# A Cross-Lingual Induction Technique for German Adverbial Participles

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

## Abstract

We provide a detailed comparison of strategies for implementing medium-to-low frequency phenomena such as German adverbial participles in a broad-coverage, rule-based parsing system. We show that allowing for general adverb conversion of participles in the German LFG grammar seriously affects its overall performance, due to increased spurious ambiguity. As a solution, we present a corpus-based cross-lingual induction technique that detects adverbially used participles in parallel text. In a grammar-based evaluation, we show that the automatically induced resource appropriately restricts the adverb conversion to a limited class of participles, and improves parsing quantitatively as well as qualitatively.

## 1 Introduction

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)). All three participle forms in (1) are morphologically identical.

(1) a. Das Experiment hat ihn **begeistert**.
'The experiment has enthused him.'
b. Er scheint von dem Experiment **begeistert**.
'He seems enthusiastic about the experiment.'
c. Er hat **begeistert** experimentiert.
'He has experimented in an enthusiastic way' or:
'He was enthusiastic when he experimented.'

This paper adresses the question of how to deal with medium-to-low frequency phenomena such as adverbial participles in a broad-coverage, rule-based parsing system. 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, on the basis of the German LFG grammar (Rohrer and Forst, 2006), we show that such a rule can have a strong negative on the overall performance of the parsing system, despite the fact that it produces the desired syntactic and semantic analysis for specific sentences.

This trade-off between large-scale, statistical and theoretically precise coverage is often encountered in engineering broad-coverage and, at the same time, linguistically motivated parsing systems: adding the analysis for a specific phenomenon does not necessarily improve the overall quality of the system since the rule might overgenerate and interact with completely different phenomena in unpredicted ways.

In principle, there are two ways of dealing with such an overgeneration problem in a grammar-based framework: First, one could hand-craft word lists or other linguistic constraints that restrict the adverb conversion to a certain set of participles. Second, one could try to mine corpora for this particular type of adverbs and integrate this automatically induced knowledge into the grammar (i.e. by means of pre-tagged input, word lists, etc.). In the case of adverbial participles, both ways are prone with difficulties. To our knowledge, there has not been much theoretical work on the linguistic properties of the participle adverb conversion. Moreover, since the distinction between (predicative) adjectives and adverbs is theoretically hard to establish, the standard tag set for German and, in consequence, annotated corpora for German do not explicitly capture this phenomenon. Thus, available statistical taggers and parsers for German usually conflate the syntactic structures underlying (1-b) and (1-c).

In this paper, we present a corpus-based approach to restricting the overgenerating adverb conversion for participles in German, exploiting



*Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, ACL 2010*, pages 34–42,
Uppsala, Sweden, 16 July 2010. 

parallel corpora and cross-lingual NLP induction techniques. Since adverbs are often overtly marked in other languages (i.e. the *ly*-suffix in English), adverbial participles can be straightforwadly detected on word-aligned parallel text. We describe the ingretation of the automatically induced resource of adverbial participles into the German LFG, and provide a detailed evaluation of its effect on the grammar, see Section 5.

While the use of parallel resources is rather familiar in a wide range of NLP domains, such as statistical machine translation (Koehn, 2005) or annotation projection (Yarowsky et al., 2001), our work shows that they can be exploited for very specific problems that arise in deep linguistic analysis (see Section 4). In this way, high-precision, data-oriented induction techniques can clearly improve rule-based system development through combining the benefits of high empirical accuracy and little manual effort.

## 2 A Broad-Coverage LFG for German

Lexical Functional Grammar (LFG) (Bresnan, 2000) is a constraint-based theory of grammar. It posits two levels of representation, c(onstituent)-structure and f(unctional)- structure. C-structure is represented by contextfree phrase-structure trees, and captures surface grammatical configurations. F-structures approximate basic predicate-argument and adjunct structures.

