A Corpus Study of Creating Rule-Based Enhanced Universal Dependencies for German

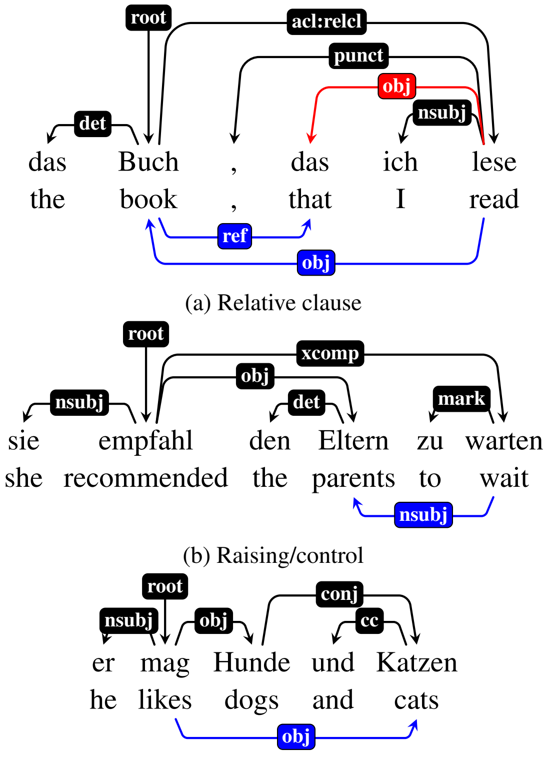
Teresa Bürkle[♦] **Stefan Grünewald**[♦★] **Annemarie Friedrich**[♦]
[♦]Bosch Center for Artificial Intelligence, Renningen, Germany
[★]Institut für Maschinelle Sprachverarbeitung, University of Stuttgart
teresa.buerkle|stefan.gruenewald|annemarie.friedrich
@de.bosch.com

Abstract

In this paper, we present a first attempt at enriching German Universal Dependencies (UD) treebanks with enhanced dependencies. Similarly to the converter for English (Schuster and Manning, 2016), we develop a rule-based system for deriving enhanced dependencies from the basic layer, covering three linguistic phenomena: relative clauses, coordination, and raising/control. For quality control, we manually correct or validate a set of 196 sentences, finding that around 90% of added relations are correct. Our data analysis reveals that difficulties arise mainly due to inconsistencies in the basic layer annotations. We show that the English system is in general applicable to German data, but that adapting to the particularities of the German treebanks and language increases precision and recall by up to 10%. Comparing the application of our converter on gold standard dependencies vs. automatic parses, we find that F1 drops by around 10% in the latter setting due to error propagation. Finally, an enhanced UD parser trained on a converted treebank performs poorly when evaluated against our annotations, indicating that more work remains to be done to create gold standard enhanced German treebanks.

1 Introduction

Universal Dependencies (UD; de Marneffe et al., 2014) are a widely used framework for syntactic dependency annotation. Aiming to provide a cross-lingual representation that focuses on relations between content words, nearly 200 UD treebanks for over 100 languages are currently available.[1] An extension of UD are *Enhanced* Universal Dependencies (Schuster and Manning, 2016), which have since become an official part of the UD guidelines (Nivre et al., 2020). Enhanced UD augments "basic" UD syntax trees by including additional dependency edges in the presence of linguistic phenomena such as coordination, relative clauses, and raising/control.



(a) Relative clause

root
xcomp
obj
nsubj
det
mark
sie empfahl den Eltern zu warten
she recommended the parents to wait
nsubj

(b) Raising/control

(c) Coordination

Figure 1: Linguistic structures in German enhanced UD.

As a result, Enhanced UD structures are not necessarily trees, but more general syntax graphs. For example, in Figure 1b, the basic dependency tree (black edges on top) is augmented by an enhanced `nsubj` link (blue edge at bottom) in order to represent that the subject of "warten", "Eltern", has been raised to the object position of "empfahl."

Enhanced UD representations have been shown to be useful for downstream NLP tasks (Schuster et al., 2017), and parsing them has experienced a rise of attention in the scope of two recent shared tasks (Bouma et al., 2020, 2021). Yet, annota-

[1] https://universaldependencies.org

Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop, pages 85–95
November, 2021. ©2021 Association for Computational Linguistics

tions for this dependency representation are still only available for a small subset of UD treebanks.[2] Many of these existing annotations were created using rule-based systems operating on the basic layer with little to no human supervision/evaluation, or via conversion from other syntactic representations (Droganova and Zeman, 2019).

In this paper, we take a first step towards extending Enhanced UD coverage to German treebanks, for which no enhanced annotations exist to date. We give details on our findings when generating Enhanced UD representations from the basic layer of German UD treebanks in a rule-based way, similarly to the work of Schuster and Manning (2016) for English. We focus on three linguistic phenomena (relative clauses, raising/control, and coordination) and present rule-based algorithms for each phenomenon that take into account the peculiarities of the German language, which has less strict rules for word order compared to English.

