Detecting Cross-lingual Semantic Similarity Using Parallel PropBanks

Shumin Wu
Computer Science Dept.
Univ. of Colorado at Boulder
shumin@colorado.edu

Jinho D. Choi
Computer Science Dept.
Univ. of Colorado at Boulder
choijd@colorado.edu

Martha Palmer
Linguistics Dept.
Univ. of Colorado at Boulder
mpalmer@colorado.edu

Abstract

This paper suggests a method for detecting cross-lingual semantic similarity using parallel PropBanks. We begin by improving word alignments for verb predicates generated by GIZA++ by using information available in parallel PropBanks. We applied the Kuhn-Munkres method to measure predicate-argument matching and improved verb predicate alignments by an F-score of 12.6%. Using the enhanced word alignments we checked the set of target verbs aligned to a specific source verb for semantic consistency. For a set of English verbs aligned to a Chinese verb, we checked if the English verbs belong to the same semantic class using an existing lexical database, WordNet. For a set of Chinese verbs aligned to an English verb we manually checked semantic similarity between the Chinese verbs within a set. Our results show that the verb sets we generated have a high correlation with semantic classes. This could potentially lead to an automatic technique for generating semantic classes for verbs.

1 Introduction

This paper discusses attempts to use alignments between English and Chinese predicate-argument structures in a parallel PropBanked corpus[1] as a basis for determining cross-lingual semantic similarity. As the foundation of many machine translation decoders (DeNeefe and Knight, 2009), word alignment has continuously played an important role in machine translation. There have been several attempts to improve word alignment, most of which have focused on tree-to-tree alignments of syntactic structures (Zhang et al., 2007; Mareček, 2009a). Our hypothesis is that the predicate-argument structure alignments can abstract away from language specific syntactic variation and provide a more robust, semantically coherent alignment across sentences.

We begin by running GIZA++ (Och and Ney, 2003), one of the most popular alignment tools, to obtain automatic word alignments between the parallel English/Chinese corpus. We then improve these alignments by using Gold Standard PropBank structures for predicates (including both verbs and nominalizations). For each Chinese and English verb predicate pairs within a parallel sentence, we examine the quality of both the predicate and argument alignment (using both Chinese to English and English to Chinese GIZA++ word alignment output) and devise a symmetrical similarity measure. From that, we pose predicate-argument mapping as a linear assignment problem (optimizing the total similarity of the mapping) and solve it with the Kuhn-Munkres method (Kuhn, 1955). With this approach, we can greatly improve recall while maintaining the high precision of the original GIZA++ alignments.

Using these new improved alignments, we then hypothesize that the set of Chinese verb predicates aligned with a specific English verb predicate will be semantically similar; or alternatively, the set of English verb predicates aligned with a specific Chinese verb predicate will also be semantically similar. To verify this hypothesis, we ran two experiments.

[1] PropBank is a corpus in which the arguments of each verb predicate are annotated with their semantic roles in relation to the predicate (Palmer et al., 2005).

For each set of English verb predicates aligned to a Chinese verb predicate, we used an existing English lexical database, WordNet (Fellbaum, 1998), to see if the English verb predicates within the set belong to the same semantic class. For each set of Chinese verb predicates aligned to an English verb predicate, since we do not currently have access to a WordNet like lexical database for Chinese, we manually measured the semantic similarity of each predicate in the set. The evaluation was done by a Chinese-English bilingual speaker. Our experiments show that it is possible to generate semantic classes by using word-alignments from the parallel corpus. This is encouraging because our approach suggests a way of automatically generating semantic classes, which is a labor intensive manual effort.

In this paper, we propose advancements on a previous technique for improving word alignment using parallel PropBanks. With our approach, we gained 12.6% with respect to the F-score over GIZA++ word alignment. Furthermore, we show that sets of verbs we extracted from the parallel corpus correlate well with synonym sets and verb classes in both English and Chinese.

