Cross-Lingual Induction for Deep Broad-Coverage Syntax: A Case Study on German Participles

Sina Zarrieß **Aoife Cahill** **Jonas Kuhn** **Christian Rohrer**
Institut für Maschinelle Sprachverarbeitung (IMS), University of Stuttgart
{zarriesa,cahillae,jonas.kuhn,rohrer}@ims.uni-stuttgart.de

Abstract

This paper is a case study on cross-lingual induction of lexical resources for deep, broad-coverage syntactic analysis of German. We use a parallel corpus to induce a classifier for German participles which can predict their syntactic category. By means of this classifier, we induce a resource of adverbial participles from a huge monolingual corpus of German. We integrate the resource into a German LFG grammar and show that it improves parsing coverage while maintaining accuracy.

1 Introduction

Parallel corpora are currently exploited in a wide range of induction scenarios, including projection of morphologic (Yarowsky et al., 2001), syntactic (Hwa et al., 2005) and semantic (Padó and Lapata, 2009) resources. In this paper, we use cross-lingual data to learn to predict whether a lexical item belongs to a specific syntactic category that cannot easily be learned from monolingual resources. In an application test scenario, we show that this prediction method can be used to obtain a lexical resource that improves deep, grammar-based parsing.

The general idea of cross-lingual induction is that linguistic annotations or structures, which are not available or explicit in a given language, can be inferred from another language where these annotations or structures are explicit or easy to obtain. Thus, this technique is very attractive for cheap acquisition of broad-coverage resources, as is proven by the approaches cited above. Moreover, this induction process can be attractive for the induction of deep (and perhaps specific) linguistic knowledge that is hard to obtain in a monolingual context. However, this latter perspective has been less prominent in the NLP community so far.

This paper investigates a cross-lingual induction method based on an exemplary problem arising in the deep syntactic analysis of German. This showcase is the syntactic flexibility of German participles, being morphologically ambiguous between verbal, adjectival and adverbial readings, and it is instructive for several reasons: first, the phenomenon is a notorious problem for linguistic analysis and annotation of German, such that standard German resources do not represent the underlying analysis. Second, in Zarrieß et al. (2010), we showed that integrating the phenomenon of adverbial participles in a naive way into a broad-coverage grammar of German leads to significant parsing problems, due to spurious ambiguities. Third, it is completely straightforward to detect adverbial participles in cross-lingual data since in other languages, e.g. English or French, adverbs are often morphologically marked.

In this paper, we use instances of adverbially translated participles in a parallel corpus to bootstrap a classifier that is able to identify an adverbially used participle based on its monolingual syntactic context. In contrast to what is commonly assumed, we show that it is possible to detect adverbial participles using only a relatively narrow context window. This classifier enables us to identify an occurence of an adverbial participle independently of its translation in a parallel corpus, going far beyond the induction methodology in Zarrieß et al. (2010). By means of the participle classifier, we can extract new types of adverbial participles from a larger corpus of German newspaper text and substantially augment the size of the resource extracted only on Europarl data. Finally, we integrate this new resource into the German LFG grammar and show that it improves coverage without negatively affecting performance.

Coling 2010: Poster Volume, pages 1426–1434,
Beijing, August 2010

The paper is structured as follows: in Section 2, we describe the linguistic and computational problems related to the parsing of adverbial participles in German. Section 3 introduces the general idea of using the translation data to find instances of different participle categories. In Section 4, we illustrate the training of the classifier, evaluating the impact of the context window and the quality of the training data obtained from cross-lingual text. In Section 5, we apply the classifier to new, monolingual data and describe the extension of the resource for adverbial participles. Section 6 evaluates the extended resource by means of parsing experiments using the German LFG grammar.

2 The Problem

In German, past perfect participles are ambiguous with respect to their morphosyntactic category. As in other languages, they can be used as part of the verbal complex (Example (1-a)) or as adjectives (Example (1-b)). Since German adjectives can generally undergo conversion into adverbs, participles can also be used adverbially (Example (1-c)). The verbal and adverbial participle forms are morphologically identical.