We assess the accuracy of the extractions made by our system with the help of expert human annotators, evaluating system performance on both gold as well as automatically parsed basic dependencies. In the gold setting, we find that around 90 % of added relations are correct, outperforming the system by Schuster and Manning (2016). On automatically parsed text, performance drops by around 10 %, indicating a reliance on correct basic dependencies. A graph parser trained on automatically extracted enhanced dependencies achieves low accuracy, indicating that more work remains to be done to automatically create high-quality enhanced German treebanks.

To the best of our knowledge, our work presents the first investigation into creating Enhanced UD annotations for German treebanks. Our system achieves accuracies which we believe may already be useful in downstream applications (e.g., to increase recall in information extraction tasks). Furthermore, we hope that our findings may serve as a starting point for the creation of high-quality Enhanced UD corpora for the German language. To facilitate further research, our code and annotations will be made freely available.[3]

2 Related Work

In this section, we briefly review work on extracting enhanced UD from the basic representation, as well as enhanced UD parsing and related graph-based meaning representations.

Extracting enhanced UD from basic UD. Much prior work has focused on automatically retrieving enhanced dependencies from the basic layer for various languages. Prior to the introduction of UD, Nyblom et al. (2013) used an SVM-based approach to retrieve enhanced Stanford dependencies for the Finnish-TDT treebank. They find that their approach works well for gold basic dependencies, but less so for automatic parses. A rule-based converter for English UD was first proposed by Schuster and Manning (2016) together with their initial formulation of the enhanced UD layer. Nivre et al. (2018) compare the above two approaches for enhancing Italian and Swedish, finding that both are adequate for creating enhanced treebanks. Candito et al. (2017) use graph-rewriting rules to create enhanced dependencies for two French treebanks, finding good accuracy in a manual evaluation.

More specific works on modeling certain phenomena within enhanced UD further include automatically reconstructing elided predicates in sentences with gapping constructions (Schuster et al., 2018) and the modeling of coordinate constructions (Grünewald et al., 2021b).

Parsing into Enhanced UD. Following a series of shared tasks focusing on basic UD (Zeman et al., 2017, 2018), the 2020 and 2021 IWPT shared tasks (Bouma et al., 2020, 2021) address the prediction of enhanced UD from raw text. While some participating systems directly parse enhanced dependency graphs using a graph-based parser (e.g. He and Choi, 2020; Wang et al., 2020; Grünewald and Friedrich, 2020), others parsed into standard UD before applying hand-written enhancement rules (e.g., Dehouck et al., 2020; Attardi et al., 2020). Heinecke (2020) employs a rule-based approach to convert from basic to enhanced UD and then optimize the result using a classifier.

Graph-based Meaning Representations. Aiming to move annotations closer to "meaning" on the meaning–form scale (Droganova and Zeman, 2019), Enhanced UD can be seen as part of a larger effort to annotate sentences with graph-based rep-

[2] The official UD website lists 41 treebanks as providing enhanced dependencies, however, manual inspection shows that in some of them the enhanced layer is identical to the basic layer. 29 treebanks for 17 languages were used for Enhanced UD parsing in the 2020 and 2021 IWPT Shared Tasks.

[3] https://github.com/boschresearch/german_enhanced_ud_converter_law_dmr_2021

resentations of their meaning. Other frameworks with the same goal include Abstract Meaning Representation (AMR; Banarescu et al., 2013), Universal Conceptual Cognitive Annotation (UCCA; Abend and Rappoport, 2013), Universal Decompositional Semantics (UDS; White et al., 2016) and broad-coverage Semantic Dependency Graphs (SDP; Oepen et al., 2016). Enhanced UD differs from these frameworks due to its very close integration with the syntactic annotation of a sentence.

3 Basic-to-Enhanced Converter

In this section, we describe the implementation of our basic-to-enhanced UD converter for German. Along the way, we discuss important differences w.r.t. the converter for English (Schuster and Manning, 2016).

3.1 Enhancing Relative Clauses

As shown in Figure 2, in basic UD, relative clauses are represented as sub-trees, with the root of this subtree being linked to the noun modified by the clause. The basic UD representation provides no direct link between a relative pronoun and its antecedent. As the relative pronoun and the antecedent can be interpreted as co-referring, the referent of the antecedent can also be considered as an argument of the relative pronoun's head (here: "kenne"). Enhanced UD marks this by means of a ref relation linking the antecedent and the relative pronouns, as well as adding an argument link (here: obj) from the relative clause's main verb to the antecedent.

