Bitext-Based Resolution of German Subject-Object Ambiguities

Florian Schwarck **Alexander Fraser** **Hinrich Schütze**
Institute for Natural Language Processing
University of Stuttgart
{koehlefn,fraser}@ims.uni-stuttgart.de

Abstract

We present a method for disambiguating syntactic subjects from syntactic objects (a frequent ambiguity) in German sentences taken from an English-German bitext. We exploit the fact that subject and object are usually easily determined in English. We show that a simple method disambiguates some subject-object ambiguities in German, while making few errors. We view this procedure as the first step in automatically acquiring (mostly) correct labeled data. We also evaluate using it to improve a state of the art statistical parser.

1 Introduction

Ambiguity of grammatical role is a problem when parsing a number of natural languages. In German, subject-object ambiguities are frequent. The sentence "Die Maus jagt die Katze" "the – mouse – chases – the – cat" exhibits such an ambiguity. Because word order is freer in German than in English, the sentence has two possible meanings: (i) The cat is chasing the mouse and (ii) the mouse is chasing the cat. We exploit the fact that such ambiguities are much less frequent in languages that possess a less flexible syntax than German. In English, the translation of the sentence "Die Maus jagt die Katze" is not ambiguous. If we have access to this translation, we can use this information to disambiguate the German sentence. The English translation is viewed as a surrogate for both contextual knowledge from the text and for world knowledge.

We present a method for disambiguating the subject and object roles in German sentences. We use an English-German *bitext* and exploit the fact that subject and object roles are rarely ambiguous in English. Using a new gold standard we created we show that our method disambiguates a significant proportion of subject-object ambiguities in German with high precision. We view this procedure as the first step in automatically acquiring (mostly) correct labeled data for training a statistical disambiguator that can be used on German text (even when no translation is available). In addition to measuring algorithm performance directly, we present experiments on improving the disambiguation of BitPar, a state of the art statistical parser.

2 Algorithm

Data and Word Alignment. We use the aligned English and German sentences in Europarl (Koehn, 2005) for our experiments. The corpus contains long and complex sentences. To establish translational correspondence between parallel sentences we use GIZA++ (Och and Ney, 2003). Its input is a tokenized parallel corpus. We lemmatized the text prior to aligning it.

Procedure. Figure 1 shows the architecture of our system. The boxes signify data sets, while the lines are processes applied to the data sets. The paper presents two applications. The first is the creation of a set of disambiguated German sentences (which involves word alignments in the upper right corner, and the use of parsers in the middle of the graphic). We also present a reranking of the N-best parses produced by BitPar (Schmid, 2004), a state of the art statistical parser (bottom of the graphic).

For processing of German we chose FSPar

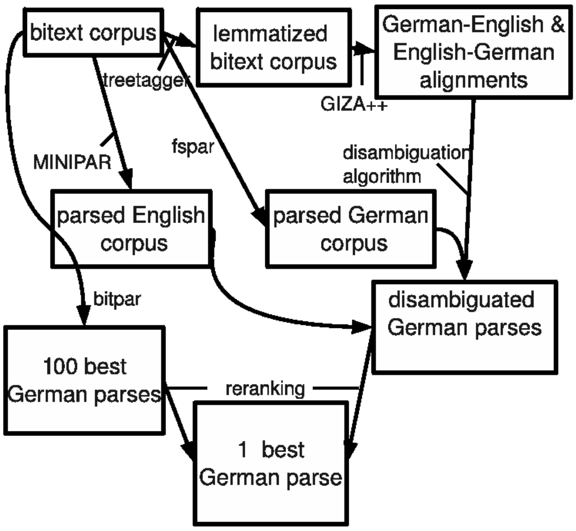

Figure 1: System Architecture

(Schiehlen, 2003), a fast shallow dependency parser. FSPar has extensive lexical knowledge which helps it to find subject-object ambiguities with high accuracy, but it does not try to resolve such ambiguities.

The key to our approach is to project syntactic roles from English text. For English parsing we used MINIPAR (Lin, 1998).

Based on FSPar's analysis, all German sentences with a subject-object ambiguity (about a third) were selected from EuroParl. The parallel English sentences were parsed with MINIPAR.

Words marked as ambiguous by FSPar were then processed using our algorithm. If an ambiguous German word was aligned to an English word that MINIPAR had (unambiguously) assigned the grammatical role of subject or object, then the syntactic role of the German word was defined by this information, see Figure 2.

