Building the AURIS Corpus of Reference and Information Structure

Christian Chiarcos, Christian Fäth, Tabea Gröger, Quentin Alastair Frey
Applied Computational Linguistics (ACoLi)
University of Augsburg, Germany
{christian.chiarcos, christian.faeth, tabea.groeger, quentin.frey}@uni-a.de

Abstract

We present AURIS, the *Augsburg corpus for Reference and Information Structure*, a multilingual corpus annotated for reference, discourse relations, and aspects of information structure. AURIS introduces an innovative use of off-the-shelf spreadsheet software for complex annotation tasks, reducing technical barriers and dependencies common in discourse annotation. Designed for classroom use, it enables linguistics and philology students to explore diverse theoretical frameworks while working in their language of choice. The paper focuses on technical design and workflows that integrate and generate pre-annotations from heterogeneous sources. Despite its low-tech approach, AURIS aligns with established standards and remains interoperable with existing projects. Preprocessing scripts support multiple languages, with an initial annotation round on German texts evaluated against TED-MDB and ParCorFull data converted into AURIS formats. This approach demonstrates that accessible tools can yield high-quality, replicable annotations for discourse and information-structure research.

Keywords: coreference, discourse, spreadsheet-based annotation

1. Background and Motivation

Discourse phenomena play a crucial role in applied linguistics (e.g., translation), theoretical linguistics (e.g., cognitive and functional grammar), and in technical contexts (e.g., natural language understanding), but despite their importance to both research and technology, corpora annotated for multiple discourse dimensions remain rare. Corpora that combine both coreference and discourse annotations include early campaigns addressing these layers independently on partially overlapping texts of the Wall Street Journal (Hirschman and Chinchor, 1998; Hovy et al., 2006; Rösiger, 2018; Carlson et al., 2003; Wolf and Gibson, 2005; Prasad et al., 2017), and modern multi-layer corpora such as the GUM Corpus for English (Zeldes, 2017), the Potsdam Commentary Corpus for German (Bourgonje and Stede, 2020), and extensions of the Prague Dependency Treebank for Czech (Nedoluzhko et al., 2016; Synková et al., 2024). What is characteristic of these annotation projects is that each involves several specialized tools for manual annotation, operating with different formats, and that these formats are later merged in complex integration steps. For example, coreference in the Potsdam Commentary Corpus was annotated with MMAX2 (Müller and Strube, 2006) – a special-purpose tool for the annotation of standoff relations applied over a text –, while three separate tools were used for discourse annotation: the classical RSTTool (Stede, 2004; O'Donnell, 2000), the in-house tool Connanno, and the PDTB Annotator (Bourgonje and Stede, 2020). These layers, together with morphosyntactic annotations, were converted into the unified yet intricate corpus format PAULA (Chiarcos et al., 2008) and published via ANNIS (Krause, 2019), a well-known corpus management tool specifically designed for multi-layer corpora. Similarly, in the GUM corpus, coreference is annotated using WebAnno (Yimam et al., 2013), respectively its successor INCEpTION (De Castilho et al., 2024), and discourse structure via rstWeb, a server-based clone of the RSTTool (Zeldes, 2016).

However, combining outputs from multiple tools poses technical challenges. Corpus formats need to be sufficiently generic (i.e., complex) to integrate heterogeneous layers without loss. While graph technologies have proven effective (Ide and Suderman, 2007; Chiarcos et al., 2008; Krause, 2019), conversion and consolidation remain complex (Chiarcos et al., 2012). Tool-specific segmentation or content edits (e.g., in tokenization or whitespace handling) can further complicate integration.

In the context of developing new courses in Computational Linguistics and Digital Humanities for students of linguistics and philologies, we designed a workflow for building a multilingual (parallel) corpus annotated for coreference, discourse structure, and information structure. The workflow emphasizes: (a) minimal technical barriers, (b) static and dynamic pre-annotation, (c) simple, well-supported formats, and (d) fast, keyboard-only annotation.

The key idea is to conduct all annotation stages with standard spreadsheet software. While this may seem unconventional given the sophistication of existing tools, we argue that this approach – when supported by tailored guidelines – proves both practical and advantageous. It enables rule-based dynamic pre-annotation, where embedded formulas automatically propose candidate annotations de-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 2035–2047
11–16 May 2026.

rived from prior manual input. This can complement the functionality of specialized tools such as INCePTION, especially for lightweight, rule-based interactive defaults in classroom settings. However, this setup also requires annotation schemes compatible with a tabular model, where each row receives at most one annotation per column. For coreference, the use of one-word-per-line TSV/CSV formats is already standard (e.g., in CorefUD, Nedoluzhko et al., 2022), for discourse structure, however, this is less established and segmentation and labeling require reconsideration.

This paper introduces the AURIS corpus, with a focus on corpus design, creation and technical workflows. Pilot studies with students have been conducted over the course of two years, and primarily served to refine annotation guidelines. In preparation of this publication, this was expanded into a coordinated annotation effort conducted by the authors and student assistants, currently focusing on (modern) German.

2. Corpus Design

AURIS is designed to build on and complement CoNLL-U corpora, and uses a custom CoNLL format as its basis. Moreover, it is designed to be used with conventional spreadsheet software. Complementary scripts perform static pre-annotation for many languages, as well as formulas embedded in the annotation spreadsheets to perform dynamic, interactive pre-annotation during annotation. Spreadsheets are treated as an annotation *interface*; the released corpus is distributed in open, repository-friendly formats (TSV/CoNLL-U/CorefUD/PAULA).

2.1. Source Material

The corpus is designed to be multilingual (parallel), but also to be grounded in the state of the art. For this reason, the initial sampling of corpus files was guided by the availability of comparable annotations, as these could be used for external evaluation. The initial release of AURIS (Tab. 1) uses the German data of the Multilingual Discourse Bank (Zeyrek et al., 2020, TED-MDB, with external annotations for discourse) and the news section of ParCorFull (Lapshinova-Koltunski et al., 2018). An added value of AURIS is to provide annotations for discourse and anaphora, whereas TED-MDB and ParCorFull only provided either.

TED-MDB is a collection of TED talks annotated in accordance with PDTB v.3 for English, German, Lithuanian, Polish, Portuguese, Russian and Turkish. Our discourse guidelines are grounded in PDTB, as well (Sect. 4), so, this data is closely comparable and can be automatically converted to our format. However, note that this data is used for evaluation only, and our annotations have been created independently. ParCorFull provides coreference annotation for news articles from the WMT17 dataset (Bojar et al., 2017) (and a different set of TED talks, not considered here), and its converted data is used for evaluation. Again, these texts have been independently annotated for coreference (as well as for discourse).

