

Universal Dependencies: Extensions for Modern and Historical German

Stefanie Dipper, Cora Haiber, Anna Maria Schröter,
Alexandra Wiemann, Maike Brinkschulte

Department of Linguistics

Fakultät für Philologie

Ruhr-Universität Bochum

{Stefanie.Dipper, Cora.Haiber, Anna.Schroeter,
Alexandra.Wiemann, Maike.Brinkschulte}@ruhr-uni-bochum.de

Abstract

In this paper we present extensions of the UD scheme for modern and historical German. The extensions relate in part to fundamental differences such as those between different kinds of arguments and modifiers. We illustrate the extensions with examples from the MHG data and discuss a number of MHG-specific constructions. At the current time, we have annotated a corpus of Middle High German with almost 29K tokens using this scheme, which to our knowledge is the first UD treebank for Middle High German. Inter-annotator agreement is very high: the annotators achieve a score of $\alpha = 0.85$. A statistical analysis of the annotations shows some interesting differences in the distribution of labels between modern and historical German.

Keywords: Universal Dependencies, Treebank of Historical German, Annotation Scheme

1. Introduction

The work reported here was carried out as part of two projects investigating the syntax of Middle High German (1050–1350 CE). Reliably-annotated treebanks do not yet exist for this period, so we have created our own corpus.¹ We chose Universal Dependencies (UD, [de Marneffe et al., 2021](#)) as our framework in order to benefit from the UD-related resources (schemas, treebanks, parsers) on the one hand, and on the other hand UD with its focus on content words is well suited for abstracting from changes in the system of function words as they often occur in diachronic syntax.

We started from the official UD schema for German, i.e. the specification for UD relations as provided on UD’s website,² and realized that for the adequate annotation of central syntactic properties of historical (and modern) German, we need to adapt and extend the annotation scheme for German in various places, because the schema contradicts in a number of cases the common syntactic analyses of German and of standard (constituent) treebanks of German. In some cases this concerns key distinctions, such as the distinction between arguments and modifiers, or between subject and object infinitival clauses.

The extensions we propose concern a range of

¹The Middle High German Treebank by [Chiarcos et al. \(2018\)](#) has been parsed automatically. The parses were not evaluated, though, so the quality of the analyses is unclear.

²<https://universaldependencies.org/de/>

Treebank	#Tokens	Annotation
UD_German-HDT (Völker et al., 2019–2023)	3.4M	conv
UD_German-GSD (Petrov et al., 2015–2023)	290K	80% conv 20% nat
UD_German-LIT (Salomoni, 2017–2023)	40K	conv
UD_German-PUD (Uszkoreit et al., 2019–2023)	21K	nat
tweeDe (Rehbein et al., 2019)	12K	nat

Table 1: Overview of existing UD treebanks for modern German. Column Annotation specifies if the UD relations have been automatically converted or natively annotated in UD style.

syntactic relations for modern and historical German. To fit these extensions in the already existing UD scheme and to make them downward compatible as much as possible, we make use of the possibility to customize general UD labels by defining subtypes using the notation *universal:extension* (see for example the already existing *aux:pass* for passive auxiliaries, [de Marneffe et al., 2021](#)). With this proposal, we hope to establish a standard for subtypes in German so that future treebanks and parsers of German can build on them.

For modern German, there are a number of UD treebanks available. However, most of them were not natively annotated using the UD-style but result from converting pre-UD treebanks automati-

ReM ID	ReM Title	Text type	Date (ms.)	Sent no	#Tokens	
1	M205A-N1	Pfaffe Konrad: 'Rolandslied' (A)	Chanson de geste	late 12th cent	1–480	5,024
2	M005-G1	Trierer Aegidius	Saint legend	1200	1–513	9,288
3	M008-G1	Straßburger Alexander	Novel about antiquity	early 13th cent	1–412	5,568
4	M335-G1	Rheinisches Marienlob	Marian praise	1250	220-670	8,862

Table 2: The four MHG data sets that have been annotated in this study. All texts are verse texts, data sets 1, 2 and 4 are from the religious domain. The dates are estimates of the date of origin of the respective manuscript. Text M005-G1 has been annotated in full, the other texts in extracts, as specified by the sentence numbers. The MHG corpus consists of 28.742 tokens (1,856 sentences) in total.

cally. In particular, this holds for the largest treebanks available for German, see Table 1 for an overview. Part of the information required by our schema is actually contained in the original treebanks and could be transferred to the UD treebanks.

The paper is structured as follows. We introduce our data in Section 2 and present the main features of our annotation scheme extensions in Section 3. In Section 4, we report the Inter-Annotator Agreement (IAA) and some statistics of the new subtypes. The main contributions of this paper are:

- Extensions of the UD annotation scheme for modern and historical German
- The first UD treebank for Middle High German (with almost 29K tokens)³

2. Data

We developed and tested the extensions using two data sets, one for modern German and one for Middle High German.

Modern German (ModG) This data set is taken from the German GSD treebank (Petrov et al., 2015–2023), which is part of the Universal Dependencies framework and consists of news, reviews and web data. Our data set is taken from the development set⁴ and, hence, comes from the part that has been manually annotated in UD style (McDonald et al., 2013). For details on the data size, see Sec. 4.

Middle High German (MHG) This data set is based on the Reference Corpus of Middle High German, version 1.0 (ReM) (Klein et al., 2016) and consists of four texts in verse form, with a focus on the religious domain, see Table 2 for more details.^{5,6} ReM is annotated at many levels, including

³<https://doi.org/10.5281/zenodo.10870745>

⁴Dev set: Reviews = s1-s500, News = s501-s799; https://github.com/UniversalDependencies/UD_German-GSD/blob/master/de_gsd-ud-dev.conllu, v2.10, commit 4d38410.

