Survey of Tools for Manual Linguistic Annotation: Supporting Diversity through Interactive Exploration

Ludovica Pannitto[1], **Kaja Dobrovoljc**[2], **Bruno Guillaume**[3]
[1] University of Bologna, Bologna - Italy,
[2] University of Ljubljana and Jozef Stefan Institute, Ljubljana - Slovenia,
[3] Université de Lorraine, CNRS, Inria, LORIA, Nancy - France
ludovica.pannitto@unibo.it, kaja.dobrovoljc@ijs.si, bruno.guillaume@loria.fr

Abstract

Manual annotation tools are core infrastructure for corpus creation, enabling the development of linguistically informed language resources relevant for both linguistic discovery and computational applications. We present a comprehensive survey of 21 tools supporting morphosyntactic and multi-word expression annotation, systematically documenting more than 50 features relevant for annotation workflows—from software architecture and usability to linguistic coverage and annotation scope. The survey results are published as an open dataset and made accessible through an interactive online platform that allows users to filter and compare tools according to their specific needs. Our initial analysis highlights a robust and open ecosystem of annotation tools, but advanced needs for complex and language-independent annotation are inconsistently addressed.

Keywords: manual annotation tools, linguistic annotation, annotation diversity, linguistic diversity, corpus creation

1. Introduction

The creation of high-quality language resources continues to rely on manual linguistic annotation (Ide and Pustejovsky, 2017; Wilcock, 2009), even as automatic methods have advanced rapidly. Manually annotated corpora provide the gold standards required for training and evaluating natural language processing systems (Stollenwerk et al., 2023), serve as indispensable benchmarks for evaluating and probing the performance of large language models (Felkner et al., 2024; Liu et al., 2023), and constitute essential data for linguistic description and discovery (Gries and Berez, 2017). To support these efforts, a wide range of tools have been developed for different annotation tasks, forming an essential infrastructure for building linguistically informed corpora across languages.

In practice, annotation projects require tools that address a range of technical and linguistic needs: active maintenance, flexible input/output formats, collaborative annotation workflows, and crucially, features that ensure applicability to diverse linguistic contexts, such as support for language-specific phenomena, scripts and writing directions. Due to the diversity of annotation goals and user needs, the tool landscape is necessarily fragmented and heterogeneous.

Existing surveys and reviews of annotation environments have provided valuable overviews that help navigate this fragmentation (Gessler, 2022b; Neves and Ševa, 2021; Borisova et al., 2024; Berberich and Kleiber, 2023), but they typically remain high-level and descriptive (e.g., listing tools with basic information on availability and supported annotation types), without systematically documenting these practical functionalities in sufficient detail. As a result, researchers often lack the information needed to make informed choices about which tools best fit their annotation workflows, particularly when working with less-resourced languages (Gessler, 2022a), unconventional data types (Dobrovoljc, 2022) or structurally complex annotation tasks (Schuster and Manning, 2016).

To address this gap, we carried out a comprehensive survey of 21 tools for manual linguistic annotation, systematically documenting a wide range of features relevant to annotation workflows—from basic technical properties to collaborative features and multilingual support. Among these, we paid particular attention to functionalities relevant for lexical and grammatical annotation, where cross-linguistic variation is greatest and tool support is often most critical. Specifically, we examine support for various aspects of the Universal Dependencies[1] (de Marneffe et al., 2021) and PARSEME[2] (Savary et al., 2017) annotation schemes, two widely adopted multilingual frameworks for morphosyntactic and multi-word expression annotation that have been recognized as crucial for reconciling language diversity with rapid progress in language technology (Savary et al., 2023). As such, it does not aim to comprehensively evaluate support for all annotation paradigms, such as phrase-structure, discourse, or semantic annotation. While some sur-

[1] https://universaldependencies.org/
[2] https://gitlab.com/parseme/corpora/wikis

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 11562–11573
11–16 May 2026.

veyed tools may support such layers, these were not systematically evaluated within the scope of this study.

Our survey results are published as an open dataset and made accessible through an interactive online inventory that allows researchers to filter and explore tools according to their specific annotation needs. In the remainder of this paper, we present the survey design and methodology (Section 2), describe the publication of results as an open dataset and interactive platform (Section 3), provide a comparative analysis of the surveyed tools (Section 4), and conclude by outlining directions for future work (Section 5).

2. Survey Design and Methodology

We conducted the survey between September and December 2024 using a structured questionnaire. This section describes the questionnaire design (Section 2.1), the data collection procedure (Section 2.2), and the set of tools covered by the survey (Section 2.3).