2 Related work

The basic approach described here is similar to the idea of semantic similarity based on triangulation between parallel corpora outlined in Resnik (2004) and Madnani et al. (2008a; 2008b), but is implemented here quite differently. It is most similar in execution to the work of (Mareček, 2009b), which improves word alignment by aligning tectogrammatical trees in a parallel English/Czech corpus. The Czech corpus is first lemmatized because of the rich morphology, and then the word alignment is "symmetrized". However, this approach does not explicitly make use of the predicate-argument structure to confirm the alignments or to suggest new ones.

Choi et al. (2009) showed how to enhance Chinese-English word alignments by exploring predicate-argument structure alignment using parallel PropBanks. They used the 'English Chinese Translation Treebank' (ECTB) for their experiments, the same corpus we use. The resulting system showed improvement over pure GIZA++ alignment, though it only took advantage of Chinese to English GIZA++ output and required careful tuning of a number of threshold parameters to balance between precision and recall.

Fung et al. (2007) demonstrated that there is poor semantic parallelism betwen Chinese-English bilingual sentences. Their technique for improving Chinese-English predicate-argument mapping ($ARG_{Chinese,i} \mapsto ARG_{English,j}$) consists of employing features from automatic syntactic parses of the Chinese and English sentences, word alignment with a bilingual lexicon, and tuning on an unannotated parallel corpus. Later, Wu and Fung (2009) used parallel PropBanks to improve MT system outputs. Given the outputs from Moses (Koehn et al., 2007), a machine translation decoder, they reordered the outputs using the predicate-argument structure annotated in parallel PropBanks. Although their work is not directly related to ours, it shows how parallel PropBanks can be used to improve MT systems in general.

3 Symmetric predicate mapping

3.1 Parallel Corpus

We apply our predicate-argument mapping method to the 'Engish Chinese Translation Treebank' (ECTB), a parallel English-Chinese corpus. In addition to Treebank syntactic structure, the corpus has also been annotated with semantic role labels in the standard PropBank style of ARG0, ARG1, etc., based on verb specific frame file definitions (Xue and Palmer, 2009). While the corpus contains both the Xinhua Chinese newswire with literal English translations (4,363 parallel sentences) and the Sinorama Chinese news magazine with non-literal English translations, we chose only the (more accurate) mapping output of the Xinhua corpus for automatically generating semantic classes.

3.2 Word alignment

The predicate-argument mapping method starts with running GIZA++ word alignment between Chinese and English parallel sentences. As the GIZA++ alignment output is asymmetrical, we use both the Chinese to English, and English to Chinese word alignment output. In GIZA++, source language words cannot share aligned target words with each other. Since Chinese words typically translate into several alternative English words, the GIZA++ Chinese to English word alignment output tends to have smaller alignment errors.

With just the GIZA++ output, only 48.1% of the Chinese predicates are aligned to words that are predicates in Engish, while only 59.2% of the English predicates are aligned to words that are predicates in Chinese (Table 1). Compared to a Chinese-English bilingual human annotator's results, GIZA++ misses 20.0% of the Chinese predicates and 19.8% of the English predicates.

Alignment	GIZA++	Human annotator
Ch.pred $\rightarrow$ En.pred	48.1%	60.1%
En.pred $\rightarrow$ Ch.pred	59.2%	73.8%
Ch.pred $\leftrightarrow$ En.pred	53.1%	66.3%

Table 1: Percentage of aligned predicates on 200 random sentences (671 Chinese predicates and 546 English predicates) in the Xinhua corpus

3.3 Measure predicate similarity

Following Choi et al. (2009)'s work, we took advantage of PropBank arguments (Palmer et al., 2005) to produce a better mapping between Chinese and English predicates. For each Chinese/English predicate pair, we measure similarity between the predicates based on the aligned words in the arguments as well as the predicates themselves. Chinese and English words that are aligned to the same argument types (ARG0, ARG1, etc) contribute more heavily to the similarity score than words aligned to different argument types. Likewise, words labelled with ARG0 (agent) and ARG1 (patient), the more dominant and less ambiguous argument types, contribute more heavily to the similarity score than words labelled with other arguments (e.g., ARG2-5, ARGM).