Along with the publication of this paper, the annotated German section of both these components of our corpus is made available in several formats, including its native XSLX and TSV formats, CorefUD and, for subsequent querying and visualization with the corpus management tools, PAULA (Chiarcos et al., 2008). In addition to ParCorFull and TED-MDB data, AURIS also features a selection of (excerpts from) novels, fairy tales and Bible excerpts in multiple languages, but as these have been partially annotated only, so far, they are not part of the initial release. Although not necessarily representative of common language, Bible excerpts have been added so that we can offer parallel texts to students of medieval studies and historical philologies. Also, coreference annotations over this data may potentially be evaluated against the English OntoNotes corpus (Hovy et al., 2006).

2.2. Spreadsheet Software

We are aware that many people use office software for annotation tasks, but also that this is generally considered bad practice in our field – and rightly so, because it has the potential to create *massive* issues with data normalization. Overall, a tool that permits free-text edits is a nightmare for post-processing. Moreover, spreadsheet software is not necessarily convenient for annotation tasks in general; it is well-suited for simple labelling tasks, but limited in that it is restricted to 1:1 labelling so that you can only annotate elements that are stored in a single cell (row). (Multi-cell annotation by merging cells is possible, but not very convenient.)

We suggest to circumvent both problems by (a) performing head-based annotation (annotate one row only, i.e., having one row represent the entire markable – in discourse annotation – or, limiting annotation to the syntactic head as identified in automated pre-processing), and (b) using cell/column protection to prohibit edits beyond annotation cells. Neither of these techniques can be perfect, of course, but possible downsides may be compensated by some additional advantages:

- Annotators can use a wide range of spreadsheet software across platforms, including MS Office, LibreOffice, and browser-based tools like Google Sheets.

	total	TED-MDB, German		ParCorFull (News, German)	
		AURIS	from original	AURIS	from original
tokens / sentences	18720 / 960	7773 / 406		10947 / 554	
referring expressions	3485	1538	n/a	1947	739
discourse					
explicit markers	256	122	n/a	134	n/a
implicit markers	669	274	(English)	395	n/a
relations	925	396	285	529	n/a

Table 1: Statistics for AURIS (German) in comparison to (the AURIS conversion of) its source corpora

- Annotation can be efficiently performed via keyboard-only input with autocompletion and shortcuts; for instance, in LibreOffice, `CTRL + Arrow` jumps to the next annotated cell.
- Modern spreadsheets share a uniform formula system that enables advanced lookup, aggregation, and data manipulation despite lacking full programming capabilities.
- Annotated data can be directly analyzed using built-in tools; contingency tables combining manual and automatic features can be processed with functions such as `CHISQ.TEST` (χ^2) or `CORREL` (Pearson's r).
- Established libraries support creation and conversion of spreadsheets, including support for pre-filled cells, color coding, cell protection, and formulas to generate default values during annotation.

2.3. Data processing

Our spreadsheets require complex pre-annotation routines. We use the following key components:

- CoNLL-Merge for merging TSV files, with automatic resolution of segmentation and spelling differences (Chiarcos and Schenk, 2019).
- Fintan/CoNLL-RDF (Fäth et al., 2020) to access and process CoNLL data streams as RDF graphs, providing fast, parallelized and flexible lookup, transformation and enrichment with SPARQL.
- XlsxWriter[1] for creating formatted Excel spreadsheets.

Dynamic, rule-based pre-annotation is implemented through formulas that predict overridable default annotations based on previous input and dynamically mark – with exclamation marks – rows requiring annotation according to the type of referring expression in the static pre-annotation. Moreover, if a definition cannot be replicated exactly, its predicted annotation is marked with a question mark. Once no exclamation or question marks remain a file is considered fully annotated.

A technologically innovative element is the use of SPARQL for extraction and pre-annotation tasks. With conventional RDF technology, this approach would not scale, but Fintan's stream-based approach allows retrieval and transformation with SPARQL *in-memory* and *on the fly*, enabling full interoperability with Semantic Web standards while still being able to be directly be integrated into conventional transformation workflows with conventional corpus formats and/or tabular data. In addition to being able to perform complex pattern matching, we would like to highlight three key features of SPARQL used in our workflow: the use of property paths for transitive search, support for (implicit and explicit) $\forall$ quantification (e.g., for querying for the non-existence of a certain feature), and aggregation over the graph in nested SELECT statements. Another feature of SPARQL is to access remote data in federated search, if provided as Linked Data/RDF. Normally, this is done at query time (i.e., for every SPARQL invocation), but in order to speed up transformation, Fintan also allows to pre-load external datasets, and can use these for lookup-based annotation. So, loading an external model is performed once only, and then merged with every segment to be processed.

The preprocessing workflow is summarized in Fig. 1: Given a paragraph-separated text file in one of the supported languages (currently, German, English, French, Italian, Portuguese and Russian), we perform syntactic parsing using UDpipe (Straka et al., 2016, with UD 2.5 models), and with one set of SPARQL scripts, this is converted into a TSV file with pre-annotation for coreference. With another set of SPARQL scripts, and using discourse marker inventories available in RDF for many languages (Chiarcos and Ionov, 2021; Chiarcos, 2022), we produce pre-annotations for discourse annotation, serialized in TSV, with one utterance per line (currently corresponding to one CoNLL-U sentence). Note that both transformation tasks require complex pattern matching over sentence graphs and cannot be directly performed by mere lookup, and we would argue that SPARQL may have some entry bias, but as a W3C standard,

[1] https://github.com/jmcnamara/XlsxWriter

it is more portable, more transparent and more sustainable than an ad-hoc script. The resulting TSV files are merged into an Excel file containing two spreadsheets with (a) source data from CoNLL-U, (b) static pre-annotations generated via SPARQL, and (c) interactive pre-annotation formulas implemented in Excel. Using nested `IF` statements, these formulas perform lookups and update dynamically in response to annotator decisions. Dynamic pre-annotation flags every (head of a) potential referring expression (e.g., definite, demonstrative or possessive NPs, proper names, or pronouns) by inserting the default value `!!!` in the `COREF` column, implemented through equality checks over the `NP_TYPE` column.

Annotators are provided with Excel files for use with any modern spreadsheet software of their choice (MS Excel, LibreOffice, Google Spreadsheets, etc.). Note that they can only edit the cells that should be annotated, and they cannot insert or split rows (as long as cell protection is not actively disabled), so that tokenization remains intact.