2.1. Questionnaire Design

In line with the goals outlined in the introduction, the questionnaire was designed to capture a wide range of general features relevant across annotation tasks, while also including specific questions addressing support for morphosyntactic and multiword expression annotation. We implemented a Google Forms questionnaire[3] consisting of 57 questions divided into eleven thematic sections, which were organized as follows[4]:

General information: Tool name, short description, website, contact information

Software: License, code repository, active maintenance status, software type (desktop/web-based), operating system support, access mode, offline usage

Data support: Input formats, integration of other resources (multimodal, external), data export options

Language support: Language independence, support for multiple writing scripts, right-to-left writing direction, encoding, interface language

Annotation layers: word and sentence segmentation, annotation types (tokens, spans, multi-tokens, relations, chains), annotation scope (within/across sentences, full documents), tagset customization

User interface: Annotation modes (keyboard, mouse, touchscreen), keyboard shortcuts, annotation display (graphical, tabular, raw format)

Workflow: Automatic pre-annotation, automatic recommendation, collaborative annotation modes, project management, user roles

Annotation analysis: Comparison of annotations, inter-annotator agreement calculation, querying capabilities, mass editing

UD-specific support: Editing capabilities for lemmas, POS tags (UPOS), morphological features (FEATS), language-specific tags (XPOS), multiword tokens, basic dependencies, enhanced dependencies, empty nodes; metadata editing; UD validation

PARSEME-specific support: Discontinuous expressions, nested expressions, overlapping expressions, consistency checking

Open-ended questions further allowed respondents to provide examples of projects and additional comments.

2.2. Data Collection

Because of the questionnaire's detailed nature, data collection proceeded in two stages. In the first stage, we contacted development teams directly to ensure accurate reporting of tool features, based on a list compiled from existing online inventories, such as the UD[5] and MALA[6] websites. In the second stage, the call was disseminated more broadly via the mailing list of the COST Action UniDive[7] – Universality, diversity and idiosyncrasy in language technology (Savary et al., 2024), which gathers a large number of corpus annotation experts working on many different languages, thereby enabling contributions from researchers and users familiar with particular tools. The survey was open from September to December 2024.

2.3. Surveyed Tools

In total, we collected 24 replies for 21 tools (3 tools were surveyed twice), ranging from general-purpose annotation environments to specialized editors for UD morphosyntacic parsing and PARSEME multi-word expression annotation.

Listed by alphabet, these include[8]: ArboratorGrew (Guibon et al., 2020), BoAT (Türk

[3] https://forms.gle/uaUtbo4zvgwJsrMZA
[4] See also Appendix A
[5] https://universaldependencies.org/tools.html
[6] https://lgessler.com/mala/
[7] https://unidive.lisn.upsaclay.fr/
[8] See also Appendix B

et al., 2019), brat (Stenetorp et al., 2012), CoNLLU-Editor (Heinecke, 2019), DgAnnotator (Attardi, n.d.), ELAN (Wittenburg et al., 2006), EXMARaLDA (Schmidt, 2012), FLAT (van Gompel and Reynaert, 2013), FLeX (Butler and Van Volkinburg, 2007), GitDox (Zhang and Zeldes, 2017), INCEpTION (Eckart De Castilho et al., 2024), Label Sleuth (Shnarch et al., 2022), Label Studio (Tkachenko et al., 2020-2025), Palmyra (AbuOdeh et al., 2024), PLURAL-conllu-annotator (Krishnamurthy et al., 2024), Q-CAT (Brank, 2023), TALEN (Mayhew and Roth, 2018), Toolbox (International, 2019), TrEd (Pajas and Fabian, 2019), UD Annotatrix (Tyers et al., 2017), and WebAnno (Eckart de Castilho et al., 2016).

3. Publication of Survey Results

The survey data are made publicly available in two complementary forms: a structured dataset for computational analysis (Section 3.1) and an interactive web-based platform for user-friendly exploration and comparison (Section 3.2). Both are hosted in an open repository (Section 3.3) designed for easy community-driven expansion.

3.1. Comparison Table

When the data collection described in Section 2 was completed, responses were exported from Google Forms and manually processed to ensure data quality and consistency. This involved verifying uncertain information, standardizing free-text entries, and merging duplicate responses for the three tools that were surveyed twice. The resulting dataset is available as a TSV file in the project repository,[9] with tools listed in rows and consolidated features in columns. The machine-readable format enables both direct inspection in spreadsheet applications and computational analysis.

3.2. Interactive Platform

To support tool discovery and selection, we developed an interactive web interface that enables users to dynamically filter, browse, and examine tools according to their specific annotation requirements.[10]

3.2.1. Interface Structure

The interface appears as a two-pane, filterable catalogue of annotation tools (Figure 1): the right panel is dedicated to the list of features, divided into the newly introduced 7 areas, while the left panel lists all the surveyed tools (on the top), reserving a space for an extended description of the tool on the bottom. The three areas are resizeable thanks to a vertical and an horizontal dividers.

3.2.2. Feature Selection

When expanding each area on the right panel (Figure 2), a full list of features appears: each option is accompanied by the number of tools that include that value. By clicking on a specific value, the matching tools on the left are brought to the top of the list and made more evident by graying the tiles for the others (Figure 3). Multiple features can be combined to further refine the list: the top of the list expands to include tools matching any of the selected features. Counts for each features are also updated showing, out of the total number of tools that include that feature, the tools compatible with the current selection.

3.2.3. Tool Details

Each tool tile bears a + symbol that toggles the display of the tool details in the bottom-left panel. In the bottom space, the name of the tool is shown with link to its homepage, together with a short description with a logo and the link to the code base (if available). Moreover, the full list of features for the tool is shown and, if provided, an extended description and further links are listed on the left. The + action also highlights corresponding features on the right panel, which appear bordered in blue (Figure 4): this allows a faster comparison between desired features and the ones that the tool supports. Finally, a reset button clears all current selection and restores the initial state of the interface.