Algorithm. If a relative clause is indicated by an acl:relcl relation, we add a ref relation from the clauses' antecedent (which is head of the acl:relcl relation) to the relative pronoun or relativizer. Because not all treebanks use the subtype relcl for this construction, an acl relation is also considered if the token that might be the relative pronoun has the UPOS PRON or the lemma "wo" ("where"). The relative pronoun is found by selecting the first child of the acl relation's dependent according to surface order which is linked to the relative clause's root via any relation except punct. The relation between the relative pronoun or relativizer and its head is deleted, and a relation from the relative pronoun's / relativizer's former head to the antecedent is added, using the relation type of the deleted relation, as in Figure 2.

The English converter (Schuster and Manning, 2016) treats relative clauses the same way, but their way of identifying the relative pronoun is slightly different to ours. The leftmost word contained in a word list of possible relative pronouns is taken as the dependent of the ref relation if any child or grandchild of the acl:relcl relation depends on a possible relative pronoun. In addition, only the specific subtype acl:relcl is considered for the identification of relative clauses.

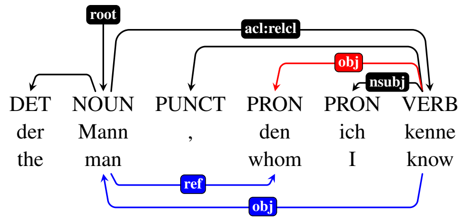


Figure 2: Basic and enhanced representation of a relative clause.

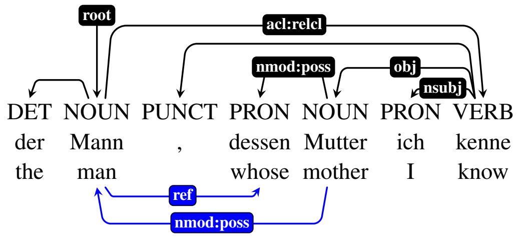

root
acl:relcl
nmod:poss
obj
nsubj
DET NOUN PUNCT PRON NOUN PRON VERB
der Mann , dessen Mutter ich kenne
the man whose mother I know
ref
nmod:poss

Figure 3: Relative clause with possessive relative pronoun.

We treat the special case of possessive relative pronouns as follows. As shown in Figure 3, if the relative pronoun is possessive, it is not a direct child of the relative clause's root, but rather attaches to a child node thereof. In German, the only two possessive relative pronouns are "deren" or "dessen." These word forms can be detected easily and unambigously. In such constructions, the relative pronoun or relativizer (here: "dessen"), refers to "Mann." Accordingly, a ref relation is inserted between "Mann" and "dessen." In contrast to the case of Figure 2, the object of "kenne" is not the pronoun's antecedent "Mann," but "Mutter" within the relative clause. Hence, an nmod:poss relation is added between "Mutter" and "Mann" instead.

In English, this type of relative clause exists with the possessive pronoun "whose" functioning as a relativizer. When analysing the differences of our converter to the output of the converter of Schuster and Manning (2016), we noted the following. In English treebanks, the relevant relation is always nmod:poss, while in some German treebanks, the relative pronoun is linked via det or det:poss. Our converter hence covers these cases as well. Further, Schuster and Manning (2016) delete the nmod:poss relation between the possessive relative pronoun and its former governor in the enhanced layer, which our converter as used in Sec. 4 does not.[4]

3.2 Enhancing Raising and Control

The linguistic phenomena called raising and control describe sentences with embedded verbs whose subjects have been "raised" to some higher position in the syntactic structure (Polinsky, 2013). In the constructions "Susan decided to start" and "Susan seemed to start," the subject of the embedded clause is shared with that of matrix clause. In the control constructing "Susan told them to start," the object "them" is also the subject of the embedded clause. While raising verbs (e.g., "seem") do not select their arguments, control verbs such as "tell" or "decide" do. In enhanced UD, as shown in Figure 4 and Figure 5, the two phenomena are represented in the same way, i.e., by making the nsubj relation explicit. The embedded clause is linked to the matrix verb via xcomp ("open clausal complement"), hence, the construction can be detected easily in the basic layer. However, resolving is non-trivial, as we will discuss in Sec. 4.

Algorithm. We first identify xcomp relations whose head node is a verb or adjective and whose dependent is a verb, noun, or adjective node. For these, we add nsubj links starting at the xcomp relation's child node according to the following criteria. We check the outgoing links of the xcomp relation's head node (the embedding predicate) for the relations iobj, obj, and nsubj in this fixed order. For the first match, we assume that child node linked using this relation is the endpoint of the newly added nsubj relation.

Raising and control constructions where the controlling or raising predicate is a noun, such as "der Versuch, etwas zu tun" ("the attempt to do something") or "die Möglichkeit, etwas zu tun" ("the opportunity to do something") often lead to the insertion of wrong links. This is the reason for only addressing cases where the contolling or raising predicate is a verb or an adjective and the controlled or embedded predicate is a verb, an adjective or a noun.

