Multilingual Aliasing for Auto-Generating Proposition Banks

Alan Akbik
IBM Research Almaden
650 Harry Road, San Jose
CA 95120, USA
akbika@us.ibm.com

Xinyu Guan
Yale University
82-90 Wall Street, New Haven
CT 06520, USA
xinyu.guan@yale.edu

Yunyao Li
IBM Research Almaden
650 Harry Road, San Jose
CA 95120, USA
yunyaoli@us.ibm.com

Abstract

Semantic Role Labeling (SRL) is the task of identifying the predicate-argument structure in sentences with semantic frame and role labels. For the English language, the Proposition Bank provides both a lexicon of all possible semantic frames and large amounts of labeled training data. In order to expand SRL beyond English, previous work investigated automatic approaches based on parallel corpora to automatically generate Proposition Banks for new target languages (TLs). However, this approach heuristically produces the frame lexicon from word alignments, leading to a range of lexicon-level errors and inconsistencies. To address these issues, we propose to manually *alias* TL verbs to existing English frames. For instance, the German verb *drehen* may evoke several meanings, including ”turn something” and ”film something”. Accordingly, we alias the former to the frame TURN.01 and the latter to a group of frames that includes FILM.01 and SHOOT.03. We execute a large-scale manual aliasing effort for three target languages and apply the new lexicons to automatically generate large Proposition Banks for Chinese, French and German with manually curated frames. We present a detailed evaluation in which we find that our proposed approach significantly increases the quality and consistency of the generated Proposition Banks. We release these resources to the research community.

1 Introduction

Semantic role labeling (SRL) is the task of labeling predicate-argument structure in sentences with shallow semantic information. The prominent labeling scheme for the English language is the Proposition Bank (Palmer et al., 2005), which provides a lexicon of possible *frames* for English verbs. Each frame corresponds to one semantic interpretation and comes with frame-specific *role* labels and descriptions. Over the past decade, large amounts of text data have been annotated based on these guidelines (Palmer et al., 2005; Hovy et al., 2006; Bonial et al., 2014). They enable the training of statistical SRL systems, which have proven useful for downstream applications such as information extraction (IE) (Fader et al., 2011), question answering (QA) (Shen and Lapata, 2007; Maqsud et al., 2014) and machine translation (Lo et al., 2013).

However, such manual efforts are known to be highly costly. Possible frames need to be manually determined, their roles individually described, and large volumes of text data annotated accordingly. For this reason, Proposition Banks do not exist for most languages.

Annotation projection. Recent research has explored the possibility of using annotation projection to automatically generate Proposition Banks from parallel corpora for new target languages (TL) (Padó and Lapata, 2009; Van der Plas et al., 2011; Akbik et al., 2015). This approach requires a large word-aligned corpus of English sentences and their TL translations. An English SRL system predicts semantic labels for the English sentences. These labels are then transferred along word alignments to automatically annotate the TL corpus. Recent work has shown that such auto-generated Proposition Banks can be used to train semantic role labelers for a wide range of languages (Akbik and Li, 2016).

Heuristically generated frame lexicon. However, a major drawback of such approaches is that the frame lexicon is heuristically produced from available alignments, which leads to a range of errors and inconsistencies. Consider the German verb *drehen*. In the English-German portion of the OPENSUB-

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 3466–3474, Osaka, Japan, December 11-17 2016.