3.3. Repository and New Contributions

Both the dataset (Section 3.1) and the interactive platform (Section 3.2) are hosted in an open GitHub repository[11]. The repository is organized into three main directories: `tools` contains individual YAML files describing each tool; `data` stores the consolidated TSV dataset (Section 3.1) and a JSON file summarizing information for all tools); and `interface` houses the HTML and companion code.

To enable community contributions, we have established an initial protocol for adding new tools to the inventory. Contributors can use the provided template to submit new tools via pull requests, following the instructions in the README file provided in the repository. After review and merge by the maintainers, an automatic integration step is run to

[9] https://github.com/UniDive/corpus-annotation-tools/blob/main/data/results_230125.tsv

[10] The interface is available at https://unidive.github.io/corpus-annotation-tools/interface/tools.html.

[11] The repository is available at https://github.com/UniDive/corpus-annotation-tools

Question	Options	Number of Tools and Percentage (%)
Response provided by developer	Yes	11 (52.4)
	No	10 (47.6)
Software type	Web-based application	14 (66.7)
	Desktop-based application	7 (33.3)
Operating system	Windows	21 (100.0)
	Mac OS	17 (80.9)
	Linux	19 (90.5)
Software availability	Run locally (user machine/server)	15 (71.4)
	Online (with sign-up)	4 (19.0)
Input formats	Raw text	16 (76.2)
	CONLL-U	14 (66.7)
	XML	12 (57.1)
Data export	Data can be exported	21 (100.0)
	Data can be stored on third-party platform	4 (19.0)
Annotation types	Tokens	20 (95.2)
	Spans	16 (76.2)
	Multi-tokens	15 (71.4)
	Relations	15 (71.4)
	Chains	8 (38.1)
Annotation mode	Mouse	18 (85.7)
	Touchscreen	3 (14.3)
	Keyboard	17 (80.9)
Keyboard annotation and shortcuts	Yes	13 (61.9)
	No	8 (38.1)
Workflow	Automatic pre-annotation	7 (33.33)
	Automatic recommendation	5 (23.8)
	Collaborative annotation (annotators working together)	4 (19.1)
	Collaborative annotation (annotators working independently)	8 (38.1)
	Project management	6 (28.6)
	User roles	8 (38.1)
Annotation analysis	Comparison of annotations	8 (38.1)
	Agreement calculation	3 (14.3)
	Querying	15 (71.4)
	Mass editing	12 (57.1)
UD-specific annotations	Lemmas	15 (71.4)
	UPOS	17 (80.9)
	XPOS	15 (71.4)
	FEATS	14 (66.67)
	Basic dependencies	15 (71.4)
	Metadata	13 (61.9)
PARSEME-specific annotations	Discontinuous expressions	5 (23.8)
	Nested expressions	7 (33.33)
	Overlapping expressions	7 (33.33)
Validation	UD validation	5 (23.8)
	PARSEME consistency check	1 (4.8)

Table 1: Distribution of the analyzed tools across different characteristics. Percentages are calculated over the total number of tools ($N = 21$).

update information in the JSON file. Consequently, the web interface can automatically display the new tool and its features.

4. Survey Findings

As the complete table of results is available in a dedicated table and explorable through an interface (See Section 3), we will highlight here the

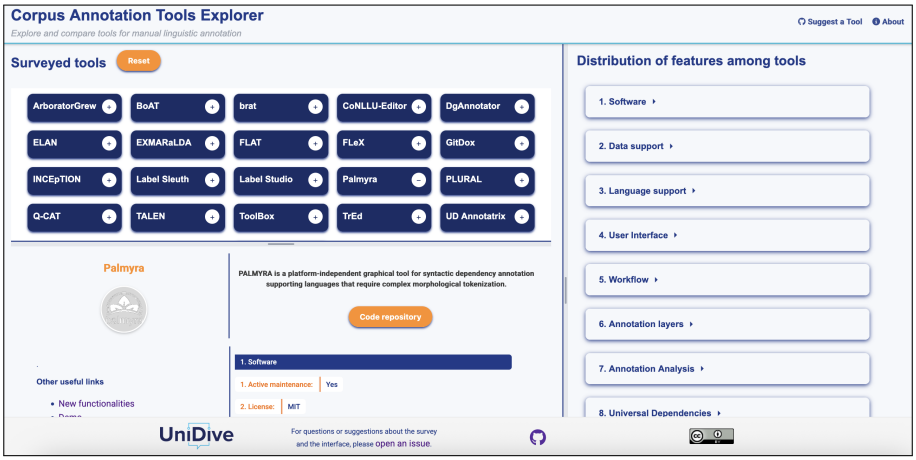

Figure 1: Full view of the web interface showing the three-panel layout with available tools (top left), detailed tool information (bottom left), and feature filters (right).

