Using Parallel Corpora for Cross-language Projection of FrameNet Annotation

Richard Johansson and Pierre Nugues
LUCAS, Department of Computer Science, Lund University
Box 118
SE-221 00 Lund, Sweden
{richard, pierre}@cs.lth.se

Abstract

We present a method for producing FrameNet-annotated text for non-English text. The method uses sentence-aligned corpora and transfers target words and frame elements using a word aligner.

The system was tested on an English-Spanish parallel corpus. On the task of target word projection, the system had a precision of 69% and a recall of 59%. For sentences with non-empty targets, it had a precision of 84% and a recall of 81% on the task of transferring frame elements. The approximate precision of the complete frame element bracketing of Spanish text was around 64%.

1 Introduction

The availability of annotated corpora such as FrameNet [1], MUC, or TimeBank [11], has played an immense role in the recent development of automatic systems for semantic processing of text. While manually annotated corpora of high quality exist for English, this is a scarce resource for smaller languages. Since the size of the training corpus is of utmost importance, this could impair significantly the quality of the corresponding language processing tools. All things being equal, the corpus size is the key factor to improve accuracy [2]. Given the annotation cost, it is unrealistic to believe that hand-annotated corpora in smaller languages will ever reach the size of their equivalent counterparts in English.

This article describes an automatic system for FrameNet annotation (target words and frame elements (FEs)) of texts in new languages. It uses an English semantic role labeler to automatically annotate the English sentences in a parallel corpus. A word aligner is then used to transfer the marked-up entities to the target language. We describe results of the system applied to an English-Spanish parallel corpus taken from the proceedings of the European Parliament [6].

2 Related Work

Parallel corpora are now available for many language pairs. English texts aligned with languages with fewer resources such as Basque, Spanish, and Chinese have been used to derive automatically treebanks and parsers. Annotated corpora are much larger in English, which means that language processing tools, including parsers, are generally performing better for this language. Hwa *et al* [5] applied a parser on the English part of a parallel corpus and projected the syntactic structures on texts in the second language. They reported results that rival commercial parsers. Parallel corpora have other applications such as in machine translation. Diab and Resnik [3] also used them to disambiguate word senses.

Yarowsky *et al* [14] describe a method for cross-language projection, using parallel corpora and a word aligner, that is applied to a range of linguistic phenomena, such as named entities and noun chunk bracketing. This technique is also used by Riloff *et al* [12] to transfer annotations for Information Extraction systems.

Padó and Lapata [10] use similar methods, and a set of filtering heuristics, to induce a dictionary of FrameNet target words (frame evoking elements).

3 Method

A fundamental question is whether it is indeed meaningful to project semantic markup of text across languages. Although the counterparts of words will exhibit subtle but significant semantic differences, we believe that a transfer makes sense, since the nature of FrameNet is rather coarse-grained. Even though the words that evoke a frame may not have exact counterparts, it is probable that the frame itself has. When applying the method, we have assumed that a target word that evokes a certain frame will have a counterpart in the target language that evokes the same frame, and that some of the FEs on one side will have counterparts with the same semantic roles on the other. In addition, we made the (obviously simplistic) assumption that the contiguous entities we project are also contiguous on the target side.

Using well-known techniques [4, 8], we trained an SVM-based semantic role labeler using 25000 randomly selected sentences from FrameNet. On a test set from FrameNet, we estimated that our labeler has a precision of 0.72 and a recall of 0.63. The result is slightly lower than the systems at Senseval [8], possibly because we used all frames from FrameNet rather than a subset, and that we did not assume that the frame is known a priori.

We used the Europarl corpus [6] and the included sentence aligner, which uses the Gale-Church algorithm. We removed those instances where one English sentence was mapped to more than one in the target language, and for each pair of sentences, a word alignment was produced using GIZA++ [9]. Figure 1 shows an example of a sentence pair with word alignment. Since we are transferring bracketing from English, the word aligner produces a mapping for each English token. This is why in the figure, only the second English token, rather than the first two, is mapped onto the first Spanish.

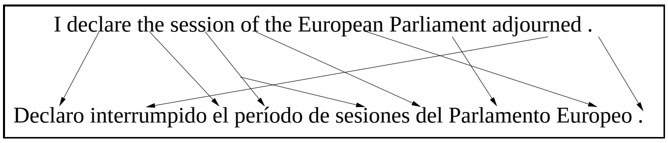

Figure 1: Word alignment example.