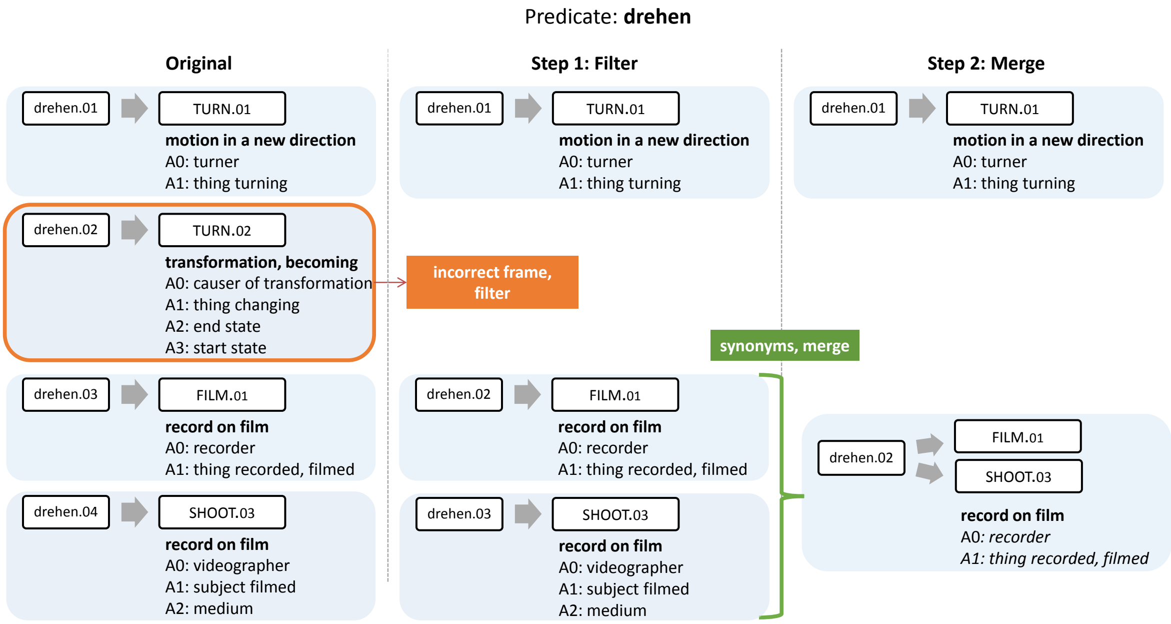

Figure 1: Illustration of merging and filtering steps over heuristically produced frame lexicon. This process reduces the number of distinct frames for the verb *drehen* from 4 to 2.

TITLES2016 parallel corpus (Lison and Tiedemann, 2016), this verb is aligned to many different English frames, four of which are illustrated in Figure 1. Previous annotation projection approaches treat each alignment as a separate and valid frame of the TL verb. In this example, it is therefore heuristically determined that *drehen* may evoke four separate frames, namely TURN.01, TURN.02, FILM.01 and SHOOT.03. See Figure 1 (left pane) for an illustration and explanation of the four frames. The example illustrates the two main problems in heuristically produced lexicons:

Incorrect frames The first problem is posed by errors in the frame lexicon. For instance, the German verb *drehen* cannot evoke the frame TURN.02 (transformation). The corresponding entry in the lexicon is therefore incorrect. This lexicon-level error has a significant impact on the generated Proposition Bank since every TL sentence with this annotation is incorrectly labeled. Therefore, as the middle pane in Figure 1 illustrates, we wish to filter out such lexicon-level errors.

Redundant frames The second problem is posed by redundancy that occurs if multiple entries for a TL verb are in fact synonyms. For instance, *drehen* is heuristically determined to evoke SHOOT.03 (record on film) and FILM.01 (record on film) as two separate meanings. However, the TL usages of *drehen* in these contexts are clearly identical. This lexicon-level error causes inconsistent annotation of the same semantics throughout the generated Proposition Bank. Therefore, as the right pane in Figure 1 illustrates, we wish to merge these two entries into a single entry comprising both frames.

In this paper, we propose to address these issues by manually curating incorrect alignments and grouping synonymous English frames with a process of filtering and merging as illustrated in Figure 1. With this approach, we effectively follow a process of *aliasing* TL verbs to English frames (Bonial et al., 2014; Jagfeld and van der Plas, 2015). Our goal is to remove lexicon-level errors and redundancies in order to generate higher quality TL Proposition Banks with consistent annotation and salient verb senses.