⁵See <https://www.linguistics.rub.de/rem/corpus/details.html> for full details of the texts.

⁶Large parts of the Old and Middle High German manuscript tradition are written in bound language,

part of speech, but not syntactic relations.

Since there are only few syntactically annotated corpora for MHG, our focus is on this part of the data set. The ModG data set serves mainly as a basis for comparison.

We performed the annotation in INCEpTION, an annotation tool that, among many other formats, supports linking elements by pointers (Klie et al., 2018). The screenshots of the annotated examples are also from INCEpTION.

3. Annotation Scheme Extensions

We see two issues that prevent the full usability of the UD-DE scheme (called UD-DE in the following) and UD-annotated treebanks such as the German GSD treebank for German users: inconsistent terminology and lack of distinctions.⁷

Inconsistent terminology Certain labels of UD-DE contradict the traditional analyses for German syntax, as evidenced, e.g., by the schemes used by the two major treebanks for German, Tübingen Treebank of Written German (TüBa-D/Z) (Telljohann et al., 2017) and TIGER (Brants et al., 2004). For example, UD-DE only accepts accusative objects as so-called direct or indirect objects (*obj*, *iobj*) and analyses dative and genitive objects as “oblique” non-core arguments (*obl:arg*). Prepositional phrases – whether arguments or modifiers – are marked as *obl*.

In contrast, Zeman (2017) argues for distinguishing between different kinds of objects according to their form: accusative objects (*obj*) vs. dative/genitive objects (*iobj*) vs. prepositional objects (*obl:arg*), thus closely following the traditional analyses of German syntax.

Lack of distinctions The lack of a more fine-grained specification in UD-DE leads to some very

while pure prose occurs less frequently. Bound language often uses syntactically marked word orders but is syntactically still more regular than lyric poetry (such as minnesong) or strophic epic poetry (such as the Nibelungenlied).

⁷We refer to version 2.13 of the UD-DE scheme, available at <http://hdl.handle.net/11234/1-5287>.

UD	Extension	Description	Example
(i) Dependents of verbs			
obl	:arg	prepositional object	[<i>uon den ungnaden</i>] <i>loste</i> ‘redeemed from the toil’
	:loc/:dir/:tmp	locative/directional/ temporal phrase	<i>man <u>furten</u> in [<i>in der stunde</i>]. <i>witene hiene vnd hare.</i> ‘they led him far to and fro in the hour’</i>
	:mod	modifier	<i>her <u>gab sich an sine gewalt</u>. [<i>mit bitteren trehenen</i>] ‘he went into his power with bitter tears’</i>
xcomp	:compar	comparative phrase	<i>da he <u>gedenet steit</u> [<i>alse ein snar</i>] ‘there he stands stretched like a string’</i>
	:subj	infinitival subject	ModG: <i>Den Beschäftigten ist es nicht weiter <u>zuzumuten</u>, [<i>ohne Bezahlung weiterzuarbeiten</i>] ‘employees cannot be expected to continue working without pay’</i>
advmod	:pred	predicative xcomp	<i>al <u>widermude</u>. dat bitter machde din suze gemude</i> ‘all the adversity that made your sweet mood bitter’
	:arg	adverbial object	<i>her bat daz sie sich des wereten [...] vnd immer [dar vmme] <u>rungen</u></i> ‘he asked them to fight back about it and always struggle for it’
	:loc/:dir/:tmp	locative/directional/ temporal adverb	<i>stille her danen <u>quam</u></i> ‘silently he came from there’
(ii) Dependents of nouns			
nmod	:arg	argument of deverbal/ deadjectival noun	ModG: <i>Eine Auseinandersetzung ist ja immer auch ein <u>Eingehen auf die Figuren</u></i> ‘An argument is always also a response to the characters’
	:loc/:dir/:tmp	loc/dir/temp. modifier of deverb/adj. noun	ModG: <i>durch Hilfen zum <u>Bleiben</u> [<i>im Lande</i>] <u>zu bewegen</u></i> ‘to encourage them to stay in the country by providing assistance’
	:part	partitive	<i>Vol [<i>dines heiligen geistes salueN</i>]. ‘full of the ointment of your holy spirit’</i>
	:det	non-adjacent quantifier	<i><u>uater</u> vnd muoter lach beide</i> ‘father and mother were both lying’
advmod	:nmod	adverbial modifier	<i>also manige <u>raste</u></i> ‘many a rest’
amod	:pred	depictive	<i>dat <u>si got louende</u> danne gingen</i> ‘that they went away praising God’
(iii) Other dependents			
aux	:cop	copula-like constructions	<i>sie wart <u>berehaft</u></i> ‘she became fertile’
	:fut	future tense	ModG: <i>wir werden auch die Kosten <u>übernehmen</u></i> ‘we will also bear the costs’
	:pass	passive voice	<i>so wart sie uon deme guoten man. harte wole <u>intfangen</u>.</i> ‘so she was very well received by the good man’
	:sub	subjunctive	ModG: <i>Ich würde nicht <u>empfehlen</u> ...</i> ‘I would not recommend’
compound	:case	part of pronominal adverb	<i><u>da</u> her sich mite solde neren</i> ‘with this he should feed himself’

Table 3: Extensions proposed for the German UD. In the examples, the dependent is printed in bold and its head is underlined. Boundaries [...] are provided for phrasal dependents. Unmarked examples are from MHG.

general labels like *xcomp*, which covers quite different kinds of phrases, including infinitival subject and object clauses as well as predicative phrases, or *advmod*, which subsumes all adverbs which modify a predicate or another modifier word.