Once annotation is completed, reviewed and confirmed, the annotated file is deposited in an export folder, both spreadsheets are merged (with CoNLL-Merge) with a CoNLL-U file into an in-memory representation that connects reference, discourse and syntax annotations, and automatically converted into several output formats, i.e., several TSV formats with partial, complementary annotation for discourse (TSV with one utterance per line) or coreference (TSV or CorefUD files), as well as into PAULA XML (Chiarcos et al., 2008). Unlike TSV formats, PAULA XML allows to preserve AURIS annotations in a lossless multi-layer representation. It is then used to publish the annotations via the corpus management system ANNIS (Krause, 2019).

For evaluation, we convert ParCorFull and TED-MDB annotations into AURIS formats in a similar way.

3. Reference and Anaphora

The AURIS coreference guidelines are based on PoCoS (Krasavina and Chiarcos, 2007), with the following modifications: Similar to CorefUD (Nedoluzhko et al., 2022), our annotation is designed as an *additional annotation layer* on top of CoNLL-U dependency syntax. The PoCoS *maximum span* principle is thus abandoned in favor of annotating *only the syntactic head* of each markable. Because dependency syntax is not easily accessible via Excel, the extraction of markables is relegated to (static) pre-annotation, where every (head of a) referring expression is automatically assigned its `NP_TYPE`. Only these elements are to be annotated. Annotators are instructed to delete incorrectly pre-annotated referring expressions and to document this in a specific `COMMENT` column as these may be indicative of either a parsing error or an error in the SPARQL script used for extraction. Similarly, for adding referring expressions missed by pre-annotation.

The spreadsheet for coreference annotation uses different colors for columns where annotation should take place (Fig. 4, see Appendix). Columns without color are not to be annotated and are protected. The first column features the word, followed by several columns with static pre-annotation (including a number of hidden columns whose values are not shown, but evaluated by Excel formulas as part of dynamic pre-annotation). Reference annotation according to the PoCoS guidelines operates on three key columns: **COREF**, **REF**, and **COMMENT**:

- **COREF**: user-defined abbreviation for the discourse referent, consistent across the entire coreference chain. If an expression is non-referential, the field remains empty (not "_").
- **REF**: referentiality type (e.g., `OLD`, `NEW`, `GROUP`, `CAT`, `BOUND`, `SIT`, `GEN`, `EXPL`, `PRED`, `IDIOM`, `other`). This classification extends PoCoS by integrating detailed categories of referring expressions and non-referential forms.
- **COMMENT**: optional notes for uncertainty, ambiguous antecedents, or design decisions; ambiguity can be explicitly marked (e.g., `AMBIG:COREF(...)`).

PoCoS defines a taxonomy of referring expressions, which is automatically identified from syntax annotation in pre-annotation and stored in the `NP_TYPE` column. Like the `WORD` column, this column is protected and can thus not be changed (but commented on in the `COMMENT` column) by the annotator. From the `NP_TYPE`, the `COREF` column is pre-populated (dynamically annotated) with `!!!` for every (head of a) referring expression that *could* be an anaphor. These, together with their antecedents, are to be annotated. By restricting annotation to syntactic heads and using a compact, spreadsheet-friendly representation, the approach streamlines annotation while retaining the semantic precision of PoCoS. This adaptation thus supports multi-layer discourse annotation over existing syntactic corpora without redundancy.

In the `COREF` column, coreference is annotated by assigning every referring expression either a new or a previously used identifier. There are no constraints as to naming conventions for this identifier, but it is recommended that it can be easily memorized and typed. Automated pre-annotation only detects nominal expressions, but the guidelines also require to annotate events as antecedents if these are subsequently referred to with a nominal

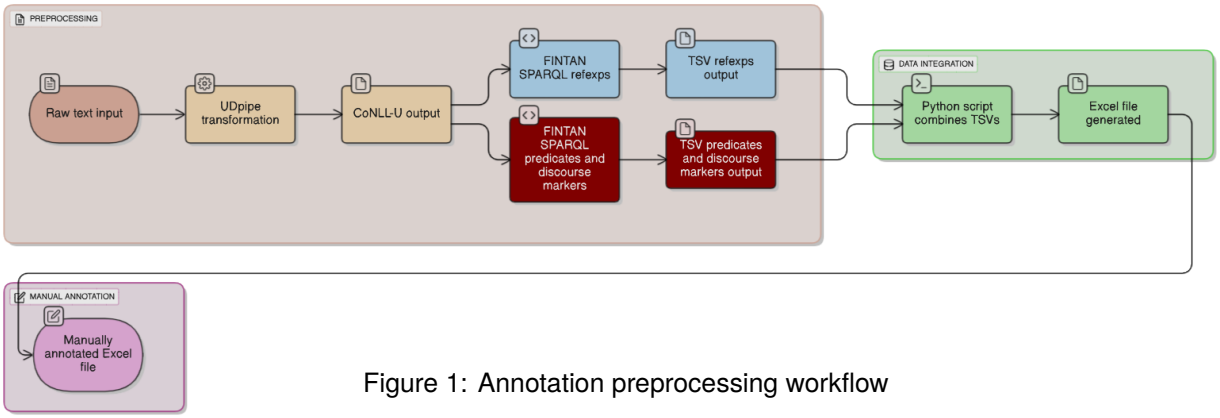

Figure 1: Annotation preprocessing workflow

or pronominal expression. For these, also the syntactic head is to be annotated, i.e., the syntactic head of the clause or sentence that contains the antecedent. If these are coordinated main clauses, the (syntactic head of the) first main clause is to be annotated (in accordance with the treatment of coordination in the Universal Dependencies). If the antecedent is not conveyed in a single sentence, but in multiple sentences, the last sentence is to be annotated as antecedent. In such cases, as well as in cases in which annotators feel inconfident or perceive ambiguities, they are instructed to leave a comment in the `COMMENT` column.

For the `REF` column, the scheme classifies referring expressions depending on their anchoring in discourse. If they are previously mentioned, they are `OLD`, if they are not, but physically present, they are `SIT`(uationally evoked), and if they are neither, they are annotated as `NEW`. As soon as a `COREF` id is provided in a row, dynamic pre-annotation suggests candidate values for `REF` based on a `VLOOKUP` for a previous mention (`?OLD`, `?NEW`), but annotators must confirm or correct them.

With AURIS, we provide a new layer of coreference annotation over the news section of ParCorFull (German) and over TED-MDB, performed by one annotator. In addition to that, experimental annotations have been applied to other texts in German, English, Italian and French, but so far, these have not been paired with discourse annotations and are not included in the initial release of the AURIS corpus.