Contributions Our contributions are: 1) We propose a method for manually curating a heuristically determined frame lexicon and discuss our curation guidelines. 2) We execute our method over large-scale parallel data for three target languages (Chinese, French and German) to automatically generate Proposition Banks with curated frame lexicons. 3) We present an experimental evaluation in which we find that our proposed approach significantly increases the quality of automatically generated Proposition Banks and greatly reduces redundancy. 4) We analyze the verb coverage of the generated lexicons and

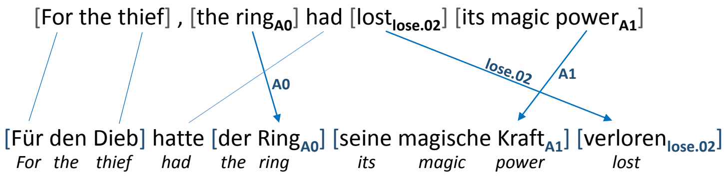

Figure 2: Example of annotation projection for an English-German sentence pair. English frame (LOSE.02) and role labels (A0, A1) are projected onto aligned German words.

conduct a comparison of our work against manual efforts to create Proposition Banks. 5) Finally, we release all resources to the research community for the training of multilingual SRL and the study of crosslingual semantics[1].

2 Related Work

Annotation projection Annotation projection takes as input a word-aligned parallel corpus of English sentences and their target language (TL) translations. An English semantic role labeler is used to predict semantic labels for the English sentences. These labels are then projected onto aligned TL words, automatically producing a TL corpus annotated with English frame and role labels. Refer to Figure 2 for illustration.

The use of annotation projection to train parsers for new languages was first introduced in the context of learning a PoS tagger (Yarowsky et al., 2001). Initial work on projecting semantic labels used FrameNet (Padó and Lapata, 2009), but subsequent work has focused on PropBank annotation due to its broader coverage and the availability of high quality semantic role labelers for English (Van der Plas et al., 2011; van der Plas et al., 2014). Recent work has focused on increasing the accuracy of projected labels by scaling up projection to larger corpora and retraining SRL models (Van der Plas et al., 2011; van der Plas et al., 2014) as well as using filtering techniques to block labels most likely affected by translation shift (Akbik et al., 2015). The latter found that the largest portion of errors in generated PropBanks results from incorrectly predicted labels for English sentences, which are then projected onto TL sentences, thereby propagating this error.

Consistency of the frame lexicon. However, so far, previous work has not investigated the overall correctness and consistency of the frame lexicon. All previous annotation projection efforts treat each distinct global alignment as a different sense of each verb, which, as argued in section 1, is not the case. In this paper, we specifically address this issue.

Aliasing Our work is similar to recent efforts in *aliasing*, in which existing English verb frames are reused for new types of frame evoking elements (Bonial et al., 2014). One advantage of this approach is that it reduces the effort required to define new frames. More importantly, this ensures consistent annotation for different syntactic elements that evoke the same semantics. An example is the verb frame FEAR.01 that is reused for the noun *fear* (as in *I have a fear of spiders*) and the adjective *afraid* (as in *I am afraid of spiders*). Recent work has proposed a method to automatically identify aliases for verbal complex predicates using a distributional model over parallel corpora (Jagfeld and van der Plas, 2015).

Unlike previous works that exclusively consider English, we consider a multilingual setting in which we alias English frames to verbs in other languages. We also allow multiple aliases for each TL verb. We pursue this approach not only to define a frame lexicon, but also to increase the quality and consistency of Proposition Banks generated with annotation projection.

3 Method

Our approach curates a frame lexicon of a Proposition Bank generated with annotation projection. The approach has two curation steps: filtering (section 3.1) and merging (section 3.2). We then make a final pass to add human readable explanations to the curated frame lexicon (section 3.3).

[1]Please contact the first author of this paper for access to the data.

