

**PBML**



---

**The Prague Bulletin of Mathematical Linguistics**

**NUMBER 121 JUNE 2024**

---

**EDITORIAL BOARD**

**Editor-in-Chief**

Jan Hajič

**Editorial staff**

Martin Popel

**Editorial Assistant**

Jana Hamřlová

**Editorial board**

Nicoletta Calzolari, Pisa  
Walther von Hahn, Hamburg  
Jan Hajič, Prague  
Eva Hajičová, Prague  
Erhard Hinrichs, Tübingen  
Philipp Koehn, Edinburgh  
Jaroslav Peregrin, Prague  
Patrice Pognan, Paris  
Alexandr Rosen, Prague  
Hans Uszkoreit, Saarbrücken

Published twice a year by Charles University (Prague, Czech Republic)

Editorial office and subscription inquiries:

ÚFAL MFF UK, Malostranské náměstí 25, 118 00, Prague 1, Czech Republic

E-mail: [pbml@ufal.mff.cuni.cz](mailto:pbml@ufal.mff.cuni.cz)

ISSN 0032-6585

eISSN 1804-0462





---

The Prague Bulletin of Mathematical Linguistics  
NUMBER 121 JUNE 2024

---

## CONTENTS

### Articles

<b>Revamping the SLTev Tool for Evaluation of Spoken Language Translation</b> <i>Michelle Elizabeth, Ondřej Bojar</i>	5
<b>Let's Do It Orderly: A Proposal for a Better Taxonomy of Adverbs in Universal Dependencies, and Beyond</b> <i>Flavio Massimiliano Cecchini</i>	15
<b>Instructions for Authors</b>	66





## Revamping the SLTev Tool for Evaluation of Spoken Language Translation

Michelle Elizabeth, Ondřej Bojar

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

---

### Abstract

This article describes recent improvements of SLTev, a tool for automatic evaluation of machine translation, speech recognition and speech translation systems. The changes include the implementation of the COMET score for evaluation of machine translation and spoken language translation outputs as well as a fix for the problematic delay calculation for repeated words which favoured longer segments. Additionally, the system outputs of the IWSLT 2022 shared task have also been evaluated using SLTev and a comparison study with another speech evaluation toolkit, SimulEval, has been done.

---

### 1. Introduction

Spoken Language Translation or SLT is a prominent task in NLP. In the so-called cascaded approach, it involves translation of speech into various languages and combines automatic speech recognition in the source language and then machine translation into the target languages. On the end-to-end approaches, the intermediate transcription in the source language is not explicitly considered. As with any application in NLP, it is necessary to evaluate the results produced with suitable metrics.

SLTev<sup>1</sup> or Spoken Language Translation Evaluation tool (Ansari et al., 2021) performs the evaluation of the outputs of spoken language translations by reporting the quality, latency, and stability of a candidate output based on its time-stamped transcript and reference translation into a target language.

---

<sup>1</sup><https://github.com/ELITR/SLTev>

This article gives an overview of the work done to improve the SLTEv tool. Our main contributions include the addition of a new metric, COMET for evaluating machine translation and the fix for delay calculation so that it takes into account the time-stamps of repeated tokens. Also, an evaluation of the system outputs of IWSLT 2022 (Anastasopoulos et al., 2022) was done using the SLTEv tool, comparing the scores also to the results of SimulEval, the tool used officially for IWSLT 2022.

We first look at the SLTEv tool and the metrics it uses for evaluation. Then we discuss the issues and challenges existing in the current implementation of the tool and the work done to mitigate these issues and improve the tool. Finally, we evaluate the IWSLT 2022 outputs using SLTEv.

## 2. Background

The SLTEv tool is an open source tool for evaluating SLT outputs against reference translations and time-stamped source transcripts. It was developed as part of the European Live Translator (ELITR) project (Franceschini et al., 2020) and provided three metrics namely SacreBLEU (Post, 2018) for measuring quality, Flicker for measuring the stability and Delay for measuring the latency of SLT outputs.

Translation quality is estimated using the SacreBLEU tool applied in three different ways within SLTEv. The first one considers all completed segments as a single joint segment and compares it with the reference which is also considered as a single concatenated segment. The second variant uses mwerSegmenter (Matusov et al., 2005) to compare candidate and segmented reference outputs. The final variant relies on time-span quality and divides the whole document into chunks or segments of a fixed duration which are then separately evaluated using BLEU and also averaged for the score of the whole document.

Flicker assesses the amount of intermediate output updates which can distract the user by counting the number of words after the first difference between two consecutive output updates. Flicker is reported in two variants: average revision count per second and normalised revision count which are described by Ansari et al. (2021).

The final measure is delay which measures the difference between the time that a target word was displayed and an estimate of when it should have been displayed given the source transcript. The delay is calculated using two approaches. The first one is proportional which estimates the timing of each source word based on partial segments in the golden transcript. These times are then passed to the words in the reference translation proportionally along the sequence of words. The second approach uses automatic word alignment between the source and reference translations to account for word order differences across languages.

## 2.1. SLTev vs. SimulEval

Another toolkit that is similar to SLTev and has been developed for evaluating simultaneous translation is SimulEval (Ma et al., 2020). It has been used as the evaluation toolkit for the IWSLT Shared Task since its first edition in 2020.

SimulEval is based on a client-server scheme in which the server sends the source input when requested by the client, receives the translation for evaluation from the client and reports various metrics pertaining to translation quality and latency. The client is composed of an agent and a state. The agent is responsible for executing the system’s policy and the state tracks the necessary information for executing the policy when generating the translation. SLTev on the other hand uses a time-stamped golden transcript in the source language, a reference translation and candidate output in the target language to evaluate the translation quality, latency and stability off-line, i.e. from logs and without running the system again.

SimulEval reports BLEU, TER and METEOR for evaluating translation quality and has adapted Average Proportion, Average Lagging and Differentiable Average Lagging for speech translation. However, it does not support any assessment of output; the evaluated systems are not permitted to their older outputs in any way. SLTev reports stability using the flicker metric along with measuring translation quality and latency using SacreBLEU and delay respectively as described previously. An evaluation of the IWSLT 2022 using SLTev and a comparison of the results with the official SimulEval results is reported later in Section 3.3.

## 3. Improvements to the Current Implementation of SLTev

This section describes some of the issues that the current implementation of SLTev had and the work done to improve the tool.

### 3.1. Delay Computation

In the existing implementation of delay computation, there was an issue in how the time-stamps were assigned to repeated words in a segment. This problem has been reported by Amrhein and Haddow (2022). The following example can be used to explain the problem:

P 13.18 O

P 14.18 O horror,

P 15.18 O horror, terror, horror

C 16.18 O horror, horror, horror.

SLTev assigned the time stamp of 14.18 to all occurrences of the word “horror”, i.e. it assigned the token the time-stamp of its first occurrence even though later updates actually discarded some of these occurrences. When translating longer segments easily consisting of multiple sentences, the likelihood of encountering tokens that were

previously seen increases. In such cases, all of these tokens would be assigned the time-stamp of their first occurrence. Hence, this tends to favour longer segments in the translation process.

### 3.2. COMET

COMET (Rei et al., 2020), which stands for Cross-lingual Optimized Metric for Evaluation of Translation, is a popular neural framework for training multilingual machine translation evaluation models. Typically, COMET models are trained with the objective of predicting quality scores for translations. These scores are usually normalized through a z-score transformation and serve as a valuable metric for ranking translations and systems based on their quality.

The COMET library has several evaluation models and we use the default model Unbabel/wmt22-comet-da (Rei et al., 2022). This model utilizes a reference-based regression methodology and is constructed using the XLM-R framework. It has undergone training on direct assessments from WMT17 to WMT20, offering scores within the 0 to 1 range. A score of 1 indicates a perfect translation.

The Unbabel/wmt22-comet-da model is available on HuggingFace and can be downloaded. A list of dictionaries containing the source, candidate translation and the reference is given as input to the model. It generates scores for each set of source, candidate, reference triplet and also reports an overall system score.

For SLTev, the segmented candidate sentences were concatenated together to form one single segment. The same was done for the source and reference segments in order to generate an overall score for the document. The generated system score which is reported by the model in the range  $[0, 1]$  has been scaled to  $[0, 100]$  in order to be consistent with the SacreBLEU reporting in SLTev.

One issue that was observed during the implementation was that internet connection was necessary in order to download the model to the local system. Currently, this situation is being handled in a way that does not disrupt the flow of the evaluation by handling the exception where the download has failed and moving on to the next metrics.

### 3.3. Evaluation of the IWSLT 2022 System Outputs

The SLTev tool was used to run an evaluation of the outputs by the models submitted to the IWSLT 2022 Simultaneous Speech Translation task. The language pair was English to German and the outputs of five systems namely CUNI-KIT (Polák et al., 2022), FBK (Gaido et al., 2022), HW-TSC (Wang et al., 2022), NAIST (Fukuda et al., 2022) and UPV (Iranzo-Sánchez et al., 2022) were evaluated. Each system has produced outputs for three latency regimes — high, medium and low — determined by a maximum latency threshold measured by Average Lagging on the Must-C tst-COMMON set.



Model	Delay Without Partials	Delay With Partials	SacreBLEU	COMET	Flicker
CUNI-KIT.high	61.24	62.49	<b>25.59</b>	<b>66.57</b>	0.00
CUNI-KIT.medium	71.39	71.84	23.32	64.98	0.00
CUNI-KIT.low	79.72	79.93	17.76	60.01	0.00
FBK.high	90.25	91.91	18.57	54.12	0.00
FBK.medium	58.51	60.20	15.85	51.22	0.00
FBK.low	51.07	52.23	8.62	41.18	0.00
HW-TSC.high	50.31	50.91	9.88	35.09	0.00
HW-TSC.medium	60.24	60.85	9.82	35.3	0.00
HW-TSC.low	65.44	66.1	8.31	33.94	0.01
NAIST.high	77.82	79.37	9.00	38.33	0.00
NAIST.medium	<b>27.79</b>	<b>28.73</b>	9.16	38.00	0.00
NAIST.low	38.15	38.34	7.03	39.75	0.00
UPV.high	238.31	241.12	22.88	62.29	0.00
UPV.medium	148.25	150.31	19.49	59.97	0.00
UPV.low	176.52	179.75	12.81	52.06	0.00

Table 1. SLTeV Evaluation of IWSLT 2022 System Outputs

Latency	High	Medium	Low
Pearson Correlation	0.89	0.864	0.891

Table 2. Pearson Correlation of IWSLT 2022 System Outputs with respect to SimulEval and SLTeV Results

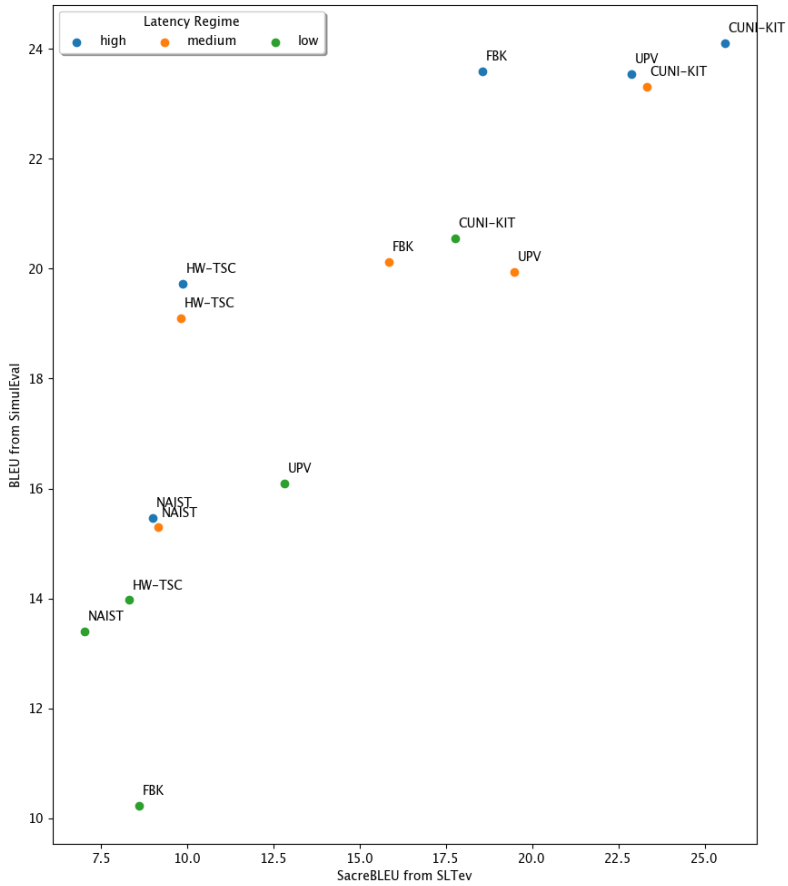


Figure 1. Scatterplot showing the BLEU scores reported by SLTev and SimulEval for IWSLT 2022 systems for three latency regimes

IWSLT 2022 used the SimulEval toolkit (Ma et al., 2020) for evaluating the quality and latency of the submissions. The metrics they used were BLEU for measuring translation quality and average proportion (AP), average lagging (AL) and differentiable average lagging (DAL) for measuring translation latency. Using SLTev, the delay measurements calculated using the partial segments and the one which considers only the completed segments, SacreBLEU, COMET and Flicker have been measured. We can see that the translation quality is highest in each system for the high latency regime except for NAIST which has the best BLEU score for medium latency regime. The CUNI-KIT system in the high latency regime has the best translation quality in terms of both BLEU and COMET. NAIST (medium) has the least average delay whereas UPV systems have the highest delays among all the systems. The metrics are reported in Table 1.

The BLEU scores reported by SimulEval and SLTev in the three latency regimes — high, medium and low — have also been reported as a scatter plot in Figure 1. We can observe that the general trend is that SLTev has scored the systems lower than the scores reported by SimulEval except for CUNI-KIT in the high latency regime. HW-TSC and NAIST have been scored much lower by SLTev than by SimulEval, a difference in the range of approximately 5–10 points. This can also be seen for FBK though not to the extent of HW-TSC and NAIST. CUNI-KIT and UPV have similar scores reported by SLTev and SimulEval. Table 2 reports the Pearson correlation for scores of the systems in the three latency regimes.

#### 4. Conclusion and Future Work

SLTev is a comprehensive tool for evaluating the quality of spoken language translation. We wish it became the standard toolkit with a wide adoption.