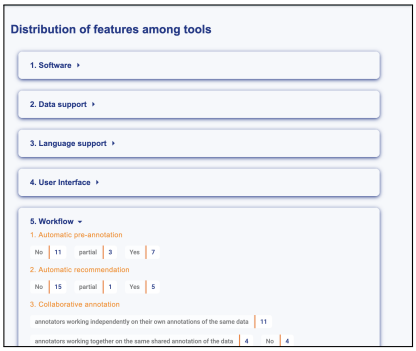

Figure 2: Right panel showing expandable feature categories.

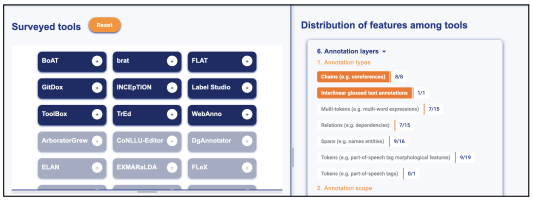

Figure 3: Feature selection: matching tools are highlighted while others are grayed out (left panel).

main observations derived from a critical overlook at responses. An overlook of the main examined features is given in Table 1.

4.1. Software Infrastructure

As far as general software features are concerned, 17 of the 21 surveyed tools are actively maintained, and 18 provide public code repositories. While these repositories are not uniformly structured (e.g., not all provide pointers to the latest stable release or well-structured documentation, echoing findings by (Borisova et al., 2024)), the prevalence of open repositories is encouraging. 18 tools are released

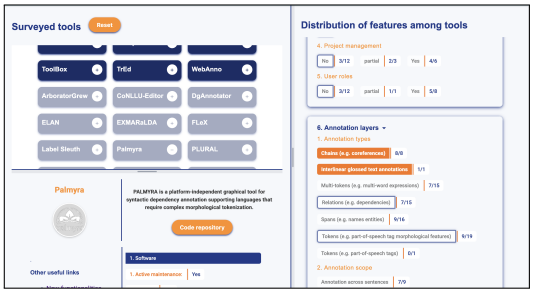

Figure 4: Tool details with its features highlighted in blue (right panel).

under open-source licenses such as Apache, MIT, or GPL, representing a positive trend for tool sustainability and workflow reproducibility. Most surveyed tools (14 of 21) are web-based, while seven are desktop-based. However, only five web-based tools are freely available through public instances; most require self-hosting on local infrastructure. This flexibility supports data privacy and customization but limits out-of-the-box usability for collaborative work.

4.2. Data Support

Concerning data support, most surveyed tools accept `CoNLL-U` (14) or raw text (16) input, and many also handle XML, with several additionally supporting TSV, ensuring compatibility with widely used corpus pipelines. A smaller subset of tools (5) offers integration of multimedia and multimodal data, such as audio, video, and images, which is crucial for spoken and multimodal corpora but remains relatively rare. All tools enable data export, though only four include built-in synchronization with third-party storage or version control platforms such as Git. While this indicates a solid foundation for interoperability, limited support for non-textual data and

integrated data management remains a key area for future improvement.

4.3. Language Support

With respect to language support, all surveyed tools report being language-independent and compatible with `utf-8` encoding, enabling the annotation of texts in diverse scripts. Support for left-to-right writing is universal, while right-to-left scripts are explicitly supported by 17 tools, though several respondents note partial or display-only functionality. Interface localization remains limited: nearly all tools are available only in English, with a few offering additional options such as French, German, or Chinese. Overall, while the technical foundations for multilingual annotation are in place, true multilingual usability, particularly in user interfaces and script handling, remains uneven.

4.4. Annotation Types

Looking at the annotation types supported, all surveyed tools (except LabelSleuth) enable token-level annotation, and most also support additional layers such as multi-token spans, and syntactic relations. Around two thirds allow token-level (re)segmentation, while just over half support sentence-level segmentation. Most tools are designed for sentence-internal annotation, treating the sentence as the primary unit of analysis. However, 11 tools support annotation across sentence boundaries or at the document level, with speech-oriented ELAN and EXMARaLDA offering the most flexibility by imposing no structural restrictions. Nearly all tools allow customization of tagsets, which is essential for adapting annotation schemes to new languages and domains.

4.5. User Experience and Workflow

With respect to user experience and workflow, most tools provide intuitive graphical interfaces supporting both keyboard and mouse input, with several also offering tabular or raw-text views. A few include touchscreen compatibility. Workflow automation remains limited: only seven tools offer automatic pre-annotation and just five provide recommendation systems that learn from annotator decisions. Collaborative annotation and project management functions are supported to a similar extent, often requiring local hosting. Quality-control options – such as multi-annotation comparison (curation), inter-annotator agreement calculation, annotation querying, and mass editing – are less consistently implemented, although they remain essential for ensuring annotation reliability.

4.6. Framework-Specific Support

Finally, with respect to framework-specific support, most tools are compatible with Universal Dependencies (UD) data structures, offering annotation and editing for lemmas, part-of-speech tags, morphological features, and basic dependencies. However, more advanced UD layers – such as enhanced dependencies and empty nodes – are supported by only a minority of tools. Similarly, support for PARSEME-related phenomena such as discontinuous, nested, or overlapping multiword expressions remains limited. Integration of validation or consistency-checking tools, including the UD validator and PARSEME checker, is rare, meaning that developers still rely on external pipelines to ensure compliance with these major frameworks — a clear point of attention for future tool development.