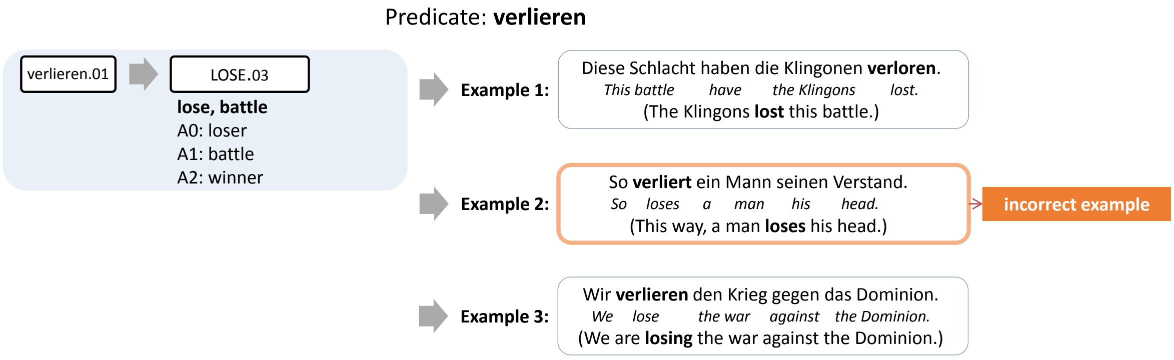

Figure 3: Filtering task example. A curator is shown one lexicon entry, consisting of a TL verb (*verlieren*) and an English frame (LOSE.03, as in *lose a battle*). In addition, the curator is shown five example sentences (only three displayed in this image). While examples 1 and 3 are correct, example 2 does not evoke the *lose a battle* sense.

3.1 First Curation Task: Filtering

The first task is to identify all incorrect frames for TL verbs. For each entry in the lexicon, curators must make a binary decision on whether the entry is correct or not. In order to make this decision, curators are presented with the following information: 1) The TL verb. 2) A description of the English frame and its roles. 3) A sample of TL sentences annotated with this frame. Refer to Figure 3 for illustration.

Given this information, curators must answer two questions (detailed below). If the answers to both questions are *yes*, the entry is considered valid. If one of the questions is answered with *no*, this entry must be removed from the lexicon.

Q1: Is the English frame a valid sense for the TL verb? The first question concerns the semantic validity of the English frame for the TL verb. To answer this question, curators only consider the English frame description. If the description refers to semantics that the TL verb clearly cannot evoke, the answer to this question is *no*. We encountered such a case in section 1 with the verb *drehen* that cannot evoke frame TURN.02. In all other cases, the answer is *yes*. Notably, we do not ask if an English frame is a perfect fit in semantics. At this point in the process, we are only interested in filtering out clear errors.

Q2: Does the TL verb accurately reflect the English frame description in the sample sentences? Even if an entry is valid in principle, it may still be subject to errors in practice. We find that some entries are correct judging from their description, but are never correctly detected in the corpus due to errors made by the English SRL. This problem disproportionally affects frames for which only limited English training data is available. For this reason, we require the curator to inspect a sample of 5 TL sentences per entry and determine whether they are correctly labeled. Refer to Figure 3 examples for both cases: Sentence 1 and 3 correctly invoke LOSE.03 (lose a battle), whereas Sentence 2 evokes LOSE.02 (lose an item). If a majority of example sentence is incorrectly labeled, this question must be answered with *no*.

3.2 Second Curation Task: Merging

The second task addresses the issue of redundancy caused by multiple entries for TL verbs that evoke the same semantics. For each pair of entries for the same TL verb, a curator must decide whether they are synonymous and need to be merged into a single entry. This task therefore effectively decides the semantic granularity of the lexicon entries for each TL verb.

We base merging decisions on the annotation guidelines of the English Proposition Bank, which specify that new frames need to be created to reflect different syntactic usages of a verb. In addition, new frames are created for broadly different meanings of a verb even if the syntactic subcategorization is the same (Palmer et al., 2005).

For each merging decision, we present curators with the following information: 1) The TL verb. 2) The two frames and their descriptions. 3) A set of TL sample sentences for each frame. The latter is the most important since sample sentences illustrate how the TL verb is used in related contexts when labeled with a specific frame. Refer to Figure 4 for example.

