UD for German Poetry

Stefanie Dipper and Ronja Laarmann-Quante
Ruhr University Bochum, Germany
Faculty of Philology
Department of Linguistics
{firstname.lastname}@ruhr-uni-bochum.de

Abstract

This article deals with the syntactic analysis of German-language poetry from different centuries. We use Universal Dependencies (UD) as our syntactic framework. We discuss particular challenges of the poems in terms of tokenization, sentence boundary recognition and special syntactic constructions. Our annotated pilot corpus currently consists of 20 poems with a total of 2,162 tokens, which originate from the PoeTree.de corpus. We present some statistics on our annotations and also evaluate the automatic UD annotation from PoeTree.de using our annotations.

1 Introduction

For many years, poetry played a rather subordinate role in the field of Natural Language Processing. With the emergence of the field of Digital Humanities, the interest in computational methods for philological questions has increased considerably.

It is often assumed that the particular repetitive structures and patterns of poems serve, among other things, to support memorization. These structures prominently include meter and the rhyme scheme. Many computational approaches to the analysis of poetry have focused on these features, such as the work by Bobenhausen and Hammerich (2015); Haider (2023, 2021); Delente and Renault (2017), who automatically analyze phonological features like meter and rhyme in English, German and French poetry (also see the overview in De Sisto et al., 2024).

On a more abstract level, syntactic patterns can also support memorization, e.g. by using the same syntactic structure several times. Lee and Kong (2012) investigate and compare such constructions in poems of two Chinese poets under the name 'parallel couplets', which refers to semantic or syntactic correspondences between two lines.

Another syntactic phenomenon that is specific to poems is unusual word order, which often results from metrical constraints. A special case is the enjambment, in which the elements of a syntactic phrase are separated by spreading them over two lines, whereby special retarding effects can be achieved. Enjambment has been examined by Ruiz Fabo et al. (2017) and Hussein et al. (2018).

Syntactic annotations of poems are also interesting for other reasons. For instance, DeHass (2024) uses them to compare paratactic vs. hypotactic style in Latin colloquial texts, poetry, and prose. Syntax annotations can in general facilitate access to the content of a text and, e.g., allow for easy retrieval of (syntactically and semantically) related units such as entities and their properties, or events and their participants, as suggested by Bamman (2020) for literary texts.

For a long time, such studies could only be carried out on a rather small amount of data that has been manually annotated by experts. Examples of such manually-created poetry treebanks are the York-Helsinki parsed corpus of Old English poetry (YCOEP)[1] (Pintzuk and Leendert, 2001), the Ancient Greek and Latin Dependency Treebank (AGLDT)[2] (Bamman and Crane, 2011), which comprises a subset of poetic texts, and the treebank of Classical Chinese poems (Lee and Kong, 2012).

Only recently has there been work on the *automatic* syntax analysis of poems and the first poetry treebanks have been generated automatically, such as the diachronic treebank of Spanish Sonnets (Ruiz Fabo et al., 2017), the treebank of Classical Arabic poetry (Al-Ghamdi et al., 2021), or the PoeTree treebank with poems in ten different languages (Plecháč et al., 2024). All three treebanks were automatically annotated by generic parsers

[1] https://www-users.york.ac.uk/~lang18/pcorpus.html

[2] http://perseusdl.github.io/treebank_data/

Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities, pages 177–188
November 16, 2024.

that were not specifically trained for poems because no training data was available (see Section 4.2 for evaluation results).

Our work aims to contribute to the creation of training and evaluation data for the syntactic analysis of poetry. Our paper deals with the analysis of German-language poems from different centuries. We use the framework of Universal Dependencies (UD, de Marneffe et al., 2021) and discuss some difficulties for annotation that arise due to the peculiarities of poems. We manually annotate a pilot corpus of 20 poems with 2,162 tokens, following the guidelines proposed by Dipper et al. (2024), who define a set of customized labels for German.

Our data comes from the PoeTree corpus, that has been automatically annotated with UD trees by UDPipe (Straka, 2018). We evaluate the UD trees by means of our manually created annotations. For the evaluation, we map the customized labels back to UD labels. The main contributions of this paper are:

- A pilot UD treebank of currently 20 German-language poems from different centuries, which are available under a free license.[3]
- An in-depth discussion of the specific challenges of (German) poetry.
- A first evaluation of the automatic analyses from UDPipe.

2 Data

The data for our study comes from the PoeTree corpus, which consists of more than 330,000 poems with 89,000,000 tokens from 10 European languages (Plecháč et al., 2023; Plecháč et al., 2024).[4] All poems have been annotated automatically with UD-style dependencies using UDPipe 2.0 (Straka, 2018). However, only the annotations of the Czech-language subcorpus have already been evaluated (Cinková et al., 2024) (see Section 4.2).

The German-language sub-corpus of the PoeTree corpus, called PoeTree.de (Bobenhausen and Hammerich, 2015; Haider, 2021) consists of 74,000 poems. The automatic annotations comprise lemma, inflection features, universal part-of-speech (POS) tags, language-specific POS according to the STTS tagset (Schiller et al., 1999) and UD-style dependency relations.

[3]The corpus is available at https://gitlab.ruhr-uni-bochum.de/vamos-cl/ud-for-german-poetry.

[4]Version 0.0.2, https://versologie.cz/poetree/.

	Mean ± SD	Total
Tokens	108.1 ± 92.8	2,162
Lines	15.7 ± 15.1	314
Stanzas	4.9 ± 5.7	97

Table 1: Statistics on the annotated German sub-corpus: mean number and standard deviation of tokens, lines and stanzas per poem.