The work done reported in this article is just the beginning, there is more room for improvement. The implementation of COMET score can be enhanced further by reporting segment-level scores as well. It would also be beneficial to figure out how to download the COMET model available in HuggingFace locally when installing SLTev and not having to rely on a stable Internet connection to generate the score. The bug fix for delay computation should give more accurate results and will no longer favour longer segments since the time-stamps of repeated tokens are being accurately calculated. Additional metrics relevant to translation can be added including average lagging and chrF3 (Popović, 2015).

The tool can be made more versatile by making it platform independent. Right now, it relies on `mwerSegmenter` which can only be run on Linux systems. Word-error-rate-based segmentation is thus not preformed for quality evaluation when used on other platforms. The readability and reusability of the code can be improved by using more Pythonic constructions. Also, writing and maintaining unit tests, implementing a proper error handling module and detailed logging are some other ways in which the tool can be made more user and developer friendly.

## Bibliography

- Amrhein, Chantal and Barry Haddow. Don't Discard Fixed-Window Audio Segmentation in Speech-to-Text Translation. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 203–219, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics. URL <https://aclanthology.org/2022.wmt-1.13>.
- Anastasopoulos, Antonios, Loïc Barrault, Luisa Bentivogli, Marcelly Zanon Boito, Ondřej Bojar, Roldano Cattoni, Anna Currey, Georgiana Dinu, Kevin Duh, Maha Elbayad, Clara Emmanuel, Yannick Estève, Marcello Federico, Christian Federmann, Souhir Gahbiche, Hongyu Gong, Roman Grundkiewicz, Barry Haddow, Benjamin Hsu, Dávid Javorský, Věra Kloudová, Surafel Lakew, Xutai Ma, Prashant Mathur, Paul McNamee, Kenton Murray, Maria Nädejde, Satoshi Nakamura, Matteo Negri, Jan Niehues, Xing Niu, John Ortega, Juan Pino, Elizabeth Salesky, Jiatong Shi, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Marco Turchi, Yogesh Virkar, Alexander Waibel, Changhan Wang, and Shinji Watanabe. Findings of the IWSLT 2022 Evaluation Campaign. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 98–157, Dublin, Ireland (in-person and online), May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.iwslt-1.10. URL <https://aclanthology.org/2022.iwslt-1.10>.
- Ansari, Ebrahim, Ondřej Bojar, Barry Haddow, and Mohammad Mahmoudi. SLTEV: Comprehensive Evaluation of Spoken Language Translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 71–79, Online, April 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.eacl-demos.9. URL <https://aclanthology.org/2021.eacl-demos.9>.
- Franceschini, Dario, Chiara Canton, Ivan Simonini, Armin Schweinfurth, Adelheid Glott, Sebastian Stüker, Thai-Son Nguyen, Felix Schneider, Thanh-Le Ha, Alex Waibel, Barry Haddow, Philip Williams, Rico Sennrich, Ondřej Bojar, Sangeet Sagar, Dominik Macháček, and Otakar Smrž. Removing European Language Barriers with Innovative Machine Translation Technology. In *Proceedings of the 1st International Workshop on Language Technology Platforms*, pages 44–49, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-64-1. URL <https://aclanthology.org/2020.iwlt-1.7>.
- Fukuda, Ryo, Yuka Ko, Yasumasa Kano, Kosuke Doi, Hirotaka Tokuyama, Sakriani Sakti, Katsuhito Sudoh, and Satoshi Nakamura. NAIST Simultaneous Speech-to-Text Translation System for IWSLT 2022. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 286–292, Dublin, Ireland (in-person and online), May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.iwslt-1.25. URL <https://aclanthology.org/2022.iwslt-1.25>.
- Gaido, Marco, Sara Papi, Dennis Fucci, Giuseppe Fiameni, Matteo Negri, and Marco Turchi. Efficient yet Competitive Speech Translation: FBK@IWSLT2022. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 177–189, Dublin, Ireland (in-person and online), May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.iwslt-1.13. URL <https://aclanthology.org/2022.iwslt-1.13>.
- Iranzo-Sánchez, Javier, Javier Jorge Cano, Alejandro Pérez-González-de Martos, Adrián Giménez Pastor, Gonçal Garcés Díaz-Munío, Pau Baquero-Arnal, Joan Albert Silvestre-Cerdà, Jorge Civera Saiz, Albert Sanchis, and Alfons Juan. MLLP-VRAIN UPV systems for

- the IWSLT 2022 Simultaneous Speech Translation and Speech-to-Speech Translation tasks. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 255–264, Dublin, Ireland (in-person and online), May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.iwslt-1.22. URL <https://aclanthology.org/2022.iwslt-1.22>.
- Ma, Xutai, Mohammad Javad Dousti, Changhan Wang, Jiatao Gu, and Juan Pino. Simuleval: An evaluation toolkit for simultaneous translation. In *Proceedings of the EMNLP*, 2020. doi: 10.18653/v1/2020.emnlp-demos.19.
- Matusov, Evgeny, Gregor Leusch, Oliver Bender, and Hermann Ney. Evaluating Machine Translation Output with Automatic Sentence Segmentation. In *Proceedings of the Second International Workshop on Spoken Language Translation*, Pittsburgh, Pennsylvania, USA, October 24-25 2005. URL <https://aclanthology.org/2005.iwslt-1.19>.
- Polák, Peter, Ngoc-Quan Pham, Tuan Nam Nguyen, Danni Liu, Carlos Mullov, Jan Niehues, Ondřej Bojar, and Alexander Waibel. CUNI-KIT System for Simultaneous Speech Translation Task at IWSLT 2022. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 277–285, Dublin, Ireland (in-person and online), May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.iwslt-1.24. URL <https://aclanthology.org/2022.iwslt-1.24>.
- Popović, Maja. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-3049. URL <https://aclanthology.org/W15-3049>.
- Post, Matt. A Call for Clarity in Reporting BLEU Scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels, October 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6319. URL <https://www.aclweb.org/anthology/W18-6319>.
- Rei, Ricardo, Craig Stewart, Ana C Farinha, and Alon Lavie. COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.213. URL <https://aclanthology.org/2020.emnlp-main.213>.
- Rei, Ricardo, José G. C. de Souza, Duarte Alves, Chrysoula Zerva, Ana C Farinha, Taisiya Glushkova, Alon Lavie, Luisa Coheur, and André F. T. Martins. COMET-22: Unbabel-IST 2022 Submission for the Metrics Shared Task. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 578–585, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics. URL <https://aclanthology.org/2022.wmt-1.52>.
- Wang, Minghan, Jiaxin Guo, Yinglu Li, Xiaosong Qiao, Yuxia Wang, Zongyao Li, Chang Su, Yimeng Chen, Min Zhang, Shimin Tao, Hao Yang, and Ying Qin. The HW-TSC’s Simultaneous Speech Translation System for IWSLT 2022 Evaluation. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 247–254, Dublin, Ireland (in-person and online), May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.iwslt-1.21. URL <https://aclanthology.org/2022.iwslt-1.21>.

**Address for correspondence:**

Ondřej Bojar

bojar@ufal.mff.cuni.cz

Institute of Formal and Applied Linguistics

Faculty of Mathematics and Physics,

Charles University

Malostranské náměstí 25

118 00 Praha 1, Czech Republic



## Let's Do It Orderly: A Proposal for a Better Taxonomy of Adverbs in Universal Dependencies, and Beyond

Flavio Massimiliano Cecchini

KU Leuven

---

### Abstract

With this paper, we hope to highlight the critical issues and “bad practices” that are currently observed in the representation of adverbs in the annotation framework of Universal Dependencies (where they are ideally identified with the universal part of speech ADV), which themselves more generally mirror the conspicuous lack of systematic definitions of this word class in traditional grammars. This fact, on the one side, hampers a useful and meaningful linguistic description of adverbs in Universal Dependencies’ treebanks and beyond, and, on the other side, has unfortunate consequences for linguistic research, such as too confused, impractical results when, for example, querying a treebank for all ADV-tagged words. Therefore, with the aid of minimal data analysis and numerous examples, this contribution tries to raise awareness about this issue, and proposes a revised, more typologically grounded and general framework for the classification of adverbs in Universal Dependencies, and more broadly advocates a more flexible representation of the interplay between word classes and syntactic functions, through the comprehensive concept of “transposition” or “transfer”. While grounded in the Universal Dependencies formalism, the scope of the discussion of this paper is by no means limited to it, and might be of interest to any practitioner of (computationally-oriented) linguistic annotation.

---

### 1. Introduction

Despite the historical and consolidated use in (Western) linguistics of the notion of “adverb” as a part of speech, reaching at least as far as grammatical studies in Ancient Rome,<sup>1</sup> surprisingly this word class still seems to lack a clear and systematic

---

<sup>1</sup>See e. g. Pinkster (1972, §4).

general definition in literature. On the contrary, at a practical level it often appears as a confusing concoction of different morphosyntactic elements. This might sometimes even be pointed out in general works, as by Schachter and Shopen (2007, p. 19f):

“The label *adverb* is often applied to several different sets of words in a language, sets that do not necessarily have as much in common with one another, either notionally or grammatically.”

or also by Anderson et al. (2003) s. v. *Adverb* (*Grammar*) in the *SL Glossary of Linguistic Terms* (which is referred to by the Universal Dependencies’ guidelines):

“An adverb is a lexical category whose members have the same syntactic distribution [...]”

followed immediately later, contradictorily, by:

“The general class *adverb* is a mixture of very different kinds of words, which cover a wide range of semantic concepts and whose syntactic distribution is disparate.”

On this note, even in the introduction to a comprehensive treatment of the more general concept of adverbial construction such as van der Auwera and Ó Baoill (1998), we find the following “statement of disengagement” (van der Auwera, 1998, §3):

“The notion of adverb or adverbial has not figured prominently in discussions of linguistic typology. This has at least three reasons. First, the category itself is elusive: it is not clear what the defining or prototypical features of adverbs and adverbials are [...] [The book] does not attempt to throw any new light on the definition of adverb and adverbial, but merely presupposes a classical understanding of «adverb» as the word-level adverbial expression, and of «adverbial» as a syntactically optional modifier of primarily nonnominal constituents.”

And so it might well be the case that some works primarily dealing with adverbs do not even define them explicitly, cf. Cinque (1999).

This situation is most evident in language-specific traditions, where, even though this fact is often highlighted,<sup>2</sup> all these heterogeneous elements are then still kept together, as if by a sort of rationalisation for want of alternatives. Others still, apparently mostly with respect to non-European languages, might eschew the topic completely.<sup>3</sup> We claim that this state of disarray is reflected by current annotation practices in Universal Dependencies, too, where the problem becomes acute because of the desire for a typologically valid formalism. We recall that Universal Dependencies

---

<sup>2</sup>As e. g. for German in Barz et al. (2009, §837): “Adverbs represent a heterogeneous and therefore difficultly definable word class” (translated from German by the author).

<sup>3</sup>For example, in the comprehensive presentation of Oceanic languages in Lynch et al. (2011), hardly any of the 43 grammar sketches ventures into describing adverbs as a standalone category, which are only shortly mentioned marginally to other “lexical morpheme classes” when portraying Proto-Oceanic (p. 63: “There may also have been a small class of underived adverbs”).



“is at the same time a framework for crosslinguistically consistent morphosyntactic annotation, an open community effort to create morphosyntactically annotated corpora for many languages” (de Marneffe et al., 2021, §1)

and, at the time of writing, its latest release <sup>4</sup> v2.13 (November 2023) contains 259 treebanks for 148 languages and 31 language families.

In the following we propose, if not the ultimate definition of what an adverb is, at least an overview of the problem and directions for an improved annotation of this word class (or the corresponding syntactic function, as it will be discussed) in Universal Dependencies. Section 2 introduces the Universal Dependencies (henceforth also UD) formalism for readers who might not be so familiar with it, as it is the main setting of this paper. Section 3 describes the problem generally and more specifically in UD; Section 4 brings the perspective of this work into focus; Section 5 puts forward a gradient of increasingly “innovative” solutions by also, inevitably, going beyond merely adverbs; Section 6 concludes the paper.

**A note about data** The reference for the cited data is the latest release of the UD project, v2.13, published on November 15<sup>th</sup>, 2023 (Zeman et al., 2023). The languages and treebanks thereof considered in this paper are, ordered by their number of syntactic words:

- German HDT (Borges Völker et al., 2019; 3 455 580 words)
- Latin, all five treebanks, i. e.
  - IT-TB (Cecchini et al., 2018; Passarotti, 2019; 450 517 words)
  - LLCT (Cecchini et al., 2020a; 242 411 words)
  - PROIEL (cf. Eckhoff et al., 2018; 205 566 words);
  - UDante (Cecchini et al., 2020b; 55 519 words)
  - Perseus (cf. Bamman and Crane, 2011, and also Gamba and Zeman, 2023; 29 221 words)
- English EWT (Silveira et al., 2014; 254 862 words)
- Bulgarian БТБ (cf. Osenova and Simov, 2004; 156 149 words)
- Greek (modern) GDT (Prokopidis and Papageorgiou, 2017; 63 441 words)

A further set of five other languages, less or possibly not at all (marked with \*) familiar to the author, is selected randomly with the aim of widening the typological perspective, especially in §5:

- Armenian\* (western) ArMTDP (Yavrumyan, 2019; 122 907 words)
- Faroese FarPAHC (40 484 words)
- Thai\* PUD (cf. Zeman et al., 2017; 22 322 words)
- K’iche’\* IU (Tyers and Henderson, 2021; 10 013 words)

---

<sup>4</sup><http://hdl.handle.net/11234/1-5287>.

– Amharic\* ATT (Seyoum et al., 2018; 10 010 words)

For all details, the respective documentation pages, reachable via the main UD website,<sup>5</sup> can be referred to. The selected treebanks all currently conform to the validation standards<sup>6</sup> of UD or are considered to be “legacy” with minor errors, mostly related to recent amendments in the guidelines.<sup>7</sup> When more than one treebank is present for a language, the largest one is chosen. All examples in the text refer to the sentence identifiers (`sent_id` comment line) in the CoNLL-U files. For all tags of parts of speech and dependency relations, the reference are the respective entries in the UD guidelines,<sup>8</sup> which we sometimes cite in this paper. In the examples, the focused element is highlighted in bold, while other elements of interest are underlined; forms are in italics, while lemmas are in small capitals. All words are transcribed using recognised standards.