We labeled 50 English sentences and transferred the annotation to the Spanish sentences. A set of FEs was produced for each word in the English sentence that acted as a target word in at least one annotated sentence in the FrameNet corpus. 240 target words were found on the English side. Since we did not assume any knowledge of the frame, we used the available semantic roles for all possible interpretations of the target word as features for the classifier. We ignored some common auxiliary verbs: *be*, *have*, *do*, and *get*.

For each entity (target word or FE), we found the target-language counterpart using the maximal span of all the words within the bracketing. We added the constraint that FEs should not cross the target word (in that case, we just used the part that was to the right of the target).

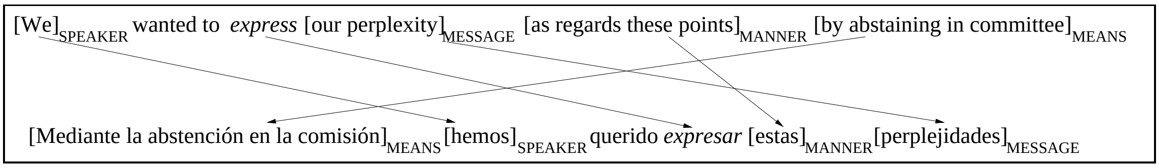

Figure 2: An example of automatic transfer of FEs and target in a sentence from the European Parliament corpus.

Figure 2 shows an automatically annotated sentence in English and its counterpart in Spanish. The example exhibits two sources of errors: first, incorrect English annotation (the MANNER role, caused by a parser error); secondly, two transfer errors ("we" is mapped onto the auxiliary verb "hemos", and "these" is incorrectly mapped onto "estas").

4 Results

We evaluated the system for three cases: transfer of target words, transfer of FEs, and the complete system. For all cases, evaluation was done by manual inspection. We ignored punctuation and articles when checking the boundaries.

In some cases, there was no Spanish counterpart for an entity on the English side. For FEs, the most common reason for this is that Spanish does not have a mandatory subject pronoun as in English (as in Figure 1). In addition, since the translations are not literal and sentence-by-sentence, the target sentence may not express exactly the same information. In the tables below, these cases are listed as N/A.

4.1 Target Word Transfer

We first measured how well target words were projected across languages. Table 1 shows the results. Since our corpus is rather large, and since we perform no FE transfer for target words that could not be transferred, precision is more important than recall for this task in order to produce a high-quality annotation.

Spurious target words were sometimes a problem, especially for the verbs "take" and "make", which seem to occur as a part of multiword units, or as support verbs for noun predicatives, more often than in their concrete sense. When such words were transferred, they were listed as "noise" in Table 1. This problem could often be side-stepped, since the word aligner frequently found no counterparts of these words on the Spanish side.

Correct transfer	102
N/A	53
Overlapping	23
No overlap	9
Lost in transfer	40
Noise	13
Precision	0.69
Recall	0.59

Table 1: Results of target word transfer.

Word sense ambiguity of the target word was a frequent problem. FrameNet often does not cover all senses of a target word (sometimes not even the most common one). We did not have time to try the recent 1.2 release of FrameNet, but we expect that the issue of sense coverage will be less of a problem in the new release. The FrameNet annotators state that the new release has been influenced by their recent annotation of running text.

4.2 FE Transfer

Table 2 shows the results of the transfer of FEs for non-empty targets.

A few errors were caused by the alignment of English personal pronouns with Spanish auxiliary verbs (such as in Figure 2). These cases are listed as "pronoun to auxiliary" in Table 2. Since these cases are restricted and easily detected, we did not include them among the errors when computing precision and recall.

Correct transfer	129
N/A	33
Pronoun to auxiliary	7
Overlapping	9
Lost in transfer	5
No overlap	16
Precision	0.84
Recall	0.81

Table 2: Results of FE transfer for sentences with non-empty targets.

4.3 Full Annotation

We finally made a tentative study on how well the final result turns out. Since we lacked the lexicographical expertise to produce a FrameNet-annotated gold standard in the short time span available, we manually inspected the FEs and labeled them as Acceptable or not. This allowed us to measure the precision of the annotation. Because of the sometimes subtle differences between different frames and different semantic roles in the same frame, the result may be somewhat inexact.