We randomly selected 20 poems from PoeTree.de, see Table 1 for an overview and Table 6 in Appendix A for more details. Standard deviations are very large and show that the poems differ greatly with regard to their size. For annotating dependency relations, we use the manual annotation tool INCEpTION (Klie et al., 2018). Each poem was annotated once, by one of the authors, and difficult cases were discussed together.

3 Poetry-Specific Issues

In this section, we discuss selected special features of poetry that pose a challenge for automatic language processing. The focus is on German-language poetry; many of the challenges, however, also arise for poetry in other languages.

3.1 Word forms and tokenization

Poetry belongs to non-standard language data in several respects. For example, capitalization is often handled differently than in standard language, e.g., some poems are written entirely in lower case. In the poems of our corpus, the first word of a line is usually capitalized, whereas in standardized spelling in German only sentence beginnings and nouns are capitalized, so that the unusual capitalization sometimes results in incorrect POS tags. For instance, in (1) the verb *schreibe* 'write' occurs twice. The first (lowercase) instance is correctly tagged in the corpus with VVFIN (finite verb), the second (capitalized) instance is incorrectly tagged with NN (noun).

(1) *Da sitz ich am Tisch und schreibe,*
Schreibe wie mir es glückt,
'I sit at the table and write, write the way I manage.' Source: 00-1734-0000-0002-9F08-4#0[5]

[5]ID as provided in the PoeTree.de corpus.

Due to the meter, word forms can be shortened (elision) or lengthened (epenthesis). Elisions are often marked by an apostrophe, which can be problematic for tokenization. For example, the elision in *heil'gen* 'holy' in (2), which stands for *heiligen*, has been split into three tokens *heil*, apostrophe and *gen* in PoeTree.de.[6] We annotate such partial tokens with the UD relation `goeswith`.

(2) *Liebste, laß in Dir die Schauer*
Weben dieser heil'gen Nacht,
'Dearest, let the shivers of this holy night
weave in you,' Source: dta.poem.21583

Another difficulty is the UD treatment of contracted prepositions plus articles, which are very common in German, e.g. *am* ≈ *an dem* 'at the'. In UD style (and in the PoeTree data), such contractions are split and treated as two words, each annotated with its own UD relation (cf. Grünewald and Friedrich, 2020).[7] In (manual) annotation of poetry, this procedure is problematic: on the one hand, the meter is no longer correct due to the inserted syllable, and on the other hand, there are differences in meaning between the contracted and the split form, i.e., one form cannot be replaced equivalently by the other (see, e.g., Cieschinger, 2016). Moreover, split forms are highly marked in German and make up less than 10% of the occurrences in a newspaper and a web corpus (Cieschinger, 2016, p. 6), i.e., the UD tokenization applied in PoeTree.de makes it harder to understand the poems. Annotators often have to re-merge the preposition and the article mentally when annotating in order to understand the meaning of a poem correctly. For example, the phrase *aufs Neue* 'anew' – shortened to *aufs neu'* in the poem line shown in (3) – is rendered as *auf das neu'* 'on the new' in PoeTree.de. The line with the split form could mean something like 'always hope for the new', which is clearly not the correct meaning and does not fit the context.

(3) *Und was dir fehlschlug, hoffe stets aufs neu'*
'And what you have failed, always hope anew'
Source: 00-1734-0000-0002-B719-B#0

[6]The partial tokens exist as independent words: *heil* means 'sound' and *gen* is an obsolete preposition meaning 'towards'. Presumably this leads to the incorrect tokenization.

[7]Also see the distinction between tokenization and word segmentation in the UD Guidelines `https://universaldependencies.org/u/overview/tokenization.html`: Tokens are defined orthographically in UD and are marked by whitespace. Words are defined syntactically and, consequently, clitics are split off and contractions are undone.

Note that the original form *aufs* 'on the' is recorded in the 'multiword' entry in the JSON format of the PoeTree.de treebank and could be used in the dependency annotation. However, a new, complex dependency label would then have to be introduced for the contracted forms (e.g. `case+det`). This would deviate from the UD guidelines, though, and the automatically generated analyses of PoeTree could no longer be evaluated directly, which is why we have annotated the split forms.[8]

3.2 Sentence boundaries

Another feature is the non-standard punctuation of poetry, which, together with the unusual capitalization, means that sentence boundaries are often not correctly determined and, for example, arguments or modifiers are not located in the same sentence segment[9] as their head. This causes problems for syntax analysis, since dependencies are usually only marked within a sentence. In our manual annotations we have treated incorrect sentence segments as follows:

1. If there are several sentences in one sentence segment, link them by the relation `parataxis`.

2. If two incomplete fragments are distributed over two sentence segments, each fragment is annotated as usually, but only those relations that apply within the segment are annotated. Note that such fragments are not linked to other parts of the sentence with the relation `parataxis`, so that there can be several roots within a sentence segment in our corpus.

An example is given in (4), which consists of one sentence. The double slashes "//" indicate the (incorrect) sentence boundaries as defined in PoeTree.de. This means that several central relations are missing, see the annotations shown in Figure 1: *Meinem Herzen* 'to my heart' is governed by the verb *reiche* 'give' as a dative object, *von der Schlange* 'from the snake' belongs to the verb *gebissen* 'bitten'.