For evaluation, we compared our release data for the news section of the German ParCorFull corpus with ParCorFull annotations, which we converted from its original MMAX2 source (not the CorefUD edition derived from it) by (a) converting MMAX2 to a TSV representation using a generic MMAX2 converter we developed in-house, (b) applying CoNLL-Merge to align it with the AURIS CoNLL-U parses, (c) identifying and classifying markables with the AURIS pipeline, and (d) exporting the resulting AURIS data to CorefUD (and other formats). We ran the official CorefUD scorer for the CRAC Shared Tasks on Multilingual Coreference Resolution,[2] using (our CorefUD exports of) ParCorFull and AURIS as key and system respectively. Note that we evaluate only for the German news files of ParCorFull, so, our scores are not directly comparable to those reported in CRAC shared tasks.

Table 2 summarizes the evaluation results. Overall, we found that recall exceeded precision substantially, with recall scores ranking from 43.71% (ceafe) to 72.97% (mor). These numbers are relatively low, but it is to be noted that this is an evaluation across different annotation schemes. The internal evaluation for inter-annotator agreement between individual annotators working with AURIS guidelines is still in preparation, and expected to yield better results. A qualitative evaluation of the data in comparison with ParCorFull annotations revealed that their deviations can be clearly traced back to different annotation philosophies. Whereas AURIS aims for a complete annotation, ParCorFull annotation (in the MMAX release, at least) seems to be focused on selected, specific referents. Furthermore, the criteria for coreference seem to be different. Whereas ParCorFull aims to annotate entity identity, the primary annotation criterion in AURIS (as inherited from the PoCoS schema) is substituability, i.e., contextual equivalence. In AURIS and PoCoS, the primary criterion is that a referring expression can be added to a chain if and only if it can be substituted with every other referring expression in the same chain without a change of meaning in the textual context. This means that certain cases of metonymy are covered by AURIS coreference, but not necessarily by ParCorFull coreference.

Along with coreference, the spreadsheet also con-

[2]See https://ufal.mff.cuni.cz/corefud/crac25.

Metric	Recall	Precision	F1
muc	69.81	31.65	43.56
bcub	61.83	21.94	32.38
ceafe	43.71	23.05	30.18
ceafm	67.37	31.69	43.10
blanc	61.91	23.64	33.32
lea	57.59	19.59	29.23
mor	72.97	34.32	46.68

Table 2: Coreference evaluation against ParCorFull, German news texts

tains the columns IS for information status, and CB for the backward-looking center. We briefly report on these columns because they serve to illustrate the role of dynamic pre-annotation. However, they have not been evaluated against external data, nor for either inter-annotator agreement or any systematic bias that might arise as a result of dynamic pre-annotation.

For IS, we apply the annotation guidelines of Gundel et al. (2006) and the theoretical framework of the Givenness Hierarchy (Gundel et al., 1993), because (a) these have been developed for and applied to many, typologically diverse languages, (b) the coding manual is short and formalized in a way that it can also be annotated by laypersons, and (c) the criteria formulated by Gundel et al. (2006) are organized in a decision tree and, to a large extent, they can be automatically checked if coreference annotation is already present in a text and the text contains static pre-annotations for grammatical roles, cf. Chiarcos (2025). These rules are implemented as Excel formulas using the VLOOKUP operation to evaluate the preceding and following context, so that (approximative, ?-marked) IS labels can be provided by dynamic pre-annotation. As an example, the label IN_FOCUS is to be given to a referring expression (i.e., a word/row with a non-empty value for COREF), if the referent (i.e., the identifier in the COREF column) is used earlier in the same utterance. If not, annotate IN_FOCUS if it is used in both directly preceding utterances. If not, annotate IN_FOCUS if it was the subject (from static pre-annotation for grammatical roles, in the GR column of the preceding utterance, etc.

The column CB is used to define the 'backward-looking center' of every utterance in accordance with Centering Theory (Grosz et al., 1995), i.e., the referring expression that represents the most salient referent per utterance. We employ CB as a technical operationalization of the notion of topic, i.e., the referent that an utterance is construed about (Lambrecht, 1994), because its criteria can be automatically checked, to a large extent. Salience in Centering is defined by grammatical roles (and the syntactic embedding) of the antecedent, and the salience ranking is restricted to referents mentioned in the preceding utterance. If no referent of the current utterance has been mentioned in the preceding utterance, we predict no backward-looking center. Using (manual) COREF annotation, static pre-annotations for grammatical roles and the depth of syntactic embedding (column GR) and an aggregation function to count the number of sentence boundaries (empty rows, Excel COUNTBLANK) between the current expression and its last mention, these criteria can be automatically checked: We predict the value CB for the subject referent of the (first main clause of the) preceding utterance (which – if it occurs – is always highest-ranking), and (if no CB can be established for the sentence) ?CB for every referent mentioned in the preceding utterance as a candidate CB from which the annotator can choose.

4. Discourse

For discourse annotation, we focus on the annotation of intersentential discourse relations and a style of shallow discourse annotation as also adopted by the Penn Discourse Treebank (Prasad et al., 2008). For segmentation, we thus rely on CoNLL-U sentence boundaries. The AURIS discourse annotation scheme implements ISO SemAF DR-Core (Bunt and Prasad, 2016), complemented with background information and examples drawn from PDTB v.2 (Prasad et al., 2008). Using SemAF ensures interoperability not only with corpora with existing PDTB-style annotations, but also other forms of discourse annotation, as Bunt and Prasad (2016) defines a cross-theoretical mapping between relations from different theories and frameworks.

All annotation is carried out in spreadsheet software with static pre-annotations (see Fig. 5 in Appendix), ensuring accessibility and minimizing technical overhead for annotators. In order to provide an overview over the global context, annotators operate with a separate sheet (within a single Excel file) where one row represent sentences (utterances) rather than words. For each sentence, annotators specify one MARKER (the discourse cue syntactically associated with the current sentence), a RELATION (the discourse relation signaled by this marker), a TARGET (the numerical ID of the related sentence), and an optional COMMENT. We apply the original PDTB v.2 distinction between internal and external arguments, in that the internal argument syntactically associated with a discourse marker (the 'anchor') is the current row, whereas the external argument (the 'target') is identified by its row number.