(1) a. Sie haben das Experiment **wiederholt**.
'They have repeated the experiment.'
b. Das **wiederholte** Experiment war erfolgreich.
'The repeated experiment was succesful.'
c. Sie haben das Experiment **wiederholt** abgebrochen.
'They cancelled the experiment repeatedly.'

Moreover, German adjectival modifiers can be generally used as predicatives that can be either selected by a verb (Example (2-a)) or that can occur as free predicatives (Example (2-b)).

(2) a. Er scheint **begeistert** von dem Experiment.
'He seems enthusiastic about the experiment.'
b. Er hat **begeistert** experimentiert.
'He has experimented enthusiastic.'

Since predicative adjectives are not inflected, the surface form of a German participle is ambiguous between a verbal, predicative or adverbial use.

2.1 Participles in the German LFG

In order to account for sentences like (1-c), an intuitive approach would be to generally allow for adverb conversion of participles in the grammar. However, in Zarrieß et al. (2010), we show that such a rule can have a strong negative effect on the overall performance of the parsing system, despite the fact that it produces the desired syntactic and semantic analysis for specific sentences. This problem was illustrated using a German LFG grammar (Rohrer and Forst, 2006) constructed as part of the ParGram project (Butt et al., 2002). The grammar is implemented in the XLE, a grammar development environment which includes a very efficient LFG parser and a stochastic disambiguation component which is based on a log-linear probability model (Riezler et al., 2002).

In Zarrieß et al. (2010), we found that the naive implementation of adverbial participles in the German LFG, i.e. in terms of a general grammar rule that allows for participles-adverb conversion, leads to spurious ambiguities that mislead the disambiguation component of the grammar. Moreover, the rule increases the number of timeouts, i.e. sentences that cannot be parsed in a predefined amount of time (20 seconds). Therefore, we observe a drop in parsing accuracy although grammar coverage is improved. As a solution, we induced a lexical resource of adverbial participles based on their adverbial translations in a parallel corpus. This resource, comprising 46 participle types, restricts the adverb conversion such that most of the spurious ambiguities are eliminated.

To assess the impact of specific rules in a broad-coverage grammar, possibly targeting medium-to-low frequency phenomena, we have established a fine-grained evaluation methodology. The challenge posed by these low-frequent phenomena is typically two-fold: on the one hand, if one takes into account the disambiguation component of the grammar and pursues an evaluation of the most probable parses on a general test set, the new grammr rule cannot be expected to show a positive effect since the phenomenon is not likely to occur very often in the test set. On the other hand, if one is interested in a linguistically precise grammar, it is very unsatisfactory to reduce grammar coverage to statistically frequent phenomena. Therefore, we combined a coverage-oriented evaluation on specialised testsuites with a quantitative evaluation including disambiguation, making sure that

the increased coverage does not lead to an overall drop in accuracy. The evaluation methodolgy will also be applied to evaluate the impact of the new participle resource, see Section 6.

2.2 The Standard Flat Analysis of Modifiers

The fact that German adjectival modifiers can generally undergo conversion into adverbs without overt morphological marking is a notorious problem for the syntactic analysis of German: there are no theoretically established tests to distinguish predicative adjectives and adverbials, see Geuder (2004). For this reason, the standard German tag set assigns a uniform tag ("ADJD") to modifiers that are morphologically ambiguous between an adjectival and adverbial reading. Moreover, in the German treebank TIGER (Brants et al., 2002) the resulting syntactic differences between the two readings are annotated by the same flat structure that does not disambiguate the sentence.

Despite certain theoretical problems related to the analysis of German modifiers, their interpretation in real corpus sentences is often unambiguous for native speakers. As an example, consider example (3) from the TIGER treebank. In the sentence, the participle *unterschrieben* (*signed*) clearly functions as a predicative modifier of the sentence's subject. The other, theoretically possible reading where the participle would modify the verb *send* is semantically not acceptable. However, in TIGER, the participle is analysed as an ADJD modifier attached under the VP node which is the general analysis for adjectival and adverbial modifiers.

(3) Die sollte **unterschrieben** an die Leitung
It should signed to the administration
zurückgesandt werden.
sent back be.
'It should be sent back signed to the administation.'