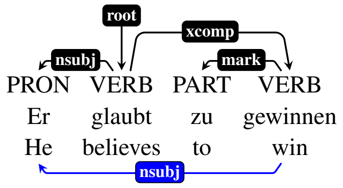


Figure 4: Raising construction example.

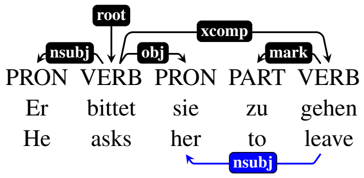


Figure 5: Object control example.

3.3 Enhancing Coordinate Constructions

In UD, coordinate constructions are represented by a conj relation between the coordinated tokens (see, e.g., Figure 9). Generally, the first conjunct of a coordinate construction according to surface order is the head, with all other conjuncts attached to it via conj relations. Other dependency relations that are semantically shared by all conjuncts (e. g., nsubj or obj relations) are also attached to the first one. Since the first conjunct can also have dependents that are **not** shared by the other conjunct(s), this means that the dependency structures do not capture the information whether a dependent of the first conjunct is also a dependent of the other conjuncts.

For enhancing conjunctions, we distinguish two cases: (a) conjoined verbs, and (b) other types of conjunctions.

Conjoined Verbs. When two verbs are conjoined, the basic UD representation lacks an explicit link between the second verb and complements that both verbs share semantically, such as subjects or objects.

Algorithm. If the second conjunct does not have an outgoing nsubj link, we add an nsubj link between the second conjunct and all nodes that are linked to the first conjunct via nsubj links.

Regarding shared objects, we distinguish three cases. If the second conjunct ("schreibt" ("writes")) comes after the first conjunct ("liest" ("reads")) and before the first conjunct's object ("Bücher" ("books")) according to surface order, as in Figure 6, we propagate an obj link between the sec-

[4] This oversight is fixed in the released version of our converter.

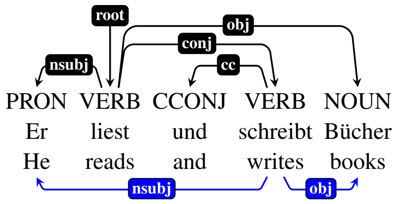

root
obj
conj
nsubj
cc
PRON VERB CCONJ VERB NOUN
Er liest und schreibt Bücher
He reads and writes books
nsubj
obj

Figure 6: Conjoined verbs that unambiguously share a subject and an object

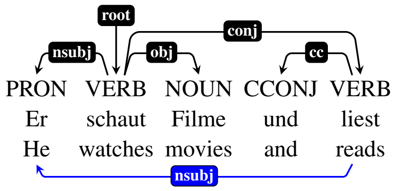

root
conj
nsubj
obj
cc
PRON VERB NOUN CCONJ VERB
Er schaut Filme und liest
He watches movies and reads
nsubj

Figure 7: Conjoined verbs that unambiguously share a subject, but no object

ond conjunct and this object. If the first conjunct's object comes after the first conjunct but before the second conjunct in surface order, as exemplified by Figure 7, no obj link is inserted. In these scenarios, it is possible to decide with certainty whether the first conjunct's object is semantically shared by both conjuncts, as it is always the case in English due to the strict surface order.

However, if the object comes before both conjuncts according to surface order, as in Figure 8, we face an ambiguity. In spite of both sentences in Figure 8 having the same dependency relations and POS tags, in the first sentence, the first conjunct's object ("Film") is semantically not shared with the second conjunct, while in the second sentence, both conjuncts share the object ("Bücher"). Therefore, in the second case, the enhanced UD representation should contain an additional obj link. However, we do not propagate in such cases in order to maximize precision rather than recall.

Other types of conjunctions. In case of coordinated noun phrases, adjectives, or modifiers, the token with an incoming conj relation (the "second conjunct"), is not explicitly linked to the governor or dependents of the first conjunct. That the two conjuncts fill the same roles in the sentence's argument structure is made explicit in the enhanced layer.

Algorithm. First, we identify the conj relation and the tokens it interlinks. We add a direct link between the second conjunct and the governor of the first conjunct. The type of the added relation is copied over from the relation linking the first conjunct to its governor.

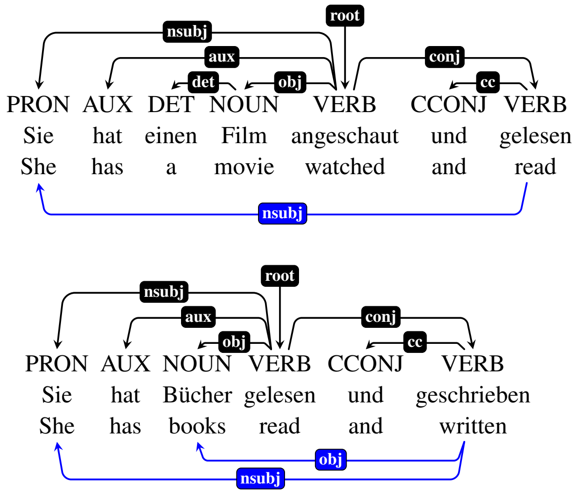

nsubj
root
aux
conj
det
obj
cc
PRON AUX DET NOUN VERB CCONJ VERB
Sie hat einen Film angeschaut und gelesen
She has a movie watched and read
nsubj
root
nsubj
aux
conj
obj
cc
PRON AUX NOUN VERB CCONJ VERB
Sie hat Bücher gelesen und geschrieben
She has books read and written
obj
nsubj