The experiments reported in this paper use the German LFG grammar 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. Within the spectrum of appraoches to natural language parsing, XLE can be considered a hybrid system combining a hand-crafted grammar with a number of automatic ambiguity management techniques: (i) c-structure pruning where, based on information from statstically obtained parses, some trees are ruled out before f-structure unification (Cahill et al., 2007), (ii) an Optimaly Theory-style constraint mechanism for filtering and ranking competing analyses (Frank et al., 2001), and (iii) a stochastic disambiguation component which is based on a log-linear probability model (Riezler et al., 2002) and works on the packed representations.

The German LFG grammar integrates a morphological component which is a variant of DMOR1 (Becker, 2001). This means that the (internal) lexicon does not comprise entries for surface word forms, but entries for specific morphological tags, see (Dipper, 2003).

## 3 Participles in the German LFG

### 3.1 Analysis

The morphosyntactic ambiguity of German participles presents a notorious difficulty for theoretical and computational analysis. The reason is that adjectives (i.e. adjectival participles) do not only occur as attributive modifiers (shown in (1-a)), but can also be used as predicatives (see (2-b)). These predicatives have exactly the same form as verbal or adverbial participles (compare the three sentences in (2)). Predicatives do appear either as arguments of verbs like *seem* or as free adjuncts such that they are not even syntactically distinguishable from adverbs. The sentence in (2-c) is thus ambiguous as to whether the participle is an adverb modifying the main verb, or a predicative which modifies the subject. Especially in the case of modifiers refering to a psychological state, the two underlying readings are hard to tell apart (Geuder, 2004). It is due to the lack of reliable semantic tests that the standard German tag set (Schiller et al., 1995) assigns the tag "ADJD" to predicative adjectives as well as adverbs.

(2) a. Das Experiment hat ihn **begeistert**.
'The experiment has enthused him.'
b. Er scheint von dem Experiment **begeistert**.
'He seems enthusiastic about the experiment.'
c. Er hat **begeistert** experimentiert.
'He has experimented in an enthusiastic way' or:
'He was enthusiastic when he experimented.'

For performance reasons, the German LFG does not cover free predicatives at the moment. In the context of our crosslingual induction approach, the distinction between predicatives and adverbs is rather straigtforward since we base our experiments on languages that have morphologically distinct forms for these categories. In the following, we will thus limit the discussion to adverbial participles and ignore the complexities related to predicative participles.

In the German LFG, the treatment of a given participle form is closely tight to the morphological analysis encoded in DMOR. In particular, adverbial participles can have different degrees of lexicalisation. For *bestimmt* (*probably*) in (3-a), which is completely lexicalised, the morphology

proposes two analyses: (i) a participle tag of the verbal lemma *bestimmen* (*determine*) and (ii) an adverb tag for the lemma *bestimmt*. In this case, the LFG parsing algorithm will figure out which morphological analysis yields a syntactically well-formed analysis. For *gezielt* (*purposeful*) in (3-b), DMOR outputs, besides the participle analysis, an adjective tag for the lemma. However, the grammar can turn it into an adverb by a general adverb conversion rule for adjectives. The difficult case for the German LFG grammar is illustrated in (3-c) by means of the adverbial participle *wiederholt* (*repeatedly*). This participle is neither lexicalised as an adverb nor as an adjective, but it still can be used as an adverb.

(3) a. **Bestimmt** ist dieser Mann sehr traurig.
Probably is the man very sad.
b. Der Mann hat **gezielt** gehandelt.
The man has acted purposefully.
c. Der Mann hat **wiederholt** geweint.
The man has repeatedly cried.

To cover sentences like (3-c), the grammar needs to include a rule that allows adverb conversion for participles. Unfortunately, this rule is very costly in terms of the overall performance of the grammar, as is shown in the following section.

### 3.2 Assessing the Effect of Participle Ambiguity on the German LFG

In this section, we want to illustrate the effect of one specific grammar rule, i.e. the rule that generally allows for conversion of participles into adverbs. We perform a contrastive evaluation of two versions of the grammar: (i) the *No-Part-Adv* version which does not allow for adverb conversion (except for the lexicalised participles from DMOR), (ii) the *All-Part-Adv* version which allows every participle to be analysed as adverb. Otherwise, the two versions of the grammar are completely identical.