The relation inventory adopts the organization of PDTB v.2 relation into four top-level categories **ADVERSATIVE** (for PDTB v.2 COMPARISON, corresponds to PDTB v.3 COMPARISON without the

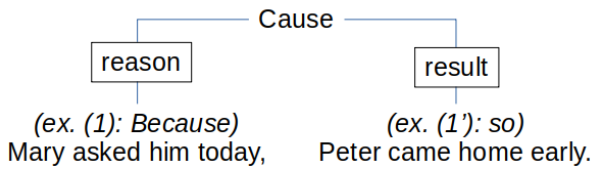

Figure 2: ISO SemAF analysis for a **Cause** relation with two different argument roles

Similarity relation), **CONTINGENCY**, **TEMPORAL**, and **EXPANSION**. These are, however, only used to structure the documentation, not actually to be annotated. In addition, SemAF **DIALOG** relations have been added as a separate top-level category. To account for the PDTB v.3 relation Hypophora (for rhetorical questions in monologs, which does not have a counterpart in SemAF), Hypophora has also been added as an additional EXPANSION relation. We also follow PDTB in applying a *structural differentiation* between arguments, whereas SemAF provides a *functional differentiation* in that (in asymmetric relations) one argument receives one role and the other the other role. As an example, the **Cause** relation in the **CONTINGENCY** cluster has the argument roles **reason** and **result**. In our guidelines, we annotate the argument role of the internal argument, only, because the SemAF relation and the role of the external argument can be reliably inferred. Consider the following examples:

(1) *[Because] Mary asked him today,*
(2) *Peter came home early.*

We annotate segment (1) as the internal argument for marker *because*, with the relation **reason** and segment (2) as target. Because **reason** can only occur as an argument role of SemAF **Cause**, this entails that this relation holds and that the role of the target is **result**. But consider (1') and (2'):

(1') *Mary asked him today,*
(2') *[so] Peter came home early.*

Here, we annotate (2') with marker *so* and relation **result**. Internal argument and external argument are switched, but the inferred **Cause** relation remains identical. SemAF argument roles thus introduce a third layer in the relation taxonomy: For asymmetric relations, each argument role is added as a sub-relation, the corresponding SemAF relation serves as the parent in the taxonomy and can be inferred. The corresponding ISO SemAF analysis (for both examples) is illustrated in Fig. 2.

Discourse pre-annotation involves two components, the pre-annotation of discourse markers, and the identification of the main predicate (provided to annotators as a supportive feature to better narrow down the main statement of each utterance). Discourse marker pre-annotation is implemented in SPARQL with Fintan, using language-specific discourse marker inventories (Chiarcos and Ionov, 2021) as external knowledge component. Predicate identification is implemented in SPARQL over CoNLL sentence graphs, where we traverse the graph to find (the predicate of) the structurally most prominent embedded statement, if the syntactic root is an attribution verb or the nominal/adjectival head of a copular clause.

We evaluate against the German section of TED-MDB, but although we could rely on an established mapping from PDTB/TED-MDB to SemAF/AURIS relations (Bunt and Prasad, 2016), the conversion is not trivial, e.g., TED-MDB allows annotations between arbitrary spans, but AURIS annotations only hold between pre-defined utterances. We thus used CoNLL-Merge to enforce the AURIS utterance splits onto TED-MDB and looked only into intersentential relations, i.e., TED-MDB annotations connecting (elements of) two distinct AURIS utterances. Where a TED-MDB span overlapped with more than one AURIS utterance, we identified span with the AURIS utterance with greater overlap. In five cases, a single resulting utterance had relations with more than one target, for which we selected the closest preceding target as the preferred annotation. This reflects an annotation instruction in AURIS.

Following Zeyrek et al. (2020), we calculated the agreement for whether or not the same spans were connected, but – as TED-MDB comes without the requirement to annotate all utterances – in two different versions, (a) for agreement over all annotated spans, we reached an accuracy of 56.82%, and (b) for agreement over only spans also annotated in TED-MDB, we reached 78.95%. The setting and the numbers are not directly comparable, and they also do not report accuracy, but the f-score for the discourse relation spotting between two rounds of annotation (Zeyrek et al., 2020, Tab.4) seem to be at a similar level, ranging from 70% (Russian) to 88% (Polish).

For cases in which the same pair of utterances has been annotated in AURIS and TED-MDB, we also evaluated the identification of AURIS discourse relations, cf. Zeyrek et al. (2020). The corpus is too small to draw definite conclusions, but overall, 54.67% of intersentential relations in AURIS and the converted TED-MDB corpus are identical, and it seems that errors are mostly concentrated in the 'long tail': Out of 45 occurrencies of relations with relation frequency lower than 10 in the converted German TED-MDB, only 6 (13.3%) were annotated in the same way in AURIS. On the other hand, for higher-frequency relations, the agreement reported in Tab. 3 reaches reasonable 65.0% (117/180). We consider this a good number in the light of an increased level of noise expected from adopting a different annotation workflow and annotation phi-

Relation	TED-MDB	TP	P	R	F1	Top Confusions
specific	34	29	0.85	0.51	0.64	conjunction (3), reason (1), instance (1)
conjunction	32	25	0.78	0.51	0.62	specific (4), similarity (1), result (1)
result	22	15	0.77	0.81	0.79	conjunction (2), specific (1), reason (1)
instance	20	11	0.55	0.92	0.69	specific (4), conjunction (3), hypophora (2)
reason	15	11	0.73	0.92	0.81	specific (2), restatement (1), result (1)
contra-expectation	12	11	0.92	0.73	0.81	contrast (1)
restatement	12	5	0.42	0.83	0.56	specific (5), favoured (1), conjunction (1)
after	12	0	0.00	0.00	0.00	before (9), specific (1), result (1), conjunction (1)
contrast	11	6	0.55	0.86	0.67	conjunction (3), contra-expectation (2)
broad	10	4	0.40	1.00	0.57	specific (2), conjunction (2), result (2)

Table 3: Evaluating AURIS discourse relation annotations against ISO-SemAF conversion of TED-MDB, relations with at least 10 attestations in the TED-MDB dataset, reporting true positives (TP), precision (P), recall (R) and F-measure (F1) as well as most frequent confusions.

losophy, and from the conversion process from the PDTB-style annotations of TED-MDB to AURIS. In addition to the fact that this hints at a considerable need to reconsider some of the less frequent ISO SemAF relations and the annotation guidelines, a number of observations stand out that may lead to a subsequent refinement of the annotation manual and a better alignment between TED-MDB/PDTB and AURIS:

- In many (and for the case of *after*, all) cases, the annotated TED-MDB relation is the inverse of the AURIS relation.
- Annotators seem to have difficulties to disentangle *instance* and *specific*, and, also, *conjunction*, *similarity* and *specific* appear to be frequently confused. In AURIS, the *specific* role of SemAF Elaboration is the most frequent relation, probably due to its high degree of underspecification (Knott et al., 2008).
- AURIS annotates the relation *hypophora* as pointing from the question to the answer, because this is where the marker (e.g., a question mark, or a WH-form) resides. TED-MDB, however, annotates *hypophora* as a relation pointing from the answer to the question. For this reason, there is never any match between both relations in both annotation schemes.