Sentence (4) (also taken from TIGER) illustrates the case of an adverbial participle. In this example, the reading where *angemessen* (*adequately*) modifies the main verb is the only one that is semantically plausible. In the treebank, the participle is tagged as ADJD and analysed as a modifier in the VP.

(4) Der menschliche Geist läßt sich rechnerisch nicht
The human mind lets itself computationally not
angemessen simulieren.
adequately simulate.
'The human mind cannot be adequately simulated in a computational way.'

The flat annotation strategy adopted for modifiers in the standard German tag set and in the treebank TIGER entails that instances of adverbs (and adverbial participles) cannot be extracted from automatically tagged, or parsed, text. Therefore, it would be very hard to obtain training material from German resources to train a system that automatically identifies adverbially used participles. However, the intuition corroborated by the examples presented in this section is that the structures can actually be disambiguated in many corpus sentences.

In the following sections, we show how we exploit parallel text to obtain training material for learning to predict occurences of adverbial participles, without any manual effort. Moreover, by means of this technique, we can substantially extend the grammatical resource for adverbial participles compared to the resource that can be directly extracted from the parallel text.

3 Participles in the Parallel Corpus

The intuition of the cross-lingual induction approach is that adverbial participles can easily be extracted from parallel corpora since in other languages (such as English or French) adverbs are often morphologically marked and easily labelled by statistical PoS taggers. As an example, consider sentence (5) extracted from Europarl, where the German participle *verstärkt* is translated by an English adverb (*increasingly*).

(5) a. Nicht ohne Grund sprechen wir **verstärkt**
Not without reason speak we increasingly
vom Europa der Regionen.
of a Europe of the Regions.
b. It is not without reason that we **increasingly** speak in terms of a Europe of the Regions.

The idea is to project specific morphological information about adverbs which is overt in languages like English onto German where adverbs cannot be directly extracted from tagged data. While this idea might seem intuitively straightfor-

ward, we also know that translation pairs in parallel data are not always lingusitically parallel, and as a consequence, word-alignment is not always reliable. To assess the impact of non-parallelism in adverbial translations of German participles, we manually annotated a sample of 300 translations. This data also constitutes the basis for the experiments reported in Section 4.

3.1 Data

Our experiments are based on the same data as in (Zarrieß et al., 2010). For convenience, we provide a short description here.

We limit our investigations to non-lexicalised participles occuring in the Europarl corpus and not yet recorded as adverbs in the lexicon of the German LFG grammar (5054 participle types in total). Given the participle candidates, we extract the set of sentences that exhibit a word alignment between a German participle and an English, French or Dutch adverb. The word alignments have been obtained with GIZA++. The extraction yields 27784 German-English sentence pairs considering all alignment links, and 5191 sentence pairs considering only bidirectional alignments between a participle and an English adverb.

3.2 Systematic Non-Parallelism

For data exploration and evaluation, we annotated 300 participle alignments out of the 5191 German-English sentences (with a bidirectional participle-adverb alignment). We distinguish the following annotation categories: (i) parallel translation, adverb information can be projected, (ii) incorrect alignment, (iii) correct alignment, but translation is a multi-word expression, (iv) correct alignment, but translation is a paraphrase (possibly involving a translation shift).

Parallel Cases In our annotated sample of English adverb - German participle pairs, 43%[1] of the translation instances are parallel in the sense that the overt adverb information from the English side can be projected onto the German participle. This means that if we base the induction technique on word-alignments alone, its precision would be relatively low.

Non-Parallel Cases Taking a closer look at the non-parallel cases in our sample (57% of the translation pairs), we find that 47% of this set are due to incorrect word alignments. The remaining 53% thus reflect regular cases of non-parallel translations. A typical configuration which makes up 30% of the the non-parallel cases is exemplified in (6) where the German main verb *vorlegen* is translated by the English multiword expression *put forward*.

(6) a. Wir haben eine Reihe von Vorschlägen **vorgelegt**.
b. We have **put forward** a number of proposals.