The comparison between the *All-Part-Adv* and *No-Part-Adv* grammar version pursues two major goals: On the one hand, we want to assess their overall quantitative performance on representative gold standard data, as it is common practice for statistical parsing systems. On the other hand, we are interested in getting a detailed picture of the quality of the grammar for parsing adverbial participles. These two goals do not necessarily go together since we know that the phenomenon is not very frequent in the data which we use for evaluation. Therefore, we do not only report accuracy on gold standard data in the following, but also focus on error analysis and describe ways of qualtitatively assessing the grammar performance.

For evaluation, we use the TIGER treebank (Brants et al., 2002). We report grammar performance on the development set which consists of the first 5000 TIGER sentences, and statistical accuracy on the standard heldout set which comprises 371 sentences.

**Quantitative Evaluation** We first want to assess the quantitative impact of the phenomenon of adverbial participles in our evaluation data. We parse the heldout set storing all possible analyses obtained by both grammars, in order to compare the upperbound score that the both versions can optimally achieve (i.e. independently of the disambiguation quality). Then, we run the XLE evaluation in the "oracle" mode which means that the disambiguation compares all system analyses for a given sentence to its gold analysis, and chooses the best system analysis for computing accuracy. The upperbound f-score for both grammar versions is almost identical (at about 83.6%). This suggests that the phenomenon of adverbial participles does not occur in the heldout set.

If we run the grammar versions on a larger set of sentences, the difference in coverage becomes more obvious. In Table 1, we report the absolute number of parsed sentences, starred sentences (only receiving a partial or fragment parse), and the timeouts [1] on our standard TIGER development set. Not very surprisingly, the coverage of the *All-Part-Adv* version seems to be broader. However, this does not necessarily mean that the 40 additionally covered sentences all exhibit adverbial participles (see below). Moreover, Table 2 gives a first indication of the fact that the extended coverage comes at a price: the *All-Part-Adv* version massively increases the number of ambiguities per sentence. Related to this, in the *All-Part-Adv* version, the number of timeouts increases by 16% and parsing speed goes down by 6% compared to the *No-Part-Adv* version.

To assess the effect of the massively increased ambiguity rate and the bigger proportion of timeouts in *All-Part-Adv*, we perform a statistical evaluation of the two versions of the grammar against the heldout set, i.e. we compute f-score based

[1] Sentences whose parsing can not be finished in predefined amount of time, the maximally allowed parse time is set to 20 seconds.

| Grammar | Parsed Sent. | Starred Sent. | Time-outs | Time in sec |
|---|---|---|---|---|
| No-Part-Adv | 4301 | 608 | 90 | 6853 |
| All-Part-Adv | 4339 | 555 | 105 | 7265 |

Table 1: Coverage-based evaluation on the TIGER development set (sentences 1-5000), 4999 sentences total

| Sent. length | Av. ambiguities per sent. *No-Part-Adv* | *All-Part-Adv* | Av. Incr. |
|---|---|---|---|
| 1-10 | 2.95 | 3.3 | 11% |
| 11-20 | 24.99 | 36.09 | 44% |
| 21-30 | 250.4 | 343.76 | 37% |
| 31-40 | 1929.06 | 2972.847 | 54% |
| 41-50 | 173970.0 | 663310.4 | 429% |

Table 2: Average number of ambiguities per sentence

on the parses that the XLE disambiguation selects as the most probable parse. Both versions use the same disambiguation model which results in a slightly biased comparison but still reflects the effect of increased ambiguity on the disambiguation component. In Table 3, we can see that the *All-Part-Adv* version performs significantly worse than the grammar version which does not capture adverbial participles. The spurious ambiguities and timeouts produced in *All-Part-Adv* have such a strong negative impact on the disambiguation component that it can not be outweighed by the extended coverage of the grammar.