5. Summary and Outlook

We introduced AURIS, the *Augsburg corpus for Reference and Information Structure*, a novel corpus for the annotation of reference, discourse relations, and aspects of information structure, grounded in existing standards and annotation guidelines, but somewhat special in being designed for annotation with spreadsheet software. This may be a radical step, as prolific annotation tools are available for many aspects, but it also eliminates possible technical entry barriers to discourse annotation by occasional annotators such as students in a philology class. But also, we are not tied to any specific piece of (and, in particular, no proprietary) software, and can nevertheless benefit from an existing, and widely used technical ecosystem, with off-the-shelf solutions for Desktop-based annotation on different platforms, including OS-independent, browser-based solutions that collaborative editing (e.g., using OneDrive, SharePoint or Google Spreadsheets). As for the risk that annotators tinker around with the original data, this can be controlled by integrated cell protection. While multi-file spreadsheet annotations can be easily converted to TSV formats and processed by downstream applications or statistical tools, they cannot be directly queried. We thus provide an export to PAULA XML to facilitate search, visualization and querying with ANNIS, as visualized in Fig. 3 with a sample response to a query to retrieve conjunction relations between two utterances:

```
node & node &
#1 ->discourse[RELATION="conjunction"] #2
```

As Fig. 3 shows, we configured ANNIS to provide a grid view visualization of word- and phrase-level annotations (for AURIS annotations for coreference and information status), a conventional coreference view (minimized in the figure) for links, and a second coreference view is used to visualize discourse annotations. Here, two segments that are connected by a discourse relation are underlined in the same color, and mouseover events provide access to the annotation of incoming and outgoing relations. Although this is operational both for querying and visualization, this is preliminary solution only and to be revised with the next release of the corpus.

We see the systematic application of off-the-shelf software for relatively complex types of annotation

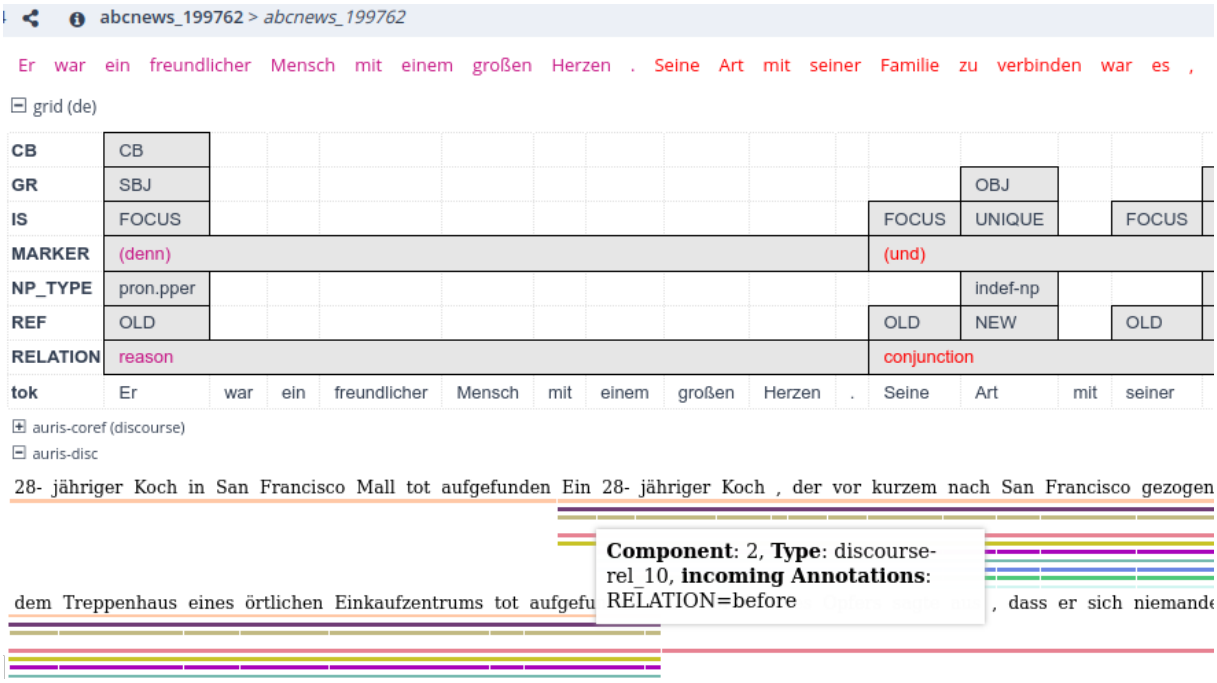

Figure 3: Screenshot from the AURIS corpus as published via ANNIS.

as the core contribution of this paper, as this eliminates the entry bias and technical dependencies that corpora with discourse annotation are typically characterized by, and certain features of common spreadsheet software actually exceed the capabilities of more advanced tools; this includes cell protection to preserve annotation consistency (lacking in otherwise comparable tier-based annotation tools such as EXMARaLDA (Schmidt and Wörner, 2009)), dynamic pre-annotation (tools such as INCEpTION (De Castilho et al., 2024) offer static pre-annotation, only), special key combinations for fast navigation and keyboard-only annotation (often lacking in graphical annotation tools), off-the-shelf libraries in many programming languages for creating and reading such data, autocompletion, etc.

We suggest that, in combination with advanced pre- and postprocessing components as described in the paper, spreadsheet-based annotations can provide another component in the emerging ecosystem of resources, tools and standards developed for the study, annotation and technical processing of discourse, reference and information structure, and one that – together with more advanced techniques and tools for annotation by experts, or, specifically trained annotators – may facilitate the creation of much-needed data in class.

The AURIS corpus and scripts developed for data processing are published under open licenses and available from https://github.com/acoli-repo/AURIS. The corpus can be queries online under http://corpora.philhist.uni-augsburg.de/.