All code and data used for this work (especially in §3.2, 4.4, and 4.4.2) are made available in a public GitHub repository,<sup>9</sup> so as to make our results reproducible, and to allow for updates with respect to future UD releases.

## 2. Overview of the Universal Dependencies formalism

Even if the issue discussed in this paper is of general linguistic interest, it is true that some points (especially in §3.1, 3.2, and §4.4.2) and the concrete design of the proposal of a new taxonomy (as in §5) are particularly rooted in the formalism of the Universal Dependencies project. It is thus in order to give a short introduction to its basic principles here, highlighting those that are of most interest to this work, and to supply a brief explanation of some of the UD labels which are most relevant to the present discussion. Apart from the documentation that can be found on the project’s website, the main cited introductory papers are Nivre et al. (2020) and de Marneffe et al. (2021). Moreover, in Table 1 we give a synopsis and a brief comment on UD labels, focusing on those most mentioned and used throughout this article: there, the acronym *POS* stands for “part of speech”, the abbreviation *deprel* for “(syntactic) dependency relation”, while *feature* is short for “morpholexical feature”.

At its core, UD is

“a project that is developing cross-linguistically consistent treebank annotation for many languages” (Nivre et al., 2020, §1)

---

<sup>5</sup><https://universaldependencies.org/>

<sup>6</sup><http://quest.ms.mff.cuni.cz/udvalidator/cgi-bin/unidep/validation-report.pl>

<sup>7</sup><https://universaldependencies.org/changes.html>.

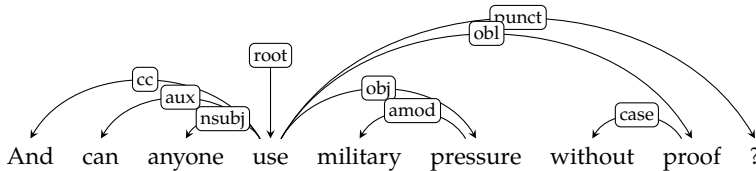
<sup>8</sup><https://universaldependencies.org/guidelines.html>

<sup>9</sup><https://github.com/Stormur/OrderlyAdverbs>

With regard to syntax, it makes use of a dependency-based paradigm with a lexicalist approach, taking (syntactic) words as its base units (so words depend on each other, and syntactic trees do not include nodes for phrases, collocations, or similar) and giving content words primacy over function words. All of this can be exemplified by the simple sentence

- (1) *And can anyone use military pressure without proof?*  
 – English EWT weblog - blogspot.com\_grandpasgripes\_20060413051000\_ENG\_-20060413\_051000-0011

which is analysed as follows:



Here we observe that:

- the sentence is headed by the predicate, here realised by the verbal phrase *can use*;
- in this English treebank, the modal verb form *can* is considered an auxiliary, and so inside this phrase depends on *use* as *aux* (‘auxiliary’), making *use* the head of the whole sentence, or *root*;
- the co-ordinating conjunction *And* introducing the sentence depends on the predication, as *cc* (‘co-ordinating conjunction’);
- in the phrase *military pressure*, the modifier *military* depends on the modified noun as *amod* (‘adjectival modifier’), and thus *pressure* is the head depending on the predicate as *obj* (‘object’);
- in the phrase *without proof*, it is the preposition (function word) *without* that depends on the noun (content word) *proof* as *case* (‘case marking’), and thus *proof* depends on the predicate as head of this *obl* (‘oblique’) argument;<sup>10</sup>
- finally, we note that punctuation is assigned a purely formal “relation” *punct*, attached so as not to create non-projective (i. e. discontinuous) configurations, usually to the head of the phrase it appears the nearest, here the *root use*.

<sup>10</sup>The dichotomy between *core* (basically *A*, *P* and *s* in the terminology described e. g. by Haspelmath, 2011) and *oblique* arguments (all the others), as opposed for example to that between (obligatory) *complements* and (facultative) *adjuncts*, is another basic tenet of *UD*. We are not dealing with this complex topic here, but point to Andrews (2007) as one of the main introductory works on the subject.

Some of these annotation choices contrast with other widespread formalisms, such as those found in the Prague Dependency Treebank<sup>11</sup> Hajič (1998); Hajič et al. (2017, 2020), or those assumed in Osborne (2019), where e.g. adpositions and auxiliaries head their respective phrases.

A complete annotation in UD considers, besides syntactic relations, also a lexico-semantic layer with lemmatisation, a morphosyntactic layer with part-of-speech tagging, and a morphological layer with the assignment of morpholexical traits in feature-value pairs. So, in our sample sentence we will have e.g. the lemma *and* for the form *And* (case-normalised), the part of speech ADP ‘adposition’ for *without*, a morphological feature Number with value Sing (‘singular’) for *proof*<sup>12</sup> and a lexical one PronType (‘pronominal type’<sup>13</sup>) with value Ind (‘indefinite’) for *anyone*.

UD’s system acts according to two important distinctions, defining two kinds of “main axes” along which parts of speech (cf. Tables 1, 2 and paragraph below), but also dependency relations, can be laid out. The first one is a “typology of phrasal units” (de Marneffe et al., 2021, §2.1.2) identifying “nominals”, “clauses” and “modifiers”: so a nominal is an element which refers to an entity and is expected to appear as the argument (subject, object, oblique...) of a predicate, and, in terms of parts of speech, we can think of nouns (NOUN in UD) as the prototypical nominals, with the corresponding relations nsubj ‘nominal subject’, obj ‘object’, and so on; similarly verbs (VERB) will be the prototypical roots of clauses, with relations such as csubj ‘clausal subject’ and ccomp ‘clausal complement’, and adjectives (ADJ) the prototypical modifiers, depending as amod ‘adjectival modifier’. The second distinction regards grammaticalisation (de Marneffe et al., 2021, §2.2.1), which implies a cline between two poles. On the one pole, we have autosemantic or content words,<sup>14</sup> i. e. words with an independent meaning. On the other pole, we find synsemantic, grammatical or function words, i. e. elements which do not have an independent meaning and often serve as grammatical markers specifying the syntactic roles or grammatical categories of other elements, and, if this is the case, in UD’s system they always depend on those elements by means of special relations. Some word classes are easily recognised as tending to one of the two poles: so (proper) nouns (NOUN and PROPN), full verbs (VERB)

---

<sup>11</sup>For the latest version, refer to Hajič et al. (2020).

<sup>12</sup>We note that it might be argued that English actually does not morphologically express a singular number in nouns, as opposed to a marked plural, e.g. *house-∅* vs. *house-s*, but this English treebank chooses to annotate it, probably recognising a paradigmatic contrast.

<sup>13</sup>Actually extending also to other parts of speech than just PRON ‘pronoun’.

<sup>14</sup>Sometimes also “lexical words”, but we will avoid this terminology in this sense, as it is too loaded with different meanings, and thus ambiguous; cf. Haspelmath (2024). Here, in particular, it overlaps with the notion of lexical word as opposed to e.g. symbols or punctuation marks, or of lexical features as opposed to morphological ones.

Label	Description	Layer	Grammaticalisation	Comment
ADJ	adjectives	POS	autosemantic	The prototypical modifiers
ADP	adpositions	POS	synsemantic	Comprising prepositions, postpositions and circumpositions
ADV	adverbs	POS	underdefined	The main topic of this paper
CCONJ	co-ordinating conjunctions	POS	synsemantic	Distinct from subordinating conjunctions SCONJ
DET	determiners	POS	synsemantic	Synsemantic counterparts of adjectives, often deictic
NUM	numerals	POS	synsemantic	In UD substantially coinciding with “cardinal number”, and as such effectively a subclass of DET
PART	particles	POS	synsemantic	Residual function words; underdefined category
PRON	pronouns	POS	synsemantic	Synsemantic counterpart of nouns, referring deictically to entities instead of naming them
SCONJ	subordinating conjunctions	POS	synsemantic	Distinct from co-ordinating conjunctions CCONJ
advcl	adverbial clause modifier	deprel	autosemantic	Clause acting as a non-core modifier of the predicate; clause equivalent of advmod
advmod	adverbial modifier	deprel	underdefined	Formally required for ADV; see §3.1
advmod:emph	emphasising word	deprel	synsemantic	Subtype of advmod used by some treebanks; see §5.1.2
aux	auxiliary	deprel	synsemantic	Required for grammatical verbal elements in periphrastic verbal predicates
case	case marking	deprel	synsemantic	Required by adpositions associated to nominals, which are equated to morphological case markers
cc	co-ordinating conjunction	deprel	synsemantic	Required by co-ordinating conjunctions
det	determiner	deprel	synsemantic	Required by modifying determiners
discourse	discourse element	deprel	underdefined	Relation of pragmatic nature
mark	marker	deprel	synsemantic	Required by subordinating conjunctions and adpositions introducing an argument clause
obl	oblique nominal	deprel	autosemantic	Any argument of the predicate which is not the subject or an object
PronType=Dem	demonstrative pronominal type	feature	lexical	Element with deictic function, e.g. <i>that, this</i>
PronType=Int	interrogative pronominal type	feature	lexical	Element appearing in questions, e.g. <i>what, who</i>
PronType=Rel	relative pronominal type	feature	lexical	Element with anaphoric function appearing in so-called relative clauses, e.g. <i>which, whose</i>
VerbForm	form of verb or deverbative	feature	morphological	“External”, non-verbal syntactic role of a verb form still acting as a predicate

Table 1. Synopsis and brief explanation of the main UD labels used throughout this paper.

and adjectives (ADJ) are instances of content words; conversely, auxiliaries (AUX), adpositions (ADP), conjunctions (CCONJ and SCONJ) and generic particles (PART) supply examples of prototypical function words. While this distinction is clear in much of the literature about UD, as in de Marneffe et al. (2021) *in primis*, there remain some word classes whose status of grammaticalisation is more controversial: determiners (DET), numerals (NUM), and pronouns (PRON). In the framework of this work, with good reasons but possibly against the general consensus (especially for what concerns pronouns, or pro-forms in general), we consider all these three classes to have a synsemantic nature, as discussed in the following paragraph. This is especially reflected in Tables 1 and 2, and in Section 3.2.

**Grammaticalisation status of DET, NUM and PRON in UD and in this paper** The main issue with the three UD word classes DET, NUM and PRON is that they can occur (though at different degrees across languages) in the same contexts as NOUNS, i. e. as heads of arguments of the predicate. While this might favour the identification of their members as autosemantic, there are indeed some fundamental reasons to do otherwise. One is internal to UD’s system as defined in the guidelines: DET and NUM have dedicated dependency relations, respectively *det* and *nummod*. The former is classified under “Function Words”;<sup>15</sup> for NUM it is said:

“A numeral is a word, functioning most typically as a determiner, adjective or pronoun [...] Note that cardinal numerals are covered by NUM whether they are used as determiners or not [...] Other words functioning as determiners (including quantifiers such as *many* and *few*) are tagged DET.”

This goes to show how both DET and NUM are understood as synsemantic word classes in UD’s system, and moreover NUM as a subclass of DET. However, a different treatment seems to be in place for the class PRON, judging e. g. from de Marneffe et al. (2021, §2.2.1):

“We expect the words that enter into the main syntactic relations to be autosemantic [...] typically verbs, nouns, adjectives, or adverbs, as well as corresponding pro-forms with a contextually determined referential meaning.”

But this is at odds with the classification of determiners and numerals as synsemantic: in fact, we argue that their being “contextually determined” is an aspect of synsemanticity, which is common to determiners and pronouns (and pro-forms in general): literally, not possessing an independent meaning, be this from a syntactic (ADP, AUX, CCONJ, SCONJ, PART) or a semantic (DET, NUM, PRON) point of view. In this paper, we will use this broader, at least with respect to de Marneffe et al. (2021), notion of

---

<sup>15</sup><https://universaldependencies.org/u/dep/index.html>

synsemanticity. We notice that this stance is apparently supported by passages of UD's guidelines such as (about PRONs):

“[i]t is not always crystal clear where pronouns end and determiners start.”

Moreover, words of the PRON class can depend on others with the synsemantic relation *det*. In fact, we do expect word classes of the same type of “phrasal unit” to have very similar if not identical distributions, as it happens for ADJs and DETs (usually conflated by traditional grammars).<sup>16</sup> In any case, we contend that the main point of the present work is not affected by sharing or not this position: the mixed nature of the ADV class will be made clear in Section 3.2.

If, then, DET, NUM and PRON are acknowledged as having a synsemantic nature, with regard to parts of speech the two main axes of type of phrasal unit and grammaticalisation concur to define three main “dyads” in UD's system: NOUN vs. PRON for nominals, VERB vs. AUX for clauses, and ADJ vs. DET for modifiers. Temporarily disregarding ADV, the other 10 labels in UD's part-of-speech system represent either subcategories of the main ones (PROPN ‘proper nouns’ with respect to NOUN, and NUM ‘numerals’ with respect to DET), purely grammatical elements belonging to no particular type of phrasal unit (ADP ‘adpositions’, CCONJ ‘co-ordinating conjunctions’, SCONJ subordinating conjunctions, PART ‘particles’), non-categorisable elements (INTJ ‘interjections’ and X ‘other’), and non-lexical elements (PUNCT ‘punctuation marks’ and SYM ‘symbols’). In the rest of this work, we argue that the position of the category ADV ‘adverbs’ and the relation *advmod* ‘adverbial modifier’ inside UD's schema is under-defined with respect to many layers, in a way that undermines the soundness of the whole system, so that some reorganisation needs to take place.

### 3. The “adverb” issue: from the beginnings to Universal Dependencies

The notion of “adverb”, though being a staple of part-of-speech tagsets,<sup>17</sup> remains problematic when it comes to pinning down a precise and useful definition of it, particularly in view of (automated) linguistic annotation. Also, the problem encompasses and interweaves at the same time the definitions of parts of speech and dependency relations, as will be examined in Section 4. All of this makes it difficult to even find truly comprehensive examinations of this word class in literature. We do find some, like Ramat and Ricca (1994), whose endeavour is partly shared by the present work, and where the definition of “prototypical adverb” closely resembles our class of “core” adverbs (§4.3): but even in this case, we claim that the levels of form and function are (possibly due to a lack of an annotation formalism of reference,

<sup>16</sup>We further notice that other sources describing UD's system, such as Zeman (2018, §3.4, Table 3.15), are also very explicit as to the synsemantic position of PRONs as opposed to NOUNs.