Given this information, curators must answer two questions (explained below). If the answer to any

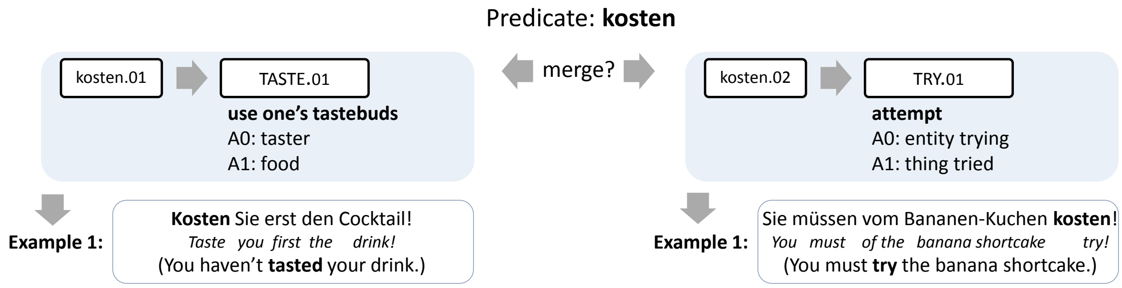

Figure 4: Information presented to curator for merge decisions: Two frames, their descriptions and example sentences.

of these questions is *no*, the two entries should not be merged.

Q1: Are the two entries usage-synonyms? We define *usage-synonyms* as target language usage synonyms. To illustrate the difference to regular synonyms, consider the example in Figure 4 in which curators must decide whether TASTE.01 and TRY.01 should be merged. While the two English frames are clearly not synonymous, their target language usages are. As lexicon entries for the German verb *kosten*, they are both solely used in the context of tasting food and are therefore usage-synonyms, illustrated by the sample sentences for each frame in Figure 4. If two entries are clearly not usage-synonyms, the answer to this question is *no*. In all other cases, the answer is *yes*.

Q2: Do the two entries represent syntactically different usages? We found a number of cases in which curators disagreed on whether two entries are usage-synonyms or not. An example of this were entries which partially overlapping semantics, such as the frame pair WRAP.01 (*enclose*) and PACK.01 (*fill, load*). To address this, we created a guideline to compare syntactic usage of TL verbs. We ask curators to build the *dictionary expansion* for both entries, which we define as the default syntactic expansion that one might find in a dictionary. An English example for the verb *turn* is *to turn something* for TURN.01 and *to turn into something* for TURN.02. However, we ask curators to create this form for TL verbs[2]. If the TL dictionary expansion is different, the answer to this question is *no*.

3.3 Final Pass: Dictionary Forms and Comments

After curators complete both tasks, we rerun annotation projection using the created lexicon to filter out incorrect entries and merge redundant entries. This produces a Proposition Bank with manually curated TL frames. To complete the curation process, we ask curators to inspect each entry in the dictionary and add comments or explanations, as well as dictionary expansions. This information is intended for human consumption. The purpose of this annotation is to make apparent the distinctions between multiple entries for the same TL verb and explain our aliasing decisions. The entire curation process thus produces an annotated Proposition Bank with salient, manually curated frames for each TL verb.

4 Evaluation

We present a set of large-scale experiments over three languages to evaluate our proposed approach. We first evaluate the curation process itself in terms of curator agreement scores, the required effort and the impact on the frame lexicons. We then present a detailed analysis of the auto-generated Proposition Banks in which we evaluate their quality in terms of precision, recall and F1 score both before and after curation. We further evaluate the curated Proposition Banks with regards to verb coverage and conduct a qualitative comparison against the manually created, official Chinese Proposition Bank (Xue and Palmer, 2005). Based on this analysis, we discuss the challenges and potential of combining large-scale annotation projection and manual aliasing to generate Proposition Banks for new target languages.

Experimental setup. We pre-generate Proposition Banks with the approach described in Akbik et. al (2015) for Chinese, French and German using parallel text from the OPENSUBTITLES2016

[2] In the example discussed in Figure 4, the dictionary expansion of *kosten* the same for both entries: *etwas kosten* ("*to taste/try something*"). This indicates that both entries take a direct object and are syntactically similar. They must therefore be merged only if the entries are also usage-synonyms.