Acknowledgments

The research described in this paper was conducted since 2023 at the Chair of Applied Computational Linguistics (ACoLi) at the University of Augsburg and has been partially supported with internal research funding of the Department of Philology and History of the University of Augsburg (Typ A) awarded to Christian Fäth and Mahmud Uz-Zaman. We would like to thank our student assistents, notably Viktoriia Remishevska and Margarita Arapi who conducted parts of the annotation, our colleagues, in particular Mahmud Uz-Zaman, Sonja Zeman and Claudia Claridge, with whom we explored the application and refinement of AURIS guidelines and workflows in class, and a large number of PhD, M.A. and B.A. students who discussed and applied these guidelines in class and produced annotations.

The student annotations were particularly important for the development and refinement of the AURIS guidelines, but because they were conducted against earlier versions of the guidelines, these annotations are not included in the current release. The current release for German has been primarily annotated by Tabea Gröger and reviewed by Christian Chiarcos. For English, experimental release data has been created by Viktoriia Remishevska and reviewed by Tabea Gröger, a pilot study for AURIS coreference in Italian was conducted by Giulia Mantovani.

Limitations and ethical considerations

We present a way to facilitate the annotation of discourse, reference and information structure with off-the-shelf tools and a minimal technical entry barrier. We are not aware of any ethical complications that could arise from this contribution. In fact, our approach leverages the access to and the participation in research for students and scholars from, say, developing countries in that we eliminate some of the constraints that the conventional technologies used for linguistic annotation have, in that annotation is possible without dependence from proprietary software, any particular operating system, without permanent/reliable internet connection, and in a distributed, decentralized manner.

The approach does, however, come with a number of limitations and possible down-sides. For example, the annotation of links with numerical indices (as for `TARGET`) comes with a risk of typing errors which can only be reliably spotted if the same text is annotated multiple times. Also, annotating spreadsheets in a decentralized fashion does not directly provide us with is any kind of version control, but in our usage so far, this is nevertheless accomplished by using GitHub and teaching it to participating students. Further, there is no direct support for moderating and merging annotations. However, this is accomplished by using existing solutions for merging and comparing CoNLL formats, for which we provide and use a replication of CoNLL-Merge in Python. Finally, the heavy reliance on pre-annotation and limited options to correct parser or segmentation errors may lead to an increased curation effort, as, at the moment, students are advised to keep track of such issues using the `COMMENT` column and let the annotation moderator work it out. And finally, dynamic pre-annotation may create a bias in the data which still needs to be evaluated. Indeed, these considerations have been one motivation to build initial annotations on top of corpora with comparable annotations for either discourse or anaphora, and our annotation results for German have been evaluated against AURIS-converted editions of TED-MDB and the news section of ParCorFull, with encouraging results.

Yet another downside is that spreadsheet data can contain executable code and may thus be blocked by certified research infrastructure repositories. While this is the case, indeed, for example for CLARIN.SI,[3] the anticipated use of spreadsheets is not for data exchange and publication, but only as a tool for creating the data. The actual release data comes in different formats, including the export into two CSV (TSV) files that contain all annotated data but lacks all interactive components (i.e., formulas and layout instructions). As this is otherwise identical in content with the spreadsheets it is generated from, this is recommended as an archival format instead of the original Excel files. For continuing or revising annotations, this data can be re-imported into Excel, LibreOffice, etc. – although preprocessing scripts and layout need to be copied from another Excel file.

[3] https://www.clarin.si/repository/xmlui/page/data, accessed 2026-03-06.

6. Bibliographical References

Ondřej Bojar, Jindřich Helcl, Tom Kocmi, Jindřich Libovickỳ, and Tomáš Musil. 2017. Results of the WMT17 Neural MT training task. In *Proceedings of the Second Conference on Machine Translation*, pages 525–533, Copenhagen, Denmark. Association for Computational Linguistics.

Peter Bourgonje and Manfred Stede. 2020. The Potsdam Commentary Corpus 2.2: Extending annotations for shallow discourse parsing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1061–1066, Marseille, France. European Language Resources Association.

Harry Bunt and Rashmi Prasad. 2016. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. In *Proceedings 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)*, pages 645–652, Portoroz, Slovenia.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In Jan van Kuppevelt and Ronnie W Smith, editors, *Current and new Directions in Discourse and Dialogue*, pages 85–112. Kluwer Academic Publishers, Dordrecht.

Christian Chiarcos. 2022. Inducing discourse marker inventories from lexical knowledge graphs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2401–2412, Marseille, France. European Language Resources Association.

Christian Chiarcos. 2025. Revisiting the Givenness Hierarchy. A corpus-based evaluation. In *Proceedings of the Joint Sixth workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025) and eighth workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2025), held in conjunction with EMNLP-2025*, Suzhou, China.

Christian Chiarcos, Stefanie Dipper, Michael Götze, Ulf Leser, Anke Lüdeling, Julia Ritz, and Manfred Stede. 2008. A flexible framework for integrating annotations from different tools and tag sets. *TAL (Traitement Automatique des Langues)*, 49(2):217–246.

Christian Chiarcos and Maxim Ionov. 2021. Linking discourse marker inventories. In *3rd Conference on Language, Data and Knowledge (LDK 2021)*, pages 40–1. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Christian Chiarcos, Julia Ritz, and Manfred Stede. 2012. By all these lovely tokens... Merging conflicting tokenizations. *Language Resources and Evaluation*, 46(1):53–74.

Christian Chiarcos and Niko Schenk. 2019. CoNLL-Merge: Efficient harmonization of concurrent tokenization and textual variation. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*, pages 7–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Richard Eckart De Castilho, Jan-Christoph Klie, and Iryna Gurevych. 2024. Integrating INCEpTION into larger annotation processes. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 110–121, Miami, Florida, USA. Association for Computational Linguistics.

Christian Fäth, Christian Chiarcos, Björn Ebbrecht, and Maxim Ionov. 2020. Fintan - Flexible, Integrated Transformation and Annotation eNgineering. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 7212–7221, Marseille, France. European Language Resources Association.

Barbara J Grosz, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2):203–225.

Jeanette Gundel, Nancy Hedberg, Ron Zacharski, Ann Mulkern, Tonya Custis, Bonnie Swierzbin, Amel Khalfoui, Linda Humnick, Bryan Gordon, Mamadou Bassene, and Shana Watters. 2006. Coding protocol for statuses on the Givenness Hierarchy (Gundel, Hedberg and Zacharski 1993). Technical report, University of Minnesota.

Jeanette K Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, pages 274–307.