Moreover, (modifier) prepositional phrases and adverbs can serve very different semantic functions and can be further distinguished according to the type of modification. Basic types are specifications of the time or the place of an event. These differences also have an effect on word order: in general, a modifier specifying time precedes a modifier specifying the location in modern

German. Similarly, the distinction between locative vs. directional specifications is overtly represented by different cases of alternating prepositions in German: dative case indicates a location, and accusative case, a direction. Therefore a semantic differentiation between these kinds of modifiers also makes sense from a syntactic point of view.

We introduce our scheme in two ways: First we explain the new subtypes of the most important functions. Then we illustrate with some selected cases how certain syntactic constructions in German have changed over the centuries and how we

can capture this with our scheme.

3.1. New subtypes

The current version of UD-DE defines ten subtypes: (i) arguments: *nsubj:pass*, *csubj:pass*, *obl:arg*, *obl:agent*; (ii) modifiers: *acl:relcl*, *advcl:relcl*; (iii) others: *expl:pv*, *aux:pass*, *det:poss*, *compound:prt*. We adopt six of these subtypes,⁸ but add a range of more subtypes, in particular for the following fundamental functions: *obl*, *xcomp*, *nmod*, *advmod*, *aux*. All extensions proposed for the German UD are summarized in Table 3 and illustrated primarily with examples from Middle High German. We describe the major extensions in detail below.

obl UD-DE applies *obl* to all kinds of prepositional phrases (and even used to include dative and genitive objects under this label in earlier versions).

However, the distinction between prepositional phrases that function as objects vs. as modifiers is a fundamental one: The preposition within an object phrase loses its original meaning completely and is therefore called “semantically empty”. Accordingly, the choice of the preposition is driven by the governing verb rather than by the preposition’s semantics. In contrast, the preposition within a modifier determines the type of relation to the verb (e.g., a spatial or temporal indication, or an instrument or manner). Traditionally, German grammar and German treebanks distinguish between these two types,⁹ we mark prepositional objects as *obl:arg*.

With some specific verbs, locative and temporal specifications function similar to objects in that they are obligatory, just like objects and unlike modifiers, e.g., with verbs like *legen* ‘put (somewhere)’ or *dauern* ‘last (for some time)’. In such cases, it is not easy to make the distinction between object and modifier status, so we define extra labels for them: *obl:loc* (for locative specifications) *obl:dir* (for directions) and *obl:tmp* (for temporal specifications).¹⁰

⁸We don’t use *nsubj:pass*, *csubj:pass* but record passive voice on the auxiliary only, by *aux:pass*.

Since possessive pronouns function like ordinary determiners and can be identified by their POS tag, we don’t use *det:poss*.

Finally, *obl:agent* is covered by *obl:mod* in our tagset.

⁹In the TIGER and Tüba/DZ treebanks: OP or OPP as object, respectively; MO or MOD as modifier (Albert et al., 2003; Telljohann et al., 2017)

¹⁰The UD page <https://universaldependencies.org/ext-dep-index.html> lists all language-specific extensions that occur in UD data. Unfortunately, there are no standards as to how to name the subtypes, even though quite a lot of them seem to be rather widespread. Locative adverbials, e.g., are called *advmod:lmod*, *adv-*

Finally, prepositional phrases specifying an instrument or manner are annotated as *obl:mod*; comparative prepositional phrases are annotated as *obl:compar*.

xcomp UD-DE uses *xcomp* for an “open” predicative or clausal complement without an overt subject. An important distinction that is lost here is whether the clausal complement functions as a subject clause or an object clause.¹¹ Hence, we introduce the subtype *xcomp:subj*, and reserve *xcomp* for the most common case, namely object clause. We mark predicative complements as *xcomp:pred*.¹²

advmod UD-DE uses *advmod* for any kind of adverbials. In fact, however, adverbs can differ significantly in their syntactic roles. Certain adverbials function similarly to prepositional phrases and can perform similar syntactic functions: In addition to their typical role as modifiers, they can also serve as locative, directional, or temporal complements (*advmod:loc/dir/tmp*) in cases where such a complement is required by the verb; cf. the example *legen* ‘put (somewhere)’, where the complement is either in the form of a prepositional phrase (*auf den Tisch legen* ‘put on the table’, *obl:dir*) or an adverb (*dorthin legen* ‘put there’, *advmod:dir*).

A special type of adverbs are pronominal adverbs, which are fused from a pronominal part and a (postponed) preposition, e.g. *auf das/dem* = *darauf* ‘on it; thereon’. Pronominal adverbs very often function as prepositional objects, e.g. *darauf warten* ‘wait for it’. Analogous to prepositional phrase objects, we annotate such cases with the label *advmod:arg*.¹³

nmod UD-DE annotates nominal and prepositional phrases when they modify a noun, all as *nmod*. Even in this domain, however, there are complements similar to arguments: In the case of deverbal or deadjectival nouns,¹⁴ the nouns can

mod:loc, *advmod:locy* in different languages, temporal adverbials are called *tmod* in a range of languages.

¹¹Note that this is something different than the issue of controlled subjects *within* the clause, which is represented by enhanced dependencies in UD.

¹²As it turned out, infinitive subjects do not occur in MHG (in the data we have annotated so far), but they do in ModG, see Sec. 4.3.

¹³In fact, according to our annotations, there are only a few instances of *advmod:arg*, see Sec. 4.3. The main reason for this is that such adverbs often serve as correlates to argument clauses (similar to expletive *es* or *it* in German and English). We annotate such uses accordingly as *expl*.