An example for the general paraphrase or translation shift category is given in Sentence (7). Here, the translational correspondence between *gekommen* (*arrived*) and the adverb *now* is due to language-specific, idiomatic realisations of an identical underlying semantic concept. The paraphrase translations make up 23% of the non-parallel cases in the annotated sample.

(7) a. Die Zeit ist noch nicht **gekommen** .
That time is yet not arrived.
b. That time is not **now** .

Furthermore, it is noticeable that the cross-lingual approach seems to inherently factor out the ambiguity between predicative and adverbial participles. In our annotated sample, there are no predicative participles that have been translated by an English adverb.

3.3 Filtering Mechanisms

The data analysis in the previous section, showing only 43% of parallel cases in English adverb translations for German participles, mainly confirms other studies in annotation projection which find that translational correspondences only allow for projection of linguistic analyses in a more or less limited proportion (Yarowsky et al., 2001; Hwa et al., 2005; Mihalcea et al., 2007).

In previous studies on annotation projection, quite distinct filtering methods have been proposed: in Yarowsky et al. (2001), projection errors are mainly attributed to word alignment errors and filtered based on translation probabilities.

[1] The diverging figures we report in Zarrieß et al. (2010) were due to a small bug in the script and it does not affect the overall interpretation of the data.

Hwa et al. (2005) find that errors in the projection of syntactic relations are also due to systematic grammatical divergences between languages and propose correcting these errors by means of specific, manually designed filters. Bouma et al. (2008) make similar observations to Hwa et al. (2005), but try to replace manual correction rules by filters from additional languages.

In Zarrieß et al. (2010), we compared a number of filtering techniques on our participle data. The 300 annotated translation instances are used as a test set for evaluation. In particular, we have established that a combination of syntactic dependency-based filters and multilingual filters can very accurately separate non-parallel translations from parallel ones where the adverb information can be projected. In Section 4, we show that these filtering techniques are also very useful for removing noise from the training material that we use to build a classifier.

4 Bootstrapping a German Participle Classifier from Crosslingual Data

In the previous section, we have seen that German adverbial participles can be easily found in crosslingual text by looking at their translations in a language that morphologically marks adverbials. In previous work, we exploited this observation by directly extracting types of adverbial participles based on word alignment links and the filtering mechanisms mentioned in Section 3. However, this method is very closely tied to data in the parallel corpus, which only comprises around 5000 participle-adverb translations in total, which results in 46 types of adverbial participles after filtering. Thus, we have no means of telling whether we would discover new types of adverbial participles in other corpora, from different domains to Europarl. As this corpus is rather small and genre specific, it even seems very likely that one could find additional adverbial participles in a bigger corpus. Moreover, we cannot be sure that certain adverbial participles have systematically diverging translations in other languages, due to crosslingual lexicalisation differences. Generally, it is not clear whether we have learned something general about the syntactic phenomenon of adverbial participles in German or whether we have just extracted a small, corpus-dependent subset of the class of adverbial participles.

In this section, we use instances of adverbially translated participles as training material for a classifier that learns to predict adverbial participles based on their monolingual syntactic context. Thus, we exploit the translations in the parallel corpus as a means of obtaining "annotated" or disambiguated training data without any manual effort. During training, we only consider the monolingual context of the participle, such that the final application of the classifier is not dependent on cross-lingual data anymore.

4.1 Context-based Identification of Adverbial Participles

Given the general linguistic problems related to adverbial participles (see Section 2), one could assume that it is very difficult to identify them in a given context. To assess the general difficulty of this syntactic problem, we run a first experiment comparing a grammar-based identification method against a classifier that only considers relatively narrow morpho-syntactic context. For evaluation, we use the 300 annotated participle instances described in Section 3. This test set divides into 172 negative instances, i.e. non-adverbial participles, and 128 positive instances. We report accuracy of the identification method, as well as precision and recall relating to the number of correctly predicted adverbial participles.

For the grammar-based identification, we use the German LFG which integrates the lexical resource for adverbial participles established in (Zarrieß et al., 2010). We parse the 300 Europarl sentences and check whether the most probable parse proposed by the grammar analyses the respective participle as an adverb or not. The grammar obtains a complete parse for 199 sentences out of the test set and we only consider these in the evaluation. The results are given in Table 1.