**Qualitative Evaluation** The fact that the *All-Part-Adv* version generally increases parse ambiguity suggests that it produces a lot of undesired analyses for constructions not related to adverbial participles. To assess this assumption, we drew a random sample of 20 sentences out of the additionally covered 41 sentences and checked manually whether these contained an adverbial participle: Only 40% of these sentences are actually correctly analysed. In all other cases, the grammar lacks an analysis for a completely different phenomenon (mostly related to coordination), but obtains an (incorrect) analysis on the basis of the adverb conversion rule.

| Grammar | Prec. | Rec. | F-Sc. | Time in sec |
|---|---|---|---|---|
| All-Part-Adv | 83.80 | 76.71 | 80.1 | 666.55 |
| No-Part-Adv | 84.25 | 78.3 | 81.17 | 632.21 |

Table 3: Evaluation on the TIGER heldout set, 371 sentences total

As an example, Figure 1 presents two c-structure analyses for the sentence in (4) in the *All-Part-Adv* grammar. In the second c-structure (CS2), the participle *kritisiert* (*criticised*) is analysed as adverb modifing the main verb *haben* (*have*). This results in a very strange underlying f-structure, meaning something like *the Greens possess the SPD in a criticising manner*.

(4) Die Grünen haben die SPD kritisiert.
The Greens have the SPD criticised.
"The Greens have criticised the SPD"

### 3.3 Interim Conclusion

This section has illustrated an exemplary dilemma for parsing systems that aim broad-coverage and linguisitically motivated analyses at the same time. Since these systems need to explicitly address and represent ambiguities that purely statistical systems are able to conflate or ignore, their performance is not automatically improved by adding a specific rule for a specific phenomenon. Interestingly, the negative consequences affecting the quantitative (statistical) as well as the qualitative (linguistic) dimension of the grammar seem to be closely related: The overgenerating adverb conversion rule empirically leads to linguistically unmotivated analyses which causes problems for the disambiguation component. In the rest of the paper, we show how the adverbial analysis of participles can be reasonably constrained on the basis of a lexical resource induced from a parallel corpus.

## 4 Cross-Lingual Induction of Adverbial Participles

The intuition of the cross-lingual induction approach is that adverbial participles can be easily 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 the sentence in (5), extracted from Europarl, where the German participle *verstärkt* is translated by unambiguous adverbs in English and French (*increasingly* and *davantage*).

(5) a. Nach der Osterweiterung stehen die Zeichen **verstärkt** auf Liberalisierung.
b. Following enlargement towards the east, the emphasis is **increasingly** on liberalisation.

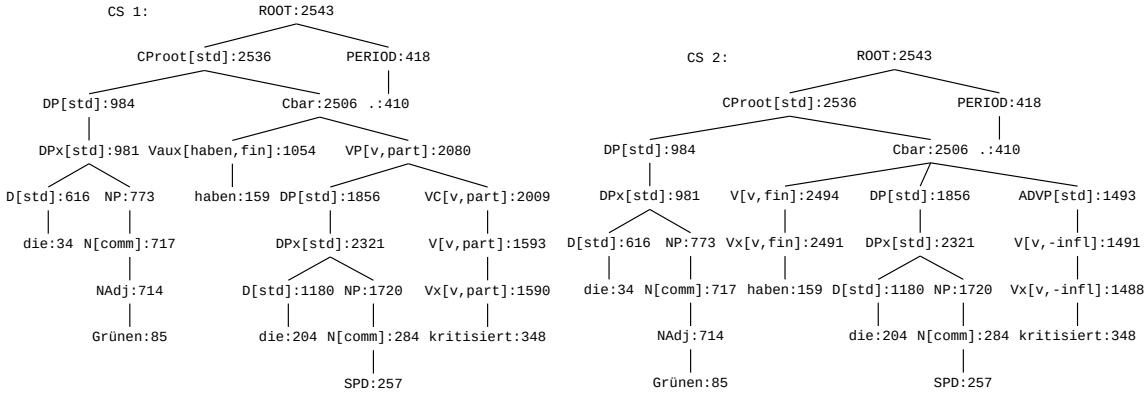


Figure 1: Two c-structures for sentence (4), obtained by the grammar *All-Part-Adv* - CS1 is correct, CS2 is semantically very strange

c. Après l' élargissement à l' Est, la tendance sera **davantage** à la libéralisation.