5. Conclusion

We have conducted a comprehensive survey of 21 tools for manual corpus annotation, with particular focus on morphosyntactic and multi-word expression annotation. Unlike previous surveys that typically document only basic tool properties, our questionnaire systematically captured over 50 features ranging from technical properties to workflow support and language-related aspects. The survey results are published as an open dataset and made accessible through an innovative interactive platform that enables users to filter and compare tools according to their specific needs.

Our initial analysis reveals encouraging trends: the majority of surveyed tools are open-source, web-based, actively maintained, and support flexible annotation workflows and collaborative work. However, substantial gaps remain in accommodating the full diversity of annotation needs, including the representation of diverse language phenomena. While most tools report being language-independent, many provide only partial support for features required to ensure true cross-linguistic applicability, such as handling non-textual data types, diverse writing directions, language-specific segmentation, and complex linguistic structures. Additionally, complete integration into the release pipelines of standard multilingual annotation frameworks such as UD and PARSEME is not yet fully implemented by any tool.

Our survey does not aim to provide an exhaustive comparison of tools across all possible linguistic annotation paradigms. In particular, support for alternative syntactic formalisms or document-level and discourse-oriented layers was not systematically evaluated. Some tools included in the survey may support these annotation types, either natively or via extensible schemas, but such capabilities were outside the primary scope of the present study.

Our focus on UD and PARSEME reflects their wide adoption and the need for structured comparison within these communities; however, we acknowledge that many annotation projects operate within different theoretical and representational traditions. Extending the comparative framework developed here to cover additional syntactic, semantic, and discourse annotation paradigms constitutes an important direction for future work.

Moreover, the survey focuses on tools designed for *manual* linguistic annotation, i.e., interactive environments that support human annotators in creating and curating structured linguistic data. We do not treat large language models (LLMs) themselves as manual annotation tools, although they are increasingly being used to generate pre-annotations or suggestions. At the time of writing, the integration of LLMs into annotation workflows is rapidly evolving, and different tools support such integration to varying degrees (e.g., via automatic pre-annotation, recommendation systems, or external pipelines). A systematic comparison of how annotation platforms incorporate LLM-based assistance could also constitute a valuable direction for future work and extending the comparative framework proposed here to explicitly capture these capabilities will be increasingly needed.

We hope to see this work continue to evolve as a living community resource, with an established protocol already in place to support new contributions from users and developers alike. In addition, complementary evaluation approaches could be explored, such as the development of dedicated test datasets that capture diverse annotation needs still insufficiently supported, providing a more systematic basis for benchmarking manual annotation tools.

We would also like to complement our approach with dedicated assessment of usability or user experience, that can further take into consideration the interface design, workflow efficiency, learning curve, or overall user satisfaction. Such an assessment, however, would require a substantially different experimental design and lies beyond the scope of the present survey.

6. Acknowledgements

We thank the tool developers and all community members who contributed to this survey by completing the questionnaire. This work received support from the CA21167 COST action UniDive, funded by COST (European Cooperation in Science and Technology), and from the ARIS research projects no. Z6-4617 and GC-002. We would also like to thank Dr. Valentina Rossetti for her support during the interface development phase. Language editing was supported by generative AI tools.

7. Bibliographical References

Muhammed AbuOdeh, Long Phan, Ahmed Farouk Zakaria Elshabrawy, and Nizar Habash. 2024. Palmyra 3.0: A user-friendly cloud-based platform for morphology and dependency syntax annotation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 12585–12591, Torino, Italia. ELRA and ICCL.

Giuseppe Attardi. n.d. Dependency grammar annotator. http://medialab.di.unipi.it/Project/QA/Parser/DgAnnotator/. Software tool. No published paper available.

Kristin Berberich and Ingo Kleiber. 2023. Tools for corpus linguistics. https://corpus-analysis.com/. Accessed: 2025-10-15.

Ekaterina Borisova, Raia Abu Ahmad, Leyla Garcia-Castro, Ricardo Usbeck, and Georg Rehm. 2024. Surveying the FAIRness of annotation tools: Difficult to find, difficult to reuse. In *Proceedings of the 18th Linguistic Annotation Workshop (LAW-XVIII)*, pages 29–45, St. Julians, Malta. Association for Computational Linguistics.

Janez Brank. 2023. Q-CAT corpus annotation tool 1.5. Slovenian language resource repository CLARIN.SI.

Lynnika Butler and Heather Van Volkinburg. 2007. FieldWorks Language Explorer (FLEx). *Technology Review*, 1(1):1.

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

Kaja Dobrovoljc. 2022. Spoken language treebanks in Universal Dependencies: an overview. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1798–1806, Marseille, France. European Language Resources Association.

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.

Richard Eckart de Castilho, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann. 2016. A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 76–84, Osaka, Japan. The COLING 2016 Organizing Committee.

Virginia Felkner, Jennifer Thompson, and Jonathan May. 2024. GPT is not an annotator: The necessity of human annotation in fairness benchmark construction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14104–14115, Bangkok, Thailand. Association for Computational Linguistics.

Luke Gessler. 2022a. Closing the NLP gap: Documentary linguistics and NLP need a shared software infrastructure. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 119–126, Dublin, Ireland. Association for Computational Linguistics.