The high precision and accuracy of the grammar-based identification of adverbial participles suggests that in a lot of sentences, the adverbial analysis is the only possible reading, i.e. the only analysis that makes the sentence grammatical. But of course, we have substantially restricted the adverb participle-conversion in the grammar,

Training Data	Precision	Recall	Accuracy
Grammar	97.3	90.12	94.97
Classifier Unigram	87.10	84.38	87.92
Classifier Bigram	88.28	88.28	89.93
Classifier Trigram	89.60	87.5	90.27

Table 1: Evaluation on 300 participle instances from Europarl

so that it does not propose adverbial analyses for participles that are very unlikely to function as modifiers of verbs.

For the classifier-based identification, we use the adverbially translated participle tokens in our Europarl data (5191 tokens in total) as training material. We remove the 300 test instances from this training set, and then divide it into a set of positive and negative instances. To do this, we use the filtering mechanisms already proposed in Zarrieß et al. (2010). These filters apply on the type level, such that we first identify the positive types (46 total) and then use all instances of these types in the 4891 sentences as positive instances of adverbial participles (1978 instances). The remaining sentences are used as negative instances.

For the training of the classifier, we use maximum-entropy classification, which is also commonly used for the general task of tagging (Ratnaparkhi, 1996). In particular, we use the open source TADM tool for parameter estimation (Malouf, 2002). The tags of the words surrounding the participles are used as features in the classification task. We explore different sizes of the context window, where the trigram window is the most succesful (see Table 1). Beyond the trigram window, the results of the classifier start decreasing again, probably because of too many misleading features. Generally, this experiment shows that the grammar-based identification is more precise, but that the classifier still performs surprisingly well. Compared to the results from the grammar-based identification, the high accuracy of the classifier suggests that even the narrow syntactic contexts of adverbial vs. non-adverbial participles are quite distinct.

4.2 Designing Training Data for Participle Classification

There are several questions related to the design of the training data that we use to build our classifier. First, it is not clear how many negative instances are helpful for learning the adverbial - non-adverbial distinction. In the above experiment, we simply use the instances that do not pass the cross-lingual filters. In this section, we experiment with an augmented set of negative instances that was also obtained by extracting German participle that are bi-directionally aligned to an English participle in Europarl. This is based on the assumption that these participles are very likely to be verbal. Second, it is not clear whether we really need the filtering mechanisms proposed in Zarrieß et al. (2010) and whether we could improve the classifier by training it on a larger set of positive instances. Therefore, we also experiment with two further sets of positive instances: one where we used all participles (not necessarily bidirectionally) aligned to an adverb, one where we only use the bidirectional alignments. The results obtained for the different sizes of positive and negative instance sets are given in Table 2.

The picture that emerges from the results in Table 2 is very clear: the stricter the filtering of the training material (i.e. the positive instances) is, the better the performance of the classifier. The fact that we (potentially) loose certain positive instances in the filtering does not negatively impact on the classifier which substantially benefits from the fact that noise gets removed. Moreover, we find that if the training material is appropriately filtered, adding further negative instances does not help improving the accuracy. By contrast, if we train on a noisy set of positive instances, the classifier benefits from a larger set of negative instances. However, the positive effect that we get from augmenting the non-filtered training data is still weaker than the positive effect we get from the filtering.

5 Induction of Adverbial Participles on Monolingual Data

Given the classifier from Section 4 that predicts the syntactic category of a participle instance

Training Data	Pos. Instances	Neg. Instances	Precision	Recall	Accuracy
Non-Filtered Instances (all alignments)	27.184	10.000	43.10	100	43.10
Non-Filtered Instances (all alignments)	27.184	50.000	74.38	92.97	83.22
Non-Filtered Instances (symm. alignments)	4891	10.000	78.08	89.06	84.56
Non-Filtered Instances (symm. alignments)	4891	50.000	82.31	83.59	85.23
Filtered Instances	1978	10.000	**91.60**	**85.16**	**90.27**
Filtered Instances	1978	50.000	90.83	77.34	86.91

Table 2: Evaluation on 300 participle instances from Europarl

based on its monolingual syntactic context, we can now detect new instances or types of adverbial participles in any PoS-tagged German corpus. In this section, we investigate whether the classifier can be used to augment the resource of adverbial participles directly induced from Europarl with new types.