Figure 8: Conjoined verb constructions that are ambiguous regarding the question whether they semantically share the object or not

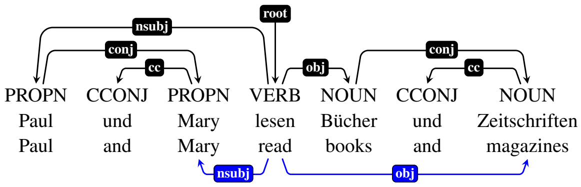

root
nsubj
conj
cc
obj
conj
cc
PROPN CCONJ PROPN VERB NOUN CCONJ NOUN
Paul und Mary lesen Bücher und Zeitschriften
Paul and Mary read books and magazines
nsubj
obj

Figure 9: Conjunction of noun phrases and objects, whith both conjuncts sharing the verb

4 Analysis and Evaluation

In this section, we first describe our manual corpus study to measure the performance of our converter. We compute precision, recall and F1 on enhanced links and draw comparisons to the English converter by Schuster and Manning (2016). In addition, we evaluate the performance of the converter in automatic parsing settings.

4.1 Manually Annotated Gold Standard

We create a test set containing sentences from the two German treebanks GSD and PUD, against which the converter is then evaluated. The test set consists of three subsets, one for each of the three phenomena we address. The subset for coordinate constructions contains 50 sentences containing at least one conj relation between two tokens, while the subsets for relative clauses and raising and control structures contain 80 sentences containing respectively an acl or subtype of it or an xcomp relation. The more specific relation subtype acl:relcl would have been more suitable for extracting relative clauses, however, the GSD treebank only uses the generic type acl.

These subsets were created by first collecting all sentences for each phenomenon from each treebank and then selecting every n/25th sentence for the coordination subset and every n/40th sentence for the relative clause and raising and control subsets, with n being the total number of sentences found for this phenomenon in this treebank. After removing duplicates and six sentences with severe errors in the basic layer, this results in 196 annotated sentences which are the basis for our evaluation.

Annotation. The test set was then converted to include enhanced UD representations by our system and manually corrected by three annotators, two computational linguists with extensive linguistic training, and one undergraduate student with basic linguistic training. Each annotator corrected a subset of the test set, discussing sentences they were unsure about with the other annotators.

Minor errors occuring in the basic layer (e.g., a wrong dependency type assigned or an inverted head and child node of a relation) were corrected, while more severe errors affecting multiple dependency relations or tokens lead to an exclusion of the sentence from the evaluation set. In total, 18 sentences were found to contain errors in the basic layer, and six were removed from the evaluation set due to severe errors.

4.2 Performance of Converter in Gold Setting

Rather than starting out with the original treebanks, we create the input for the converter by removing the enhanced layer from the manually corrected gold set. In other words, our evaluation takes into account the minor fixes we applied to the basic layer. Arguably, this procedure does not result in a completely realistic estimate how well our system would perform on the still somewhat noisy treebanks; however, we consider improving the consistency and quality of the current German treebanks to be future work and decided to abstract away from this problem in our study in order to make it less dependent on the idiosyncrasies of the current treebank versions.

We then run our converter on the basic trees for these sentences and compare the enhanced dependency relations in the converted set to the enhanced layer of the gold set. In order to make a fair comparison, we freeze the code of our converter before starting out with our annotation study. The results reported here are thus an estimate how far we can get with a rule-based approach (which, of course, was based on extensive data analysis, but not targeted to our evaluation set).

In Table 1, we report precision, recall, and F1 for all links in the enhanced layer, as well as for several relevant relations. In addition to the overall score, calculated by evaluating all sentences in our test set, we compute scores for each phenomenon by evaluating all sentences in which this phenomenon occurs. This results in the evaluation set for a phenomenon containing more sentences than the originally sampled test set for this phenomenon did, because sentences from the other phenomena's subsets can be included. For example, a sentence originally selected for the relative clause set is included in the evaluation set for conjunctions if it also contains a conj relation. This results in the evaluation for coordination constructions being based on 117 sentences, the evaluation for relative clauses being based on 105 sentences and the evaluation for raising and control being based on 98 sentences.