Luke Gessler. 2022b. MALA: The map of applications for linguistic annotation. Software available from https://github.com/lgessler/mala.

Stefan Th. Gries and Andrea L. Berez. 2017. Linguistic annotation in/for corpus linguistics. *Handbook of linguistic annotation*, pages 379–409.

Gaël Guibon, Marine Courtin, Kim Gerdes, and Bruno Guillaume. 2020. When collaborative treebank curation meets graph grammars. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5291–5300, Marseille, France. European Language Resources Association.

Johannes Heinecke. 2019. ConlluEditor: a fully graphical editor for universal dependencies treebank files. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 87–93, Paris, France. Association for Computational Linguistics.

Nancy Ide and James Pustejovsky. 2017. *Handbook of Linguistic Annotation*, 1st edition. Springer Publishing Company, Incorporated.

SIL International. 2019. Field linguist's toolbox (version 1.6.4). Computer software.

Parameswari Krishnamurthy, Mitansh Kayathwal, Chirag Jain, and Anushka Agrawal. 2024. UD Annotator. https://plural.iiit.ac.in/conllu-annotator. Developed under the Plural Project at IIIT Hyderabad. Copyright © 2024 Parameswari Krishnamurthy. All rights reserved.

Yixin Liu, Alex Fabbri, Pengfei Liu, Yilun Zhao, Linyong Nan, Ruilin Han, Simeng Han, Shafiq Joty, Chien-Sheng Wu, Caiming Xiong, and Dragomir Radev. 2023. Revisiting the gold standard: Grounding summarization evaluation with robust human evaluation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4140–4170, Toronto, Canada. Association for Computational Linguistics.

Stephen Mayhew and Dan Roth. 2018. Talen: Tool for annotation of low-resource entities. In *ACL System Demonstrations*.

Mariana Neves and Jurica Ševa. 2021. An extensive review of tools for manual annotation of documents. *Brief. Bioinform.*, 22(1):146–163.

Petr Pajas and Peter Fabian. 2019. Tree Editor TrEd. https://ufal.mff.cuni.cz/tred/. Software tool. Last released version 2.5236 (Feb 28, 2019).

Agata Savary, Cherifa Ben Khelil, Carlos Ramisch, Voula Giouli, Verginica Barbu Mititelu, Najet Hadj Mohamed, Cvetana Krstev, Chaya Liebeskind, Hongzhi Xu, Sara Stymne, Tunga Güngör, Thomas Pickard, Bruno Guillaume, Eduard Bejček, Archna Bhatia, Marie Candito, Polona Gantar, Uxoa Iñurrieta, Albert Gatt, Jolanta Kovalevskaite, Timm Lichte, Nikola Ljubešić, Johanna Monti, Carla Parra Escartín, Mehrnoush Shamsfard, Ivelina Stoyanova, Veronika Vincze, and Abigail Walsh. 2023. PARSEME corpus release 1.3. In *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*, pages 24–35, Dubrovnik, Croatia. Association for Computational Linguistics.

Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, and Antoine Doucet. 2017. The PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain. Association for Computational Linguistics.

Agata Savary, Daniel Zeman, Verginica Barbu Mititelu, Anabela Barreiro, Olesea Caftanatov, Marie-Catherine de Marneffe, Kaja Dobrovoljc, Gülşen Eryiğit, Voula Giouli, Bruno Guillaume, Stella Markantonatou, Nurit Melnik, Joakim Nivre, Atul Kr. Ojha, Carlos Ramisch, Abigail Walsh, Beata Wójtowicz, and Alina Wróblewska. 2024. UniDive: A COST action on universality, diversity and idiosyncrasy in language technology. In *Proceedings of the 3rd Annual Meeting of the Special*

Interest Group on Under-resourced Languages @ LREC-COLING 2024, pages 372–382, Torino, Italia. ELRA and ICCL.

Thomas Schmidt. 2012. EXMARaLDA and the FOLK tools — two toolsets for transcribing and annotating spoken language. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 236–240, Istanbul, Turkey. European Language Resources Association (ELRA).

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

Eyal Shnarch, Alon Halfon, Ariel Gera, Marina Danilevsky, Yannis Katsis, Leshem Choshen, Martin Santillan Cooper, Dina Epelboim, Zheng Zhang, Dakuo Wang, Lucy Yip, Liat Ein-Dor, Lena Dankin, Ilya Shnayderman, Ranit Aharonov, Yunyao Li, Naftali Liberman, Philip Levin Slesarev, Gwilym Newton, Shila Ofek-Koifman, Noam Slonim, and Yoav Katz. 2022. Label Sleuth: From unlabeled text to a classifier in a few hours. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France. Association for Computational Linguistics.

Felix Stollenwerk, Joey Öhman, Danila Petrelli, Emma Wallerö, Fredrik Olsson, Camilla Bengtsson, Andreas Horndahl, and Gabriela Zarzar Gandler. 2023. Text annotation handbook: A practical guide for machine learning projects.

Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. 2020-2025. Label Studio: Data labeling software. Open source software available from https://github.com/HumanSignal/label-studio.