LANGUAGE	PARALLEL CORPUS	PROPOSITION BANK TYPE	#VERBS	#FRAMES	#SENTENCES	EVALUATION P	R	F_1
Chinese	OPENSUBTITLES	PROJECTED	1,094	1,472	87,953	0.87	0.94	0.91
	(9 million sentences)	CURATED	942	1,003	68,829	0.93	0.96	**0.94**
French	OPENSUBTITLES	PROJECTED	1,323	2,249	175,636	0.82	0.94	0.87
	(15 million sentences)	CURATED	1,208	1,370	130,579	0.91	0.94	**0.93**
German	OPENSUBTITLES	PROJECTED	1,552	2,441	191,816	0.83	0.92	0.87
	(13 million sentences)	CURATED	1,532	1,717	150,949	0.90	0.93	**0.91**

Table 1: Annotation projection statistics for all three target languages: Number of parallel sentences available for each language, total number of covered verbs in auto-generated PropBanks as well as the total number of frames. The number of frames is higher because many verbs evoke more than one frame.

project (Lison and Tiedemann, 2016). This data is automatically mined from movie subtitles and thus covers a large array of topics (dramas, documentaries, science fiction etc.), reflecting verb usage in common speech. We execute annotation projection over 9-15 million parallel sentences for each target language, generating lexicons that cover over 1,000 verbs, respectively, as well as labeled corpora spanning over 100,000 sentences. For a full breakdown of our annotation projection numbers, refer to Table 1 (the uncurated PropBanks are marked as "PROJECTED").

These generated PropBanks are the starting point for our curation process. We had two persons each curate the French lexicon, while Chinese and German were curated by one person each. On average, for each language and each person about 60 person hours were required for the full curation process for all lexicon entries. All curators in our experiment have expert knowledge in semantic role labeling.

Using the curated lexicons, we generate the final generated Proposition Banks, marked as "CURATED" in Table 1. Following earlier evaluation practice, we randomly selected 100 sentences from each Proposition Bank before and after curation. We manually evaluated these sentences in order to get an understanding of precision, recall and F_1-score for the generated resources.

4.1 Evaluation Results

Refer to Table 1 for an overview of all three generated Proposition Banks before and after curation with our proposed process. We make a number of observations:

Curation significantly improves Proposition Bank quality. We find that incorporating the curated frame lexicons into annotation projection significantly boosts overall quality. For German, we estimate an F_1 of 0.91 (↑ 4pp), for French 0.93 (↑ 6pp), and for Chinese 0.91 (↑ 4pp). This indicates that a large number of errors are caused by incorrect entries in the frame lexicon, which can be handled globally at moderate effort using our proposed approach.

Curation reduces the amount of labeled data and covered verbs. We also note that filtering incorrect entries reduces the number of annotated sentences in the generated resource, since all affected projections are removed. For Chinese, French and German, a total of 18k, 20k and 36k sentences are affected by incorrect annotations and are therefore filtered out. Our approach therefore generates slightly smaller Proposition Banks with a higher overall quality. We also note that for some TL verbs all entries in the lexicon were deemed incorrect. These verbs are therefore no longer covered in the curated PropBanks. This reduces the amount of verbs included in the TL lexicons by 20 for German, 115 for French and 152 for Chinese, which seems to correspond to their linguistic distance to English: German, the closest relative to English, has the largest number of covered verbs while Chinese has the lowest.

Curation significantly reduces redundancy. We note that our approach significantly reduces the number of distinct frame entries for each language. Whereas in uncurated versions, every alignment is interpreted as a distinct verb frame, the filtering and merging process removes hundreds of incorrect and redundant entries. The curated Chinese, French and German PropBanks evoke 1,003 (↓ 32%), 1,370 (↓ 39%), and 1,717 (↓ 30%) frames respectively. We note that the highest difference is observed for French, for which the largest number of parallel sentences was available. It seems that greater amounts of parallel data lead to greater redundancies and more incorrect alignments.