In the following, we describe experiments on Europarl where we automatically extract and filter adverbially translated German participles.

### 4.1 Data

We base our experiments on the German, English, French and Dutch part of the Europarl corpus. We automatically word-aligned the German part to each of the others with the GIZA++ tool (Och and Ney, 2003). Note that, due to divergences in sentence alignment and tokenisation, the three word-alignments are not completely synchronised. Moreover, each of the 4 languages has been automatically PoS tagged using the TreeTagger (Schmid, 1994). In addition, the German and English parts have been parsed with MaltParser (Nivre et al., 2006).

Since we want to limit our investigation to those participles that are not already recorded as lexicalised adjective or adverb in the DMOR morphology, we first have to generate the set of participle candidates from the tagged Europarl data. We extract all distinct words (types) from the German part that have been either tagged as ADJD (predicative or adverbial modifier), 6089 types in total, or as VVPP (past perfect participle), 5469 types in total. We intersect this set of potential participles with the set of DMOR participles that only have a verbal lemma. The resulting intersection (5054 types in total) constitutes the set of all German participles in Europarl that are not recorded as lexicalised in the DMOR morphology .

Given the participle candidates, we now extract the set of sentences that exhibit a word alignment between a German participle and an English, French or Dutch adverb. The extraction yields 5191 German-English sentence pairs, 2570 German-French, and 4129 German-Dutch sentence pairs. The German-English pairs comprise 1070 types of potentially adverbial participles. The types found in the German-French and German-Dutch part form a proper subset of the types extracted from the German-English pairs. Thus, the additional languages will not increase the recall of the induction. However, we will show that they are extremely useful for filtering incorrect or uninteresting participle alignments.

For data exploration and evaluation, we annotated 300 participle alignments out of the 5191 German-English sentences as to whether the English adverbial really points to an adverbial participle on the German side (and/or the word-alignment was correct). Throughout the entire set of annotated sentences, this ratio between the parallel cases (where an English adverbial correctly indicates a German adverbial) and all adverbially translated participles is at about 30%. This means that if we base the induction on word-alignments alone, its precision would be relatively low.

The remaining 60% translation pairs do not only reflect word alignment errors, but also cases where we find a proper participle in the German sentence that has a correct adverbial translation for other reasons. A typical configuration is exemplified in (6) where the German main verb *vorlegen* is translated as the verb-adverb combination *put forward*.

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

These sentence pairs are cases of free or para-

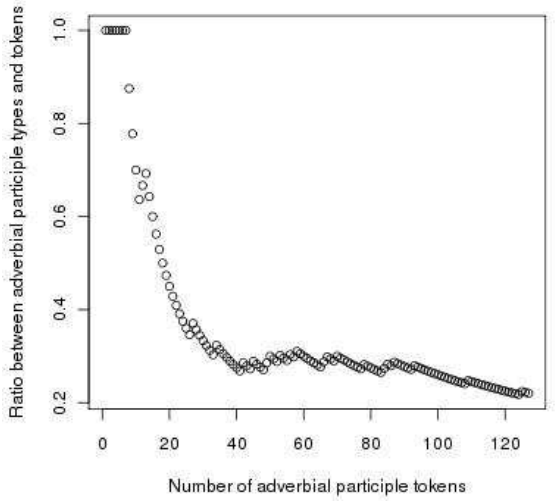


Figure 2: Type/token ratio for adverbial participles

phrasing translations. Ideally, we want our induction method to filter such type of configurations.

The 300 annotated sentences comprise 121 token instances of German adverbially used participles that have an adverbial translation in English. However, these 121 tokens reduce to 24 participle types. The graph in Figure 2 displays the type/token-ratio for an increasing number of instances in our gold standard. The curve exponentially decays from about 10 tokens onward and suggests that from about 30 tokens onward, the number of unseen types is relatively low. This can be interpreted as evidence in favour of the hypothesis that the number of adverbially used participles is actually fairly limited and can be integrated into the grammar in terms of a hard-coded resource.