[8]In some UD treebanks, e.g. in the German-LIT treebank (Salomoni, 2017, 2021) or in the NArabizi treebank (Seddah et al., 2020), such contracted forms are preserved in their original form and composite POS tags are used, e.g. P+DET in NArabizi. However, at the dependency level, these forms are annotated by the label `case` only, so that the information on the determiner is not represented in the dependency relations.

[9]We refer to the (potentially incorrect) sentences as defined in the PoeTree.de corpus as "sentence segments".

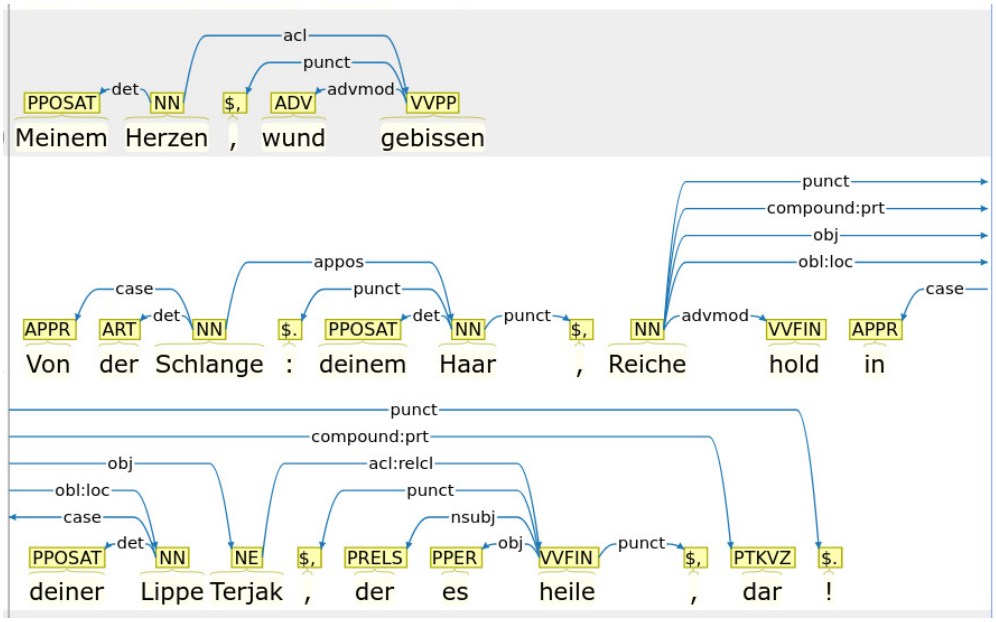


Figure 1: Annotation of the fragments of Ex. (4); screenshot of INCEpTION.

(4) *Meinem Herzen, wund gebissen //*
Von der Schlange: deinem Haar,
Reiche hold in deiner Lippe
Terjak, der es heile, dar!
'To my heart, bitten sore from the snake, [i.e.] your hair, give in your lip Terjak [a medical concoction], which heals it!'
Source: 00-1734-0000-0003-2E39-A#0

In Example (4) the fragments each form a complete subtree, with the roots *gebissen* 'bitten' and *Schlange* 'snake'. In other cases, however, the heads of the fragments can also be in the other sentence segment, so that individual words then remain unrelated. For example, in (5) the verb head *tragen* 'bear' only occurs in the second sentence segment, so that its dependents in the first segment – the comma, the conjunction *dass* 'that' and the prepositional phrase *auf der Wiese* 'in the meadow' – remain unrelated, see the annotations in Figure 2.

(5) *Erhebe dich, dass auf der Wiese //*
Durch deines Wuchses hohes Streben
Zipressenbäume Früchte tragen
'Arise, so that in the meadow through your growth's high aspiration cypress trees bear fruit' Source: 00-1734-0000-0003-28BA-1#0

A final example is (6), in which the repeated instances of *Geduld!* 'patience' are separated by a (false) sentence boundary due to the exclamation mark.

(6) *Geduld! //*
Geduld! – die ew'gen Sterne gehn
Doch ihren Pfad.
'Patience! Patience! – the eternal stars go but their path.' Source: 00-1734-0000-0002-B719-B#0

3.3 Poetry-specific constructions

Repetitions Repetitions are a typical feature of poetry. We have already seen examples in (1) and (6), another example is (7), where *mein Herz* 'my heart' is repeated three times in a row. We annotate such repetitions like lists, but with a special label, `list:rep`.[10]

(7) *Das Meer hat seine Perlen,*
Der Himmel hat seine Sterne,
Aber mein Herz, mein Herz,
Mein Herz hat seine Liebe.
'The sea has its pearls, the sky has its stars, but my heart, my heart, my heart has its love'
Source: dta.poem.10555

Anacoluthon Sentence interruptions (*anacolutha*) are also a typical feature. In Example (8), the first and third lines each begin with a free relative clause (*Die noch schlafen* 'who still sleep', *Die noch keimen* 'who still germinate'), which is not integrated into the surrounding clauses and whose referent therefore remains vague.

[10]This label is a customized UD label using the notation *universal:customized* (see for example the already existing *aux:pass* for passive auxiliaries, de Marneffe et al., 2021).