Since GIZA++ alignment output is asymmetrical, the default similarity scoring output is also asymmetrical. To symmetrize the similarity score, instead of directly symmetrizing GIZA++ alignment, we run the similarity scoring method twice, once for Chinese to English (Sim_{CE}) and once for English to Chinese (Sim_{EC}). A symmetric similarity score is then derived by taking the weighted harmonic mean between the 2 scores:

$$Sim_{SYM} = (1+\beta^2) \cdot \frac{Sim_{CE} \cdot Sim_{EC}}{\beta^2 \cdot Sim_{CE} + Sim_{EC}} \quad (1)$$

Because of the generally lower alignment error rate of the Chinese to English GIZA++ alignment output, we choose $\beta \approx 1.2$ to bias towards the Chinese to English similarity score. By not directly symmetrizing GIZA++ alignment, we avoid the potential issues of alignment sparseness (with the intersection approach) and mapping ambiguity (with the grow-diag-final or union approach, where an aligned chunk can potentially span across multiple arguments or predicates), as well as making unintended trade-offs between alignment precision and recall.

3.4 One-to-one mapping

To find the best predicate-argument mapping between Chinese and English sentences, we assume each predicate in a Chinese or English sentence can only map to one predicate in the target sentence. This assumption is mostly valid for the Xinhua news corpus, though occasionally, a predicate from one sentence may align more naturally to two predicates in the taget sentence. This typically occurs with verb conjunctions. For example the Chinese phrase "观光 旅游" is often translated to the single English verb "travel".

With the one-to-one mapping constraint, the optimal mapping may be considered to be the mapping that maximizes the sum of the similarity measure of the predicates and arguments in the mapping. Let P_C and P_E denote the sets of predicates in Chinese and English respectively, with $G(P_C, P_E) = \{g : P_C \mapsto P_E\}$ as the set of possible mappings between the two predicate sets, then the optimal mapping is:

$$\begin{aligned} g^* &= \arg\max_{g \in G} Sim_{SYM}(P_C, P_E) \\ &= \arg\max_{g \in G} \sum_{i,j \in g} Sim_{SYM}(P_{C,i}, P_{E,j}) \quad (2) \end{aligned}$$

To turn this into a classic linear assignment problem, we define $Cost(P_C, P_E) = max(Sim_{SYM}) - Sim_{SYM}(P_C, P_E)$, and (2) becomes:

$$g^* = \arg\min_{g \in G} \sum_{i,j \in g} Cost(P_{C,i}, P_{E,j}) \quad (3)$$

(3) can be solved in polynomial time with the *Kuhn-Munkres* algorithm (Kuhn (1955)).

Method	precision	recall	f-score
GIZA++	84.2%	67.5%	74.9%
SPM	87.0%	88.1%	87.5%

Table 2: Predicate mapping accuracy on 200 random sentences (671 Chinese predicates and 546 English predicates) in the Xinhua corpus

3.5 Result

For evaluation, we compared the mapping output of our symmetric predicate mapping (SPM) method and the pure GIZA++ alignment based method against the mapping found by the Chinese-English bilingual speaker. As table 2 shows, on 200 sentences of the Xinhua corpus, the SPM method improved the recall of the GIZA++ method from 67.5% to 88.1% while managing a slight advantage on precision (87.0% vs 84.2%). Examining the mapping output revealed that the SPM method can often recover predicate mappings from argument similarity even when the verb has been mis-aligned, while the one-to-one mapping constraint and linear assignment optimization were effective at disambiguating multiple potential mapping candidates.