5.1 Data Extraction

We run our extraction experiment on the Huge German Corpus (HGC), a corpus of 200 million words of newspaper and other text. This corpus has been tagged with TreeTagger (Schmid, 1994). For each of the 5054 participle candidates, we extract all instances from the HGC which have not been tagged as finite verbs (at most 2000 tokens per participle). For each participle token, we also extract its syntactic context in terms of the 3 preceding and the 3 following tags. For classification, we use only those participles that have more than 50 instances in the corpus (2953 types).

In contrast to the cross-lingual filtering mechanisms developed in Zarrieß et al. (2010) which operate on the type-level, the classifier makes a prediction for every token of a given participle candidate. Thus, for each of the participle candidates, we obtain a percentage of instances that have been classified as adverbs. As we would expect, the percentage of adverbial instances is very low for most of the participles in our candidate set: for 75% of the 2953 types, the percentage is below 5%. This result confirms our initial intuition that the property of being used as an adverb is strongly lexically restricted to a certain class of participles.

5.2 Evaluation

Since we know that the classifier has an accuracy of 90% on the Europarl data, we only consider participles as candidates for adverbs where the classifier predicted more than 14% adverbial instances. This leaves us with a set of 210 participles, which comprises 13 of the original 46 participles extracted from Europarl, meaning we have discovered 197 new adverbial participle types.

We performed a manual evaluation of 50 randomly selected types out of the set of 197 new participle types. Therefore, we looked at the instances and their context which the classifier predicted to be adverbial. If there was at least one adverbial instance among these, the participle type was evaluated as correctly annotated by the classifier. By this means, we find that 76% of the participles were correctly classified.

This evaluation suggests that the accuracy of our classifier which we trained and tested on Europarl data is lower on the HGC data. The reason for this drop in performance will be explained in the following Section 5.3. However, assuming an accuracy of 76%, we have discovered 150 new types of adverbial participles. We argue that this is a very satisfactory result given that we have not invested any manual effort into the annotation or extraction of adverbial participles. This results also makes clear that the previous resource we induced on Europarl data, comprising only 46 participle types, was a very limited one.

5.3 Error Analysis

Taking a closer look at the 12 participle candidates that the classifier incorrectly labels as adverbial, we observe that their adverbially classified instances are mostly instances of a predicative use. This means that our Europarl training data does not contain enough evidence to learn the distinction between adverbial and predicative participles. This is not surprising since the set of negative instances used for training the classifier mainly comprises verbal instances of participles. Moreover, the syntactic contexts and constructions in which some predicatives and adverbials are used

Grammar	Prec.	Rec.	F-Sc.	Time in sec
46 Part-Adv	84.12	78.2	81.05	665
243 Part-Adv	84.12	77.67	80.76	665

Table 3: Evaluation on 371 TIGER sentences

Grammar	Parsed Sent.	Starred Sent.	Time-outs	Time in sec
No Part-Adv	665	315	64	3033
46 Part-Adv	710	269	65	3118
243 Part-Adv	767	208	69	3151

Table 4: Performance on the specialised TIGER test set (1044 sentences)

are very similar. Thus, in future work, we will have to include more data on predicatives (which is more difficult to obtain) and analyse the syntactic contexts in more detail.

6 Assessing the Impact of Resource Coverage on Grammar-based Parsing

In this section, we evaluate the classifier-based induction of adverbial participles from a grammar-based perspective. We integrate the entire set of induced adverbial participles (46 from Europarl and 197 from the HGC) into the German LFG grammar. As a consequence, the grammar allows the adverb conversion for 243 lexical participle types. We use the evaluation methodolgy explained in Section 2.