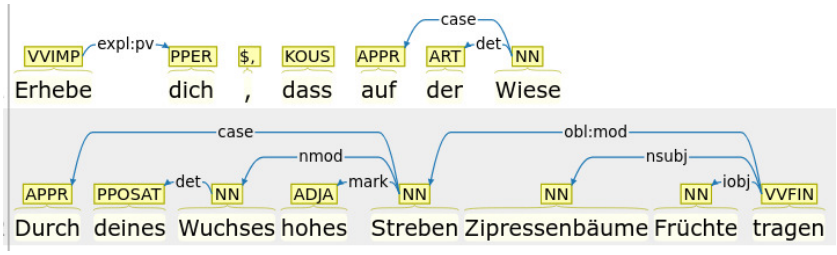


Figure 2: Partial annotation of the fragments of Ex. (5); screenshot of INCEpTION.

(8) *Die noch schlafen, aus den Wäldern*
Rauscht's wie leiser Vogelsang,
Die noch keimen, von den Feldern
Blüht's wie Duft das Thal entlang.
'Who still sleep, from the woods it rustles like soft birdsong, who still germinate, from the fields it blossoms like fragrance along the valley.' Source: dta.poem.21583

We annotate such interruptions with the label `parataxis`, as proposed by Paccosi et al. (2023) (see their example (3)). However, `parataxis` captures a rather heterogeneous class of relations, such as unconnected sentences and also parentheses. We currently also annotate cases in which a sentence boundary is missing (see Section 3.2) with this label. Additional subtypes for differentiation should possibly be defined here.

News		Reviews		Poetry	
punct	16.2	punct	13.5	punct	16.9
det	13.5	advmod	12.5	det	14.4
nsubj	8.6	det	11.0	nsubj	10.2
case	8.6	nsubj	9.3	advmod	8.6
advmod	7.1	case	8.1	case	6.9
obl	5.9	obl	6.0	obl	6.3
nmod	4.8	obj	4.6	obj	5.2
amod	4.7	conj	4.4	conj	4.4
flat	3.8	cc	4.3	cc	3.9
aux	3.7	aux	4.0	amod	3.2
Total	1,772		1,241		1,962

Table 2: The 10 most frequent dependency labels in the poetry, news and reviews data, along with their relative frequencies (ignoring root and unspecified labels). The last line specifies the total number of labels in the respective data set.

4 Results

In this section, we present selected statistics from our annotations. In the second part, we evaluate the automatic annotations of PoeTree.de.

4.1 Selected statistics

We first compare the distributions of the different labels with the two corpora for modern German from Dipper et al. (2024), which were annotated according to the same UD scheme. These corpora originate from the GSD treebank and contain modern news and reviews. Since all three corpora are rather small,[11] we truncate all extended labels (e.g. `obl:arg`) and map them to the basic labels (`obl`).

Table 2 shows the 10 most frequent labels in the three text types. Six of the labels (`punct`, `det`, `nsubj`, `advmod`, `case`, `obl`) occur in all three text types and their distributions are rather similar. Coordinations (labels `conj` and `cc`) are typical for reviews and poetry, while they make up less than 3% of the labels in news. The label `flat`, which is mainly used to annotate complex proper names, is typical for news; in reviews it occurs with 1.1%, in poetry not at all. The label `aux`, which is used to annotate modal verbs and auxiliaries, occurs in poetry with 2.3%, i.e. less often than in reviews and news.

We also compare the distributions of all labels using Spearman's rank correlation coefficient r, which compares the ranks of the labels in the three text types, and the Jensen-Shannon distance, which is a symmetric and normalized variant of the Kullback-Leibler divergence and compares the probabilities of the labels with each other.[12]

Table 3 shows that with both measures, the text types news and reviews are very similar to each other (high correlation, small distance), while news and poetry are clearly less similar to each other. The reviews occupy a position between the other two text types. The data sets can therefore be arranged as follows based on their similarities:

[11] News: 100 sentences with 1,872 tokens; reviews: 100 sentences with 1,341 tokens.

[12] Spearman's rank correlation coefficient and Jensen-Shannon distance were calculated with SciPy, `https://docs.scipy.org`.

	News	Reviews	Poetry
News	1		
Reviews	0.94	1	
Poetry	0.76	0.83	1

	News	Reviews	Poetry
News	0		
Reviews	0.145	0	
Poetry	0.213	0.173	0

Table 3: Pairwise Spearman's rank correlation coefficient r (top) and Jensen-Shannon distance (bottom) between the three data sets.

Metric	PT.de	GSD	PT.cz	PDT
UAS	79.6 ± 8.7	82.8	85.0	95.0
LAS	68.9 ± 9.1	78.2	79.7	93.6
CLAS	59.2 ± 10.7	–	–	–

Table 4: Mean (and standard deviation) for different evaluation metrics. Column PT.de shows the results of our evaluation, whereas columns GSD, PT.cz and PDT show the results of evaluations from others (see the main text for details).

news – reviews – poetry. This result can possibly be attributed by the fact that poetry has more conceptually oral characteristics and is therefore closer to reviews than to news.

4.2 Evaluation of PoeTree.de

We compare the automatically created dependency relations that come with the PoeTree.de corpus with our manually created gold-standard annotations. As a preprocessing step, we again map the extended labels to the basic labels. Furthermore, we remove punctuation marks (i.e. tokens tagged as `$.`, `$(` or `$,` according to the STTS tagset) for two reasons: Firstly, the label `punct` with which they are attached to their heads is the most frequent one (see Table 2) but of little interest for the analysis of syntactic structure. Therefore, removing them makes the evaluation more informative. Secondly, for long sequences of `parataxis`, we systematically deviated from the UD guidelines for the label `punct` in that we attached it to the head of the last clause rather than the root node of the whole sentence to avoid overly long dependency arcs, thereby facilitating the manual annotation process.