4 Data-driven semantic classes

4.1 Deriving Chinese semantic classes

For each English verb, we have a set of Chinese verbs aligned to it by either GIZA++ or by our SPM approach. Because we do not yet have access to Chinese verb class resources, we relied on a Chinese-English bilingual speaker to score the verb mapping on a scale of “0” to “3”.

Score “3” indicates the Chinese verb is a direct (possibly dictionary) translation of the English verb. An example is the Engligh verb *believe* mapping to Chinese verbs {认为, 相信}. Here the 2 Chinese verbs correspond to 2 different senses of *believe*, 认为 means *to have an opinion*, while 相信 means *to have faith or accept as truth*.

Score “2” usually indicates the Chinese verb has a hypernym/hyponym relationship with the English verb, or it may be interchangeable in most usages. Examples of these include *decide* mapping to 评出 (‘judge’), a hyponym of *decide* in WordNet), and *reduce* mapping to 压缩 (‘compress’).

Score “1” usually indicates the verbs are often used in the same context, but otherwise have very little semantic relation. An example of this is *issue* mapping to 出台 (used metaphorically for *introduce*).

Score “0” indicates no relation was found (possible mis-alignment by the system). An interesting example is *pay* mapping to 关注 (‘pay close attention’). While this is not a correct translation, the underlying semantic role mapping of the predicates may have been correct, as *pay* was likely used as a light verb in this context.

4.2 Deriving English semantic classes

For each Chinese verb, we have a set of English verbs aligned to it by either GIZA++ or our enhanced predicate matching approach. For example, a Chinese verb, 下降 (‘decrease’), is aligned to a set of English verbs {*decrease*, *drop*, *fall*} by our approach. The intuition behind automatically deriving English semantic classes is simple: if the alignments are correct, English verbs within a set should have similar meanings or have some kind of semantic relations with one another.

We used semantic relations from WordNet (version 3.0): WordNet is an English lexical database that groups nouns, verbs, adjectives and adverbs into sets of synonyms (Fellbaum, 1998). WordNet also provides taxonomy information such as hypernyms, hyponyms, etc. By using WordNet, it is usually possible to find a least common hypernym[2] of all English verbs in a set. For the previous example, we can derive the following taxonomy using WordNet (Figure 1).

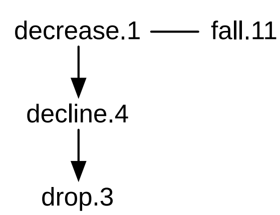

Figure 1: Taxonomy of {*decrease*, *drop*, *fall*}

‘decrease.1’ stands for the 1st verb-sense of the verb *decrease* in WordNet. According to Figure 1, *decrease* is an indirect hypernym of *drop* and a syn-

[2]Least common hypernym: the lowest hypernym that is an ancestor of all words in a set.

onym of *fall*. Therefore, every verb in this set belongs to the semantic class, 'decrease.1'.

Since WordNet defines very find-grained semantic relations, it is sometimes useful (or rather necessary) to merge some of the senses to derive one semantic class per set. For example, the Chinese verb 主办 ('host'), is aligned to the set of English verbs {*sponsor*, *hold*} that can be represented as in Figure 2.

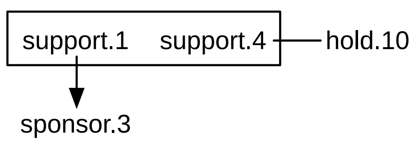

Figure 2: Taxonomy of {*sponsor*, *hold*}

'support.1' is a direct hypernym of *sponsor* whereas 'support.4' is a synonym of *hold*. For our purpose, it is appropriate to merge the two senses, 'support.1' and 'support.4' and label the set as 'support.1,4'.

It is possible that there is no hypernym that is an ancestor of all verbs in a set. For example, a verb, 出现 ('appear'), is aligned to {*appear*, *occur*, *emerge*, *exhibit*}, which forms the taxonomy in Figure 3.