First, we conduct an accuracy-oriented evaluation on the standard TIGER test set. We compare against the German LFG that only integrates the small participle resource from Europarl. The results are given in Table 3. The difference between the 46 Part-Adv and 243 Part-Adv resource is not statistically signficant. Thus, the larger participle resource has no overall negative effect on the parsing performance. As established by an automatic upperbound evaluation in Zarrieß et al. (2010), we cannot not expect to find a positive effect in this evaluation because the phenomenon does not occur in the standard test set.

To show that the augmented resource indeed improves the coverage of the grammar, we built a specialised testsuite of 1044 TIGER sentences that contain an instance of a participle from the resource. Since this testsuite comprises sentences from the training set, we can only report a coverage-oriented evaluation here, see Table 4. The 243 Part-Adv increases the coverage by 8% on the specialised testsuite.

Moreover, we manually evaluated 20 sentences covered by the 243-Part-Adv grammar and not by 46-Part-Adv as to whether they contain a correctly analysed adverbial participle. In two sentences, the grammar obtained an adverbial analysis for clearly predicative modifiers, based on the enlarged resource. In three different sentences, it was difficult to decide whether the participle acts as an adverb or a predicative. In the remaining 15 sentences, the grammar established the the correct analysis of a clearly adverbially used participle.

7 Conclusion

We have proposed a cross-lingual induction method to automatically obtain data on adverbial participles in German. We exploited this cross-lingual data as training material for a classifier that learns to predict the syntactic category of a participle from its monolingual syntactic context. Since this category is usually not annotated in German resources and hard to describe in theory, the finding that adverbial participles can be predicted relatively precisely is of general interest for theoretic and computational approaches to the syntactic analysis of German.

We showed that, in order to obtain an accurate participle classifier, the quality of the training material induced from the parallel corpus is of crucial importance. By applying the filtering techniques from Zarrieß et al. (2010), the accuracy of the classifier increases between 5% and 7%. In future work, we plan to include more data on predicative participles to learn a more accurate distinction between predicative and adverbial participles.

Finally, we used the participle classifier to extract a lexical resource of adverbial participles for the German LFG grammar. In comparison to the relatively small resource of 46 types that can be directly induced from Europarl, we discovered a large number of new participle types (197 types in total). In a parsing experiment, we showed that this much bigger resource does not negatively impact on parsing performance and improves grammar coverage.

References

Bouma, Gerlof, Jonas Kuhn, Bettina Schrader, and Kathrin Spreyer. 2008. Parallel LFG Grammars on Parallel Corpora: A Base for Practical Triangulation. In Butt, Miriam and Tracy Holloway King, editors, *Proceedings of the LFG08 Conference*, pages 169–189, Sydney, Australia. CSLI Publications, Stanford.

Brants, Sabine, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. 2002. The TIGER Treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*.

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The Parallel Grammar Project.

Geuder, Wilhelm. 2004. Depictives and transparent adverbs. In Austin, J. R., S. Engelbrecht, and G. Rauh, editors, *Adverbials. The Interplay of Meaning, Context, and Syntactic Structure*, pages 131–166. Benjamins.

Hwa, Rebecca, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Nat. Lang. Eng.*, 11(3):311–325.

Malouf, Robert. 2002. A comparison of algorithms for maximum entropy parameter estimation. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pages 49–55.

Mihalcea, Rada, Carmen Banea, and Jan Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the Association for Computational Linguistics (ACL 2007)*, pages 976–983, Prague.

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

Ratnaparkhi, Adwait. 1996. A maximum entropy model for part-of-speech tagging. In *Proceedings of EMNLP 96*, pages 133–142.

Riezler, Stefan, Tracy Holloway King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell, and Mark Johnson. 2002. Parsing the Wall Street Journal using a Lexical-Functional Grammar and Discriminative Estimation Techniques . In *Proceedings of ACL 2002*.

Rohrer, Christian and Martin Forst. 2006. Improving coverage and parsing quality of a large-scale LFG for German. In *Proceedings of LREC-2006*.

Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*.

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

Zarrieß, Sina, Aoife Cahill, Jonas Kuhn, and Christian Rohrer. 2010. A Cross-Lingual Induction Technique for German Adverbial Participles. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, ACL 2010*, pages 34–42, Uppsala, Sweden.