### 4.2 Filtering

The data analysis in the previous section has shown that approximately one third of the English adverb alignments actually point to an adverbial participle on the German side. This means that we have to rigorously filter the data that we extract on the basis of word-alignments in order to obtain a high quality resource for our grammar. In this section, we will investigate several filtering methods and evaluate them on our annotated sentence pairs.

**Frequency-based filtering** As a first attempt, we filtered the non-parallel cases in our set of participle-adverb translations by means of the relative frequency of the adverb translations. For each participle candidate, we counted the number of tokens that exhibit an adverbial alignment on the English side, and divided this number by its total number of occurrences in the German Europarl. The best f-score of the ADV-FREQ filter (see Table 4) is achieved by the 0.05 threshold, but generally, the precision of the frequency filters is too low for high-quality resource induction. The reason for the poor performance of the frequency-based filters seems to be that some German verbs are systematically translated as verb - adverb combinations as in (6). For these participles, the relative frequency of adverbial alignments is not a good indicator for their adverbial use in German.

**Multilingual Filtering** Similar to filters used in annotation projection where noisy word-alignments are "cleaned" with the help of additional languages (Bouma et al., 2008), we have implemented a filter that only selects those participles as adverbials which also exhibit a certain amount of adverbial translations in the French and Dutch Europarl. We count the total number of adverbial translations of a given participle on the French side and divide it by the number of English adverbial translations. For French, the best f-score is achieved at a threshold of $>0.1$ (filter FR). For Dutch, the best f-score is achieved at a threshold of $>0.05$ (filter NL). The exact precision and recall values are given in Table 4.

**Syntax-based Filtering** The intuition behind the filters presented in this section is that adverbial translations which are due to cross-lingual divergences can be identified on the basis of their syntactic contexts. Information about these contexts can be extracted from the dependency analyses produced by MaltParser for the German and English data. On the German side, we want to exclude those participle instances for which the German parser has found an auxiliary head, since this configuration points to a normal partciple context in German. The filter is called G-HEAD in Table 4. It filters all types which have an auxiliary head in more than 40% of their adverbial translation configurations. On the English side, we exclude all translations where the adverb has a verbal head which is also aligned to the German partciple. The filter is called E-HEAD in Table 4. It excludes all participle types which exhibit the E-HEAD configuration in more than 50% of the cases.

| filter | prec. | rec. | f-sc. |
|---|---|---|---|
| ADV-FREQ | 0.38 | 0.75 | 0.51 |
| FR | 0.48 | 0.76 | 0.58 |
| NL | 0.33 | 0.73 | 0.45 |
| G-HEAD | 0.65 | 0.8 | 0.71 |
| E-HEAD | 0.4 | 0.8 | 0.53 |
| COMBINED-1 | 0.61 | 0.8 | 0.69 |
| COMBINED-2 | 0.86 | 0.76 | 0.81 |

Table 4: Performance of filters on the set of gold adverbial participle types

**Combined Token-level Filtering** So far, we have shown that multilingual and syntactic information is useful to filter non-parallel participle translations. We have found that the precision of the syntactic filters can still be increased by combining it with the multilingual filters. COMBINED-1 in Table 4 refers to the filter which only includes those participle types which have at least one adverbial translation on the English target side such that (i) the adverbial translation is paralleled on the French or Dutch target side for the same German participle token and (ii) the German participle token does not have an auxiliary head. If we combine this token-level filtering with the syntactic type-level filtering G-HEAD and E-HEAD (the filter called COMBINED-2 in Table 4), the precision increases by about 25% with little loss in recall.