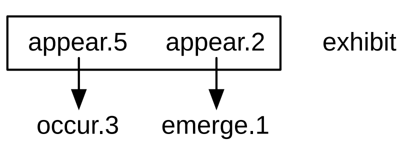

Figure 3: Taxonomy of {*appear*, *occur*, *emerge*, *exhibit*}

By merging 'appear.2' and 'appear.5', 'appear.2,5' becomes the least common hypernym of *occur* and *emerge*. However, according to WordNet, there is no link between *exhibit* and 'appear.2,5' so that *exhibit* is not connected to the rest. Thus, we end up having two semantic classes for this set, 'appear.2,5' and 'exhibit'.

By using the taxonomy, we can evaluate how well each set of English verbs corresponds to WordNet semantic relations. There already exist English lexical databases such as WordNet, VerbNet (Kipper et al., 2006), FrameNet (Baker et al., 1998), but this approach suggests an automatic way of deriving new semantic classes and validating and extending preexisting classes, which can be applied to other languages. Verifying that English semantic classes can be derived correctly using a parallel corpus gives us an idea about how well this approach may work on other languages that do not already have similar lexical databases.

5 Experiments

We used only the SPM method on the Xinhua English-Chinese parallel corpus to derive semantic classes in English and Chinese, since the SPM method performed better than the GIZA++ mapping method on both precision and recall. The large recall advantage of the SPM method was evident in the derived semantic classes. For example, the Chinese verb 下降 ('decrease'), is aligned only to the English verbs {*decrease*, *drop*} by GIZA++ but to {*decrease*, *drop*, *fall*} by our SPM method.

5.1 English to Chinese semantic classes

For evaluating the English to Chinese semantic class mapping , we chose 50 diversely-mapped (to Chinese) English verbs from the Xinhua corpus. We excluded English light (or predominantly used as) verbs (*be*, *have*, *take*, etc), Chinese light verbs (是, 有, etc), as well as Chinese verbs that only occured once for each English verb. The total number of verbs comprising all sets is 218 ($\sum_{i=1}^{50} |s_i| = 218$, where s_i is a verb set). On average, each English verb is mapped to 4.36 Chinese verbs. For example, the English verb *show*, is mapped to the Chinese verb set {表明, 呈现, 显示, 呈, 展示} (they all translate to *show* in English).

Figure 4 shows the score distribution of the Chinese verbs as evaluated by the Chinese-English bilingual speaker. Out of 218 verbs, 179 verbs (82.1%) received a "3", while 14 verbs (6.4%) received a "0". In general, verbs that scored lower appear less frequently; Figure 5 shows the score distribution of the Chinese verbs based on mapping frequency. Out of 2,457 verb appearances, 2,313 appearances (94.1%) received a "3", while only 41 appearances (1.7%) received a "0".

5.1.1 Analysis

While some of the generated Chinese semantic classes contain more than a few members, we found that often a number of the verbs share the same Chinese character that actually expresses the bulk of the meaning. For example, 展示 and 显示 share the