French	Source	#Al.	Error Class
rentrer	go	9,070	CP: *go home*
ressembler	look	6,160	CP: *look like*
naître	bear	5,937	LE: *be born*
pouvoir	be	4,541	CP: *be able to*
asseoir	sit	4,300	LE: *s'assoir*
adorer	love	3,391	DI
pouvoir	get	3,325	CP: *get to do something*
rentrer	come	2,813	CP: *come home*
appartenir	belong	2,745	DI
enfermer	lock	2,143	DI

German	Source	#Al.	Error Class
möchten	want	26,996	DI
sollen	suppose	12,849	CP: *be supposed to*
sollen	want	12,619	CP: *want sb. to do sth.*
herausfinden	find	7,756	CP: *find out*
beschützen	protect	6,416	DI
fällen	like	5,846	LE
mitnehmen	take	5,604	CP: *take along*
kennenlernen	meet	5,254	CP: *get to know sb..*
übernehmen	take	4,029	CP: *take over*
erledigen	do	3,914	CP: *get sth. done*

Table 2: Top 10 unlabeled verbs for French and German, with English source verbs and alignment counts in the parallel corpus. Error classes are lemmatiation error (LE), dictionary incomplete (DI) and complex predicates (CP).

4.2 Curation Guidelines

In order to assess our curation guidelines, we asked two curators to independently execute the curation process for the French lexicon. We produce two independently curated French Proposition Banks using the two lexicons. We compared both versions and found that curators had agreed on 1,317 out of 1,370 (96%) of all curated entries. This speaks to the deterministic nature of our guidelines.

One contributing factor to our high agreement score is that we discussed and determined representative cases of disagreement in early iterations of the curation process. In the filtering task, for instance, we encountered initial disagreements on theoretically correct entries that were incorrectly recognized in practice (see section 3.1). In the merging task, we also had initial difficulties concerning semantically related frames that were neither clear usage-synonyms nor clearly distinct (see section 3.2). As previously illustrated, we defined deterministic rules for these cases.

Disagreements. Some disagreement remained in the merging task: Some entries have syntactically similar usage and partially overlapping semantics. This often affects merging questions in which one entry is a highly specialized usage of the other. This frequently involves slang language. For example, the French verb *secher* (to dry) can also be used in spoken language to indicate "cutting class" as in not attending class in school. Due to the lack of a more appropriate English frame, this sense is aligned to the frame CUT.01 (slice or injure). Such cases required further discussion by curators.

4.3 Qualitative Evaluation of Curated Proposition Banks

The curated Proposition Banks give us the opportunity to gain insights on verb coverage and to compare annotation projection against manual annotation efforts. We present results of a qualitative inspection of common TL verbs that are absent in generated frame lexicons and a qualitative comparison of our generated resource for Chinese against the official Chinese Proposition Bank (Xue and Palmer, 2005).

Verb coverage. From the parallel corpora, we retrieve the most common TL verbs that are not contained in our frame lexicon. We manually inspect the 100 most common unlabeled verbs to determine reasons for lack of coverage. We find that lack of coverage can be traced back to one of three reasons: 1) The translation dictionary we use to filter out translation shift errors is incomplete (DI). 2) The lemmatizer is unable to correctly lemmatize certain verbs (LE). 3) The TL verb meaning can only be rendered in English as a complex predicate (CP). We list the top 10 most common unlabeled verbs for French and German in Table 2.

Complex phrasal constructions. A crucial error class are TL verbs that cannot be expressed with a single verb in English. Consider the French verb *rentrer*, which frequently expresses the meaning of "returning home", rendered in English with the LVCs *go home* or *come home*. Another example is *pouvoir*, which in English needs to be rendered with the adverb *able* as in *be able to*. A German example for this phenomenon is *sollen*, rendered as *to be supposed to* or *to want someone to do something*. This represents a limitation of our current verb-based projection approach. However, there are ongoing efforts to expand the English Proposition Bank with frames for complex predicates. We believe that this will allow us to address this source of verb coverage loss in future work.

Comparison to Chinese PropBank We randomly sample 100 Chinese verbs from our lexicon and compare all entries against the official Chinese Proposition Bank. We find an encouragingly high agreement, which 94.6% of our lexicon entries corresponding to senses in the Chinese PropBank. We also find 7 lexicon entries that do not exist in the Chinese PropBank, indicating that annotation projection with manual frame curation may be used to increase coverage of existing Proposition Banks. We also find 4 instances in which valid entries for a verb in the two PropBanks are complimentary. However, we point out that we conduct this study only in one direction. The official Chinese Proposition Bank contains frames for all types of frame evoking elements, including verbs, nouns and complex predicates (Xue, 2006; Xue and Palmer, 2009). Their coverage is therefore significantly larger than our current approach. Nevertheless, we find the significant overlap in both frame lexicons encouraging.