### 4.3 Analysis

Based on the filtering techniques described in the previous section, we can finally induce a list of 46 German adverbial participles from Europarl. The fact that this participle class seems fairly delimited in our data raises the theoretical question whether the adverb conversion is licensed by any linguistic, i.e. lexical-semantic, properties of these participles. However, we observe that the automatically induced list comprises very diverse types of adverbs, as well as very distinct types of underlying verbs. Thus, besides adverbs that clearly modify events (see sentence (5)), we also found adverbs that are more likely to modify adjectives (sentence (7-a)), or propositions (sentence (7-b)).

(7) a. Es ist eine **verdammt** gefährliche Situation.
'It is a damned dangerous situation.'
b. Wir machen einen Bericht über den Bericht des Rechnungshofes , **zugegeben**.
'We are drafting a report about the report of the Court of Auditors , admittedly.'

A more fine-grained classification and analysis of adverbial participles is left for future research.

## 5 Grammar-based Evaluation

The resource of participles licensing adverbial use, whose induction was described in the previous section, can be straightforwardly integrated into the German LFG. By explicitly enumerating the participles in the adverb lexicon, the grammar can apply the standard adverb macros to them. To assess the effect of the filtering, we built two new versions of the grammar: (i) *Euro-Part-Adv*, its adverb lexicon comprises all adverbially translated participles found in Europarl (1091 types) and (ii) *Filt-Part-Adv*, its adverb lexicon comprises only the syntactically and multilingually filtered participles found in Europarl (46 types).

Although we have seen in section 3.2 that adverbial participles do not seem to occur in the TIGER heldout set, we also know that it is important to assess the effect of ambiguity rate on the overall grammar performance. Therefore, we computed the accuracy of the most probable parses produced by the *Euro-Part-Adv* and *Filt-Part-Adv* on the heldout set. As is shown in Table 5, the *Euro-Part-Adv* performs significantly worse than *Filt-Part-Adv*. This suggests that the non-filtered participle resource is not constrained enough and still produces a lot of spurious ambiguites that mislead the disambiguation component. The coverage values in Table 6 further corroborate the observation that the unfiltered participle resource behaves similar to the unrestricted adverb conversion in *All-Part-Adv* (see Section 3.2). The coverage of the filtered vs. the unfiltered version on the development set is identical, however the timeouts in *Euro-Part-Adv* increase by 17% and parsing time by 8%.

By contrast, there is no significant difference in f-score between the *No-Part-Adv* version presented in Section 3.2 and the *Filt-Part-Adv* version. Thus, we can, at least, assume that the filtered participles resources has restricted the massive overgeneration caused by the general adverb conversion rule such that the overall performance of the original grammar is not negatively affected.

To evaluate the participle resource as to whether it could have a positive qualtitative effect on parsing TIGER at all, we built a specialised testsuite which comprises only sentences containing a non-lexicalised participle, which has an adverbial translation in Europarl and is tagged as ADJD

| Grammar | Prec. | Rec. | F-Sc. | Time in sec |
|---|---|---|---|---|
| Euro-Part-Adv | 82.32 | 75.78 | 78.91 | 701 |
| Filt-Part-Adv | 84.12 | 78.2 | 81.05 | 665 |

Table 5: Evaluation on the TIGER heldout set, 371 sentences total

| Grammar | Parsed Sent. | Starred Sent. | Time-outs | Time in sec |
|---|---|---|---|---|
| Euro-Part-Adv | 4304 | 588 | 107 | 7359 |
| Filt-Part-Adv | 4304 | 604 | 91 | 6791 |

Table 6: Performance on the TIGER development set (sentences 1-5000), 4999 sentences total

in TIGER. The sentences were extracted from the whole TIGER corpus yielding a set of 139 sentences. In this quality-oriented evaluation, we only contrast the *No-Part-Adv* version with the filtered *Filt-Part-Adv* version since the unfiltered version leads to worse overall performance. As can be seen in Table 7, the *No-Part-Adv* can only completely cover 36% of the specialised testsuite which is much lower than its average complete coverage on the development set (86%). This suggests that a substantial number of the extracted ADJD participles are actually used as adverbial in the specialised testsuite.