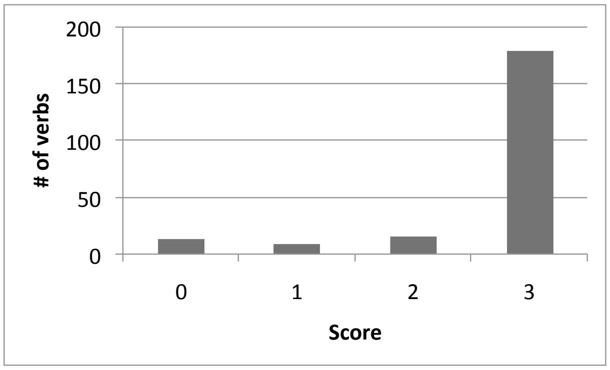

Figure 4: Chinese verb output scored by bi-lingual speaker

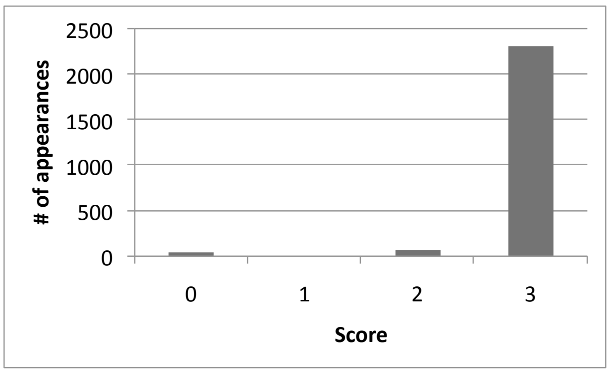

Figure 5: Chinese verb output score scaled by frequency of appearance

character 示 ('show'), 建设, 兴建 and 筹建 share the character 建 ('construct'). It's possible that if we can generate Chinese semantic classes based on the root characters, we can expand the semantic class membership to include words containing the root characters, even if they do not appear in the parallel corpus.

5.2 Chinese to English semantic classes

For evaluating the Chinese to English semantic class mapping, we started with the 50 English verbs used for the English to Chinese semantic class mapping, generated all Chinese verbs found in the mapping, and then used the SPM approach to retrive the English verb sets. Singleton verb sets and verbs that only appear once in the set are ignored. This resulted in 53 English verb sets. The total number of verbs comprising the sets is 127 ($\sum_{i=1}^{53} |s_i| = 127$, where s_i is a verb set).

To verify how each English verb set (derived from the parallel corpus using our SPM word alignment) corresponds to WordNet, we use three kinds of measurements. First, for each verb in a set, we measure the height between the verb and the least common hypernym for the set. For example in Figure 1, the height between 'decrease.1' and *decrease* is 0, *drop* is 2, and *fall* is 0 (synsets have a height of 0). This shows how closely related each verb is to the least common hypernym.

Second, for each set, we count how many sense-mergings are needed to derive the least common hypernym. In Figure 2, there is only one sense-merging needed to derive the hypernym, 'support.1,4'. This gives an idea of how fine or coarse-grained our semantic classes are compared to WordNet synsets. Finally, for each set, we count the number of derived semantic classes. In Figure 3, there are two semantic classes, 'appear.2,5' and 'exhibit', derived from the set. This shows how reliable each set is as a semantic class. Note that we tried to use verb frequency counts to filter out members that should not be in the set, but this showed no improvement.

Figure 6 shows the height between each verb and its lowest common hypernym. Out of 127 verbs, 78 of them have a height of 0, 37 of them have a height of 1, and so on. Less than 10% of the verbs have a height of 2 or above. This implies that most verbs in each set are closely related to the least common hypernyms of the sets.

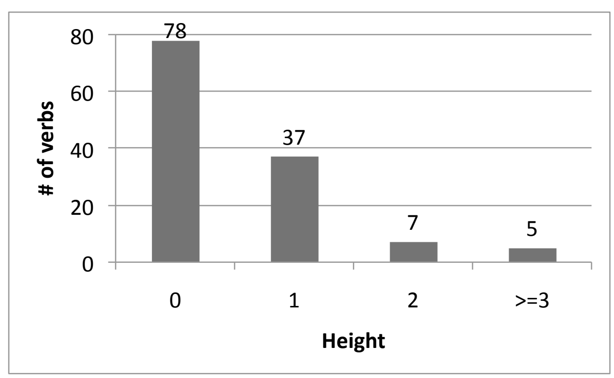

Figure 6: Height with respect to the number of verbs

Figure 7 shows the number of sense-mergings with respect to the number of verbs. Out of 53 verb sets, 42 of them required no sense-merging, 10 of them required 1 sense-merging, and so on. Clearly, most sets did not require any sense merging. This implies that the semantic classes derived by our approach are almost as fine-grained as WordNet synsets.