UAS, LAS, CLAS For each poem, we calculate the following evaluation metrics: Unlabeled Attachement Score (UAS), which measures the percentage of tokens that are assigned the correct head; Labeled Attachement Score (LAS), which is the percentage of tokens that are assigned the correct head and the correct dependency label; Content-Word Labeled Attachment Score (CLAS, Zeman et al., 2017), which calculates LAS only for content words, ignoring function words. We use the evaluation script of the 2018 CoNLL shared task.[13]

[13] https://universaldependencies.org/conll18/evaluation.html.

Table 4 shows the mean (and standard deviation) for each metric (see column PT.de), Figure 3 displays the distribution of scores across all poems. Overall, the results are rather poor. Straka (2018) reports clearly better results when applying UDPipe to the German-GSD corpus, which contains newspaper, reviews and web texts (see column GSD). The LAS scores in particular show a large gap.

Interestingly, when UDPipe is applied to a sample of 29 poems (6,591 tokens) from the Czech PoeTree subcorpus, it clearly outperforms both evaluations of German data, as reported in Plecháč et al. (2024) (see column PT.cz). However, the UDPipe parser seems in general to perform better on Czech data than on German data, as shown by column PDT in Table 4, which displays the evaluation results for UDPipe on the Prague Dependency Treebank (PDT), which consists of Czech newswire texts. Here the results are even higher.

The performance drop between news or web texts and other genres are in line with evaluation results e.g. on aesthetic writings (Salomoni, 2017) or fiction (Jelínek, 2017), not only for parsing but also for POS tagging (Haider, 2021) and other NLP tasks (see the overview in Bamman, 2020).[14]

The boxplot for UAS in Figure 3 shows that there is a very clear downward outlier, with UAS=53.8. This text also yields poor values for the other measures: LAS=50.0 and CLAS=37.8. Figure 4 shows an excerpt of this text with both annotations, the manual (top) and the automatic ones (bottom). The excerpt contains a very long coordination of adjectives, each of which is again modified by a preceding adverb. The head noun (*Poeten* 'poets') was

[14] The Arabic parser used by Al-Ghamdi et al. (2021) (see Section 1) achieves scores of UAS = 81.52 and LAS = 75.25. For the treebank of Spanish Sonnets, Ruiz Fabo et al. (2017) do not provide an evaluation of the automatic parses.

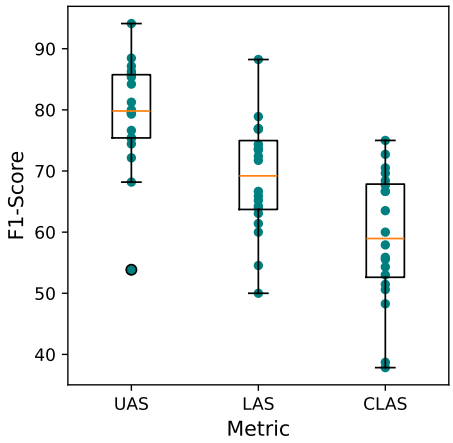


Figure 3: Distribution of UAS, LAS and CLAS scores.

Manual	PoeTree	F1	count
advcl	ccomp	0.39	7
parataxis	conj	0.28	26
obl	nmod	0.28	25
expl	obj	0.15	13
iobj	obj	0.13	11
nmod	obl	0.12	10
iobj	obl	0.11	7
expl	nsubj	0.09	12
root	appos	0.08	9
obl	obj	0.08	10

Table 5: Top ten most often confused dependency labels.

introduced in the previous line and is elided in the next lines. Therefore, in the manual annotation, the first of the adjectives (*begrabenen* 'buried (ones)') is treated as a substitute head, from which all further `conj` relations start. The system annotation, on the other hand, selects the last of the coordinated adjectives (*beglückenden* 'enchanting (ones)') as the head. This results in a large number of mismatches with regard to the structure and thus the low UAS value.[15]

Label confusion We use the F1 score as defined in Dipper et al. (2024) to rank the confusions between labels in a meaningful way, taking into account how often they were confused but also how often they occurred overall.[16] We only report confusions that occurred more than five times.

Table 5 shows the corresponding confusions. For example, seven expressions were manually annotated with `advcl` and automatically with `ccomp`. This distinction in particular also proved to be problematic in the study by Dipper et al. (2024), who annotated Middle High German texts. As we have seen, sentences are often strung together without an overt conjunction. Here it is often difficult to distinguish between pure parataxis and coordination without a conjunction. Other problematic labels concern the arguments of a verb, namely `obj`, `iobj`, `expl`. The confusion with `iobj` is partly due to the different criteria: while in the traditional UD-Treebanks for German `iobj` is only used for ditransitive verbs, the guidelines of Dipper et al. (2024) follow the suggestion of Zeman (2017), according to which `iobj` is used for all dative objects.

5 Conclusion

We presented an annotation study of German-language poetry of different centuries with UD-style syntactic relations. We showed that currently, the automatic analysis of poetry still seems problematic. At all levels – tokenization, sentence boundary determination, tagging, parsing – poems exhibit special properties that lead to faulty analyses. In this paper, however, only parsing was evaluated quantitatively.

One could argue that some or even many of the challenges discussed also apply to other non-standard language data, e.g. word repetition, anacoluthon and elision occur in spoken language, or deviations from standard spelling is typical of social media data. However, poets use these phenomena intentionally and deliberately in their poems, whereas in other data they often occur due to time pressure or attention deficits.

