Cross-lingual Parse Disambiguation based on Semantic Correspondence

Lea Frermann
Department of Computational Linguistics
Saarland University
frermann@coli.uni-saarland.de

Francis Bond
Linguistics and Multilingual Studies
Nanyang Technological University
bond@ieee.org

Abstract

We present a system for cross-lingual parse disambiguation, exploiting the assumption that the meaning of a sentence remains unchanged during translation and the fact that different languages have different ambiguities. We simultaneously reduce ambiguity in multiple languages in a fully automatic way. Evaluation shows that the system reliably discards dispreferred parses from the raw parser output, which results in a pre-selection that can speed up manual treebanking.

1 Introduction

Treebanks, sets of parsed sentences annotated with a syntactic structure, are an important resource in NLP. The manual construction of treebanks, where a human annotator selects a gold parse from all parses returned by a parser, is a tedious and error prone process. We present a system for simultaneous and accurate partial parse disambiguation of multiple languages. Using the pre-selected set of parses returned by the system, the treebanking process for multiple languages can be sped up.

The system operates on an aligned parallel corpus. The languages of the parallel corpus are considered as mutual semantic tags: As the meaning of a sentence stays constant during translation, we are able to resolve ambiguities which exist in only one of the languages by only accepting those interpretations which are licensed by the other language.

In particular, we select one language as the target language, translate the other language’s semantics for every parse into the target language and thus align maximally similar semantic representations.

The parses with the most overlapping semantics are selected as preferred parses.

As an example consider the English sentence They closed the shop at five, which has the following two interpretations due to PP attachment ambiguity:

1. “At five, they closed the shop”
   close(they, shop); at(close, 5)

2. “The shop at five was closed by them”
   close(they, shop); at/shop, 5

The Japanese translation is also ambiguous, but in a completely different way: it has the possibility of a zero pronoun (we show the translated semantics).

3. 彼らは5時に店を閉めた
   kare ra wa 5 ji ni mise wo shime ta
   “At 5 o’clock, they closed the shop.”
   close(they, shop); at(close, 5)

4. “At 5 o’clock, as for them, someone closed the shop.”
   close(φ, shop); at(close, 5)
   topic(they, close)

We show the semantic representation of the ambiguity with each sentence. Both languages are disambiguated by the other language as only the English interpretation (1) is supported in Japanese, and only the Japanese interpretation (3) leads to a grammatical English sentence.

2 Related Work

There is no group using exactly the same approach as ours: automated parallel parse disambiguation on the basis of semantic analyses. Zhechev and

---

1In fact it has four, as they can be either plural or the androgynous singular, this is also disambiguated by the Japanese.
Way (2008) automatically generate parallel treebanks for training of statistical machine translation (SMT) systems through sub-tree alignment. We do not aim to carry out the complete treebanking process, but to optimize speed and precision of manual creation of high-quality treebanks.

Wu (1997) and others have tried to simultaneously learn grammars from bilingual texts. Burkett and Klein (2008) induce node-alignments of syntactic trees with a log-linear model, in order to guide bilingual parsing. Chen et al. (2011) translate an existing treebank using an SMT system and then project parse results from the treebank to the other language. This results in a very noisy treebank, that they then clean. These approaches align at the syntactic level (using CFGs and dependencies respectively).

In contrast to the above approaches, we assume the existence of grammars and use a semantic representation as the appropriate level for cross-lingual processing. We compare semantic sub-structures, as those are more straightforwardly comparable across different languages. As a consequence, our system is applicable to any combination of languages. The input is plain parallel text, neither side needs to be treebanked.

3 Materials and Methods

We use grammars within the grammatical framework of head-driven phrase-structure grammar (HPSG Pollard and Sag (1994)), with the semantic representation of minimal recursion semantics (MRS; Copestake et al. (2005)). We use two large-scale HPSG grammars and a Japanese-English machine translation system, all of which were developed in the DELPH-IN framework:2 The English Resource Grammar (ERG; Flickinger (2000)) is used for English parsing, and Jacy (Bender and Siegel, 2004) for parsing Japanese. For Japanese to English translation we use Jaen, a semantic-transfer based machine translation system (Bond et al., 2011).

3.1 Semantic Interface and Alignment

For the alignment, we convert the MRS structures into simplified elementary dependency graphs (EDGs), which abstract away information about grammatical properties of relations and scopal information. Preliminary experiments showed that the former kind of information did not contribute to disambiguation performance, as number is typically underspecified in Japanese. As we only consider local information in the alignment, scopal information can be ignored as well. An example EDG is displayed in Figure 1.

An EDG consists of a bag of elementary predicates (EPs) which are themselves composed of relations. Each line in Figure 1 corresponds to one EP. Relations are the elementary building blocks of the EDG, and loosely correspond to words of the surface string. EPs consist either of atomic relations (corresponding to quantifiers), or a predicate-argument structure which is composed of several relations. During alignment, we only consider non-atomic EPs, as quantifiers should be considered as grammatical properties of (lexical) relations, which we chose to ignore.