¹⁴We define deverbal and deadjectival nouns morphologically: Deverbal nouns are nominal infinitives ((*das*) *Reisen* ‘traveling’), derivatives with *-ung* (*Erstattung* ‘refund’) or with *Ge-...-e* (*Gerede* ‘talk’). De-

inherit the argument roles of the underlying verbs (or adjectives). In this case, an accusative object of a verb is realized in the form of a genitive complement to the noun, and prepositional objects remain as such. We annotate such arguments with *nmod:arg*. In addition, there are again locative, directional, or temporal complements (*nmod:loc/dir/tmp*).

A fairly common phenomenon in MHG are quantifiers that refer to a noun in a similar way to a determiner, but are not adjacent to that noun and are therefore often referred to as “floating quantifiers”. The referent noun is generally easily identifiable and we analyze such quantifiers as dependents of their referent noun, with the special label *nmod:det* (to distinguish them from normal determiners).

aux The category *aux* is used in UD-DE for auxiliaries and modal verbs, for passive auxiliaries there is *aux:pass*, which we also use. We introduce additional subtypes because auxiliaries in German are generally very ambiguous: *sein* ‘be’ and *werden* ‘be/become’ can be used in copula constructions, in the passive, and for perfect and future tenses, respectively. In our scheme, *aux* covers only perfect and modal verbs, *werden*, *bleiben* ‘become, remain’ in copula constructions are annotated with *aux:cop*, future *werden* ‘will’ as *aux:fut*. For subjunctive *würden*, *täten* ‘would’, we use *aux:sub*.

3.2. Constructions specific to MHG

In this section, we want to address a selection of cases of linguistic historical interest, illustrated by examples from the annotated MHG data set. The dependency analyses of the respective examples are all displayed in Fig. 1.

to-Infinitive (1) shows the MHG construction *begunden ze lebene* ‘began to live’ (lit. ‘began to living’) consists of a verb plus a gerund functioning as a prepositional object and is analyzed accordingly, as *obl:arg*, see Fig. 1. This construction later developed into the sequence *zu* ‘to’ plus infinitive, i.e., the preposition *zu* became the infinitive marker *zu*, and the prepositional object (*lebene* ‘living’) became an infinitive (*leben* ‘live’). Accordingly in the later construction, *zu* is no longer annotated with *case*, but with *mark*, and the infinitive functions as *xcomp* instead of *obl:arg*. For determining whether the old construction is still in use in a text at hand, we check whether there is overt case inflection on the object.

adjectival nouns are derivatives with *-heit*, *-keit*, *-igkeit* (*Zufriedenheit* ‘satisfaction’). Semantically lexicalized forms are excluded.

- (1) *nach sinen rate begunden ze lebene*
 according to his advice began to living.DAT
 ‘began to live by his advice’

Dislocation There are significantly more instances of dislocation in early texts than in modern texts, presumably because of the more oral style in many of these texts. The dislocated constituent can be, e.g., a preposed (or postposed) nominal phrase or a complement or relative clause, which is followed by a resumptive pronoun. In these cases, the relevant semantic information is clearly encoded by the dislocated constituent and not by the semantically empty resumptive pronoun. Hence, we consider it important to link the preposed constituent properly with the rest of the sentence. For this purpose, we analyse the preposed constituent as the actual argument and mark the resumptive pronoun as a kind of expletive, because it “doubles” the preposed constituent.

As an example, consider (2) along with Fig. 1: the preposed constituent is linked by the argument relation *csubj* for “clausal subject”, the resumptive pronoun *die* ‘they’ is annotated with *expl*. (This example also illustrates the annotation of locative (*dar* ‘there’) and temporal modifiers (*eo* ‘before’).)

- (2) *die dir eo clageten sere die wurden do z ware der gotelichen werke uro glad*
 who there before complained a lot they became then indeed the divine works glad
 ‘those who complained there before became glad indeed of the works of God’

Note that there are also cases for which we use the UD label *dislocated*: when a nominal phrase precedes or follows the sentence and there is a clear syntactic distance between the nominal phrase and the resumptive pronoun, e.g., a pause in speech (which is appropriately graphemically marked) or other constituents intervene, as in (3): the resumptive pronoun *her* ‘he’ is not adjacent to the preceding nominal phrase *der gotis bote here* ‘the noble messenger of God’.

- (3) *der gotis bote here. mineme trechtene her des lob gab.*
 Lord he his praise gave
 ‘the noble messenger of God, to my Lord he gave his [= the Lord’s] praise’

Partitive A very frequent construction in MHG is the partitive, a substance noun in genitive case that follows a quantity expression. In the MHG construction the phrase with the quantity is the