Utku Türk, Furkan Atmaca, Şaziye Betül Özateş, Abdullatif Köksal, Balkiz Ozturk Basaran, Tunga Gungor, and Arzucan Özgür. 2019. Turkish treebanking: Unifying and constructing efforts. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 166–177, Florence, Italy. Association for Computational Linguistics.

Francis M. Tyers, Mariya Sheyanova, and Jonathan North Washington. 2017. UD annotatrix: An annotation tool for Universal Dependencies. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 10–17, Prague, Czech Republic.

Maarten van Gompel and Martin Reynaert. 2013. Folia: A practical xml format for linguistic annotation – a descriptive and comparative study. *Computational Linguistics in the Netherlands Journal*, 3:63–81.

Graham Wilcock. 2009. *Introduction to linguistic annotation and text analytics*. Springer International Publishing, Cham.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Shuo Zhang and Amir Zeldes. 2017. Gitdox: A linked version controlled online xml editor for manuscript transcription. In *FLAIRS*, pages 619–623.

A. Full Survey Questionnaire

- Tool name: *Please provide the official name of the corpus annotation tool.*
- Short description: *Please provide a one-sentence description of the tool (e.g. the tagline advertised by the tool developers).*
- Website: *Please provide a URL to the homepage of the tool. If this does not exist, provide a link to an alternative relevant website (e.g. Git or other repository).*
- Other useful links: *Please provide URL links to other useful core documentation (e.g. online platform, user documentation, tutorials).*
- Your name
- Your email address
- Have you been directly involved in the development of the tool (e.g. as a developer or consultant)? (*yes/no/other options*)

Software

- License: *Please specify the license information for the tool (e.g. Apache 2.0)*
- Code repository (for open license): *Please provide the link to the code repository (GitGub, GitLab or equivalent) where the code is available for adaptation and evolution.*
- Active maintenance: Is the tool actively maintained and some user support is available if needed? (*yes/no/other options*)
- Software type (Desktop-based/Web-based/other options)
- Operating system (for desktop based applications): *If the tool can be installed as a desktop-based application, which operating systems are supported.*
- Software availability (for web-based applications):
  - *The software is freely available online (with sign up option and/or social logins)*
  - *The software is available upon request (the user should ask to get an account)*
  - *The software should be run by the user (local machine or own server)*
- Offline usage (for web-based applications): *Can the tool be used offline with a delayed synchronisation with the server?* (*yes/no/other options*)

Data support

- Input format: *Please specify the list of supported formats when importing a corpus.* (*Options including: Raw text/PDF/CONLL-U/CUPT/XML/TSV/other*
- Integration of other resources: *Does the tool support integration of data types other than text?* (*options including Audio/Video/Image/PDF/External databases (e.g. dictionaries)/other*)
- Data export: *Which of the following options for (raw, annotated, curated) data export apply?*
  - *The data can be exported.*
  - *The data can be stored with a third party storage platform (e.g. Git repository)*
  - *other*

Language support

- Language-independence: *Is the tool generally considered to be language-independent, i.e. can it be configured to annotate data in many different languages?* (*yes/no/other options*)
- Writing script: *Does the tool support multiple writing scripts (e.g. Latin, Cyrillic, Arabic)?* (*yes/no/other options*)
- Writing direction: *Does the tool support right-to-left languages (e.g. Arabic, Hebrew)?* (*yes/no/other options*)
- Encoding: Can the tool handle Unicode and special symbols? (*yes/no/other options*)
- Language of the interface: What languages are available for the tool's interface? (*English/other options*)

Annotation layers

- Word segmentation: *Does the tool allow to change token segmentation (i.e. merge or split words)?* (*yes/no/other options*)
- Sentence segmentation: *Does the tool allow to change sentence segmentation (i.e. merge or split sentences)?* (*yes/no/other options*)
- Annotation types: *What types of annotation does the tool support?*
  - *Tokens (e.g. part-of-speech tag, morphological features)*
  - *Spans (e.g. names entities)*
  - *Multi-tokens (e.g. multi-word expressions)*
  - *Relations (e.g. dependencies)*
  - *Chains (e.g. coreferences)*
  - *other*
- Annotation scope: *What level of annotation does the tool support?*
  - *Annotation inside sentences*
  - *Annotation across sentences*
  - *Annotation of full sentences / paragraphs / documents*
  - *other*
- Tagset customization: *Does the tool enable the user to modify the list of annotation tags (labels) or define their own?* (*yes/no/other options*)

User interface

- Annotation mode
  - *Keyboard*
  - *Mouse (e.g. drag & drop)*
  - *Touchscreen*
  - *other*
- Keyboard annotation and shortcuts: *Does the tool allow for keyboard shortcuts (macros) to be defined, i.e. any shorthand action that speeds up the selection of tags?* (*yes/no/other options*)
- Annotation display: *What kind of visualisations are available for annotated data?* (*options Image/Table/Raw format (e.g. CONLL-U)/other*)