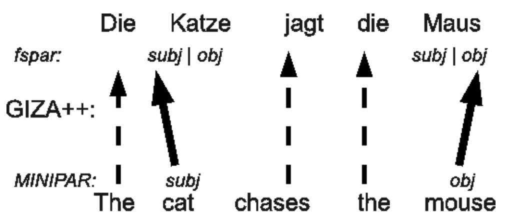

Figure 2: Disambiguation Algorithm

We used standard heuristics for improving word alignment (Och and Ney, 2003; Koehn et al., 2003), but there were many misalignments of ambiguous German words. In order for the procedure to work, we require that the German word to be disambiguated be aligned to the English subject or object. For this reason, we implemented *second guessing* based on a dictionary that lists for every German word the 10 most frequently aligned English words (found using the word alignment of all of Europarl). If an ambiguous German word was either unaligned or not aligned to the English subject or object, it was checked whether a dictionary translation was part of the parallel sentence and marked as subject or object by MINIPAR. If so, this dictionary word was used for disambiguation.

3 Evaluation

Gold Standard. We had access to a small set of gold standard parses (Padó and Lapata, 2009), but decided to create a larger corpus. We found that FSPar had acceptable performance for finding subject-object ambiguities[1]. The syntactic roles of words marked as ambiguous by FSPar were annotated. Four annotators annotated the syntactic roles in 4000 sentences using a graphical user interface (GUI). The GUI showed the ambiguous words in context and gave the annotator four different subject-object labels to choose from for each ambiguous word: *subject*, *object*, *expletive es* and *none*. Because the syntactic expletive "es" (English gloss: 'it') is frequent in German, as in "es scheint zu regnen" 'it appears to be raining,' we created a separate label for expletive "es", which is not treated as a subject.[2] The statistics are shown in table 1.

1000 sentences were annotated by all four annotators. Inter-annotator agreement was sufficient ($\kappa = 0.77$ on average (Carletta, 1996)).

Evaluation Measures. The output of our algorithm labels each word that FSPar classified as ambiguous with one of the three possible labels *subject*,

[1]FSPar has a very high precision in detecting subject-object ambiguities, as can be seen in Table 1 (approximately 0.955, the sum of two left columns divided by sum of all cells). We tried to get an idea of recall using the smaller gold standard. We made conservative assumptions about recall errors which we manually checked on a small sample, details are omitted. Using these assumptions led to an estimate for recall of 0.733, but true recall is likely higher.

[2]German "es" is also frequently used as a non-expletive, where it can take a syntactic role.

	subj	obj	expl_es	none
Annotator1	4152	3210	115	150
Annotator2	4472	3359	92	226
Annotator3	4444	3584	42	155
Annotator4	4027	3595	9	650

Table 1: Annotator decisions on the full gold standard

		DE2EN	Refined	GDFA	Intersection
	P	0.8412	0.8381	0.8353	0.8551
nosg	R	0.4436	0.3856	0.3932	0.3380
	F_1	0.5809	0.5282	0.5347	0.4845
	P	0.7404	0.7307	0.7310	0.7240
sg	R	0.5564	0.4873	0.4946	0.4528
	F_1	0.6353	0.5847	0.5900	0.5571
	P	0.9239	0.9203	0.9192	0.9277
filter-nosg	R	0.3940	0.3397	0.3461	0.2984
	F_1	0.5524	0.4962	0.5028	0.4515
	P	0.8458	0.8358	0.8369	0.8290
filter-sg	R	0.4839	0.4213	0.4279	0.3898
	F_1	0.6156	0.5602	0.5662	0.5302

Table 2: Precision, Recall and F_1 of the algorithm.

object and *no decision*[3]. We use the standard evaluation metrics **Precision** (P, the percentage of subject and object labelings in our hypothesis that are correct), **Recall** (R, the percentage of subject and object labelings in the gold standard that are correctly labeled in the hypothesis), and balanced **F** (F_1).

4 Experiments

Algorithm Performance. Table 2 shows the performance of our algorithm when evaluated against the manual annotation[4]. The lines *nosg*, *sg*, *filter-nosg* and *filter-sg* denote different configurations of the algorithm: Second guessing (section 2) was ("sg") or was not ("nosg") applied and filtering was ("filter") or was not applied. The filter increases precision by only keeping labels of subjects and objects that occur in the default order (e.g., the subject is to the left of the object in the main clause). As an aid to the user, FSPar presents such a determination of default order depending on its classification of clause type[5]. The columns indicate the heuristic postprocessing we applied to GIZA++'s alignment. *DE2EN* is the 1-to-N alignment calculated using German as the source language and English as the target language (i.e., each English word is linked to exactly zero or one German words).

As we see in table 2, with the most strict setup, *filter-nosg*, the algorithm resolves subject-object ambiguities with a precision of more than 92% but the best recall is only 39.4%, obtained using *DE2EN*. Second guessing increases recall but leads to losses in precision. The best precision result without the filter is 85.5%.