A major problem is that incorrectly determined sentence boundaries tear apart dependency-related phrases. In difficult passages, the correct syntactic relationships often only become clear on closer

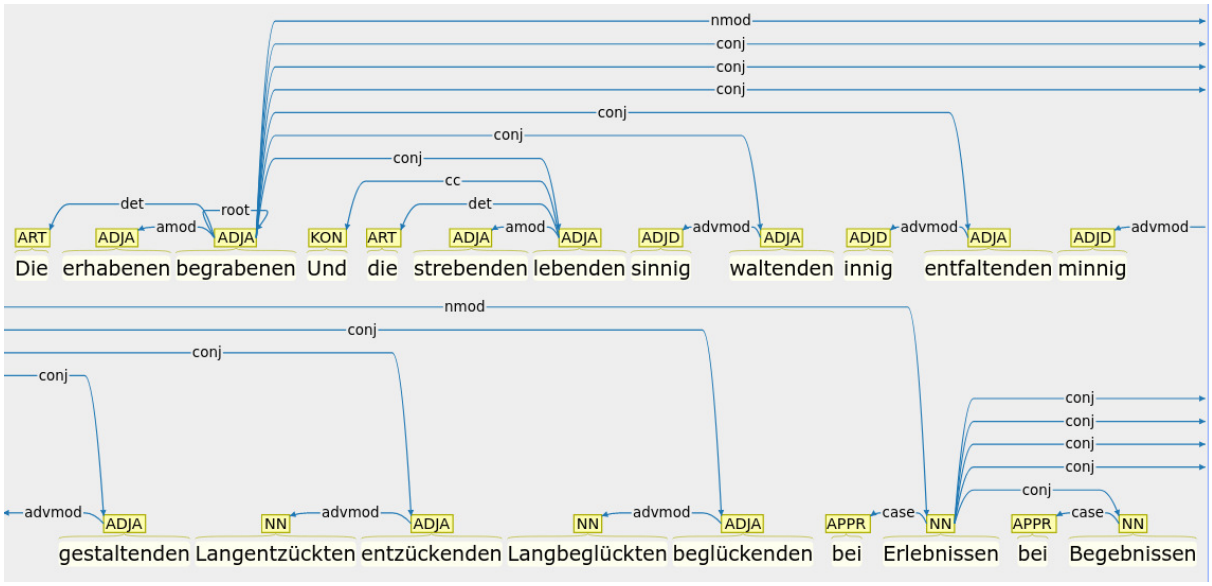
[15] In this excerpt, there are also two incorrect words: the forms *Langentzückten* and *Langbeglückten* – which do not exist in German – should actually be *klangentzückten* 'sound-delighted' and *sangbeglückten* 'song-enchanted'. A translation of the excerpt is: 'The exalted buried ones, and the striving living ones, the sensible ruling ones, the intimately unfolding ones, the lovingly shaping ones, the sound-delighted delightful ones, the song-enchanted enchanting ones, at experiences, at events'.

[16] F1 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 as the manual annotation and a_2 as the system annotation, and l_1, l_2 as the labels annotated by the human annotator and the system, respectively. Possible values are between 0 and 1, where 1 means perfect agreement if $l1 = l2$, and 0 means perfect disagreement if $l1 \neq l2$.

Manual annotation



System annotation

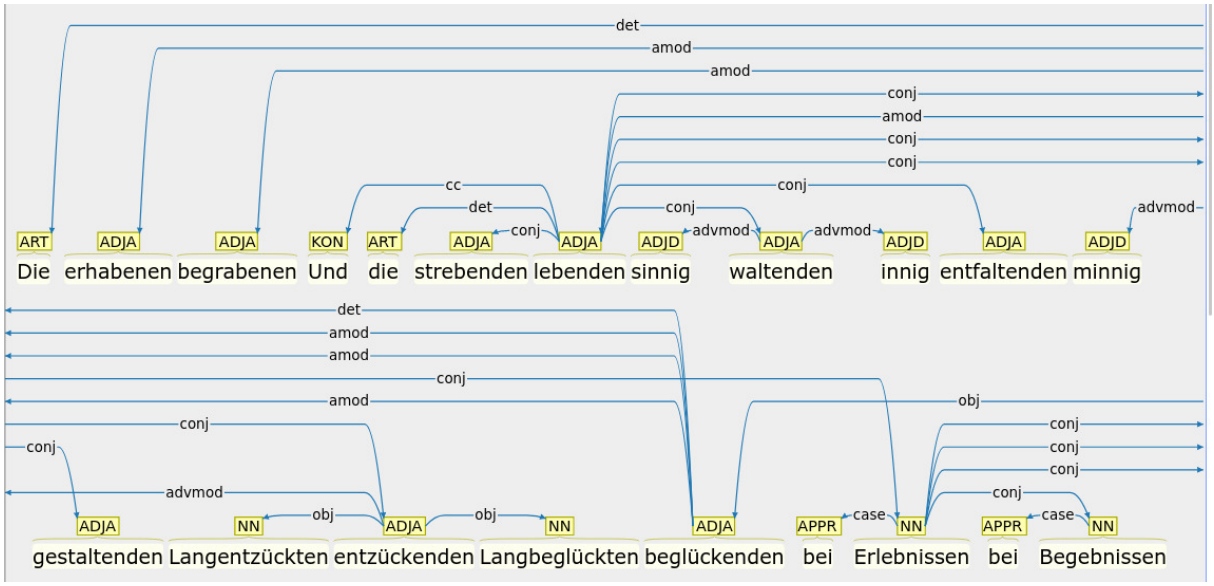


Figure 4: Excerpt of the outlier poem (ID 00-1734-0000-0003-7048-A#0u) with manual (top) and automatic annotations (bottom).

analysis and then often have a major influence on the interpretation. Therefore, we would like to propose that the sentence boundaries should be determined downstream and a dependency analysis should be carried out first, on the basis of which the sentence boundaries would then be determined (which of course would require a major modification of current dependency parsers). Alternatively, both tasks, parsing and sentence boundary detection, could be performed in parallel. In fact, this issue could concern other kinds of non-standard data as well, such as data produced by learners or children or in social media, which often does not adhere to standard punctuation rules.