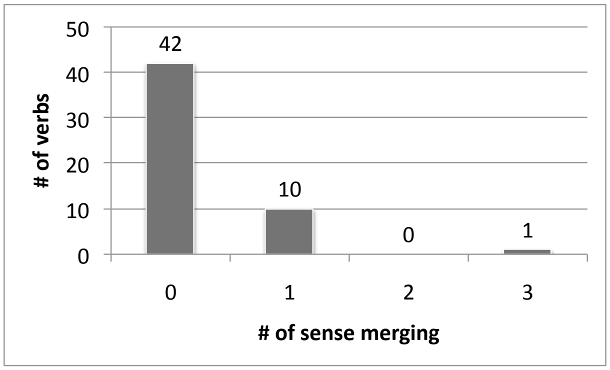

Figure 7: Sense-merging with respect to the number of verbs

Figure 8 shows the number of semantic classes found in each verb set. Out of 53 verb sets, 43 of them have one semantic class, 8 of them have two semantic classes, and so on. This shows that the verb sets we found can be considered as self-contained, semantically-coherent classes.

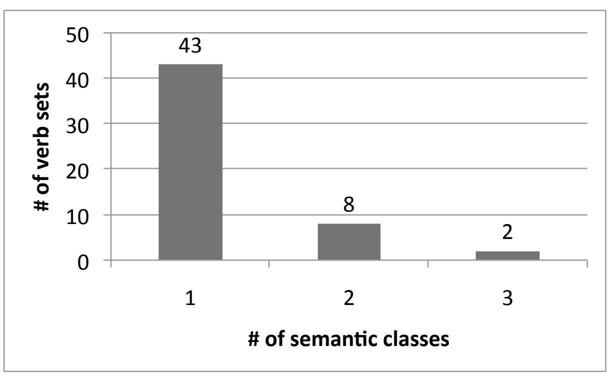

Figure 8: Number of semantic classes in verb sets

5.2.1 Analysis

Although some verbs seem more distant from their least common hypernyms than the others, it does not necessarily mean that their meanings are less closely related to the hypernyms. For example, a Chinese verb, 制定 ('formulate'), is aligned to a set of English verbs {*formulate*, *draft*, *draw*}, which forms the following taxonomy (Figure 9).

'create.1' is the least common hypernym of all verbs in the set. Although *draft* and *draw* are more

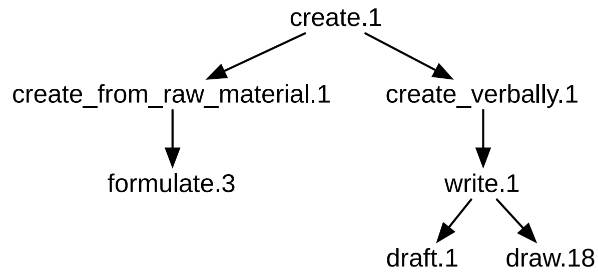

Figure 9: Taxonomy of {*formulate*, *draft*, *draw*}

distant from 'create.1' than *formulate* is, the meanings of 'draft.1' and 'draw.18' are no less related to 'create.1' than the meaning of 'formulate.3'.

Some verb sets that have multiple WordNet semantic classes can be considered as one self-contained semantic class as well. For example in Figure 3, although WordNet did not find any link between 'appear.2,5' and 'exhibit', they share the same semantic meaning at some level: they both can be thought as an action of 'showing'. For a Chinese verb, 保障 ('ensure'), aligned to a set of English verbs {*ensure*, *guarantee*, *safeguard*}, WordNet found *ensure* and *guarantee* as synsets but did not find any link between them and *safeguard*. Nonetheless, it is obvious that *safeguard* has some semantic similarity with the other two verbs.

6 Conclusion and future work

We proposed advancements on a previous technique for improving word alignment using parallel PropBanks. These advancements are based on exploring symmetric predicate similarity using PropBank predicate-argument structures. We formulated the optimal one-to-one mapping problem as a linear assignment problem. With this approach, we gained 12.6% with respect to the F-score over GIZA++ word alignment. Using this method, we demonstrated automatic generation of English to Chinese and Chinese to English semantic class mappings. For English to Chinese mappings, we verified the accuracy of the semantic classes with a human annotator; while for Chinese to English mappings, we took advantage of WordNet resources to show the tight clustering of members in the semantic classes.