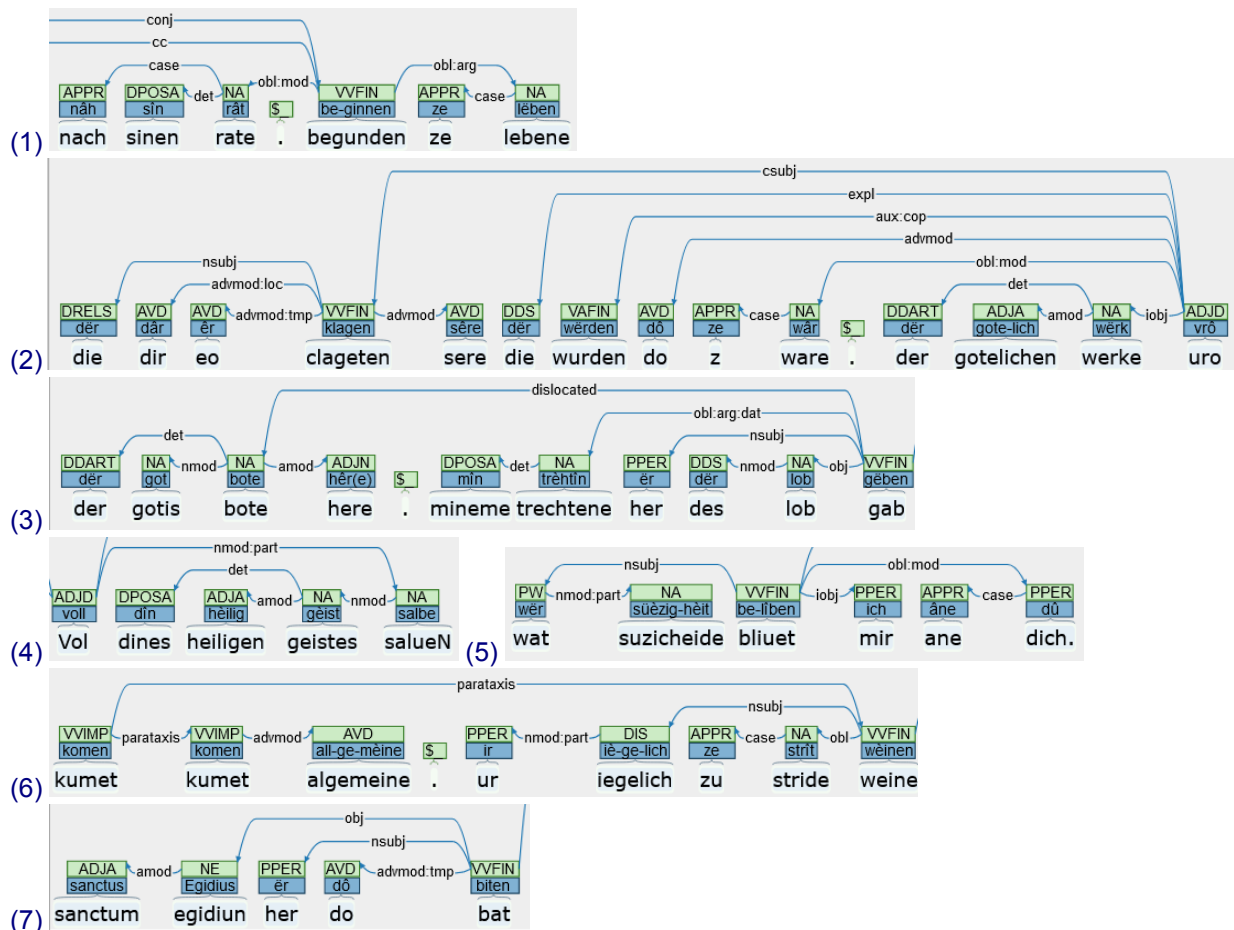


Figure 1: Annotations of the MHG examples (1)–(7) discussed in Sec. 3.2.

head, taking over the argument function, cf. (4)–(6). The genitive phrase, which functions as a kind of modifier, does not agree with the head and may – unlike in modern German – even precede it, cf. the expression *ur iegelich* ‘each of you’ in (6), where the genitive part “of you” (*ur*) is preposed. In corresponding constructions in modern German, the quantity expression functions as the determiner and the substance expression is the head noun, similar to English *much sweetness*.

- (4) *Vol dines heiligen geistes saluen*
 full your holy spirit ointment.GEN.PL
 ‘full of the ointment of your holy spirit’
- (5) *wat suzicheide bliuet mir ane dich.*
 what sweetness.GEN.SG remains me
 without you
 ‘what sweetness remains to me without you’
- (6) *kumet kumet algemeine.*
 come.IMP come.IMP all
ur iegelich zu stride weine.
 you.GEN.PL each to quarrel cry.IMP
 ‘Come, come all, every one of you cry to the quarrel’

Proper nouns In our religious MHG texts, many people are called *sanct* ‘holy’. In later times, some of them were recognized as saints by the Catholic church. We assume that the Latin adjective *sanctus* acts like a real adjective and modifier in early times and therefore annotate it with *amod* in our texts, see (7) and Fig. 1. In later times the adjective becomes part of the proper name and would then be annotated with the *flat* relation.

- (7) *sanctum egidiun her do bat*
 holy/saint Aegidius he then asked
 ‘he then asked Saint Aegidius’

4. Results

In this section, we present the results for inter-annotator agreement (IAA) and some statistics from the annotations that show the distribution of the new subtypes.

4.1. Inter-annotator agreement

To measure IAA, it is common to use chance-corrected evaluation measures. For word-based annotations like part of speech, there are a number of canonical measures that are usually applied, such as Cohen’s κ (Cohen, 1960) (cf. Car-

	Data set	Annotators	#Sent	#Tokens	Alpha	UAS	LAS	LAcc
1	News	A1, A2	50	884	.99	.95	.91	.93
2	News	A1, A3	50	988	.96	.90	.83	.90
3	Reviews	A1, A2	50	662	.98	.95	.89	.91
4	Reviews	A1, A3	50	679	.96	.92	.85	.90
5	M005-G1	A4, A5	513	9,288	.86	.91	.81	.84
6	M008-G1	A4, A6	209	2,766	.72	.79	.65	.72
7	M335-G1	A4, A5	251	4,144	.93	.92	.85	.88
	MHG total		973	16,198	.85	.89	.79	.83

Table 4: The data sets that has been double corrected (ModG, 1–4) or double annotated (MHG, 5–7) for calculating inter-annotator agreement. A1–A6 are the annotators; Alpha, UAS, LAS and Label Accuracy (LAcc) display their respective agreement scores. MHG total shows the micro-averages for the entire MHG data set.

letta, 1996). For dependency relations, however, there is no established chance-corrected measure so far. Instead, measures without chance correction are typically used, such as Labeled and Unlabeled Attachment Scores (LAS, UAS), which are the standard evaluation metrics in dependency parsing.