<sup>17</sup>We refer to Zeman (2018, §3) for an overview, where a tag for “adverb” always appears, starting from RB in the Penn Treebank's tagset (Marcus et al., 1993, §2.2).

as here UD) ultimately mixed as in the kind of contextual annotation that we criticise (cf. §4.1), in a sense projecting syntactic function back to the word class of any (lexical) element realising it. This is the only way we see how, e. g. in Latin, *merito* ‘having deserved (it)’, i. e. ‘with merit’ can be considered an adverb, while in “*maximo merito* ‘with the highest merit’ there is no adverbial lexeme!” (Ramat and Ricca, 1994, §1, p. 290). Here, the derivational status of some elements is mentioned, but finally disregarded: a form like *merito* can only be labelled as invariable if it is arbitrarily detached from the paradigm it belongs to (in this case that of the verb *mereo* ‘to deserve’; cf. §4.4.2). Further, in that work the focus seems rather to lie on a semantic classification of adverbial elements, with which we are not directly concerned here.

Trying to go to the roots of the issue at hand, since in Western linguistics the term *adverb* itself (as a calque from Greek *epírrhēma*, literally ‘(what is) beside the verb’) and its theoretical definition originate from the Ancient Roman grammatical tradition,<sup>18</sup> it can help our understanding of the current situation to look at how adverbs are treated in modern literature about Latin. Latin is known for an extensive grammatical tradition, significantly influenced by the earlier Greek one (cf. Law, 2003, §4.3ff), and itself having influenced most other Western grammatical traditions, which themselves nowadays contribute to shaping UD. This has as a consequence that what we find in dedicated literature about Latin is also mirrored by the grammatical literature of many other languages. One can thus read<sup>19</sup> there as a constant about adverbs:

“Under the label «adverb», morphologically and **syntactically** very different terms are stowed”<sup>20</sup> (Marouzeau, 1949, p. 11)

“absence of formulation of criteria” (Pinkster, 1972, p. 33)

“[...] building up a very heterogeneous set, formally as well as **functionally**” (Ricca, 2010, p. 109)

“Adverbs constitute a very heterogeneous word class, both in respect to their forms and in respect to their meanings and **syntactic** functions” (Pinkster, 2015, p. 65)

This vagueness is exemplarily reflected in the UD guidelines for the part of speech ADV:

“Adverbs are words that typically modify verbs for such categories as time, place, direction or manner. They may also modify adjectives and other adverbs, as in *very briefly* or *arguably wrong*.”

Which is more or less equivalent to what we read in papyri of the first century AD:

<sup>18</sup>On the topic see e. g. Pinkster (1972, §3)

<sup>19</sup>All cited sources are standard works about adverbs in Latin, or Latin in general, and will be cited again in this work; Bos (1967); Van Laer and Fruyt (2009) can also be added to the list for further reference. Boldfaces in quotations are by the author.

<sup>20</sup>Translated from French by the author.



“The adverb is an uninflected word form which is placed before or after the verb and not compounded with it, signifying quantity, quality, time, place, negation, agreement, prohibition, exhortation, interrogation, exclamation, comparison or doubt.” (cf. Law, 2003, p. 57)

Problems arise when we inspect more closely the elements given as examples in UD’s guidelines: one immediately notices that some which are usually tagged as ADV in the treebanks, such as *up*, *where* and others, can hardly be said to modify adjectives (?an *up* good book, ?a *where* good book) or adverbs,<sup>21</sup> as opposed to *exactly* (or any other deadjectival adverb in *-ly*; cf. §4.4), which we can observe as the modifier (see §4.1 for a definition) of an adjective (ADJ) in a copula in

- (2) *I’m not **exactly** certain*  
 – English-EWT email-enronsent35\_01-0081

but also of a head noun (NOUN) in a noun (here, prepositional) phrase in<sup>22</sup>

- (3) *I’m going to go at **exactly** the same time*  
 – English EWT email-enronsent34\_01-0007.

This latter possibility is not mentioned in the guidelines, which later even feel the need to single out a specific type of adverb, “pronominal adverbs” (§5.1.4).

This very brief overview shows that the allegedly unitary ADV class is actually forced to contain (morpho)syntactically very different elements, analogously to what is reported for Latin. This contrasts, for example, with UD’s definition of another class of modifiers, ADJ (“Adjectives are words that typically modify nouns and specify their properties or attributes.”), whose behaviour is decidedly more uniform, whereas variation rather takes place at a semantic level<sup>23</sup> (with which UD is not really concerned). We thus advocate the need for a rethinking of ADV in UD, which possibly involves other parts of speech, too, in order to achieve a more coherent, linguistically representative annotation. We now proceed to highlight more UD-specific critical points of the current ADV class.

---

<sup>21</sup>In fact, in no English UD treebank (as of v2.13) words with lemma WHERE or UP are annotated as modifying (i. e. with dependency relation advmod, see §3.1) an adjective (ADJ), or an adverb (ADV) which is not part of a predicate, or is itself a so-called “pronominal adverb” (*here*, *there*, *front*; see below and §5.1.4), with the single exception of *close* (in the phrase *get up close* in English-EWT answers-20111108094831AAn0jgr\_ans-0014), which however oscillates between adjectival and adverbial status.

<sup>22</sup>The fact that the modified element here is *time* and not the predicate *am going to go* is unambiguously signalled by the position of *exactly* inside the prepositional phrase, i. e. after *at*, and outside the noun phrase, i. e. before the determiner *the*.

<sup>23</sup>For example, adjectives are known to distribute over a semantic hierarchy, from more “nounlike” to more “verblike”, which can govern among other things the expression of degree or the use of a copula rather than verbal morphology (cf. Stassen, 2003, §5.2, p. 168ff).

### 3.1. Syntactic relations and part of speech of adverbs in UD: redundancy

Formally, even if not completely explicitly, UD requires ADVs in their modifying function to be paired with the *advmod* dependency relation;<sup>24</sup> conversely, *advmod* should only be assigned to words tagged as ADV. This is substantiated by the warning

“«*advmod*» should be «ADV» but it is \_”

issued by the UD validator script.<sup>25</sup> Nonetheless, in the same script we read that also parts of speech ADJ, CCONJ, DET, PART and SYM are admitted. So, on the one side UD enforces redundancy, similarly as for DET & *det* (part of speech and relation for determiners; see Table 1 in §2), but on the other side it “tolerates” some annotation practices that are apparently common across treebanks,<sup>26</sup> introducing a covert and unexpected behaviour. In general, we argue that a similar redundancy is not desirable, and that combinations of non-ADV words and the *advmod* dependency relation could be naturally included in UD’s framework: such a possibility would avoid obscuring transparent patterns such as the English ADJ *EXACT* producing the ADV *EXACTLY* by means of the productive<sup>27</sup> *-ly* suffix, as a consequence of suppressing morphological features not pertinent to the imposed class of ADV in the annotation (§4.4). This alternation might be seen as a rather inconsistent annotation practice, against the UD *desideratum* of “rapid, consistent annotation” (de Marneffe et al., 2021, §5, point 3). Sections 4.1 and 4.4 deal with this issue in-depth, and with the repercussions of the fixed ADV/*advmod* coupling on the representation of lexicon.

#### 3.1.1. Subtypes of *advmod*

Beyond the basic *advmod* relation, 47 languages in UD define various subtypes of this dependency label. Some appear rather widespread, like *emph* (27 languages, including Latin with *IT-TV*, *LLCT*, UDante, and Western Armenian), others more specific, e. g. *cau*, *deg*, *eval*, *foc*, *freq*, *lfrom*, *lmp*, *lto*, *mmod*, shared by a Uralic block (Erzya, Komi Permyak & Zyrian, Moksha, Skolt Sami; cf. Partanen et al., 2018). Some are mostly semantic in nature, e. g. *tto* ‘till when’ vs. *tfrom* ‘from when’ (Hungarian), but others point to actual (morpho)syntactic distinctions (cf. classification in §5): it is the case of *emph* for an “emphasizing word, intensifier” and *neg* for negations (see §5.1.2). We argue that if some words always select a relation subtype on the grounds of distinctive syntactic behaviour (and not merely semantics), and vice versa this subtyped

<sup>24</sup>Barring ellipses and co-ordinative constructions.

<sup>25</sup>Here and in the following we refer to lines 1475–1483 of the script `validate.py`, as retrieved from <https://github.com/UniversalDependencies/tools> (as of January 19<sup>th</sup>, 2024).

<sup>26</sup>Cf. §4.4, and also the issue <https://github.com/UniversalDependencies/docs/issues/617>.

<sup>27</sup>Considering the English EWT corpus, out of 709 unique lemmas of words tagged as ADV, more than half, 381, are derived from adjectival bases (including determiners and participial forms) by means of the *-ly* suffix.

	Nominals (reference)	Clauses (predication)	Modifiers (modification)	Modifiers?	?	?	Undefined
Autosemantic	NOUN+PROPN	VERB	ADJ	<b>ADV</b>	ADP	SCONJ, CCONJ	PART
Synsemantic	PRON	AUX	DET+NUM				

Table 2. UD system of parts of speech and its main axes; cf. §2.

relation is always realised by such words, then these words actually identify their own (sub)class (i. e. part of speech) and/or dependency relation, at least on some axis (cf. Table 2 and overview in §2).

### 3.2. Asymmetry of UD’s part-of-speech system and undesirable consequences

Table 2, following de Marneffe et al. (2021, §2.1.2), themselves referring to Croft (1991), shows UD’s part-of-speech system in the framework of this work along its main axes of syntactic functions (or “types of phrasal unit”) and autosemantic/synsemantic dichotomy (de Marneffe et al., 2021, §2.2.1), as discussed in Section 2. Table 2 does not include the classes SYM (symbols) and PUNCT (punctuation marks), as they do not represent linguistic entities; INTJ (interjections) is also excluded, since it lies outside of morphosyntax (cf. Anderson et al., 2003 s. v.), just like X, which is a residual non-class; subclasses are grouped together with the main ones.

Apart from the three main dyads corresponding to the types of phrasal units (corresponding on their autosemantic side to parts of speech NOUN, VERB and ADJ) and a trail of synsemantic parts of speech (ADP, CCONJ/SCONJ, PART) with no defined overarching functions, ADV stands ambiguously in the middle. This class is indeed not specified for synsemanticity or autosemanticity, but actually contains elements of both kinds, cf. VERY and EXACTLY as seen in the beginning of this Section: the former is an intensifier for members of the ADJ class<sup>28</sup> with no true autonomous meaning, comparable to what MOST (tagged as ADV for more than half of its occurrences, 102 out of 186, in English EWT) is with respect to Degree=Sup, while the latter can depend on members of the VERB class or predicates in general, and derives its meaning from its autosemantic ADJ base EXACT.<sup>29</sup> This goes to show that, as of now, ADV straddles a basic distinction of UD: again, we observe the catchall tendency derived from traditional

<sup>28</sup>Apart from a single spurious annotation as root, in

(i) (like VERY not in a good way)  
 – English EWT answers -20111107194625AAAUBXm\_ans-0011

where way ought to be the head, and very acts as a modifier of the whole noun phrase, like not.

<sup>29</sup>And this independently from the fact that its scope be only the predicate or the sentence as a whole.

grammars.

This ambiguity can be shown more generally with a small experiment: if, for a given treebank, the two classes of a dyad, say ADJ and DET, are conflated into a single one, which in this case we might call MOD (for “modifier”), we see the synsemantic elements consistently making up the majority in a ranking by frequency. Parallely, we see this happening for the ADV class, too (in the following lists, synsemantic elements are underlined):<sup>30</sup>

– German HDT

**MOD** DER ‘the’, EIN ‘a’, DIESER ‘this’, NEU ‘new’, SEIN ‘his/its’, IHR ‘her’, ALLE ‘all’, ANDERER ‘other’, KEIN ‘no (one)’, DEUTSCH ‘German’

**ADV** AUCH ‘also’, NOCH ‘still’, SO ‘so’, NUR ‘only’, ABER ‘but’, BEREITS ‘already’, MEHR ‘more’, DAMIT ‘therewith’, SCHON ‘already’, RUND ‘circa’

– Latin (IT-TB only)

**MOD** HIC ‘this’, OMNIS ‘all’, IPSE ‘same’, ALIUS ‘other’, SUUS ‘her/his/its own’, ALIIQUI ‘any’, DIVINUS ‘divine’, ILLE ‘that’, UNUS ‘a/one’, QUIDAM ‘a certain’

**ADV** IGITUR ‘therefore’, ETIAM ‘also’, UNDE ‘whence’, SIC ‘so’, ERGO ‘therefore’, TAMEN ‘notwithstanding’, SUPRA ‘above’, ITA ‘thus’, MAGIS ‘more’, SOLUM ‘only’

– English EWT

**MOD** THE, A, THIS, GOOD, ALL, SOME, ANY, GREAT, OTHER, NO

**ADV** SO, JUST, WHEN, VERY, ALSO, HOW, NOW, EVEN, THEN, THERE

– Bulgarian BTB

**MOD** TOZI ‘same’, NOV ‘new’, VSEKI ‘every’, EDIN ‘one’, DRUG ‘other’, GOJAM ‘big’, BALGARSKI ‘Bulgarian’, KAKAV ‘what’, NAŠ ‘our’, MOJ ‘my’

**ADV** MNOGO ‘much’, OŠTE ‘still’, VČERA ‘yesterday’, SAMO ‘but’, TAKA ‘so’, KOGATO ‘when’, VEČE ‘already’, TAM ‘there’, ZAŠTOTO ‘since’, OBAČE ‘however’

– Greek (modern) GDT

**MOD** O ‘the’, ÉNAS ‘a/one’, AUTÓS ‘this’, EURŌPAÏKÓS ‘European’, ÓLOS ‘all’, POLÝS ‘many/much’, MEGÁLOS ‘big’, NÉOS ‘new/young’, ÁLLOS ‘other’, SÍMANTIKÓS ‘meaningful’

**ADV** ÓPŌS ‘as’, EPÍŠIS ‘moreover’, ŌS ‘like’, ÓSO ‘that much’, SCHEŤIKÁ ‘with regard to’, POLÝ ‘much’, SÍMERA ‘today’, MÓNO ‘only’, AKÓMA ‘yet’, EDŌ ‘here’

<sup>30</sup>The entries labelled as synsemantic under MOD follow their annotation as DET in their respective treebanks: however, we notice that the elements OTHER, ÓLOS ‘all’, POLÝS ‘many’, DRUG ‘other’ should also be considered determiners (of contrast and quantity), analogously to ALIUS ‘other’, ALLE ‘all’, ANDERER ‘other’. The judgment about the synsemantic nature of entries under ADV is instead by the author and will be expounded in §5.