Some possible improvements we will explore in the future include taking into account part-of-speech tags when measuring predicate similarity, special handling of light verb constructions (identifying alignment to the true predicates (Hwang et al., 2010)), improved handling of verb conjunctions (where one-to-one mappings may not hold), as well as automatically tuning the bias parameter of the symmetric word alignment score. For Chinese semantic class generation, we would like to identify the Chinese character in a verb expressing the dominant semantics (when that is the case) and expand the verb class memberships based on those characters. Another potential approach is to identify an English verb sense, retrieve all members in the same semantic class (using VerbNet, WordNet, etc), and then merge their Chinese mappings to expand the Chinese semantic class membership. As for semantic class verification, we will explore using more sophisticated WordNet similarity measures (Budanitsky and Hirst, 2006) to check the closeness of the verbs in English verb sets. We will also test the robustness of our techniques with automatic syntactic parsing and semantic role labeling.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics*.

Alexander Budanitsky and Graeme Hirst. 2006. Evaluating wordnet-based measures of semantic distance. *Computational Linguistics*, 32(1):13–47.

Jinho D. Choi, Martha Palmer, and Nianwen Xue. 2009. Using parallel propbanks to enhance word-alignments. In *Proceedings of ACL-IJCNLP workshop on Linguistic Annotation (LAW'09)*, pages 121–124.

Steve DeNeefe and Kevin Knight. 2009. Synchronous tree adjoining machine translation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP'09)*, volume 2, pages 727–736.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. 2007. Learning bilingual semantic frames: Shallow semantic parsing vs. semantic role projection. In *11th Conference on Theoretical and Methodological Issues in Machine Translation*, pages 75–84.

Jena D. Hwang, Archna Bhatia, Claire Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue, and Martha Palmer. 2010. Propbank annotation of multilingual light verb constructions. In *Proceedings of ACL'10 Workshop on Linguistic Annotation (LAW'10)*, pages 82–90.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending verbnet with novel verb classes. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL'07), demonstration session*, pages 177–180.

Harold W. Kuhn. 1955. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97.

Nitin Madnani, Philip Resnik, Bonnie Dorr, and Richard Schwartz. 2008a. Applying automatically generated semantic knowledge: A case study in machine translation. In *NSF Symposium on Semantic Knowledge Discovery, Organization and Use*.

Nitin Madnani, Philip Resnik, Bonnie Dorr, and Richard Schwartz. 2008b. Are multiple reference translations necessary? investigating the value of paraphrased reference translations in parameter optimization. In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas (AMTA'08)*.

David Mareček. 2009a. Improving word alignment using alignment of deep structures. In *Proceedings of the 12th International Conference on Text, Speech and Dialogue*, pages 56–63.

David Mareček. 2009b. Using tectogrammatical alignment in phrase-based machine translation. In *Proceedings of WDS 2009 Contributed Papers*, pages 22—27.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

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

Philip Resnik. 2004. Exploiting hidden meanings: Using bilingual text for monolingual annotation. In Alexander Gelbukh, editor, *Lecture Notes in Computer Science 2945: Computational Linguistics and Intelligent Text Processing*, pages 283–299. Springer.

Dekai Wu and Pascale Fung. 2009. Semantic roles for smt: A hybrid two-pass model. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT'09)*, pages 13–16.

Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the chinese treebank. *Nat. Lang. Eng.*, 15(1):143–172.

Min Zhang, Hongfei Jiang, Ai Ti Aw, Jun Sun, Sheng Li, and Chew Lim Tan. 2007. A tree-to-tree alignment-based model for statistical machine translation. In *Machine Translation Summit XI*.