Considering all sentences and label types, we achieve a precision of 89.5%, recall of 89.1% and F1 of 89.3%. Comparing the scores for the three phenomena implies that the converter works best on relative clauses with F1 of 94.2%. Recall of 100% for the ref relation indicates that all relative clauses in the test set were detected as such. For sentences containing conjunctions, all scores were slightly lower with an F1 of 89.3%. Raising and control is the phenomenon for which we achieved the lowest scores, with an F1 of 82.1%. Here, nsubj is the only relation type that our converter automatically propagates, while we did not restrict propagation of relation types in our manual annotation. Precision and recall both around 77% indicate that for some cases, our converter selects the wrong token as subject of the embedded or controlled predicate. This decreases both precision and recall, because the actual subject is lacking an nsubj link to the controlled or embedded predicate and at the same time a spurious nsubj link to the wrong node is created. E.g., for the sentence "wir verwenden Zeit darauf, ein Konzept zu entwickeln" ("we spend time on developing a concept,") instead of "wir" ("we"), the converter would treat "Zeit" ("time"), the object of "verwenden" ("spend"), as a subject of the embedded verb "entwickeln" ("develop").

In our evaluation, we also noticed minor implementation errors.[5] These are fixed in the released version of our converter.

In the following, we compare our converter to the one created by Schuster and Manning (2016).[6] Our aim is to check whether implementing a targeted version for German was worth the effort compared to simply running the existing system developed for English. We made one adaptation to the English converter, which originally expects only `acl:relcl` to indicate relative clauses (see Sec. 3.1). We relax this constraint to looking for `acl`, as several German treebanks do not use the `relcl` subtype.

Since the English converter by Schuster and Manning (2016) uses some relation subtypes that we did not use in the gold set, only the generic relation type was considered in the evaluation. Overall, we found that the English converter performed worse on our test set, with an overall F1 score of 79.5% (about ten percent points lower than our targeted version for German). For `obj` and `obl` relations, the reported scores diverge from our scores by respectively 28.2 and 16.8 percentage points. The main reason for this is the different treatment of relative clauses in which the relative pronoun is governed by an `obj` relation. While we add an `obj` relation between the antecedent and the token governed by the `acl:relcl` or `acl` relation in this case, the converter by Schuster and Manning (2016) inserts an `obl:relobj` relation in this case, which is treated as an `obl` relation in the evaluation due to the evaluation algorithm ignoring subtypes. There is no meaningful difference in the handling of coordination except for the usage of more specific subtypes. Therefore, the difference in scores in Table 1 stems from the other phenomena that also occur in this evaluation subset.

Regarding raising and control constructions, in contrast to our converter (as explained in Sec. 3.2), the English converter propagates links when the head of the `xcomp` relation is a noun. Furthermore, while we used the general `nsubj` relation, they used the subtype `nsubj:xsubj` for links added while treating a raising or control construction. This is in fact an extension of the UD guidelines, which do not include the `nsubj:xsubj` subtype to date. Following the official guidelines, our system outputs only `nsubj`, but our final release includes an option for adding the `xsubj` subtype.

4.3 Performance of Converter in Automatic Parsing Setting and End-to-End Graph Parsing

To evaluate our converter's performance on automatically parsed basic dependencies, we use Grünewald et al.'s (2021a) STEPS parser (along with the provided pre-trained model for German) to parse our test set and then run our converter on the output. We find that performance of our rule-based system drops by about 10% compared to the results achieved when using gold basic dependencies. This is in line with findings reported by Grünewald et al. (2021b), who also find that a pipeline-based approach leads to considerable performance decreases. However, precision and recall exceed 75% in most cases, meaning that the resulting output may still be useful in practice.

We also investigated the alternative path of directly training an enhanced UD parser on a treebank created using our converter. Using the STEPS framework, we train an end-to-end graph parser on the converted GSD treebank, use it to parse our test set, and again evaluate against our manual annotations. As shown in Table 1, F1 for the enhanced layer drops by over 40%. By contrast, the parser achieves an F1 score of 85.1% on the same evaluation set for the prediction of the *basic* layer. These results indicate that the enhanced data produced by converting the existing German treebanks is, as of yet, of limited use for training machine learning systems.

5 Conclusion and Outlook

In this paper, we have described our approach of enriching German UD treebanks with enhanced dependencies by implementing a rule-based converter. Evaluating on a set of manually annotated sentences, our system achieves an F1 score of around 90% when operating on gold basic dependencies, outperforming the English-based system by Schuster and Manning (2016). In an automatic parsing setting, accuracy is lower with an F1 score of

[5] One such example is checking whether the second conjunct already has an outgoing `nsubj` link before propagating an addional link when enhancing conjunctions.

[6] To ensure the strongest possible baseline, we here use a slightly modified version of that converter, created by Grünewald et al. (2021b), which fixes dependency propagation in certain edge cases involving coordinate constructions. For more details, see their paper.