Here some patterns emerge: we see parallel synsemantic elements (cf. §4.3) based on the same roots, even in paradigmatic relation, such as *POLÝS* ‘many’ and *POLÝ* ‘a lot’ in Greek (the latter being the neuter singular form of the former). Also, the assortment of the most frequent adverbs is remarkably similar for all languages: one finds the same kinds of elements in many languages, such as demonstratives *SIC/SO/SO/TAKA*, focalisers *ETIAM/ALSO/AUCH* (see §5.1.2), besides deictic/relational terms which are admittedly more uncertain as to their synsemantic or autosemantic nature (see §5.1.4 and §5.1.5), and which we can tentatively arrange along a gradient:

- clearly synsemantic (i. e. purely deictic, in space or time), such as *THEN, THERE*;
- mixing synsemantic and autosemantic elements (represented by a dashed line), such as *SUPRA* ‘above’ (both deictic and with meaning of a specific location);
- clearly autosemantic, such as *NOW, SÍMERA* ‘today’, *VČERA* ‘yesterday’

The few uncontestedly autosemantic terms emerging from this conflation are adjectives of nationalities of importance to the respective treebanks, such as *DEUTSCH* ‘German’, *EURŌPAĪKÓS* ‘European’, *BĀLGARSKI* ‘Bulgarian’; some universally extremely common qualifying adjectives, such as *GOOD, GREAT/MEGÁLOS/GOLJAM* (cf. also *MAGIS*, from *MAGNUS* ‘great’), *NÉOS/NOV* ‘new’; and finally text-specific *DIVINUS* ‘divine’, coming from the theological treatises in the *IT-TB*, and *SĪMANTIKÓS* ‘meaningful’. Only farther into the ranking autosemantic elements start appearing more consistently in the class *ADV*, like the deadjectival *ACTUALLY, LONG, LATE* in English *EWT* (from 35<sup>th</sup> rank onward), or *SIMPLICITER* ‘simply’ and *PERFECTE* ‘perfectly’ in Latin *IT-TB* (from 28<sup>th</sup> rank onward).

A similar distribution appears to have the side effect of creating unwanted results in a research involving the class *ADV* in any of these treebanks: most of the retrieved elements will be synsemantic, and they will possibly not correspond to the intuition of what an adverb should look like, if we take a word like *exactly* as its prototypical member. By the way, this bipartite perspective is partly shared with some dictionaries which are sometimes even used as references for the annotation: so, in Latin, among others, *IGITUR* ‘therefore’ is not an adverb for e. g. Lewis and Short (1879, s. v.) (cf. §5.1.3), as opposed to *SIMPLICITER* ‘simply’ (which is traced back to *SIMPLEX* ‘simple’). It happens so that, in the current state of things, these two subclasses of *ADV* are unfortunately not separable without resorting to external tools.<sup>31</sup> Also, this structural fuzziness between autosemantic and synsemantic elements hinders nucleus-based approaches to parsing, like the one presented by Nivre et al. (2022). In sum, these circumstances do not seem well suited from either perspective: typological representation on the one hand, and linguistic queries at different levels of proficiency on the other hand. This problem is by all means not limited to *UD*, but, given the imprint of traditional grammars, will be found in many other treebanks, or annotated corpora.

---

<sup>31</sup>As the derivational lexicon *Word Formation Latin* might be for Latin; see Litta and Passarotti (2019) and cf. §4.4.2.

The solution to this undesirable state of affairs is then to refocus the ADV class by implementing the same autosemantic/synsemantic split as for the other classes. We notice that this might entail a more or less pronounced restructuring or rethinking of other classes, too (cf. especially §5.1.5); conversely, not every adverb in a traditional sense will be retained as ADV.

#### 4. What adverbs are, and what their function is: a working definition

The asymmetry of the ADV class inside the UD system is a reflex of the historical precedence given by traditional grammars to morphology (and even semantics) over syntax: adverbs are considered indeclinable, isolated elements, and are grouped as such with other function words (see Pinkster, 1972, §4.1). This is for example the stance of the *Index Thomisticus* for Latin, with a single class for all invariable elements, as opposed to a nominal, a verbal and a mixed “participial” one (Busa, 1974–1980). This choice, however, creates internal incoherences as discussed in Section 3. In this Section, we will try to give a definition of the word class “adverb” and of the function associated with it, from which the proposal of a framework for a better annotation in UD (and beyond) will ensue.

##### 4.1. Decoupling form from function

An explicitly stated *desideratum* of UD is to avoid what might be called “contextual annotation” (de Marneffe et al., 2021, p. 262): this is the assignment of any feature, from parts of speech to dependency relations, mechanically depending on the context a word appears in. A classical example from Latin are the *pronoun-adjectives* or *adjective-pronouns* of traditional grammars, like *demonstrative pronouns*, which “are used either adjectively or substantively” (Greenough et al., 2014, §296). In UD, this statement translates into tagging an element such as *hic* ‘this (one)’ as DET<sup>32</sup> in a phrase like

- (4) *hoc opere*  
 ‘(in) this work’  
 – Latin IT-TB dev - s35

and as PRON in another phrase like

- (5) *hoc probare*  
 ‘to verify this’  
 – Latin IT-TB dev - s363

---

<sup>32</sup>Traditional grammars of Latin do not apply a distinction based on the grammaticalisation cline as UD, and thus “adjective” can be translated either as ADJ or DET in UD; *hic* is DET because it is purely deictic (proximal) in function (see Lewis and Short, 1879, s. v.), and carries no other semantic content.

The totally predictable parameter here is merely being head of the argument of a predicate or not, at the cost of losing any information about the distribution of determiners in Latin: for example, that they can head noun phrases more easily than in English, or other languages. The constraint that *advmod* be realised only by members of *ADV* (§3.1) is indeed such a form of part-of-speech–based contextual annotation in *UD*. However, we observe in *UD* a tendency to move away from it, e. g. allowing for case/mark alternation for *ADP*, the possibility for *PRON* to receive the *det* relation<sup>33</sup> (cf. Höhn, 2021, for an extension to personal pronouns), or the guideline amendment regarding deverbal connectives.<sup>34</sup> This is possible because the cornerstone of *UD* is syntax, and the spread of annotation over four main orthogonal levels (lemmatisation, parts of speech, morpholexical features, dependencies) allows for a rich investigation of typological patterns through their interplay. The first insight re *ADV* is thus to decouple the syntactic adverbial function (*advmod*) from the word classes concretely realising it each time. So, there is only to gain in terms of linguistic analysis from observing whether and how members of *ADJ* (primarily), but also *DET*, *NUM*, *PRON*, *NOUN*, *VERB* can or cannot occur as modifiers, possibly assuming a specific adverbial form, at the same time keeping the unitarity of the dependency relation *advmod*. This perspective does not go as far as to conflating all kinds of modifications (as proposed e. g. by Gerdes et al., 2018), but is more similar to the strategy of an “umbrella” *Adv* relation as seen in the Prague Dependency Treebank (Hajič, 1998, §2.2,<sup>35</sup> and Hajič et al., 2017), as opposed to locking a dependency relation to a specific part of speech.

#### 4.2. The adverbial function

If function is decoupled from form, we acknowledge that a word belonging to a given non-*ADV* part of speech can more or less “easily”, i. e. requiring a more or less explicit specific marking, carry out the function of adverbial modification: this can happen by means of special, dedicated forms which use neither nominal nor verbal strategies. To avoid falling back into contextual analysis (§4.1), definitions of parts of speech need to be semantically grounded (at least partly, as argued by Croft, 2001, §2, and for *UD* by de Marneffe et al., 2021, p. 261). This is actually already the case in *UD* when e. g. the class *ADJ* is defined as

“specify[ing] their (sc. of nouns) properties or attributes”.

So, the key feature common to the classes *VERB* (processes), *ADJ* (properties) and *ADV* seems to be relationality: they are all predicative in nature.<sup>36</sup> As variously (al-

<sup>33</sup>From lines 1462–1466 of `validate.py`: “Determiner can alternate with a pronoun”.

<sup>34</sup><https://universaldependencies.org/changes.html#deverbal-connectives>

<sup>35</sup>See s. v. *Adverbials* (and *borderline cases*) in the manual (Hajič et al., 1999).

<sup>36</sup>This fact is for example instantiated in *UD* by the assignment of a valency to members of *ADJ* and *VERB* alike, in that both classes can have oblique *obl* or adverbial *advcl* dependents (instead of *nmod/accl* as for nominals).

beit somewhat dishomogeneously) expressed in literature (e. g. Tesnière, 2015, Ch. 32; Hengeveld, 1992, §4.3; Schachter and Shopen, 2007, §1.4), it emerges that the syntactic function of adverbs is that of *modifying*<sup>37</sup> non-nominal elements, which are often explicitly listed as “verbs, adjectives and other adverbs”. As discussed in Section 3, some elements usually referred to as adverbs do modify nouns or noun phrases, so we argue that a better and more general definition is that **the adverbial function is to modify predications**, that is, to express properties of a predication or, in other terms, to *metapredicate*.<sup>38</sup>

Modification in this sense has to be taken strictly as one type of phrasal unit, as in de Marneffe et al. (2021, §2.1.2), and is different from nominal or verbal strategies. In a broader sense, every argument and every subordinate clause, every single element, “modifies” the sentence it is part of. This vague notion of modification, which seems to be at the basis of the traditional mixed-bag adverb class (cf. §1), is not very useful. In metapredication we explicitly want to include only strategies specifically used by each language for modification: as will be discussed in Section 4.4 and elsewhere, these might be shared with the classes of ADJ/DET, or be peculiar to adverbs. In general, we will consider metapredication mutually exclusive with, but parallel to, nominal oblique arguments (represented by relation *obl* in *UD*) and embedded adverbial clauses (represented by relation *advcl* in *UD*). Put in other words, we argue that the argument of a predication cannot be considered an adverb; this point of view will be further developed in Section 5. A further important consideration here is that noun phrases, under given conditions, can be considered as (parts of) predications by some elements, e. g. focalisers, and so can receive adverbial modification (§5.1.2). The syntactic function of adverbs, represented by *advmod* in *UD*, is thus expressed not in terms of parts of speech, but of relations: without further assumptions, this naturally accounts for possibly infinite chains of modifications, since metapredication is recursive, and an adverb is itself a predication. On the contrary, adjectives cannot modify other words of their class in the sense of *amod*, since this kind of modification is directed to referents, i. e. prototypically NOUNS (cf. Table 2). The specific dependency relation tag used in the annotation to represent metapredication can then vary, according to the grammaticalisation status of the adverb, as will be discussed in Section 4.5.

---

<sup>37</sup>This term being often vague and itself underdefined, cf. Pinkster (1972, §4.3).

<sup>38</sup>Schachter and Shopen (2007, §1.4) speak of “modifiers of constituents other than nouns”, but this still does not seem to take into consideration the matrix predication; for an example regarding Latin, Ricca (2010, §1), too, does not acknowledge nominal modification, while Pinkster (1972) does.



### 4.3. Autosemantic and synsemantic adverbs: core word classes

The adverbial function as described in Section 4.2 needs a part-of-speech dyad of reference (see Table 2 and §2), representing elements which are primarily (if not exclusively) associated with it (cf. §2). One member of this dyad should be clearly and intuitively ADV, which we propose for autosemantic adverbs; for the other, we put forward a repurposing, or extension, of PART, as of now a rather undefined and officially marginal class in UD's part-of-speech system. Incidentally, it could be argued that some of the elements already regularly tagged as PART, such as negations or interrogative particles, can indeed be subsumed under the label of synsemantic adverbs (§5.1.2), and in fact they are treated as such by many traditional grammars.<sup>39</sup>

The resulting dyad ADV/PART of “core” adverbs (also “true” or “primary”, cf. Pinkster, 1972, p. 65) would be mostly composed of elements not regularly or synchronically derived from other bases, and would be more or less open, specifically to each language. So, for example, Latin notably presents words like *CLAM* ‘secretly’ or *PALAM* ‘openly’,<sup>40</sup> among its exponents of the ADV class, or also *SAEPE* ‘often’ or *EMINUS* ‘from a distance’. The latter is clearly related to *MANUS* ‘hand’, but is not part of its paradigm and has followed its own morphophonetic evolution. On the synsemantic side, we have *NAM*, a discourse connective, or *NE*, an interrogative particle (see §4.4.2 for an ampler discussion about Latin). Section 5 will explore more in detail the composition of the class of synsemantic adverbs, drawing distinctions between them and other word classes.

### 4.4. Core and derived adverbs: strategies and the notion of transposition

Core adverbs ADV/PART are opposed by derived adverbs. While any element expressing metapredication would also regularly receive an adverbial dependency relation (see §4.5), its part of speech does not have to be enforced by its function: contextual annotation (§4.1) is best avoided. In the framework proposed by this paper, we deem that, especially for a multilayered annotation such as that of UD, it is much more fruitful to consider and implement the general phenomenon designated as, among other terms, “transfer” (Tesnière, 2015, P. III) or “transposition” (Spencer, 2013, §3.3). The fundamental insight here is that inflectional, paradigmatic variations are not limited to single word classes or functions (Haspelmath, 1996): this principle is in fact

<sup>39</sup>See for example Greenough et al. (2014, §217e) for Latin, where “negative particles” are part of the “classification of adverbs”.

<sup>40</sup>While these words are related respectively to the same roots of *CELO* ‘to conceal from’ and *PLANUS* ‘flat’ with an accusative-like ending (de Vaan, 2008, s. vv.), there is no clear nor productive synchronic derivation process producing them, so lexicographically they have to be considered independent; this does happen e. g. in the Word Formation Latin lexicon (Litta and Passarotti, 2019), which however applies a rather restricted notion of derivation/inflection (cf. §4.4.2).