Given the EDG representations of the translated Japanese sentence, and the original target language EDGs, we can straightforwardly align by matching substructures of different granularity.

Currently, we align at the predicate level. We are experimenting with aligning further dependency relation based tuples, which would allow us to resolve more structural ambiguities.

3.2 The Disambiguation System

Ambiguity in the analyses for both languages is reduced on the basis of the semantic analyses returned for each sentence-pair, and a reduced set of preferred analyses is returned for both languages. For each sentence-pair, we (1) parse the English and the Japanese sentence (MRS$_E$ and MRS$_J$) (2) transfer the Japanese MRS analyses to English MRSs (MRS$_{JE}$) (3) convert the top 11 translated MRSs

Figure 1: EDG for They closed the shop at five.
and the original English MRSs to EDGs (EDG_E and EDG_JE) (4) align every possible E and JE EDG combination and determine the set of best aligning analyses (5) from those, create language specific sets of preferred parses.

We are comparing semantic representations of the same language, the English text from the bilingual corpus and the English machine translation of the Japanese text. In order to increase robustness of our alignment system we not only consider complete translations, but also accept partially translated MRSs in case no complete translation could be produced. This step significantly increases the recall, while the partial MRSs proved to be informative enough for parse disambiguation.

4 Evaluation and Results

We evaluate our model on the task of parse disambiguation. We use full sentence match as evaluation metric, a challenging target.

The Tanaka corpus is used for training and testing (Tanaka, 2001). It is an open corpus of Japanese-English sentence pairs. We use version (2008-11) which contains 147,190 sentence pairs. We hold out 4,500 sentence pairs each for development and test.

For each sentence, we compare the number of theoretically possible alignments with the number of preferred alignments returned by our system. On average, ambiguity is reduced down to 30%. For English 3.76 and for Japanese 3.87 parses out of (at most) 11 analyses remain in the partially disambiguated list: both languages benefit equally from the disambiguation.

We evaluate disambiguation accuracy by counting the number of times the gold parse was present in the partially disambiguated set (full sentence match). Table 1 shows the alignment accuracy results.

The correct parse is included in the reduced set in 80% of the cases for Japanese, and for 82% of the cases in English. We match atomic relations when aligning the semantic structures, which is a very generic method applicable to the vast majority of sentence pairs. This leads to a recall score of 99%, and an F-Score of 89.7% and 88.7% for English and Japanese, respectively.

The reduced list of parser analyses can be further ranked by the parse ranking model which is included in the parsers of the respective languages (the same models with which we determined the top 11 analyses). Given this ranking, we can evaluate how often the preferred parse is ranked top in our partially disambiguated list; results are shown in the two bottom lines of Table 1.

A ranked list of possible preferred parses whose top rank corresponds with a high probability to the gold parse should further speed up the manual treebanking process.

Performance in the context of the whole pipeline

The performance of parsers and MT system strongly influences the end-to-end results of the presented system. In the results given above, this influence is ignored. We lose around 29% of our data because no parse could be produced in one or both languages, or no translation could be produced. and a further 5% of the sentences did not have the gold parse in the original set of analyses (before alignment): our system could not possibly select the correct parse in those cases.

5 Discussion

Our system builds on the output of two parsers and a machine translation system. We reduce ambiguity for all sentence pairs where a parse could be created for both languages, and for which there was at least a partial translation. For these sentences, the cross-lingual alignment component achieves a recall of above 99%, such that we do not lose any addi-

|          | English | Japanese |
|----------|---------|----------|
| Prec     | F       | Prec     | F       |
| Included | 0.820   | 0.897    | 0.804   | 0.887   |
| First Rank | 0.659   | 0.791    | 0.676   | 0.803   |
| MRR      | 0.713   | 0.829    | 0.725   | 0.837   |

Table 1: Accuracy and F-scores for disambiguation performance of our system. Recall was 99% in every case. ‘Included’: inclusion of the gold parse in the reduced set of parses or not. ‘First Rank’: ranking of the preferred parse as top in the reduced list. ‘MRR’: mean reciprocal rank of the gold parse in the list.
tional data. The parsers and the MT system include
a parse ranking system trained on human gold anno-
tations. We use these models in parsing and transla-
tion to select the top 11 analyses. Our system thus
depends on a range of existing technologies. How-
ever, these technologies are available for a range of
languages, and we use them for efficient extension
of linguistic resources.

The effectiveness of cross-lingual parse disam-
biguation on the basis of semantic alignment highly
depends on the languages of choice. Given that we
exploit the differences between languages, pairs of
less related languages should lead to better disam-
biguation performance. Furthermore, disambiguat-
ing with more than two languages should improve
performance. Some ambiguities may be shared be-
tween languages. 4

One weakness when considering the disam-
biguated sentences as training for a parse ranking
model is that the translation fails on similar kinds of
sentences, so there are some phenomena which we
get no examples of — the automatically trained tree-
bank does not have a uniform coverage of phenom-
ena. Our models may not discriminate some phe-
omena at all.