Phenomenon	**Basic Layer**	**System**	**Score**	all	advmod	amod	nmod	nsubj	obj	obl	ref
all			total	440	10	14	22	198	40	40	97
			P	89.5	100.0	100.0	95.0	82.7	97.4	95.2	95.1
	gold	ours	R	89.1	90.0	100.0	86.4	84.3	92.5	100.0	100.0
			F1	89.3	94.7	100.0	90.5	83.5	94.9	97.6	97.5
	gold	SM16	F1	79.5	57.1	100.0	90.5	82.4	66.7	80.8	89.3
	automatic	ours	F1	80.1	94.7	88.9	76.0	77.2	83.8	79.0	90.5
	end-to-end parser		F1	46.7	20.3	70.0	31.3	65.7	72.3	51.3	88.2
conjunctions			total	315	8	14	21	135	31	36	51
			P	90.0	100.0	100.0	94.7	85.2	96.7	94.7	91.1
	gold	ours	R	88.6	87.5	100.0	85.7	85.2	93.5	100.0	100.0
			F1	89.3	93.3	100.0	90.0	85.2	95.1	97.3	95.3
	gold	SM16	F1	80.8	66.7	100.0	90.0	84.9	78.4	86.7	87.3
	automatic	ours	F1	79.0	93.3	88.9	75.0	79.7	84.2	76.7	86.0
	end-to-end parser		F1	47.1	23.5	80.0	38.3	63.2	76.2	56.3	85.0
rel. clauses			total	316	7	7	11	127	31	29	97
			P	93.5	100.0	100.0	90.0	89.9	96.7	96.7	95.1
	gold	ours	R	94.9	85.7	100.0	81.8	91.3	93.5	100.0	100.0
			F1	94.2	92.3	100.0	85.7	90.6	95.1	98.3	97.5
	gold	SM16	F1	81.4	25.0	100.0	85.7	88.5	62.2	76.3	89.3
	automatic	ours	F1	83.6	92.3	84.2	70.6	80.9	88.5	80.6	90.5
	end-to-end parser		F1	51.6	26.7	76.2	31.0	68.4	73.5	53.5	88.2
raising +			total	221	2	7	13	131	14	9	33
control			P	84.6	100.0	100.0	100.0	77.3	100.0	100.0	97.1
	gold	ours	R	79.6	50.0	100.0	76.9	77.9	78.6	100.0	100.0
			F1	82.1	66.7	100.0	87.0	77.6	88.0	100.0	98.5
	gold	SM16	F1	76.2	66.7	100.0	87.0	77.2	44.4	72.0	97.1
	automatic	ours	F1	76.1	66.7	92.3	76.9	72.5	76.5	82.4	90.4
	end-to-end parser		F1	41.2	5.3	63.6	32.4	61.6	48.8	34.3	85.3

Table 1: **Annotation study: converter results**. Rows for phenomena report results for **all** links in enhanced layer for set of sentences that contain the phenomenon. **all** = all links in enhanced layer in the subset of sentences, **total** = total count of this label type in the evaluated gold sentences. Number of sentences in the evaluation subset: Coordination: 117, Relative Clause: 105, Raising and Control: 98. SM16 = Schuster and Manning (2016).

around 80 %, which may still be useful in practice. Training an enhanced parser on treebanks created using our converter yields poor results.

In our annotation study, we found that our system fails mainly due to errors in the basic layer, which have been derived semi-automatically from German treebanks annotated with different dependency frameworks. For example, an `xcomp` relation in the wrong direction leads to severe errors in the enhanced layer when following our rule-based approach. We further observed divergences in regards to the annotation schemes applied across German treebanks, for instance the usage of different relation types for the same sentence structures. We therefore conclude that measures to increase consistency across German treebanks are an important step to facilitate the conversion to enhanced dependencies.

Outlook. There remains a number of ambiguous cases for which we have not found a solution thus far, such as the propagation of objects in case of two conjoined verbs (see Sec. 3.2). Regarding raising and control structures, examples like “Wir lassen den Jungen gehen” (“We let the boy leave”) vs. “Wir lassen den Reifen montieren” (“We have the tyre mounted”) are identical regarding their basic dependency structure and POS tags, yet should be treated differently by the converter. In control structures such as “Wir entschieden, hier nicht zu schwimmen” (“We decided not to swim here,”) the subject of the embedded verb “schwimmen” (“swim”) is the same as that of “entschieden” ("decided"), while in the sentence "Wir empfehlen, hier nicht zu schwimmen" ("We recommend not to swim here,") the subject does not co-refer with the speaker. A possibility to improve here might be the usage of word lists containing control/raising verbs for which the subject of the controlled/embedded predicate can not co-refer with the speaker (as for "recommend").

To conclude, our work demonstrates that rule-based conversion is a suitable method for creating enhanced dependencies for German. However, our analysis also shows that taking language-specific phenomena into account is important to ensure correctness of the extractions. Hence, creating targeted enhanced UD converters for more languages is a promising avenue, but will likely require further manual work for each instance.