Language/treebank	advmod distribution per PoS (>5%)	Main lemmas per PoS (>5%)
Amharic ATT	PART 36.5%, ADV 30.4%, NOUN 15.5%, VERB 7.3%, PRON 5.8%	PART: 'ALƏ 74.5%, 'ÄLƏ 15.3% (other: no lemmas)
W. Armenian ArmTDP	ADV 96.8%	ADV: AL 11.5%
Bulgarian BTB	ADV 79.9%, PART 19.8%	PART: NE 98.1%
English EWT	ADV 83.4% PART 15.2%	PART: NOT 100%
Faroese FarPAHC	ADV 100.0%	(no lemmas)
German HDT	ADV 72.1%, ADJ 20.7%, PART 7.1%	ADV: AUCH 8.0%, NOCH 5.1%; PART: NICHT 100.0%
Greek GDT	ADV 84.3%, PART 14.8%	PART: DEN 82.7%, MİN 10.1%
K'iche' IU	ADV 68.9%, PART 30.6%	ADV: SIB'ALAJ 11.3%, B'İ 9.5%, ULOQ 8.5%, KAN 6.2%, KAMIK 5.2%; PART: NA 43.6%, MAN 34.9%, WI 14.5%, MA 5.2%
Latin IT-TB	ADV 62.6%, PART 32.4%	ADV: ETIAM 11.3%, UNDE 7.3%, SIC 5.9%; PART NON 100%
Latin PROIEL	ADV 94.6%	ADV: NON 21.5%, NE 7.3%
Latin Perseus	ADV 82.0%, PART 13.4%	ADV: NE 5.7%; PART: NON 92.9%
Latin UDante	ADV 72.0%, PART 20.8%	ADV: SIC 6.9%, ETIAM 5.3%; PART: NON 94.7%
Latin LLCT	ADV 79.5%, PART 8.9%, CCONJ 6.7%	ADV: UBI 11.8%, EXINDE 8.8%, TALITER 6.1%, SUPRA 5.8%, TANTUM 5.7%; PART: NON 95.9%; CCONJ: ET 98.4%
Thai PUD	ADV 73.4%, PART 17.2%, ADJ 5.6%	(no lemmas)

Table 3. Distributions of parts of speech over all words receiving the dependency relation *advmod*, with any subtype, per treebank in our sample.

already accepted in UD when participial forms, i. e. verb forms with the function of attributive predications, more commonly labelled as “verbal adjectives” (cf. Cecchini, 2021 re UD), are still usually and foremostly labelled with VERB, and not with ADJ (a choice which would exemplify again contextual annotation depending on syntactic function, cf. §4.1).

Some treebanks appear to pursue this strategy already to some extent. In Table 3, for each well-represented (>5% of total occurrences) part of speech taking *advmod* in the treebanks of our sample, the most frequent (>5% of total occurrences) lemmas are shown (if any: in some cases, lemmas are too sparse to reach this threshold). It turns out that lemmas so detected almost exclusively pertain to function words; notably, PARTs are overwhelmingly negations (cf. §3.2). However, contrary to this trend, in the German HDT as much as 20% of words receiving an *advmod* relation are tagged as ADJ. This goes hand in hand with the known fact commonly stated as German “con-

verting” adjectives directly into adverbs (Barz et al., 2009, §478),<sup>41</sup> a feature which possibly sets it apart from the “Standard Average European” type it is in the core of (cf. Haspelmath, 2001). Other treebanks look more mixed and there probably is some noise in data, but *advmod*-taking members of ADJ are seen to some extent in Latin PROIEL (4.9%) and Thai, too, alongside CCONJs in Latin LLCT and NOUNS in (Amharic).<sup>42</sup> In most other treebanks, similar regular, though marked adverbialisation strategies (such as *-(i)ter/-e* derivation in Latin, Ricca, 2010, §2.2.1: *PAR* → *pariter* ‘equally’; *-ὄς* in Greek: *SYNEPÍS* → *synepós* ‘consequently’; etc.) are obscured by redundant adverbial annotation (§3.1). It is still possible, nevertheless, to let some of these strategies emerge from data by comparing forms of ADV- and ADJ-tagged words.

The most prevalent detected strategies<sup>43</sup> are shown in Table 4, where also the corresponding strictly morphological features<sup>44</sup> of the ADJ forms are summarised (comma notation points to alternation of the values in annotation, parentheses to the values not being consistently annotated). We notice a remarkable uniformity: while bare-form adjectives in adverbial function look universal, many Indo-European languages also share the adoption of neuter, singular or plural, forms,<sup>45</sup> and this ostensibly throughout all their history (cf. Ramat, 2009; Hummel, 2013, for Latin, but also more generally). German, as already noticed, looks synchronically more committed to the bare-form strategy than other European languages, but typologically it is perfectly aligned to a universal tendency which we systematically observe in other languages, too. The neuter strategy is also well-attested in Latin, where it is especially common for adverbs formed from determiners (also cf. §4.4.2), such as *quantum* ‘as many/much’ from *quantus* ‘how many/much’, *tantum* ‘so much/only’ from

---

<sup>41</sup>However, in some contexts one might be tempted to interpret this as secondary predication (the difference being the choice of *advcl* instead of *advmod*; see s. v. *advcl:pred* in UD guidelines). Be that as it may, of course German, too, has a class of core adverbs, mostly represented by the tag *ADV*, such as *GERN/GERNE* ‘gladly’, with no synchronic derivation (Pfeifer et al., 1993, s. v.); cf. footnote 40 for Latin.

<sup>42</sup>On the contrary, the 100% for Faroese probably arises from the automatic *advmod* ⇒ *ADV* conversion rule.

<sup>43</sup>Many of which are confirmed by literature and descriptive grammars; apart from German, here we can point e. g. to Scatton (1993, p. 208f) for Bulgarian: “Certain adjectival forms function productively as adverbs: (1) masculine singular (or plural) forms of adjectives [...]; (2) neuter singular forms of qualitative adjectives [...]; (3) colloquially, definite feminine singular forms”; or to Polian (2017, §3.7) with regard to the Mayan language K’iche’: “Many words are used adverbially in Mayan languages, some of them exclusively so, but generally without a specifically adverbial morphology. Adjectives commonly develop adverbial uses (e. g. ‘good’ for ‘well’), and some nouns and verb forms also do.”.

<sup>44</sup>That is, excluding features of the lexical kind, such as *Style* or *NumType*.

<sup>45</sup>We can guess this for Faroese forms, too, knowing from dictionaries that the citation form of *langt* is *LANGUR* ‘long’ (similarly as for the related Icelandic and Norwegian), but for some reason *Gender* and *Number* are not annotated for ADJs in the treebank.

Language/ treebank	Most frequent strategy	Example (ADV form/ADJ lemma)	Coinciding ADJ-ADV pairs over total ADV forms
Amharic ATT	Bare form? -	<i>zānədəro/-</i>	37.0%
W. Armenian ArmTDP	Bare form (Degree=Pos)	<i>nuaz/NUAZ</i>	20.6%
Bulgarian BTB	Indefinite neuter singular Definite=Ind Degree=Pos Gender=Neut Number=Sing	<i>bārzo /BĀRZ</i>	21.9%
English EWT	Bare form Degree=Pos	<i>cheap/CHEAP</i>	24.2%
FarPahC	Neuter singular? Case=Acc, Dat, Nom Degree=Pos	<i>langt/-</i>	5.3%
German HDT	Bare form Degree=Pos	<i>vermutlich/VERMUTLICH</i>	19.0%
Greek GDT	Neuter plural Case=Acc, Nom Gender=Neut Number=Plur	<i>sovará /SOVARÓS</i>	20.6%
K'iche' IU	Bare form -	<i>itzel/ITZEL</i>	9.4%
Latin (all)	Neuter singular Case=Acc, Nom Gender=Neut Number=Sing	<i>amplius/AMPLUS</i> [Degree=Cmp]	27.0%
Thai PUD	Bare form? -	<i>thi sut/-</i>	31.3%

Table 4. Detection of major deadjectival strategies for adverbialisation arising from data in the languages in our sample.

TANTUS ‘that many/much’, or *quid* ‘why/how/what way [comes that...]’ from QUIS ‘who/which one’, as in

- (6) *quid propugnaculis et pinnis urbem armasse iuvabit*  
 ‘What shall it avail you [...] to have fortified yourselves with bulwarks and  
 battlements’  
 – Latin UDante Epi - 93

where *quid* cannot be interpreted as an object of *iuvabit* ‘it will aid’. In Latin, the neuter strategy is also the only possible one for synthetic comparative forms in the adverbial function.<sup>46</sup>

The practice of enforcing the part of speech ADV for words in adverbial function is surely influenced by traditional grammars which do not have articulated enough means for multilayered linguistic representation (as this is possibly not even their

<sup>46</sup>For example *amplius* expressing the neuter nominative/accusative comparative form in the paradigm of *AMPLUS* ‘ample’ beside the masculine/feminine nominative *amplior*, accusative *amplioem*, and being used in adverbial function as in ‘more amply’, while the expected regular *amplioriter* for adjectives of its inflection class is ungrammatical.

goal), and which for example do not distinguish between different grammaticalisation statuses of word classes, nor contemplate word class-changing inflection. We argue that this practice has undesirable consequences also on the representation of lexicon: in particular, its regular byproduct are unwarranted duos of words. This happens for example with the Latin form *primum*, interpreted either as an ADV ‘firstly’ with lemma PRIMUM when it has adverbial function, as in

- (7) [...] *cum **primum** pedes iuxta Sarni fluenta securus et incautus defigerem [...]*  
 ‘no **sooner** had I set foot by the streams of Arno, in all security and heedlessness’  
 – Latin UDante Epi - 35

or as an ADJ ‘first’ with lemma PRIMUS when e. g. it modifies a noun, as in

- (8) [...] *manifestum est quod celum **primum** magis recipit de luce primi [...]*  
 ‘it is manifest that the **first** heaven receives more of the light of the Prime’  
 – Latin UDante Epi - 346

But sometimes the annotation of the two words is identical in every other respect, like *weekly* in the two sentences

- (9) *I intend to provide **weekly** updates on the status of the above actions during April*  
 – English EWT email - enronsent02\_01 - 0060
- (10) *the easiest thing would be a fish - but u need to clean the tank **weekly**, feed it the right amt, etc etc x*  
 – English answers - 20111108102204AAIivYN\_ans - 0006

tagged ADJ and ADV respectively, but with the same lemma WEEKLY (and no morphological features besides a moot Degree=Pos when ADJ). These distinctions (cf. Table 4) appear questionable in a language’s vocabulary, as we are really talking about the same words, even if serving different syntactic functions, and they might be seen as a rather inconsistent annotation practice, against the UD desideratum of “rapid, consistent annotation” (de Marneffe et al., 2021, §5, point 3).

In sum, *pace* e. g. Pinkster (1972, §5), we argue that it is possible to claim that, with the goal of meaningful linguistic representation, transparently derived forms in adverbial function should be labelled just as all other forms belonging to their respective parts of speech, with all their pertinent morphological features, but possibly accompanied by a specific morphological feature marking the transposition. This feature would not, at any rate, be Case with a new value (here agreeing with Pinkster, 1972, §5), as it happens with the “adverbial case” in the *Index Thomisticus* (Busa, 1974–1980, p. xi): a transposition connects paradigms, instead of being constrained into one. Also, we argue that the counter-arguments against recognising word class-changing

inflection in the case of “adverbialisation” can easily be dealt with, if we consider the nature of the adverbial function of metapredication, as defined in Section 4.2. For instance, the fact that some properties of a part of speech seem to be blocked by an adverbial transposition, e. g. members of the NOUN class losing the ability to take adjectival modifiers (represented by relation *amod*), is directly explainable by observing that adnominal attribution cannot be applied to predications, but another adverb is needed instead, and that, at least in some languages like Latin, adverbial transposition seems to target single words and not phrases (unlike Case). The topic deserves a wider typological investigation, but, otherwise, it is actually not uncommon for adverbs to have arguments in the same way as their bases (cf. fn. 49 and Ricca, 2010, §7.2), as seen for example in

- (11) *Unfortunately for them, the general has decided to take seriously Colin Powell’s frequent boasts that it was on his nudging that the Indians made conciliatory gestures toward Islamabad.*  
 – English ewt  
 weblog-blogger.com\_dakbangla\_20041028153019\_ENG\_20041028\_153019-0025

Thus, we are probably in the presence of the peculiarity of a non-relational word class like NOUN as opposed to the others, rather than with a “strange” property of adverbial transposition. Members of the NOUN class are also remarkably less transposed into adverbs than ADJs (cf. §4.4.2 for Latin).

As already happens to verb forms by means of the morphological feature *VerbForm* (cf. Cecchini, 2021), in UD we propose to annotate marked deadjectival adverbs with a new similar feature, e. g. *AdjForm=Adv*. More in general, we can envision a generic pattern for such a feature, with name *<POS>Form*, depending on the part of speech of the base word, and values representing the external syntactic function of the transposition. For example, given the distinction for core adverbs put forward in Section 4.3, we could have *DetForm=Part* for determiners transposed into adverbs, as the previously mentioned Latin *quantum* ‘as many/much’. For participial adverbs like Latin *diligenter* ‘industriously’, as in

- (12) *Diligenter quippe notandum est quod dicit «nescit et nequit» [...]*  
 ‘It should be noted **carefully** that he says «neither knew nor could» [...]
- Latin UDante Epi - 365

glossable as *dilig-ent-er*  $\equiv$  *diligo*.IPFV.ACT.PTCP-ADV,<sup>47</sup> from *diligo* ‘to esteem highly’, so lit. ‘highly esteemingly’, this treatment should be preferred to a more opaque

<sup>47</sup>See Leipzig glossing rules at <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>.

VerbForm=Conv. At the same time, this would need a layered treatment<sup>48</sup> of this double transposition, for example as a recursive AdjForm[Part]=Adv|VerbForm=Part.<sup>49</sup>

While members of ADJ and DET, followed by adjectival forms of VERBs, are the most common and direct sources of respectively autosemantic (ADV, traditionally *manner adverbs*, Schachter and Shopen, 2007, §1.4) and synsemantic (PART) adverbs, as they are already attributive and predicative in nature, one also encounters cases involving NOUNs like

- Latin *viritim* ‘man by man’, from NOUN VIR ‘man’ (Ricca, 2010, §2.2.2), as in

(13) *Tribuniciae potestatis duodevicesimum, consul XII, trecentis et viginti milibus plebis urbanae sexagenos denarios viritim dedi*  
 ‘In the eighteenth year of my tribunician power and my twelfth consulship I gave 240 sesterces **apiece** to 320 000 members of the urban plebs.’  
 – Latin Perseus phi1221.phi007.perseus-lat1.tb.xml@47

- Italian *gattoni* ‘on all fours’ from NOUN GATTO ‘cat’ (thus literally ‘walking as a cat does’) with the derivational suffix *-oni*,
- or more productive German formations (Barz et al., 2009, §1161), such as BEISPIEL-SWEISE ‘for example’ and VORSICHTSHALBER ‘precautionally’, from NOUNs BEISPIEL ‘example’ and VORSICHT ‘caution’ respectively.