Table 3 shows the results of the complete semantic role labeling for non-empty targets. We have not attempted to label null-instantiated roles.

Acceptable label and boundaries	98
N/A	33
Pronoun to auxiliary	7
Acceptable label, overlapping	12
Incorrect label or no overlap	44
Precision	0.64

Table 3: Results of complete semantic role labeling for sentences with non-empty targets.

The precision (0.64) is consistent with the result on the FrameNet test set (0.72) multiplicated with the transfer precision (0.84), which gives the result 0.60. Extrapolating this argument to the case of recall, we would arrive at a result of $0.63 \cdot 0.81 = 0.51$. However, this figure is probably too high, since there will be FEs on the Spanish side that have no counterpart on the English side.

5 Conclusion and Future Work

We have described a method for projection of FrameNet annotation across languages using parallel corpora and a word aligner. Although the produced data for obvious reasons has inferior quality compared to manually produced data, they can be used as a seed for bootstrapping methods (as argued by Padó and Lapata [10]). In addition, we believe that the method is fully usable in a semi-automatic system. Inspection and correction is less costly than manual annotation from scratch.

We will try to improve the robustness of the methods. Since most sentences in the Parliament debates are long and structurally complicated, it might be possible to improve the data quality by selecting shorter sentences. This should make the task simpler for the parser, the English semantic role labeler, and the word aligner. Parse and alignment probability scores could also be used for selection of data of good quality.

We will create a gold standard in order to be able to estimate the recall, and get more reliable figures for the precision.

Frame assignment is still lacking. We believe that this is best solved using a joint optimization of frame and role assignment as by Thompson *et al* [13], or possibly by applying Lesk's algorithm [7] using the frame definitions. Other aspects of FrameNet annotation that should be addressed include aspectual particles of verbs and support verbs and prepositions.

In the future, we will apply this method for other kinds of semantic annotation of text. One important example is TimeML annotation of events and temporal relations [11].

We will further investigate the projection methods to see if a more sophisticated approach than the maximal span method can be applied. Although our method, which is based on the alignment of raw words, is independent of language, it would be interesting to study if the results could be improved if morpheme information is used. In addition, we would like to study if the boundaries of the projected arguments may be adjusted using a parser or chunker.

Acknowledgements

We thank ISCI for granting us an academic license of the FrameNet data.

References

[1] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of COLING-ACL'98*, pages 86–90, Montréal, Canada, 1998.

[2] Michele Banko and Eric Brill. Scaling to very very large corpora for natural language disambiguation. In *Meeting of the Association for Computational Linguistics*, pages 26–33, 2001.

[3] Mona Diab and Philip Resnik. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL-02)*, 2002.

[4] Daniel Gildea and Daniel Jurafsky. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288, 2002.

[5] Rebecca Hwa, Philip Resnik, Amy Weinberg, and Okan Kolak. Evaluating translational correspondence using annotation projection. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL-02)*, 2002.

[6] Philipp Koehn. Europarl: A multilingual corpus for evaluation of machine translation. Technical report, 2002.

[7] Michael E. Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 1986 SIGDOC Conference*, pages 24–26, 1986.

[8] Ken Litkowski. Senseval-3 task: Automatic labeling of semantic roles. In Rada Mihalcea and Phil Edmonds, editors, *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 9–12, Barcelona, Spain, July 2004. Association for Computational Linguistics.

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

[10] Sebastian Padó and Mirella Lapata. Cross-lingual bootstrapping for semantic lexicons: The case of FrameNet. In *Proceedings of AAAI-05*, Pittsburgh, PA, 2005 (to appear).

[11] James Pustejovsky, Roser Saurí, Andrea Setzer, Rob Gaizauskas, and Bob Ingria. TimeML Annotation Guidelines. Technical report, 2002.

[12] Ellen Riloff, Charles Schafer, and David Yarowsky. Inducing information extraction systems for new languages via cross-language projection. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, 2002.

[13] Cynthia A. Thompson, Roger Levy, and Chris Manning. A generative model for FrameNet semantic role labeling. In *Proceedings of the 14th European Conference on Machine Learning*, 2003.

[14] David Yarowsky, Grace Ngai, and Richard Wicentowski. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of HLT 2001, First International Conference on Human Language Technology Research*, 2001.