Improving BitPar's Subject-Object Decisions. For improving BitPar (which always tries to disambiguate subject-object), our baseline is the accuracy of the most probable parse shown in table 3, row 1.

	configuration	P	R	F_1
1	top-1 (no change)	0.8088	0.8033	0.8060
2	relabeling *nosg*	0.7998	0.8176	0.8086
3	relabeling *filter-nosg*	0.8229	0.8344	0.8286
4	reranking *nosg*	0.8082	0.8123	0.8102
5	reranking *filter-nosg*	0.8145	0.8143	0.8144

Table 3: Precision, Recall and F_1 of changing BitPar decisions, DE2EN alignment

Using the most probable parse from BitPar, we relabel a word "subject" or "object" if our system indicates to do so. With the algorithm alone we are able to improve recall (table 3, row 2). When we add the filter both precision and recall are improved (row 3). This experiment measures the improvement possible if our syntactic role information were directly integrated as a hard constraint into a parser (see section 5).

We now perform a simple reranking experiment, using BitPar's 100-best parses. For each sentence we choose the parse which agrees with as many of the subject/object decisions of the algorithm as possible (once again ignoring words where the algorithm chooses no decision). In case of ties in the number of agreements, we take the most probable parse. The results are in rows 4–5. Reranking increases F_1 by about 0.8%.

5 Related Work

Syntactic projection has been used to bootstrap treebanks in resource poor languages (Yarowsky and

[3]If *expletive es* or *none* was annotated, the system is correct if it does not make a decision.

[4]Because of problems with BitPar caused by preprocessing for FSPar, we use 11,279 sentences of the 13,000 annotated.

[5]Using this determination alone results in P 0.7728 R 0.8206 F 0.7960, very high recall but low precision.

Ngai, 2001; Hwa et al., 2005). In contrast with such work, we are addressing subject-object ambiguity in German. German parsers have no access to the contextual and world knowledge necessary to resolve this ambiguity.

Work on *projecting semantic roles* (Padó and Lapata, 2009; Fung et al., 2007) requires both syntactic parsing and semantic role labeling and is concerned with filling in the complete information in a semantic frame. Our approach is simpler and concerned only with syntactic disambiguation, not semantic projection. We focus only on difficult cases of subject-object ambiguity and although we do not always make a prediction, we obtain levels of precision that projection approaches making no use of knowledge of German syntax cannot achieve.

In *bitext parsing*, Burkett and Klein (2008) and Fraser et al. (2009) used feature functions defined on triples of (parse tree in language 1, parse tree in language 2, word alignment), combined in a log-linear model trained to maximize parse accuracy, requiring translated treebanks. We focus only on subject-object disambiguation in German, and annotated a new gold standard. We work on sentences that a partial parser has determined to be ambiguous. Fossum and Knight (2008) and Huang et al. (2009) improve English prepositional phrase attachment using features from an unparsed Chinese sentence. The latter work integrated the PP-attachment constraint (detected from the Chinese translation) directly into an English shift-reduce parser. As we have shown in the labeling experiment, integrating our subject-object disambiguation into BitPar could result in further increases beyond 100-best reranking.

6 Conclusion

We demonstrated the utility of bitext-based disambiguation of grammatical roles. We automatically created a large corpus of 164,874 disambiguated subject-object decisions with a precision of over 92%. This corpus will be of use in future research on syntactic role preferences and for the training of monolingual subject-object disambiguators. We presented a prototype application of subject-object disambiguation through a simple reranking of the 100-best list output by BitPar, and showed a possible further improvement if integrated in the parser. The new gold standard, which is publicly available, will hopefully be useful for work on both monolingual and bitext-based disambiguation.

Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft grants SFB 732 and MorphoSynt.

References

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

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22.

Victoria Fossum and Kevin Knight. 2008. Using bilingual Chinese-English word alignments to resolve PP-attachment ambiguity in English. In *AMTA*.

Alexander Fraser, Renjing Wang, and Hinrich Schütze. 2009. Rich bitext projection features for parse reranking. In *EACL*.

Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. 2007. Learning bilingual semantic frames: Shallow semantic parsing vs. semantic role projection. In *TMI*.

Liang Huang, Wenbin Jiang, and Qun Liu. 2009. Bilingually-constrained (monolingual) shift-reduce parsing. In *EMNLP*.

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Nat. Lang. Eng.*, 11(3).

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *HLT-NAACL*.

Philipp Koehn. 2005. Europarl: a parallel corpus for statistical machine translation. In *MT Summit X*.

Dekang Lin. 1998. Dependency-based evaluation of MINIPAR. In *Workshop on Eval of Parsing Systems*.

Franz J. Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1).

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

Michael Schiehlen. 2003. A cascaded finite-state parser for German. In *Research Notes (EACL)*.

Helmut Schmid. 2004. Efficient parsing of highly ambiguous context-free grammars with bit vectors. In *COLING*.

David Yarowsky and Grace Ngai. 2001. Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora. In *NAACL*.