For these, a consistent annotation with NounForm=Adv might be felt as more controversial by some. Still, their annotation as transpositions, different from case forms, is justified when faced with the data at our disposal which do include case forms in adverbial function, and would be coherent with the proposal of this work; in fact, we suggest a unifying annotation along the lines of Transposed=ADV, which could take care of all cases of transpositions, e. g. Transposed=ADJ instead of VerbForm=Part, and so on. Similar detailed issues of annotation formalism, however, go outside the scope of this paper.

<sup>48</sup><https://universaldependencies.org/u/overview/feat-layers.html>

<sup>49</sup>The effective difference between VerbForm=Conv (‘converb’, i. e. verbal form with the syntactic function of an adverb) and AdjForm=Adv plus VerbForm=Part (‘participle’, i. e. verbal form with the syntactic function of an adjective) deserves further investigations. We may suggest that the former is derived directly from a verb root and more easily maintains a verbal valence. However, we occasionally find cases like

(i) *Premissis quoque rationibus consonanter [...]*  
 lit. ‘thus **agreeingly with the** aforementioned **reasonings** [...]  
 – Latin UDante Epi-351

where *consona-nt-er* ≡ *consono*.IPFV.ACT.PTCP-ADV, from *consono* ‘to resound’, takes an argument (obl) in the dative case like the base verb. We notice that in Late Latin, with the disappearance of the so-called supine (effectively a converb, see Cecchini, 2021), this strategy seems the only one left to form adverbs from verb roots.

In the following Subsections, we present more in-depth discussions of adverbial transpositions and adverbs in general, to help complete the picture of how the word class of adverbs would look in UD annotation if implementing the framework of this work.

#### 4.4.1. Two examples of transposition: Italian and Latin

We would like to illustrate through two interesting cases how a transpositional annotation shows its advantages on all annotation levels.

The first one is a case of attraction in Italian:<sup>50</sup> in spontaneous production, when DET *TROPPO* ‘too many/much’ adverbially modifies another determiner, the former can pass from the “default” form *tropo* (singular masculine) to agreement with the latter, as in *troppe poche volte* instead of expected *tropo poche volte* ‘**too few** times’, where *troppe*, *poche* and *volte* are all feminine plural forms. If one does not want to annotate *tropo/troppe* always simply with lemma *TROPPO*, part of speech *DET* and relation *advmod*, accepting that it can function either as *det* or *advmod* in its bare form, one implicitly has to postulate the existence of four independent distinct members of *ADV TROPPO*, *TROPPIA*, *TROPPI* and *TROPPE*. This is not an economic representation and does not seem to describe correctly what is happening at the morphosyntactic level, nor to allow for a simple, unified query of this phenomenon in the treebanks without the input of prior knowledge.

The second case consists in the interplay between so-called secondary predications and adverbs in a language like Latin (cf. Pinkster, 1990, §8, especially example (77), p. 155). In a sentence like

- (14) [Britanni] ex silvis *rari* propugnabant  
 ‘The Britons made sorties from the woods **in small groups**’  
 – *De Bello Gallico*, Caesar

the form *rari*, from *ADJ RARUS* ‘thin, rare’, refers to *Britanni* ‘the Britons’, and as such agrees with it in case (nominative), gender (masculine) and number (plural). It is a secondary predicate or “optional depictive”, and according to UD’s guidelines it is annotated as *advcl*, as it forms an embedded (nominal or implicit) clause.<sup>51</sup> This sentence is contrasted by the possible

---

<sup>50</sup>For the following, cf. the discussion in <https://accademiadellacrusca.it/it/consulenza/tropo-pochi-non-tropi-pochi/1559> (only in Italian). Though based on a prescriptive point of view, it testifies to the spread of the phenomenon.

<sup>51</sup>Cf. the guidelines for Latin (by the author) at <https://universaldependencies.org/la/dep/advcl-pred.html>, whence the example is taken.



- (15) [Britanni] *ex silvis raro propugnabant*  
 ‘The Britons **seldom** made sorties from the woods’

where *raro* ‘thinly, rarely, seldom’, singular masculine/neuter ablative of *RARUS*, modifies the predicate *propugnabant* ‘they were making sorties’ using a different strategy, i. e. a conventional case form, akin to what happens for so-called absolute constructions,<sup>52</sup> since now it does not have to agree with any referent. Here we might also expect a more canonical *rare* (adverbial form in *-e*; see Ricca, 2010, §2.2.1). Annotating both instances *rari* and *raro* with the same lemma *RARUS* and part of speech *ADJ*, one with relation *advcl:pred* and the other with *advmod*, while keeping all usual morphological features, would be morphosyntactically motivated, look like the most economical representation and, moreover, allow for an easier, more “integrated” investigation of the distribution of an adjective like this one, without requiring particular prior knowledge from the querier.

#### 4.4.2. What lies under *advmod*: a focus on Latin

In this Subsection, we want to exemplify the mixed composition of elements with adverbial function and of the *ADV* class, at the same time tackling the issues of the current annotation standards, by taking a closer look at what happens in UD Latin treebanks. This brief overview is to be considered preliminary work for more thorough investigations about the treatment of Latin adverbs in linguistic annotation, it will be necessarily coarse-grained, and offers many pointers to the wider classification brought forth in Section 5.

Taking all Latin treebanks together,<sup>53</sup> we retrieve 1360 unique lemmas for words annotated as belonging to the *ADV* part of speech. After some minimal polishing of the data, which includes standard normalisation (lowercasing; neutralising the *u/v* distinction), the expunction of non-lexical category labels (e. g. *greek.expression*, *monetary*) and of compound forms (e. g. *in+quantum*, *kal. maias*), and the selection of a canonical spelling when more than one is used (e. g. *paulatim* and *paullatim* retraced to *paulatim* only), this number reduces down to 1285 unique *ADV* lemmas. These still include some noise coming from spurious annotations, but can already deliver a meaningful picture.

We observe that as many as 918, that is 71.44%, of these supposed *ADV*s can be transparently traced back to autosemantic bases, as shown in Table 5. We notice that in our data some of these lemmas sometimes directly correspond to the base of the adver-

<sup>52</sup>See UD documentation for Latin (by the author) at <https://universaldependencies.org/la/dep/advcl-abs.html> and more generally (Pinkster, 1990, §7.4.7).

<sup>53</sup>Refer to Paragraph **A note about data** in §1.

Part of speech of autosemantic base	Number of unique lemmas	Examples with bases
ADJ	608	PRINCIPALITER < <i>principalis</i> 'original' ABSURDE < <i>absurdus</i> 'discordant' PRIMO < <i>primus</i> 'first'
VERB	214	FESTINANTER < <i>festino</i> 'to hasten' IMMERITO < <i>mereo</i> 'to deserve' DISTINCTIM < <i>distinguo</i> 'to separate'
NOUN	96	FUNDITUS < <i>fundus</i> 'bottom' TABULATIM < <i>tabula</i> 'plot (of ground)' NUMERO < <i>numerus</i> 'number'

Table 5. Distribution by part of speech of autosemantic bases of ADV-tagged words in UD Latin treebanks.

Common ending	Number of unique lemmas	Distribution of parts of speech of base
-e	315	236 ADJ, 79 VERB
-ter	234	173 ADJ, 61 VERB
-o	75	35 VERB, 30 ADJ, 10 NOUN
-im	36	22 VERB, 11 NOUN, 3 ADJ
-itus	8	4 NOUN, 4 ADJ

Table 6. Presence of the different adverbial formations seen in Table 5 in UD Latin treebanks.

bial form, e. g. AEQUALIS alongside of AEQUALITER for the form *aequaliter* in Perseus, so we are actually dealing with an even lower number of effective lexemes. This shows oscillations in the manual annotation and/or conversion processes from other formats, which incidentally testifies how these adverbial forms are spontaneously felt part of an adjectival, verbal or nominal paradigm. In fact, all these formations look productive from the data at our disposal. In particular, we have rough distributions for these formations as shown in Table 6.<sup>54</sup> No such formation is isolated, and each appears at every diachronic stage of Latin. Moreover, the data are not exhaustive, since we can find still more formations of the same kind in dictionaries, e. g. RADICITUS

<sup>54</sup>We group them under common endings, refraining here from identifying the exact form of the suffixes operating on the bases.

from *radix* ‘root’ or *FRUSTATIM* from *frustum* ‘piece’ (see Lewis and Short 1879, s. vv.), showing their vitality. The distribution of the different formation strategies among parts of speech deserves further study, but it probably comes to no surprise that the most attested ones involve adjectival and verbal bases,<sup>55</sup> and that the latter invariably pass through adjectival verb forms (“participles”, marked in UD as VerbForm=Part). Examples are *FESTINANTER* from the imperfective *festinans* (so paraphrasable as ‘by hastening’), or *DISTINCTIM* from the perfective *distinctum* (so paraphrasable as ‘by having been separated (each from the other)’). At the same time, nominal bases are rarer and more limited in their strategies of adverbial transposition. This is in accordance with the fact that (adnominal) modifiers are more easily transposed into metapredications than non-relational word classes such as nouns, as discussed in Section 4.4. It is also interesting to notice that a specialised lexical resource like Word Formation Latin (Litta and Passarotti, 2019) does not include the first two cases of Table 6, but instead includes words falling under the last three: the latter are not seen as part of paradigmatic inflections, while the former are. However, this choice seems dictated by an arbitrary cut based on frequency in the data, and is at odds, among others, with lexicographic conventions (see e. g. Gaffiot, 2016; Lewis and Short, 1879, s. v. *absurde*).

Some problems of contextual annotation (cf. §4.1) also arise from the data. One example, among the formations shown in Table 6, is the ending *-o*: this is actually the regular nominal ending for the singular ablative/dative masculine/neuter, and lemmas like those shown in Table 5 do not make exception. It is all the more striking, then, that Word Formation Latin singles them out as derived forms (through “conversion”), while their interpretation as ablatives of neuter forms is straightforward: *primo* ‘as first (thing)’, *immerito* ‘(this) not having been deserved’<sup>56</sup>, *numero* ‘with number(s), i. e. precisely’. Then, there are also elements which should be better analysed as consisting of more than one word, such as *SAEPENUMERO* ‘oftentimes’ (< ADV *saepe* + NOUN *numero*) or *SUPERFICIETENUS* ‘up to the surface’ (< NOUN *superficie* + ADP *tenus*). This however opens up issues which are amply debated among linguists and in the UD community, and are too complex to be tackled here.<sup>57</sup> What we rather want to point out is that the complex of similar “adverbialising” annotation choices brings about an unwarranted proliferation of adverbial lemmas and the fragmentation of lexemes. Forms like *primo* co-occur with *primum* and *primitus*, and *distinctim* with *distincte*. There might be reasons to distinguish some of them by means of extra morphological features (for example, the *-im* formations seem to express a distributional meaning, possibly marked as NumType=Dist), but, from a paradigmatic point

<sup>55</sup>Some are also used for the corresponding synsemantic classes, such as *QUALITER* < DET *qualis* ‘of what sort’, *ISTIM* ‘thither’ < DET *iste* ‘this’.

<sup>56</sup>A so-called absolute construction; see fn. 52.

<sup>57</sup>As starting references, we can point to Haspelmath, 2017 and Lehmann, 2020, which incidentally considers many Latin examples.

of view, they all just correspond to the respective bases *PRIMUS* and *DISTINGUO*. A similar focused reduction (here, from 5 to 2 lemmas) would greatly decrease the actual number of 1285 unique lemmas, and this would improve both the representation of the lexicon and linguistic inquiries performed on the data, as discussed in Section 4.4.

Going beyond lemmas of ADVs with autosemantic bases, we can detect similar trends among the remaining 367 unique lemmas. Apart from compounds and obvious errors, many (238, the 18.52%) can actually be considered dubious annotations for forms better interpreted as pronominal ones, especially of the relative kind (e. g. *UBI* ‘where’; cf. §5.1.4), as conjunctions (e. g. *SCONJ NISI* ‘if not’, *CCONJ ET* ‘and/too’; cf. §5.1.2 and §5.1.3), numerals (mostly multiplicative forms, e. g. *SEPTIENS* ‘seven times’ < *septem* ‘seven’), determiners (e. g. *MULTO* < *multus* ‘many, much’), adpositions (e. g. *AB* ‘from’), or interjections (*UAE* ‘alas!’). Cursorily, it is interesting to observe that classes such as *DET* and *PRON*, synsemantic counterparts of *ADJ* and *NOUN*, seem to prefer inflectional strategies (especially accusative and ablative forms) instead of assuming specific adverbial forms (as represented by the *-e* and *-ter* endings in Table 5): this distinction gets lost with contextual annotation.

We are then left with a nucleus of 113 elements (8.79%). Among them, 51 represent the problematic class of “relators”, which in the treebanks typically alternate between the classes *ADP* and *ADV*, and also *SCONJ*, according to context, but which probably could be better treated as nominals (see §5.1.5). Examples are *FORIS* ‘outside’ and *POST* ‘behind’. The remaining 72 (5.6%) elements, however, satisfy the condition of metapredicating (see §4.2) and of not being synchronically derivable from some base belonging to another part of speech. Of these, 26 show a clearly synsemantic nature: in our framework (see §4.3), they would be annotated as *PART*, as they already are in some treebanks: examples are the discursive connective *NAM* or the emphatic particle *ECCE*. The core of autosemantic “true adverbs” consists then of around 46 elements: among them, *MOX* ‘soon’, *UIX* ‘with difficulty’, *CLAM* ‘secretly’, *PALAM* ‘publicly’, *FERE* ‘approximately’, *SEMPER* ‘always’, *COMMINUS* ‘at close quarters’. The unbalanced ratio between autosemantic and synsemantic elements is in line with the one seen for other part-of-speech dyads (cf. §3.2). The smaller size of the *ADV* class (46, or 97 including relators), as opposed for example to the much larger *ADJ* class (3025 unique lemmas across UD’s Latin treebanks) gives the impression of an “ancillary” word class: a function (metapredication) for which Latin, or any other language, has developed specific words, but which can usually be covered by other word classes, especially those of modifiers, deploying different strategies. This ancillarity might be one cause of the general underdefiniteness of the adverb class in traditional grammars of Latin, and in linguistics in general (cf. §1).