Our system provides large amounts of automati-
cally annotated data at the only cost of CPU time:
so far we have disambiguated 25,000 sentences: 10
times more than the existing hand annotated gold
data. Using the parser output for speeding up man-
ual treebanking is most effective if the gold parse is
reliably included in the reduced set of parses. In-
creasing precision by accepting more than only the
most overlapping parses may lead to more effective
manual treebanking.

The alignment method we propose does not make
any language-specific assumptions, nor is it limited
to align two languages only. The algorithm is very
flexible, and allows for straightforward exploration
of different numbers and combinations of languages.

6 Conclusion and Future Work

Translating a sentence into a different language
changes its surface form, but not its meaning. In
parallel corpora, one language can be viewed as a
semantic tag of the other language and vice versa,
which allows for disambiguation of phenomena
which are ambiguous in only one of the languages.

We use the above observations for cross-lingual
parse disambiguation. We experimented with the
language pair of English and Japanese, and were
able to accurately reduce ambiguity in parser anal-
yses simultaneously for both languages to 30% of
the starting ambiguity. The remaining parses can be
used as a pre-selection to speed up the manual tree-
banking process.

We started working on an extrinsic evaluation of
the presented system by training a discriminative
parse ranking model on the output of our alignment
process. Augmenting the Gold training data with
our data improves the model. Our next step will
be to evaluate the system as part of the treebanking
process, and optimize the parameters such as disam-
biguation precision vs. amount of disambiguation.

As no language-specific assumptions are hard
coded in our disambiguation system, it would be
very interesting to apply the system to different lan-
guage pairs as well as groups of more than two lan-
guages. Using a group of languages for disambigua-
tion will likely lead to increased and more accurate
disambiguation, as more constraints are imposed on
the data.

Probably the most important goal for future work
is improving the recall achieved in the complete dis-
ambiguation pipeline. Many sentence-pairs cannot
be disambiguated because either no parse can be
generated for one or both languages, or no (par-
tial) translation can be produced. Following the
idea of partial translations, partial parses may be a
valid backoff. For purposes of cross-lingual align-
ment, partial structures may contribute enough in-
formation for disambiguation. There has been work
regarding partial parsing in the HPSG community
(Zhang and Kordoni, 2008), which we would like to
explore. There is also current work on learning more
types and instances of transfer rules (Haugereid and
Bond, 2011).

Finally, we would like to investigate more align-
ment methods, such as dependency relation based
alignment which we started experimenting with, or
EDM-based metrics as presented in (Dridan and
Oepen, 2011).

---

4 For example the PP attachment ambiguity in John said that
he went on Tuesday where either the saying or the going could
have happened on Tuesday holds in both English and Japanese.
Acknowledgments

This research was supported in part by the Erasmus Mundus Action 2 program MULTI of the European Union, grant agreement number 2009-5259-5 and the the joint JSPS/NTU grant on Revealing Meaning Using Multiple Languages. We would like to thank Takayuki Kuribayashi and Dan Flickinger for their help with the treebanking.

References

Emily M. Bender and Melanie Siegel. 2004. Implementing the syntax of Japanese numeral classifiers. In *Proceedings of the IJCNLP-2004*.

Francis Bond, Stephan Oepen, Eric Nichols, Dan Flickinger, Erik Velldal, and Petter Haugereid. 2011. Deep open-source machine translation. *Machine Translation*, 25(2):87–105.

David Burkett and Dan Klein. 2008. Two languages are better than one (for syntactic parsing). In *Proceedings of EMNLP, 2008*.

Wenliang Chen, Jun’ichi Kazama, Min Zhang, Yoshimasa Tsuruoka, Yujie Zhang, Yiou Wang, Kentaro Torisawa, and Haizhou Li. 2011. SMT helps bitext dependency parsing. In *Conference on Empirical Methods in Natural Language Processing (EMNLP2011)*, pages 73–83. Edinburgh.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal recursion semantics – an introduction. *Research on Language and Computation*, 3:281–332.

Rebecca Dridan and Stephan Oepen. 2011. Parser evaluation using elementary dependency matching. In *Proceedings of IWPT*.

Dan Flickinger. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1):15–28. (Special Issue on Efficient Processing with HPSG).

Petter Haugereid and Francis Bond. 2011. Extracting transfer rules for multiword expressions from parallel corpora. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*.

Carl Pollard and Ivan A. Sag. 1994. *Head Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Yasuhito Tanaka. 2001. Compilation of a multilingual parallel corpus. In *Proceedings of PACLING 2001*.

Dekai Wu. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational Linguistics*, 23(3):377–403.

Yi Zhang and Valia Kordoni. 2008. Robust parsing with a large HPSG grammar. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*.

Ventsislav Zhechev and Andy Way. 2008. Automatic generation of parallel treebanks. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*.