In this paper, we apply a chance-corrected measure, α , that has been proposed by Skjærholt (2014), who adapts Krippendorff’s α (Krippendorff, 1970, 2019) to syntactic relations. In its basic form, α is defined as in (1), where D_o denotes the observed disagreement and D_e is the disagreement expected by chance.

$$\alpha = 1 - \frac{D_o}{D_e} \quad (1)$$

D_e is computed with the help of a distance function δ . Skjærholt (2014) proposes different definitions of δ as to calculate agreement with syntactic annotations. The one we apply here is the unmodified tree edit distance (TED) function, called δ_{plain} , which Skjærholt (2014) evaluates as the most stable distance function for computing α . TED is defined as the minimum number of operations required to transform one (syntax) tree into another.¹⁵

Data sets For estimating IAA for ModG, three annotators independently went through a set of pre-annotated sentences.¹⁶ Annotators A2 and A3 each corrected 50 different sentences from news and 50 different from reviews, Annotator A1 corrected all of them, see Table 4, data sets 1–4, for details. The annotators corrected errors in the original dependency annotations and additionally

¹⁵For computation we used the script provided by Skjærholt (2014), <https://github.com/arnsholt/syn-agreement/> and adapted it to Python 3.

¹⁶For each data set, one researcher and two student assistants carried out the annotation.

Annotators	#Sent	Alpha	UAS	LAS	LAcc
A1, TR	200	.92	.89	.74	.81
A2, TR	100	.92	.89	.75	.80
A3, TR	100	.96	.96	.85	.87

Table 5: Agreement scores between each of the three annotators (A1, A2, A3) and the original treebank (TR).

included the new subtypes, thus changing some general labels to more fine-grained ones.

For estimating IAA for MHG, three annotators independently annotated 1,164 sentences from scratch, see Table 4, data sets 5–7. Annotator A4 annotated all of them, the others parts of it.

IAA results Table 4 shows that there is high agreement between the annotators and data sets, with $\alpha \geq 0.96$ (ModG) and $\alpha \geq 0.72$ (MHG). In general, chance-corrected scores of 0.61–0.80 are often considered “substantial” and scores of >0.81 “almost perfect agreement” (Landis and Koch, 1977). Table 4 also shows that agreement between A1 and A2 is considerably higher than between A1 and A3, in particular with the news data. Annotator A3 had the shortest training, which is reflected in the agreement values. Similarly for MHG, agreement is clearly higher between A4 and A5 as compared to A6 – again, Annotator A6 had the shortest training.

Of course, the ModG data have not been annotated from scratch so it has to be taken into account how many labels and relations have been modified at all by the three annotators. To estimate how large the proportion of modified structures and labels is, we additionally measure the agreement between the three annotators and the pre-annotations of the treebank. As Table 5 shows, a considerable fraction of the existing treebank has been modified: 13–20% of the labels have been changed (LAcc), and 4–11% of the structures

Data	Label 1	Label 2	F1
MHG	advcl	ccomp	0.2956
	acl:relcl	acl	0.2747
	obl:dir	obl:loc	0.2539
	obl:arg	obl	0.1791
	flat	appos	0.1539
ModG	acl	acl:relcl	0.2400
	obl:arg	obl:mod	0.1957
	advmod:tmp	advmod	0.1942
	advmod:tmp	fixed	0.1765
	xcomp	xcomp:subj	0.1739

Table 6: Top five confusions (ignoring empty annotations) in the two data sets measured by F1.

(UAS). For annotators A1 and A2, the values are very clearly below those among themselves (see Table 4), so that we can indeed consider their IAA for Modern German as “almost perfect”.

4.2. Error analysis

An error analysis of the annotated MHG data shows that the distinction between different types of clauses is not always clear, see Table 6 for the most frequent labels of disagreement.¹⁷ This is at least partly due to the fact that linguistic tests can be applied to historical data only to a limited extent. This affects the confusion of *advcl* and *ccomp* (adverbial vs. complement clauses) and could also affect the distinction between *acl:relcl* and *acl* (relative clauses vs. other modifying clauses).

In addition, the dividing line between locative and directional elements is often not clear. (8) shows an example: the locative element *ce stete* ‘to (the) place’ could be either the destination of going or the place of receiving, as indicated in the translation. Such distinctions can often remain unclear in historical texts.

- (8) *iedoch muoste her da ce stete. intfahen*
 however had to he there to place receive
den geistlichen gewalt
 the spiritual power
 ‘However, he had to go (to the place) to receive (at the place) the spiritual power’

For ModG, the confusions mainly concern the omission of a more specific subtype. This can

¹⁷F1 is calculated as follows:

$$2 * \frac{a_1 l_1 * a_2 l_2}{a_1 l_1 + a_2 l_2}$$

with a_1, a_2 as the annotators and l_1, l_2 as the labels annotated by the respective annotator. Possible values are between 0 and 1, where 1 means perfect agreement if $l_1 = l_2$, and 0 means perfect disagreement if $l_1 \neq l_2$. Thus, the measure corresponds to the F1 score if one of the annotators is treated as the gold standard.