Acknowledgments

We thank the anonymous reviewers and Jannik Strötgen for their helpful comments regarding this work.

References

Omri Abend and Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 228–238, Sofia, Bulgaria. Association for Computational Linguistics.

Giuseppe Attardi, Daniele Sartiano, and Maria Simi. 2020. Linear neural parsing and hybrid enhancement for enhanced Universal Dependencies. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 206–214, Online. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Gosse Bouma, Djamé Seddah, and Daniel Zeman. 2020. Overview of the IWPT 2020 shared task on parsing into enhanced Universal Dependencies. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 151–161, Online. Association for Computational Linguistics.

Gosse Bouma, Djamé Seddah, and Daniel Zeman. 2021. From raw text to enhanced Universal Dependencies: The parsing shared task at IWPT 2021. In *Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021)*, pages 146–157, Online. Association for Computational Linguistics.

Marie Candito, Bruno Guillaume, Guy Perrier, and Djamé Seddah. 2017. Enhanced UD dependencies with neutralized diathesis alternation. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 42–53, Pisa,Italy. Linköping University Electronic Press.

Mathieu Dehouck, Mark Anderson, and Carlos Gómez-Rodríguez. 2020. Efficient EUD parsing. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 192–205, Online. Association for Computational Linguistics.

Kira Droganova and Daniel Zeman. 2019. Towards deep Universal Dependencies. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 144–152, Paris, France. Association for Computational Linguistics.

Stefan Grünewald and Annemarie Friedrich. 2020. RobertNLP at the IWPT 2020 shared task: Surprisingly simple enhanced UD parsing for English. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 245–252, Online. Association for Computational Linguistics.

Stefan Grünewald, Annemarie Friedrich, and Jonas Kuhn. 2021a. Applying occam's razor to transformer-based dependency parsing: What works, what doesn't, and what is really necessary. In *Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021)*, pages 131–144, Online. Association for Computational Linguistics.

Stefan Grünewald, Prisca Piccirilli, and Annemarie Friedrich. 2021b. Coordinate constructions in English enhanced Universal Dependencies: Analysis and computational modeling. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 795–809, Online. Association for Computational Linguistics.

Han He and Jinho D. Choi. 2020. Adaptation of multilingual transformer encoder for robust enhanced Universal Dependency parsing. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing* *into Enhanced Universal Dependencies*, pages 181–191, Online. Association for Computational Linguistics.

Johannes Heinecke. 2020. Hybrid enhanced Universal Dependencies parsing. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 174–180, Online. Association for Computational Linguistics.

Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4585–4592, Reykjavik, Iceland. European Language Resources Association (ELRA).

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Joakim Nivre, Paola Marongiu, Filip Ginter, Jenna Kanerva, Simonetta Montemagni, Sebastian Schuster, and Maria Simi. 2018. Enhancing Universal Dependency treebanks: A case study. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 102–107, Brussels, Belgium. Association for Computational Linguistics.

Jenna Nyblom, Samuel Kohonen, Katri Haverinen, Tapio Salakoski, and Filip Ginter. 2013. Predicting conjunct propagation and other extended Stanford dependencies. In *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*, pages 252–261, Prague, Czech Republic. Charles University in Prague, Matfyzpress, Prague, Czech Republic.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajič, Angelina Ivanova, and Zdeňka Urešová. 2016. Towards comparability of linguistic graph Banks for semantic parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3991–3995, Portorož, Slovenia. European Language Resources Association (ELRA).

Maria Polinsky. 2013. *Raising and control*. Cambridge University Press, Cambridge.

Sebastian Schuster, Éric Villemonte de La Clergerie, Marie D Candito, Benoît Sagot, Christopher D Manning, and Djamé Seddah. 2017. Paris and Stanford at EPE 2017: Downstream Evaluation of Graph-based Dependency Representations. In *EPE 2017 - The First Shared Task on Extrinsic Parser Evaluation*, Proceedings of the 2017 Shared Task on Extrinsic Parser Evaluation, pages 47–59, Pisa, Italy.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2371–2378, Portorož, Slovenia. European Language Resources Association (ELRA).

Sebastian Schuster, Joakim Nivre, and Christopher D. Manning. 2018. Sentences with gapping: Parsing and reconstructing elided predicates. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1156–1168, New Orleans, Louisiana. Association for Computational Linguistics.

Xinyu Wang, Yong Jiang, and Kewei Tu. 2020. Enhanced Universal Dependency parsing with second-order inference and mixture of training data. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 215–220, Online. Association for Computational Linguistics.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on Universal Dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas. Association for Computational Linguistics.

Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. 2018. CoNLL 2018 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–21, Brussels, Belgium. Association for Computational Linguistics.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj, and Josie Li. 2017. CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.