Overall, we conclude that for studying syntactic properties of poems based on UD annotations, automatic parses are not yet reliable enough. Further manual analyses are necessary, also to provide training data for poetry-specific structures as reviewed in this paper.

Limitations

Our study is limited in that we have only annotated a small pilot corpus. We are planning to extend the data in the future. Furthermore, for the comparison of our manual annotations with automatically created annotations we have only used the automatically created dependency relations that come with the PoeTree.de corpus. The evaluation was end-to-end, i.e. we have not yet been able to evaluate the performance of a dependency parser that is based on, e.g., gold token and sentence boundaries.

Acknowledgments

We are very grateful to the anonymous reviewers for their constructive reviews.

References

Sharefah Al-Ghamdi, Hend Al-Khalifa, and Abdulmalik Al-Salman. 2021. A dependency treebank for classical Arabic poetry. In *Proceedings of the Sixth International Conference on Dependency Linguistics (Depling, SyntaxFest 2021)*, pages 1–9, Sofia, Bulgaria. Association for Computational Linguistics.

David Bamman. 2020. LitBank: Born-literary natural language processing. In Jessica Marie Johnson, David Mimno, and Lauren Tilton, editors, *Computational Humanities, Debates in Digital Humanities*.

David Bamman and Gregory Crane. 2011. The Ancient Greek and Latin dependency treebanks. In Caroline Sporleder, Antal Van Den Bosch, and Kalliopi Zervanou, editors, *Language Technology for Cultural Heritage*, pages 79–98. Springer Berlin Heidelberg, Berlin, Heidelberg.

Klemens Bobenhausen and Benjamin Hammerich. 2015. Métrique littéraire, métrique linguistique et métrique algorithmique de l'allemand mises en jeu dans le programme *Metricalizer*. *Langages*, 199:67–87.

Maria Cieschinger. 2016. *The Contraction of Preposition and Definite Article in German. Semantic and Pragmatic Constraints*. Ph.D. thesis, University of Osnabrück.

Silvie Cinková, Petr Plecháč, and Martin Popel. 2024. Rhymes and syntax: A morpho-syntactic analysis of Czech poetry. *Primerjalna književnost*, 47(2).

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Mirella De Sisto, Laura Hernández-Lorenzo, Javier De la Rosa, Salvador Ros, and Elena González-Blanco. 2024. Understanding poetry using natural language processing tools: a survey. *Digital Scholarship in the Humanities*, 39(2):500–521.

Matthew Timothy DeHass. 2024. Parataxis in Latin colloquial and poetic texts : A treebank-based analysis. Master's thesis, University of Missouri–Columbia.

Eliane Delente and Richard Renault. 2017. Projet Anamètre : présentation, limites et avancées. In *International Conference - Plotting poetry : on mechanically-enhanced reading*, number 7 in (Littératures), pages 73–92, Bâle, Switzerland. Presses Universitaires de Liège.

Stefanie Dipper, Cora Haiber, Anna Maria Schröter, Alexandra Wiemann, and Maike Brinkschulte. 2024. Universal Dependencies: Extensions for modern and historical German. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17101–17111, Torino, Italia.

Stefan Grünewald and Annemarie Friedrich. 2020. Unifying the treatment of preposition-determiner contractions in German Universal Dependencies treebanks. In *Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020)*, pages 94–98, Barcelona, Spain (Online). Association for Computational Linguistics.

Thomas Haider. 2021. Metrical tagging in the wild: Building and annotating poetry corpora with rhythmic features. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3715–3725. Association for Computational Linguistics.

Thomas Haider. 2023. *A computational stylistics of poetry: distant reading and modeling of German and English verse*. Ph.D. thesis, University of Stuttgart.

Hussein Hussein, Burkhard Meyer-Sickendiek, and Timo Baumann. 2018. Automatic detection of enjambment in German readout poetry. In *Proceedings of Speech Prosody, 2018, Poznán*, pages 329–333.

Tomáš Jelínek. 2017. FicTree: A manually annotated treebank of Czech fiction. In *Proceedings of the 17th Conference on Information Technologies - Applications and Theory (ITAT 2017)*, pages 181–185.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, 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.

John Lee and Yin Hei Kong. 2012. A dependency treebank of classical Chinese poems. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 191–199, Montréal, Canada. Association for Computational Linguistics.

Teresa Paccosi, Alessio Palmero Aprosio, and Sara Tonelli. 2023. Adding a novel Italian treebank of marked constructions to Universal Dependencies. *Italian Journal of Computational Linguistics*, 9(1).

Susan Pintzuk and Plug Leendert. 2001. The york-helsinki parsed corpus of old english poetry (YCOEP). Oxford Text Archive.