Similar to the qualitative evaluation procedure in 3.2, we manually evaluated a random sample of 20 sentences covered by *Filt-Part-Adv* and not by *No-Part-Adv* as to whether they contain an adverbial participle that has been correctly recognised. This was the case for 90% of the sentences, the remaining 2 sentences were cases of secondary predications. An example of a relatively simple TIGER sentence that the grammar could not cover in the *No-Part-Adv* version is given in (8).

(8) Die Anti-Baby-Pillen stehen im Verdacht , **vermehrt** Thrombosen auszulösen.
"The birth control pill is suspected to **increasingly** cause thromboses."

We also manually checked a random sample of 20 sentences that the *Filt-Part-Adv* grammar could not cover, in order to see whether the grammar systematically misses certain cases of adverbial participles. In this second random sample, the percentage of sentences containing a true adverbial participle was again 90%. The grammar could not correctly analyse these because of their special syntax that is not covered by the general adverb macro (or, of course, because of difficult constructions not related to adverbial participles). An example for such a case is given in (9).

(9) Transitreisen junger Männer vom Gaza-Streifen ins Westjordanland und **umgekehrt** sind nicht gestattet.
"Transit travels from the Gaza Strip to the West Bank and **vice versa** are not allowed for young men."

| Grammar | Parsed Sent. | Starred Sent. | Time-outs | Time in sec |
|---|---|---|---|---|
| No-Part-Adv | 50 | 77 | 12 | 427 |
| Filt-Part-Adv | 92 | 39 | 8 | 366 |

Table 7: Performance on the specialised TIGER test set, 139 sentences total

The high proportion of true adverbial participle instances in our specific testsuite suggests that the data we induced from Europarl largely carries over to TIGER (despite genre differences, for instance) and constitutes a generally useful resource. Thus, we can not only say that the filtered participle resource has no negative effect on the overall performance of the German LFG, but also extends its coverage for a less frequent phenomenon in a linguistically precise way.

## 6 Conclusion

We have proposed an empirical account for detecting adverbial participles in German. Since this category is usually not annotated in German resources and hard to describe in theory, we based our method on multilingual parallel data. This data suggests that only a fairly limited class of participles actually undergo the conversion to adverbs in free text. We have described a set of linguistically motivated filters which are necessary to induce a high-precision resource for adverbial participles from parallel data. This resource has been integrated into the German LFG grammar. In contrast to the version of the grammar which does not restrict the participle - adverb conversion, the restricted version produces less spurious ambiguities which leads to better f-score on gold standard data. Moreover, by manually evaluating a specialised data set, we have established that the restricted version also extends the coverage and produces the correct analyses which can be used for further linguistic study.

## References


Tanja Becker. 2001. DMOR: Handbuch. Technical report, IMS, University of Stuttgart.

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

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

Joan Bresnan. 2000. *Lexical-Functional Syntax*. Blackwell, Oxford.

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

Aoife Cahill, John T. Maxwell III, Paul Meurer, Christian Rohrer, and Victoria Rosén. 2007. Speeding up LFG Parsing using C-Structure Pruning . In *Coling 2008: Proceedings of the workshop on Grammar Engineering Across Frameworks*, pages 33 – 40.

Stefanie Dipper. 2003. *Implementing and Documenting Large-Scale Grammars — German LFG*. Ph.D. thesis, Universität Stuttgart, IMS.

Anette Frank, Tracy Holloway King, Jonas Kuhn, and John T. Maxwell. 2001. Optimality Theory Style Constraint Ranking in Large-Scale LFG Grammars . In Peter Sells, editor, *Formal and Empirical Issues in Optimality Theoretic Syntax*, page 367–397. CSLI Publications.

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

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

Joakim Nivre, Johan Hall, and Jens Nilsson. 2006. Maltparser: A data driven parser-generator for dependency parsing. In *Proc. of LREC-2006*.

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

Stefan Riezler, 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*.

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

Anne Schiller, Simone Teufel, and Christine Thielen. 1995. Guidelines fuer das Tagging deutscher Textkorpora mit STTS. Technical report, IMS, University of Stuttgart.

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

David Yarowsky, 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*.