrence unaligned. Since there is no corresponding German word for the second *we*, projection of the *SPEAKER* role fails. In future work, this problem could be handled with explicit identification of empty categories (see Dienes and Dubey, 2003).

6 Conclusions

In this paper, we argue that parallel corpora show promise in relieving the lexical acquisition bottleneck for low density languages. We proposed semantic projection as a means of obtaining FrameNet annotations automatically without additional human effort. We examined semantic parallelism, a prerequisite for accurate projection, and showed that semantic roles can be successfully projected for predicate pairs with matching frame assignments. Similarly to previous work (Hwa et al., 2002), we find that some mileage can be gained by assuming direct correspondence between two languages. However, linguistic knowledge is key in obtaining meaningful projections. Our experiments show that the use of constituent information yields substantial improvements over relying on word alignment alone. Nevertheless, the word-based models offer a good starting point for low-density languages for which parsers are not available. Their output could be further post-processed manually or automatically using bootstrapping techniques (Riloff and Jones, 1999).

We have presented a general, flexible framework for semantic projection which can be easily applied to other languages. An important direction for future work lies in the assessment of more shallow syntactic information (i.e., chunks) which can be obtained more easily for new languages, and generally in the integration of more linguistic knowledge to guide projection. Finally, we will incorporate into our projection approach automatic semantic role annotations for the source language and investigate the potential of the projected annotations for training semantic parsers for the target language.

Acknowledgements. The authors acknowledge the support of DFG (Padó; grant PI-154/9-2) and EPSRC (Lapata; grant GR/T04540/01). Thanks to B. Kouchnir and P. Kreischer for their annotation.

References

H. C. Boas. 2002. Bilingual framenet dictionaries for machine translation. In *Proceedings of LREC 2002*, 1364–1371, Las Palmas, Canary Islands.

- L. Burnard, 1995. *The Users Reference Guide for the British National Corpus*. British National Corpus Consortium, Oxford University Computing Service, 1995.
- X. Carreras, L. Màrquez, eds. 2005. *Proceedings of the CoNLL shared task: Semantic role labelling*, 2005.
- M. Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Proceedings of ACL/EACL 1997*, 16–23, Madrid, Spain.
- P. Dienes, A. Dubey. 2003. Antecedent recovery: Experiments with a trace tagger. In *Proceedings of EMNLP 2003*, 33–40, Sapporo, Japan.
- A. Dubey. 2004. *Statistical parsing for German: Modelling syntactic properties and annotation differences*. Ph.D. thesis, Saarland University.
- K. Erk, A. Kowalski, S. Padó, M. Pinkal. 2003. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In *Proceedings of ACL 2003*, 537–544, Sapporo, Japan.
- C. J. Fillmore, C. R. Johnson, M. R. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16:235–250.
- P. Fung, B. Chen. 2004. BiFrameNet: Bilingual frame semantics resources construction by cross-lingual induction. In *Proceedings of COLING 2004*, 931–935, Geneva, Switzerland.
- D. Gildea, D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- R. Hwa, P. Resnik, A. Weinberg, O. Kolak. 2002. Evaluation translational correspondence using annotation projection. In *Proceedings of ACL 2002*, 392–399, Philadelphia, PA.
- P. Koehn. 2002. Europarl: A multilingual corpus for evaluation of machine translation. Draft.
- S. Narayanan, S. Harabagiu. 2004. Question answering based on semantic structures. In *Proceedings of COLING 2004*, 693–701, Geneva, Switzerland.
- F. J. Och, H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–52.
- S. Padó, M. Lapata. 2005. Cross-lingual bootstrapping for semantic lexicons. In *Proceedings of AAAI 2005*, Pittsburgh, PA.
- E. Riloff, R. Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of AAAI 1999*, Orlando, FL.
- D. A. Smith, N. A. Smith. 2004. Bilingual parsing with factored estimation: Using English to parse Korean. In *Proceedings of EMNLP 2004*, 49–56, Barcelona, Spain.
- M. Surdeanu, S. Harabagiu, J. Williams, P. Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proceedings of ACL 2003*, 8–15, Sapporo, Japan.
- D. Yarowsky, G. Ngai, R. Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of HLT 2001*, 161–168.