Petr Plecháč, Silvie Cinková, Robert Kolár, Artjoms Šeļa, Mirella De Sisto, Lara Nugues, Thomas Haider, and Neža Kočnik. 2024. PoeTree: Poetry treebanks in Czech, English, French, German, Hungarian, Italian, Portuguese, Russian, Slovenian and Spanish. *Research Data Journal for the Humanities and Social Sciences*, pages 1–17.

Petr Plecháč, Robert Kolár, Silvie Cinková, Artjoms Šeļa, Mirella De Sisto, Lara Nugues, Thomas Haider, N. Kočnik, Benjamin Nagy, Éliane Delente, Richard Renault, Klemens Bobenhausen, Benjamin Hammerich, Adiel Mittmann, Gábor Palkó, Péter Horváth, Borja Navarro Colorado, Pablo Ruiz Fabo, Helena Bermúdez Sabel, Kirill Korchagin, Vladimir Plungian, and Dmitri Sitchinava. 2023. PoeTree. Poetry Treebanks in Czech, English, French, German, Hungarian, Italian, Portuguese, Russian and Spanish.

Pablo Ruiz Fabo, Clara Martínez Cantón, Thierry Poibeau, and Elena González-Blanco. 2017. Enjambment detection in a large diachronic corpus of Spanish sonnets. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 27–32, Vancouver, Canada. Association for Computational Linguistics.

Alessio Salomoni. 2017. Toward a treebank collecting German aesthetic writings of the late 18th century. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it)*, pages 292–297.

Alessio Salomoni. 2021. *A UD Literary Treebank for German*. PhD Thesis, Publisher: Università degli studi di Bergamo.

Anne Schiller, Simone Teufel, Christine Stöckert, and Christine Thielen. 1999. Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). Technical report, Universitäten Stuttgart und Tübingen, `http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-1999.pdf`.

Djamé Seddah, Farah Essaidi, Amal Fethi, Matthieu Futeral, Benjamin Muller, Pedro Javier Ortiz Suárez, Benoît Sagot, and Abhishek Srivastava. 2020. Building a user-generated content North-African Arabizi treebank: Tackling hell. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1139–1150, Online. Association for Computational Linguistics.

Milan Straka. 2018. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207, Brussels, Belgium. Association for Computational Linguistics.

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.

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.

A Appendix

PoeTree ID	Title	Author	Born	Created	Publ.	#Sents	#Tokens	#Stanzas	#Lines
dta.poem.18792	Mädchen .	Goethe, Johann Wolfgang von	1749	-	1819	2	25	1	4
dta.poem.10409	XxXIV.	Heine, Heinrich	1797	-	1827	2	32	1	4
dta.poem.10555	ViI. Nachts in der Cajüte .	Heine, Heinrich	1797	-	1827	4	76	3	12
dta.poem.16173	227.	Rückert, Friedrich	1788	-	1837	6	132	6	12
dta.poem.3030	76.	Rückert, Friedrich	1788	-	1838	7	115	6	12
dta.poem.21583	Märznacht . 1884.	[anonymous]	-	-	1885	14	319	13	52
00-1734-0000-0003-28BA-1#0	135.	Ḥāfeẓ, Šams o'd-din Moḥammad[1]	1325	1357	1858	21	268	20	40
00-1734-0000-0003-2E39-A#0	10.	Ḥāfeẓ, Šams o'd-din Moḥammad[1]	1325	1357	1858	17	209	18	36
00-1734-0000-0003-F9D5-D#0	92.Schläge	Logau, Friedrich von	1604	1630	1872	4	70	1	6
00-1734-0000-0004-01CB-A#0	41.Grabschrifft einer Buhlerin	Logau, Friedrich von	1604	1630	1872	2	24	1	2
00-1734-0000-0004-05D3-0#0	3.Franckenthal Friedens-Hindernüß	Logau, Friedrich von	1604	1630	1872	2	25	1	2
00-1734-0000-0003-576E-4#0	7. Was da braust	Herder, Johann Gottfried	1744	1773	1879	3	34	1	4
00-1734-0000-0005-8B68-7#0	MinenteFußnoten	Waiblinger, Wilhelm	1804	1817	1893	1	28	1	2
00-1734-0000-0003-7048-A#0	Breslauer Schillerfest	Hoffmann von Fallersleben, August H.	1798	1819	1841	2	65	1	17
00-1734-0000-0004-A660-A#0	Ausdruck der Empfindung	Rückert, Friedrich	1798	1827	1897	2	47	2	8
00-1734-0000-0002-EE1F-6#0	Begabung	Grillparzer, Franz	1791	1831	1960	1	20	1	4
00-1734-0000-0002-B719-B#0	Geduld!	Geibel, Emanuel	1815	1833	1918	15	184	6	24
00-1734-0000-0003-6F5D-5#0	Die Schlittenfahrt mit dem Schneemann	Hoffmann von Fallersleben, August H.	1798	1836	1976	9	97	6	12
00-1734-0000-0002-9E01-9#0	Nacht und Morgen	Eichrodt, Ludwig	1827	1841	1856	9	137	3	21
00-1734-0000-0002-577F-0#0	Es ist so still geworden ...	Conradi, Hermann	1862	1876	1911	23	255	5	40

Table 6: Overview of annotated poems. Titles are reported as specified in the JSON data of PoeTree.de. The columns 'Born', 'Created' and 'Publ.' refer to the year in which the author was born, the year the poem was created and the year it was published, respectively.

[1] Translated into German by Vincenz Ritter v. Rosenzweig- Schwannau, 1858.