MHG		ModG	
Label	Freq (%)	Label	Freq (%)
nsubj	14.41	det	14.66
det	14.19	nsubj	10.52
obj	7.94	case	9.89
advmod	7.65	advmod	7.00
case	6.82	obj	4.77

Table 7: Top five most frequent dependency labels (ignoring root, punct and unspecified relations) in the two data sets.

be explained by the fact that the ModG data was pre-notated and such cases are easily overlooked. There were also problems of differentiation between prepositional objects and modifiers (which, of course, is known to be notoriously difficult). The confusion between *advmod:tmp* and *fixed* concerns fixed (but transparent) expressions such as *nicht mehr* ‘no longer’ or *nur noch* ‘only’.

4.3. Corpus statistics

In this section, we want to compare the distribution of dependency relations in the two data sets, MHG and ModG. The comparison can only be rough, though, for two reasons: the data sets differ greatly in size, and the text types (news, reviews in ModG, mostly religious texts in MHG) also differ considerably. We use the complete annotated data set MHG with 1,856 sentences (28,742 tokens, see Table 2 for details) and the 200 corrected sentences (3,213 tokens) of ModG as data basis.

As Table 7 shows, the five most frequent labels are the same in both data sets, but they differ in their order. In both data sets, there are roughly twice as many subjects as objects, however, their overall proportion is significantly lower in ModG (15% vs. 22% in MHG). There seem to be more considerably nominal phrases with articles in ModG (*det*), and also significantly more prepositional phrases (*case*).

We next look at the distribution of the newly introduced subtypes, grouped according to their main functions: *obl*, *xcomp*, *advmod*, *nmod*, *aux*, cf. Fig. 2. The bar plots compare the distributions of MHG (left) and ModG (right). The numbers above the columns show the absolute frequency of the entire class, i.e. this number corresponds to the frequency of the original UD labels without our extensions.

One can observe that the unmodified, underspecified labels (i.e. *xcomp*, *advmod* etc.) represent the vast majority in four of the five classes with both data sets, see the parts in light blue.¹⁸

¹⁸The label *xcomp*, however, is actually the specific relation *xcomp:obj*, which represents a kind of default

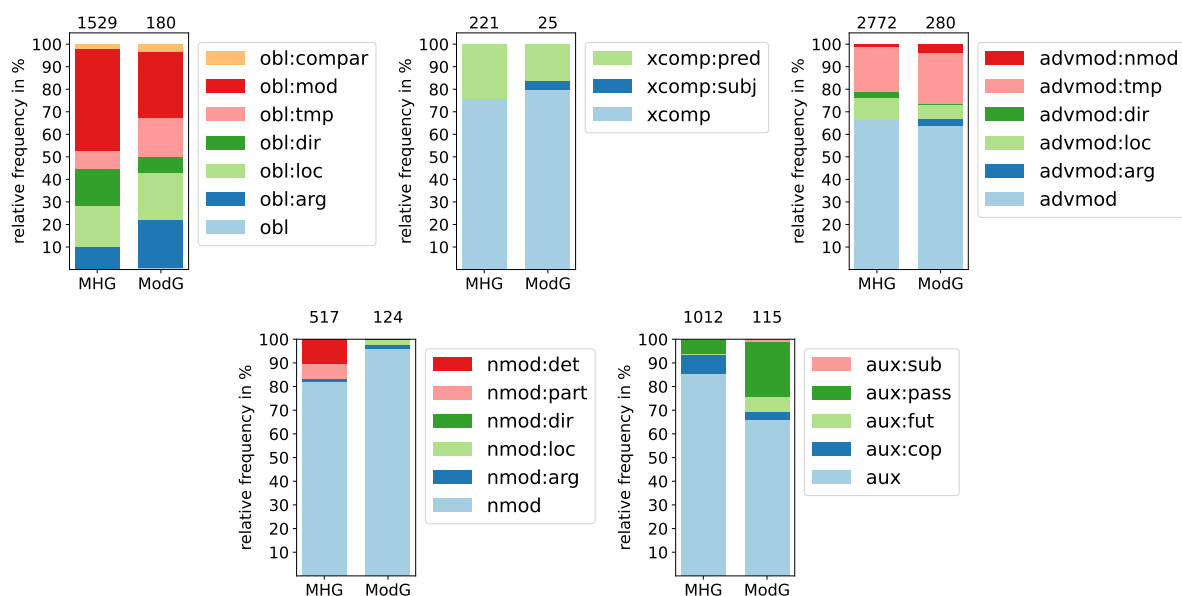


Figure 2: Distributions (in %) of the proposed subtypes of *obl*, *xcomp*, *advmod*, *nmod* and *aux* in MHG (left bars) and ModG (right bars). The figures on top of the bars specify the absolute number of occurrences.

In contrast, the subtypes are strongly represented in the area of the *obl*-functions. In ModG, the subtype *obl:arg* for prepositional objects in particular has a large share compared to MHG (22% vs. 10%); this may be related to the fact that the linguistic tests (obligatory status of the phrase; semantic transparency of the preposition) can only be determined for MHG with the help of a lexicon, so that overall annotation is more conservative. The semantic subtypes (*loc*, *dir*, *tmp*, *mod*) are strongly represented in both data sets. For *xcomp* and *advmod* the distributions look quite similar. Predicative phrases and locative and temporal adverbs occur quite frequently in both data sets.

The two remaining classes, *nmod* and *aux* show interesting differences between the two data sets: As already discussed in Sec. 3.2, the partitive (*nmod:part*) is a rather common construction in Middle High German that no longer exists in modern German – this is also evident in our data (6% vs. 0%). In addition, non-adjacent determiners (*nmod:det*) are a frequent phenomenon in MHG but not in ModG (10% vs. 0%). It is noticeable that nominal modifiers in general occur much more frequently in ModG (3.9% of the total data) than in MHG (1.8%).