Workflow

- Automatic pre-annotation: *Does the tool enable automatic pre-annotation of the data (e.g. automatic parsing)?* (*yes/no/other options*)
- Automatic recommendation: *Does the tool enable automatic annotation recommendation based on active learning from the already annotated Tools for Corpus Linguisticsdata (bootstrapping)?* (*yes/no/other options*)
- Collaborative annotation: *Does the tool support collaborative annotation with more than one annotator working on the data?*
  - *No, the tool does not support multiple annotators working on the same data.*
  - *Yes, the tool supports multiple annotators working independently on their own annotations of the same data.*
  - *Yes, the tool supports multiple annotators working together on the same shared annotation of the data.*
  - *other*
- Project management: *Does the tool provide a project management dashboard (e.g. overview of assigned documents and progress)?* (*yes/no/other options*)
- User roles: *Does the tool enable the users to be differentiated by their roles (e.g. admin, manager, annotator, curator)?* (*yes/no/other options*)

Annotation analysis

- Comparison of annotations: *Does the tool provide an interface to display the differences between annotations by different annotators?* (*yes/no/other options*)
- Agreement calculation: *Does the tool calculate inter-annotator agreement scores (e.g. Cohen's Kappa)?* (*yes/no/other options*)
- Querying: *Does the tool support performing corpus queries on top of annotations?* (*yes/no/other options*)
- Mass editing: *Does the tool enable to globally change several annotations at once (e.g. with rules)?* (*yes/no/other options*)

Annotation of Universal Dependencies

- Which of the following types of UD-specific annotations can be edited in the tool? *yes/no/unsure options for each of the following:*
  - Lemmas
  - Part-of-speech tags (UPOS)
  - Features (FEATS)
  - Language-specific tags (XPOS)
  - Multi-word tokens
  - Basic dependencies
  - Enhanced dependencies
  - Empty nodes
  - Lemmas
  - Part-of-speech tags (UPOS)
  - Features (FEATS)
  - Language-specific tags (XPOS)
  - Multi-word tokens
  - Basic dependencies
  - Enhanced dependencies
  - Empty nodes
- Metadata: *Does the tool support Metadata edition (lines with a '#' symbol)* (*yes/no/other options*)
- UD validation: *Does the tool support annotation validation using the official UD validation tool?* (*yes/no/other options*)
- Other annotation layers: *Please describe other annotation layes that are handled by the tool.*

PARSEME

- Discontinuous expressions: *Does the tool enable the annotation of multi-word expressions where sequences of words belonging to the MWE are interrupted by other, non-MWE words or tokens?* (*yes/no/other options*)
- Nested expressions: *Does the tool enable the annotation of multi-word expressions inside longer multi-word expressions?* (*yes/no/other options*)
- Overlapping expressions: *Does the tool enable the annotation of multi-word expressions that share some of the words in the sentence?* (*yes/no/other options*)
- Consistency check: *Does the tool enable the user to check whether the same combination of words is annotated consistently across the corpus (using the Parseme consistency checker or similar tool)?* (*yes/no/other options*)

Final comments

- Example projects: *Please provide examples of annotation projects using the tool.*
- Additional features: *Please provide information on any other functionality of your tool that are important for end users.*
- Final comments: *Please provide any other remarks relating to your tool or this survey that you think are important*

B. List of Surveyed Tools

1. **EXMARaLDA** (Schmidt, 2012): https://www.exmaralda.org
2. **INCEpTION** (Eckart De Castilho et al., 2024): https://inception-project.github.io
3. **Label Studio** (Tkachenko et al., 2020-2025): https://labelstud.io/
4. **TrEd** (Pajas and Fabian, 2019): https://ufal.mff.cuni.cz/tred
5. **Palmyra** (AbuOdeh et al., 2024): https://palmyra.camel-lab.com/
6. **DgAnnotator** (Attardi, n.d.): http://medialab.di.unipi.it/Project/QA/Parser/DgAnnotator/
7. **GitDox** (Zhang and Zeldes, 2017): https://gucorpling.org/gitdox/
8. **FLeX** (Butler and Van Volkinburg, 2007): https://software.sil.org/fieldworks/
9. **CoNLLU-Editor** (Heinecke, 2019): https://github.com/Orange-Opensource/conllueditor
10. **Label Sleuth** (Shnarch et al., 2022): https://www.label-sleuth.org/
11. **TALEN** (Mayhew and Roth, 2018): https://github.com/CogComp/talen
12. **ToolBox** (International, 2019): https://software.sil.org/toolbox/
13. **UD Annotatrix** (Tyers et al., 2017): https://maryszmary.github.io/ud-annotatrix/standalone/annotator.html
14. **ELAN** (Wittenburg et al., 2006): https://archive.mpi.nl/tla/elan
15. **BoAT** (Türk et al., 2019): https://github.com/furkanakkurt1335/boat
16. **Q-CAT** (Brank, 2023): http://hdl.handle.net/11356/1844
17. **PLURAL** (Krishnamurthy et al., 2024): https://plural.iiit.ac.in/conllu-annotator
18. **FLAT** (van Gompel and Reynaert, 2013): https://github.com/proycon/flat
19. **brat** (Stenetorp et al., 2012): http://brat.nlplab.org/
20. **WebAnno** (Eckart de Castilho et al., 2016): https://webanno.github.io/webanno/
21. **ArboratorGrew** (Guibon et al., 2020): https://arborator.grew.fr/