5 Discussion and Conclusion

We presented an approach for addressing lexicon-level inconsistencies in automatically generating Proposition Banks using annotation projection. Our approach manually curates the heuristically determined frame lexicon in two steps: A filtering and a merging step. We executed the approach on large-scale parallel data to generate Proposition Banks with curated frames. Our evaluation shows that our approach significantly increases the quality of the generated resources, while reducing redundancy and inconsistency in the frame lexicon.

Our evaluation also revealed TL verbs that require complex predicates in English as a natural limitation of our current verb-based approach. Accordingly, future work will investigate this issue by expanding the range of projections from verbs to other types of frame-evoking elements. We aim to expand our frame lexicon to include not only TL verbs, but also nouns, adjectives and eventually complex TL predicates.

Another avenue for future work is to investigate the use of SRL trained over projected Proposition Banks in applications. As previous work has shown information extraction and question answering to benefit from SRL, we aim to investigate multilingual applications in these tasks.

We release the Proposition Banks created with our approach in order to encourage discussion with the research community. We believe this resource to potentially be valuable for investigating crosslingual semantics and for training statistical SRL systems for Chinese, French and German.

References

Alan Akbik and Yunyao Li. 2016. Polyglot: Multilingual semantic role labeling with unified labels. In *ACL 2016, 54th Annual Meeting of the Association for Computational Linguistics: Demostration Session*, page to appear.

Alan Akbik, Laura Chiticariu, Marina Danilevsky, Yunyao Li, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2015. Generating high quality proposition banks for multilingual semantic role labeling. In *ACL 2015, 53rd Annual Meeting of the Association for Computational Linguistics*, pages 397–407.

Claire Bonial, Julia Bonn, Kathryn Conger, Jena D Hwang, and Martha Palmer. 2014. Propbank: Semantics of new predicate types. In *LREC*, pages 3013–3019.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. Ontonotes: the 90% solution. In *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*, pages 57–60. Association for Computational Linguistics.

Glorianna Jagfeld and Lonneke van der Plas. 2015. Towards a better semantic role labeling of complex predicates. In *NAACL-HLT 2015 Student Research Workshop (SRW)*, page 33.

Pierre Lison and Jörg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In *In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*.

Chi-kiu Lo, Meriem Beloucif, and Dekai Wu. 2013. Improving machine translation into chinese by tuning against chinese meant. In *Proceedings of the Tenth International Workshop on Spoken Language Translation (IWSLT 2013)*.

Umar Maqsud, Sebastian Arnold, Michael Hülfenhaus, and Alan Akbik. 2014. Nerdle: Topic-specific question answering using wikia seeds. In *COLING (Demos)*, pages 81–85.

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual annotation projection for semantic roles. *Journal of Artificial Intelligence Research*, 36(1):307–340.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *EMNLP-CoNLL*, pages 12–21.

Lonneke Van der Plas, Paola Merlo, and James Henderson. 2011. Scaling up automatic cross-lingual semantic role annotation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 299–304. Association for Computational Linguistics.

Lonneke van der Plas, Marianna Apidianaki, and Chenhua Chen. 2014. Global methods for cross-lingual semantic role and predicate labelling. In *COLING*, pages 1279–1290. ACL.

Nianwen Xue and Martha Palmer. 2005. Automatic semantic role labeling for chinese verbs. In *IJCAI*, volume 5, pages 1160–1165. Citeseer.

Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the chinese treebank. *Natural Language Engineering*, 15(01):143–172.

Nianwen Xue. 2006. Semantic role labeling of nominalized predicates in chinese. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 431–438. Association for Computational Linguistics.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the first international conference on Human language technology research*, pages 1–8. Association for Computational Linguistics.