In the class of auxiliaries, only copula-like constructions with the verbs *wërden* ‘become’ and *beliben* ‘remain’¹⁹ and passive voice occur some-

within this class, see Sec. 3.1.

¹⁹Unmarked copula constructions (*cop*) in MHG use the verbs *sîn* or *wësen*, which both mean ‘be’. At that time, these were two different lemmas with morpholog-

ically different forms, which later merged.

5. Conclusion

In this paper we presented our extended UD scheme for modern and historical German. We have shown that we have achieved high inter-annotator agreement scores and that the scheme can be reliably annotated. We make the annotated ModG and MHG data available under a free license.²⁰ We plan to annotate more MHG data, including data from other genres, using this scheme and train automatic methods on it. We are also working on an automatic parser for our scheme. We hope that our proposals will increase the acceptance and usability of UD for German-language data.

Acknowledgments

We are very grateful to the anonymous reviewers for their very constructive reviews. Many thanks also to Madeleine Landsberg Scherff and Alona Solopov for their contribution to the annotation. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1102 – Project ID 232722074; and – SFB 1475 – Project ID 441126958 / Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – SFB 1102 – Project ID 232722074; und – SFB 1475 – Projektnummer 441126958.

²⁰<https://doi.org/10.5281/zenodo.10870745>

6. Bibliographical References

- Stefanie Albert, Jan Anderssen, Regine Bader, Stephanie Becker, Tobias Bracht, Sabine Brants, Thorsten Brants, Vera Demberg, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Hagen Hirschmann, Juliane Janitzek, Carolin Kirstein, Robert Langner, Lukas Michelbacher, Oliver Plaehn, Cordula Preis, Marcus Pußel, Marco Rower, Bettina Schrader, Anne Schwartz, George Smith, and Hans Uszkoreit. 2003. *TIGER Annotationsschema*. Technical report, Universität des Saarlandes, Universität Stuttgart, Universität Potsdam.
- Sabina Brants, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith, and Hans Uszkoreit. 2004. *TIGER: Linguistic interpretation of a German corpus*. *Journal of Language and Computation*, 2:597–620.
- Jean Carletta. 1996. *Assessing agreement on classification tasks: the kappa statistic*. *Computational Linguistics*, 22(2):249–254.
- Christian Chiarcos, Benjamin Kosmehl, Christian Fäth, and Maria Sukhareva. 2018. *Analyzing Middle High German syntax with RDF and SPARQL*. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Jacob Cohen. 1960. *A coefficient of agreement for nominal scales*. *Educational and Psychological Measurement*, 20(1):37–46.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. *Universal Dependencies*. *Computational Linguistics*, 47(2):255–308.
- Jan-Christoph Klie, Michael Bugert, Beto Boulosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. *The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation*. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics.
- Klaus Krippendorff. 1970. *Estimating the reliability, systematic error and random error of interval data*. *Educational and Psychological Measurement*, 30(1):61–70.
- Klaus Krippendorff. 2019. *Content Analysis: An introduction to its methodology*, 4th edition.

J. Richard Landis and Gary G. Koch. 1977. *The measurement of observer agreement for categorical data*. *Biometrics*, 33(1).

Arne Skjærholt. 2014. *A chance-corrected measure of inter-annotator agreement for syntax*. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 934–944, Baltimore, Maryland. Association for Computational Linguistics.

Heike Telljohann, Erhard W. Hinrichs, Sandra Kübler, Heike Zinsmeister, and Kathrin Beck. 2017. *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z)*. Seminar für Sprachwissenschaft, Universität Tübingen, Germany.

Daniel Zeman. 2017. *Core arguments in Universal Dependencies*. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 287–296, Pisa, Italy.

7. Language Resource References

Thomas Klein, Klaus-Peter Wegera, Stefanie Dipper, and Claudia Wich-Reif. 2016. *Referenzkorpus Mittelhochdeutsch (1050–1350)*, Version 1.0. <https://www.linguistics.ruhr-uni-bochum.de/rem/>. ISLRN 332-536-136-099-5.

Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. *Universal Dependency annotation for multilingual parsing*. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria. Association for Computational Linguistics.

Slav Petrov and Wolfgang Seeker and Ryan McDonald and Joakim Nivre and Daniel Zeman and Adriane Boyd. 2015–2023. *UD German-GSD*. Universal Dependencies Project, distributed via LINDAT/CLARIN, Universal Dependencies, 2.13. [[link](#)].

Ines Rehbein, Josef Ruppenhofer, and Bich-Ngoc Do. 2019. *tweeDe – a Universal Dependencies treebank for German tweets*. In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 100–108, Paris, France. Association for Computational Linguistics.

Alessio Salomoni. 2017–2023. *UD German-HDT*. Universal Dependencies Project, distributed via LINDAT/CLARIN, Universal Dependencies, 2.13. [\[link\]](#).

Hans Uszkoreit and Vivien Macketanz and Aljoscha Burchardt and Georg Rehm and Kim Harris and Katrin Marheinecke and Slav Petrov and Tolga Kayadelen and Mohammed Attia and Ali Elkahky and Zhuoran Yu and Emily Pitler and Saran Lertpradit and Michael Mandl and Sebastian Bank and Martin Popel and Daniel Zeman. 2019–2023. *UD German-PUD*. Universal Dependencies Project, distributed via LINDAT/CLARIN, Universal Dependencies, 2.13. [\[link\]](#).

Emanuel Borges Völker and Felix Hennig and Arne Köhn and Maximilian Wendt. 2019–2023. *UD German-HDT*. Universal Dependencies Project, distributed via LINDAT/CLARIN, Universal Dependencies, 2.13. [\[link\]](#).