Lynette Hirschman and Nancy Chinchor. 1998. MUC-7 coreference task definition (version 3.0). In *Proceedings of the Seventh Message Understanding Conference (MUC-7)*, Fairfax, Virginia.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 57–60, New York City, USA. Association for Computational Linguistics.

Nancy Ide and Keith Suderman. 2007. GrAF: A graph-based format for linguistic annotations. In *Proceedings of the Linguistic Annotation Workshop*, pages 1–8, Prague, Czech Republic. Association for Computational Linguistics.

Alistair Knott, Jon Oberlander, Michael O'Donnell, and Chris Mellish. 2008. Beyond Elaboration: The interaction of relations and focus in coherent text. In Ted J.M. Sanders, Joost Schilperoord, and Wilbert Spooren, editors, *Text Representation. Linguistic and psycholinguistic aspects*, pages 181–196. John Benjamins Publishing Company.

Olga Krasavina and Christian Chiarcos. 2007. Pocos-potsdam coreference scheme. In *Proceedings of the Linguistic Annotation Workshop*, pages 156–163.

Thomas Krause. 2019. *ANNIS: A graph-based query system for deeply annotated text corpora*. Ph.D. thesis, HU Berlin, Germany.

Knud Lambrecht. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.

Ekaterina Lapshinova-Koltunski, Christian Hardmeier, and Pauline Krielke. 2018. ParCorFull: A parallel corpus annotated with full coreference. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 423–428, Miyazaki, Japan. European Language Resources Association (ELRA), European Language Resources Association (ELRA).

Christoph Müller and Michael Strube. 2006. Multi-level annotation of linguistic data with mmax2. In S Braun, K Kohn, and J Mukherjee eds, editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, volume 3 of *English Corpus Linguistics*, pages 197–214. Lang, Frankfurt/M.

Anna Nedoluzhko, Michal Novák, Silvie Cinková, Marie Mikulová, and Jiří Mírovský. 2016. Coreference in Prague Czech-English Dependency

Treebank. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 169–176, Portorož, Slovenia. European Language Resources Association (ELRA).

Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Amir Zeldes, and Daniel Zeman. 2022. CorefUD 1.0: Coreference meets Universal Dependencies. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4859–4872, Marseille, France. European Language Resources Association.

Michael O'Donnell. 2000. RSTTool 2.4 - A markup tool for Rhetorical Structure Theory. In *Proceedings of the First International Conference on Natural Language Generation (INLG-2000)*, pages 253–256, Mitzpe Ramon, Israel. Association for Computational Linguistics.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi, and Bonnie L. Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 2961–2968.

Rashmi Prasad, Bonnie Webber, and Aravind Joshi. 2017. The Penn Discourse Treebank: An annotated corpus of discourse relations. In *Handbook of Linguistic Annotation*, pages 1197–1217. Springer.

Ina Rösiger. 2018. BASHI: A corpus of Wall Street Journal articles annotated with bridging links. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Thomas Schmidt and Kai Wörner. 2009. EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19(4):565–582.

Manfred Stede. 2004. The Potsdam Commentary Corpus. In *Proceedings of the Workshop on Discourse Annotation*, pages 96–102, Prague.

Milan Straka, Jan Hajic, and Jana Straková. 2016. UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297.

Pavlína Synková, Jiří Mírovský, Lucie Poláková, and Magdaléna Rysová. 2024. Announcing the Prague Discourse Treebank 3.0. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1270–1279, Torino, Italia. ELRA and ICCL.

Florian Wolf and Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Computational linguistics*, 31(2):249–287.

Seid Muhie Yimam, Iryna Gurevych, Richard Eckart De Castilho, and Chris Biemann. 2013. WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL-2013): system demonstrations*, pages 1–6.

Amir Zeldes. 2016. rstWeb - A browser-based annotation interface for Rhetorical Structure Theory and discourse relations. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-2016): Demonstrations*, pages 1–5.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

Deniz Zeyrek, Amalia Mendes, Yulia Grishina, Murathan Kurfali, Samuel Gibbon, and Maciej Ogrodniczuk. 2020. TED Multilingual Discourse Bank (TED-MDB): A parallel corpus annotated in the PDTB style. *Language Resources and Evaluation*, 54(2):587–613.

Appendix

We illustrate the spreadsheet-based annotation with two samples for the annotation of coreference and information status in Fig. 4 and discourse structure in Fig. 5 in LibreOffice. Our spreadsheets have also been tested with MS Excel, Google Spreadsheets and Gnumeric on Linux, Windows and MAC, and generally worked without major problems. One difficulty observed with LibreOffice in particular was that larger files tend to load slowly. (Observed during a partial annotation of the Italian translation of the Biblical book of Genesis, containing a total 33.013 tokens.) Our immediate solution was to convert the XLSX format into the native ODS spreadsheet format first, but in general, it is advisable to focus on the annotation of smaller chunks of text. With XSLX data, Gnumeric was usually particularly fast in loading data.

WORD	GR	NP_TYPE	REF_AUTO	COREF	REF	IS	CB	COMMENT
# text = 28-jähriger Koch in San Francisco Mall tot aufgefunden								
28-								
jähriger								
Koch	SBJ	ne	?OLD	frank_galicia	NEW	REF		
in								
San	other_2	ne	?OLD	san_francisco	NEW	FAMILIAR		
Francisco								
Mall				westfield_mall	NEW	REF		
tot								
aufgefunden								
# text = Ein 28-jähriger Koch, der vor kurzem nach San Francisco gezogen ist, wurde im Treppenhaus eines örtlichen Einkaufzentrums tot aufgefunden.								
Ein								
28-								
jähriger								
Koch	SBJ	ne	?OLD	frank_galicia	OLD	FOCUS	CB	
,								
der	SBJ_3	pron	?OLD		BOUND			parser error: Relativpronomen
vor								
kurzem	other_3	indef-np.bare						

Figure 4: Reference annotation with spreadsheets (here, LibreOffice)

ID	PREDICATE	TEXT	MARKER	TARGET	RELATION	COMMENT
1	auffinden	28- jähriger Koch in San Francisco Mall tot aufgefunden	(zusammen gefasst)	2	broad	
2	auffinden	Ein 28- jähriger Koch , der vor kurzem nach San Francisco gezogen ist , wurde im in dem Treppenhaus eines örtlichen Einkaufzentrums tot aufgefunden .	(konkret)	1	specific	
3	vorstellen	Der Bruder des Opfers sagte aus , dass er sich niemanden vorstellen kann , der ihm schaden	_	2	EntRel	

Figure 5: Discourse annotation with spreadsheets (here, LibreOffice)