This in-depth look at Latin “adverbs” showcases phenomena which are present to a greater or lesser extent in every language of the world, each according to its mor-

phosyntactic characteristics. It also highlights widespread annotation approaches in UD and beyond. Our aim is again to demonstrate how the annotation framework presented in this work makes it possible to pursue typologically relevant inquiries, all the while rationalising the representation of lexicon and of the transpositional processes of the language. Many of the aspects touched upon in this Section will be discussed more in detail in Section 5.

#### 4.5. Dependency relations of adverbs

Finally, it remains to be seen which dependency relations are best suited to adverbs and words acting with adverbial function. Given the split of core adverbs into two parts of speech along the autosemantic/synsemantic dichotomy (see §4.3), we might also envision to use a separate, specific dependency relation for members of PART with syntactic adverbial function, in the same spirit of distinguishing autosemantic *amod* and synsemantic *det* for autosemantic ADJ and synsemantic DET. On the other hand, this step might not be truly necessary, as the autosemantic/synsemantic dichotomy can be deduced just from the choice of part of speech (as it happens for the dyad NOUN/PRON).

If such a difference is to be implemented, we argue that the relation *aux* comes in handy here. The justification for this would be that synsemantic adverbs do not really specify attributes of a predication as defined in Section 4.2, as much as they convey more grammatical/pragmatical nuances, similarly as to how verbal AUXs “lend” Tense, Mood, Voice, etc. to a predication. Incidentally, we note that similar solutions are already adopted by other annotation formalisms, such as PROIEL for Latin and other ancient languages.<sup>58</sup> Expanding the use of the *aux* dependency relation would also have the desirable consequence of making it typologically more interesting to track the morphosyntactic distribution of features inside a phrase, instead of keeping it redundantly locked to a specific part of speech (AUX).

### 5. A new taxonomy of adverbs, and beyond

This section deals with the internal restructuring of the word class of adverbs (ADV/PART), and especially with the repercussions that this has in relation to other parts of speech, given that the whole linguistic system is interconnected. The proposed new taxonomy is summarised in Tables 7 and 8. The former reinterprets UD’s part-of-speech system (Table 2) in view of the definitions and proposals of Section 4, and of the particular problems posed by some elements previously subsumed under the class ADV; part of it has to be considered tentative. The latter synthesises how

---

<sup>58</sup>See the PROIEL annotation manual at <http://dev.syntacticus.org/annotation-guide/>, under the Section “Auxiliary words”.

	Referents	Process predicators	Referent predicators	Predication predicators	Relators
Autosemantic	NOUN+PROPN	VERB	ADJ	ADV	REL? (ADV or PRON)
Synsemantic	PRON	AUX	DET+NUM	PART	ADP/SCONJ+CCONJ

Table 7. Possible revised part-of-speech system of UD (cf. Table 2).

non-core elements currently labelled as ADV should be redistributed among parts of speech, according to our framework.

We note that the proposals put forward in this Section can be seen as lying on a “gradient” from minor to possibly more radical (§5.1.5) interventions on the UD system of parts of speech, and/or readaptations of existing formalisms. While some of these changes might be easily implemented and, to some extent, could be considered to already exist *in embryo* among UD treebanks (such as ADJ taking the relation *adv-mod* in German, or *today* being annotated as NOUN in English), others might admittedly require more effort by the community, possibly involving a redefinition of annotation practices which are currently seen as established. As discussed in Section 3, the fundamental key point here is that if we do identify (morpho)syntactically different groups under the same label of ADV as it is currently intended, then it is not tenable to keep them together: in a way or the other, some action has to take place in order to be internally coherent with UD’s base framework. At the very least, a distinction between autosemantic and synsemantic elements inside the adverb class should be acknowledged. So, while Sections 4.3–4.5 form the most defined core of our proposal, and we think they should be seen as the cornerstone of a new, improved taxonomy of adverbs, this last Section might look sketchy in comparison. However, we note that it could hardly be any different, since, beyond a given point, many annotation choices will ultimately depend on language-specific considerations, so that here we can just hint to how the present framework can be followed down into all its ramifications, thereby leaving further definitions to future studies.

### 5.1. Better annotation strategies for current ADVs

This Section goes more into the detail of some particular groups of words that can be currently found in the ADV class, beyond the core adverbs as identified in Section 4. The treatment of some seems straightforward, but the more one reaches towards other “invariable” classes such as conjunctions and adpositions, the more notoriously difficult it seems to find a satisfying accommodation for some phenomena (cf. §5.1.4 and §5.1.5).

Currently tagged as ADV	Revised part of speech	Typical features	Typical relations	Typical subrelations	English examples
Core	ADV		advmod		<i>always</i>
Synsemantic	PART		aux discourse	emph: focalisers neg: negations	<i>very, also</i>
Derived	Same as base (auto/synsemantic)	Same as base (+ possible <PoS>Form=Adv)	advmod/aux discourse	as ADV/PART	<i>exactly, piecewise thusly, thrice</i>
Not an adverb	Same as base	Regularly as base	obl/advcl	vary	<i>tomorrow</i>
Pronominal	PRON	<PoS>Type=Loc Tim (Deixis)	obl	lmod: spatial tmod: temporal	<i>there</i>
Prepositional	REL? (or ADV or PRON)	<PoS>Type=Loc Tim	obl/case (or advmod)	lmod: spatial tmod: temporal	<i>beyond</i>

Table 8. Redistribution of current ADVs across UD’s parts of speech, as presented throughout Sections 4 and 5.

5.1.1. Not adverbs

In keeping with UD’s tripartite distinction for non-core dependents advmod/obl/advcl, according to a phrase’s head (de Marneffe et al., 2021, §2.3), some elements customarily labelled as ADV should simply be annotated by means of the oblique (obl) dependency relation. In other words, they do not belong to the ADV/PART class dyad, but rather to nominal (NOUN/PRON and possibly ADJ/DET/NUM), or even verbal (VERB and possibly ADJ/DET/NUM), classes.

Latin, having extensive morphology, is rich in such examples (some seen in §4.4.2). One of them is *forte* ‘by chance’: even if the base NOUN *FORIS* ‘chance’ is deemed defective,<sup>59</sup> it still remains a NOUN (cf. Wackernagel, 2009, L. 1 49, p. 374, on *sponte* ‘of one’s free will’). Despite this, some Latin treebanks (PROIEL and UDante) uniformly tag this form as ADV.<sup>60</sup> Analogously *primo* ‘at first’, from ADJ *PRIMUS* ‘first’, just a canonical use of the ablative case for quality or degree of difference (cf. Greenough et al., 2014, §398) in the sense of ‘as the first (thing)’ (also §4.4.2). We see similar occurrences in the German series *erstens* ‘firstly’, *zweitens* ‘secondly’, *drittens* ‘thirdly’, etc., all labelled as ADV in the German HDT treebank, but clearly related to *ERST* ‘first’, *ZWEIT* ‘second’, *DRITTE* ‘third’ as their genitive forms. In general, annotation should refrain from “adverbialising” nominal adjuncts (a case of contextual annotation; §4.1): they are

<sup>59</sup>That is, lacking expected forms of its paradigm; in the case of *FORIS*, all but the singular nominative and ablative cases are attested.

<sup>60</sup>Maybe influenced by lexical entries such as that in the Oxford Latin Dictionary (Glare, 2012, s. v.), which however admits the impossibility to distinguish an autonomous adverb.

already covered in UD by obliques.<sup>61</sup> The difference between a case like *forte* and one like *raro* (see §4.4.1) is identifiable in that the former still regularly acts as an argument of the predicate, as we expect from members of the NOUN class, while the latter changes its attribution from a nominal (generic, implicit) referent to the predicate, but is not an argument itself. The interesting link (which gets lost with a redundant adverbial annotation) is though that both instances share the same strategy, here the use of the ablative case. The recognition that some of these forms can become “lexicalised” in time, or represent conventionalised, crystallised, possibly non-compositional expressions is part of a different, more lexicographical level of annotation, which does not necessarily coincide with the morphosyntactic one.<sup>62</sup> Another cause for reflection is that similar “paradigm schisms” seem to be pursued only for non-core, oblique arguments, highlighting one of the dichotomies at the base of UD.

Finally, we note that this category might also include nominals which almost exclusively appear as oblique arguments: a typical case are temporal expressions like English *today*, Latin *cras* ‘tomorrow’. While the former has a transparent etymology, even if the second does not, they nonetheless both behave in the same way. This subcategory shows similarities to that of pro-forms, as discussed in Section 5.1.4, but it could be argued (actually *contra* 5.1.4) that these words display a semantic content which goes beyond mere deixis.

### 5.1.2. Focalisers

A particular subclass of synsemantic adverbs, this time part of the core ADV/PART dyad as defined in Section 4.3, are so-called focalisers or intensifiers, currently tied to the subtyped relation *advmod:emph* (§3.1.1), which in the schema presented here could possibly become *aux:emph* (cf. §4.5). They are peculiar in having a wider distribution than autosemantic adverbs, often including nominals in their scope. From UD’s guidelines:

“While other adverbial modifiers usually modify verbs, adjectives or adverbs, these emphasizees often modify noun phrases, including prepositional phrases.”

From Table 3 and data in §3.2 we can see them appear with much higher frequencies than other ADV-labelled elements, a Zipfian distribution typical of grammatical words: so in German *AUCH* ‘also’; in Latin *ETIAM* ‘also’ (39.7% of its occurrences modifying a non-copular nominal in *IT-IB*, the most among significantly frequent adverbs together with *solum* ‘only’), *sic* ‘so’, *taliter* ‘in such a way’, *tantum* ‘so much’. The last three

<sup>61</sup>Cf. Gamba and Zeman, 2023, p. 10 for an overview of other similar cases in need of harmonisation in Latin treebanks. This has started taking place with release v2.13.

<sup>62</sup>But see Savary et al. (2015) for a project dealing with a similar topic, possibly to be integrated into UD in the future (Savary et al., 2023).



terms all are demonstrative elements (PronType=Dem), and the last two are in fact derived from determiners DET TALIS ‘of that kind’ and DET TANTUS ‘that many/much’, making their synsemantic nature explicit. We also note that, prior to UD v2.12, in Latin Perseus ETIAM was grouped with CONJs.<sup>63</sup> In fact, co-ordinating conjunctions, also unrestricted in distribution, do regularly function as focalisers, like Latin ET ‘and’ (7.6% of its occurrences in the IT-TB) or NEQUE ‘and not’. This topic deserves further studies, but here we can just suggest that the identification of focalisers as synsemantic adverbs (PART, aux:emph) can be justified in that, even though they do not directly metapredicate (§4.2), the focus they supply to a noun phrase has a meaning only in relation to the predicate itself (differently than, say, an ADJ): e. g. in

- (16) *et vix Ytali infelices lacrimis metiuntur*  
 ‘even the Italians in their woe can scarce measure with their tears’  
 – Latin UDante Epi - 79

the *et* emphasises how *Ytali* as nominal subjects (nsubj) perform the action. Focalisers thus metapredicate indirectly through the argument of a predicate, and this way we explain also the rarer, occasional adverbial modification of nouns by autosemantic or pronominal adverbs (cf. Pinkster, 1972, §4.3.2f. for Latin).<sup>64</sup> Negations as members of PART, as mandated by the UD guidelines, also enter in this picture, in that they are a form of focus by contrast of a given constituent.

### 5.1.3. Discourse and connecting elements

Some adverbs seem to metapredicate an entire sentence, i. e. they are somehow “external” to the rest of a clause (“modal” or “evaluative” adverbs; for Latin see Ricca, 2010, §4; Pinkster, 1972, §6.2.1.1, speaks of “periphery”). Other elements are seen to connect clauses (Kortmann, 1997; Rosén, 2010), but they move more freely than conjunctions and independently of them, even if the two classes can be confused,<sup>65</sup> or a mixed, distinct category considered, like *Konjunkionaladverbs* in German for words such as TROTZDEM ‘despite’ (Barz et al., 2009, §864).

The common theme to all these elements seems to be textual coherence, expressed outside the argumental structure of a clause, and so the relation discourse might possibly suit them better than advmod (or aux). The exact nature of these connectives would be given by their adverbial subclass: ADV or another autosemantic basis

<sup>63</sup>As in Lewis and Short (1879), but not in Glare (2012). For the harmonisation in UD v2.12 see Gamba and Zeman (2023).

<sup>64</sup>We note that UD is not directly concerned, with regard to annotation, with possibly systematic semantic variations, as seen for example for EXACTLY veering towards an intensifier (§3).

<sup>65</sup>So, for Latin, in Lewis and Short (1879) the particle *nam* is listed as a conjunction, while other sources such as Glare (2012) retain it a “particle” (an even less defined term than “adverb” in this context).

for elements effectively steering the meaning of a sentence, e. g. Greek *fysiká* ‘naturally; of course’, from ADJ *φυσικός* ‘natural’, when not taken literally, as opposed to the semantically completely bleached and non-derived Latin *QUIDEM* (Kroon, 2005; already annotated as PART & discourse in some treebanks), which shows a clitic-like, Wackernagelian behaviour in fixed second position, as many other similar elements (Wackernagel, 1892; Spencer and Luís, 2012, §3.2.1). If these elements are effectively distinguishable from conjunctions, or could possibly be annotated as such (by means of cc) while retaining an adverbial part of speech, remains an open issue.

#### 5.1.4. “Pronominal adverbs”

UD’s guidelines for ADV make a peculiar distinction for “pronominal adverbs”, and a special position is given them also in Zeman (2018, §3.4, Table 3.15). Since they are defined as

“refer[ring] to circumstances in context, rather than naming them directly”,

which reads as deixis (see Anderson et al., 2003, s. v.) (i. e. a function typically associated to the class PRON and/or DET), there seem to be very good reasons for not annotating them as ADV.<sup>66</sup> Indeed, UD’s instructions are to analyse them as pronouns/determiners under all respects (e. g. they take the feature *PronType*) apart from part of speech and, consequently, dependency relation. There are actually many factors in favour of an analysis of such elements simply as pronouns (PRON):

1. a parallelism with the oblique arguments they substitute. This needs the acknowledgment that there can be oblique pronouns (as there apparently are oblique nouns, cf. §5.1.1), which give the impression of being “invariable” (as recounted by many traditional grammars) only if taken in isolation, but are possibly clearly part of paradigms, such as Latin *hic* ‘here’ descending from DET *hic* ‘this (one)’ (declined for case, gender and number as *hic, haec, hoc, huius,...*);
2. as a consequence, this would eliminate purely morphology-based asymmetries such as Latin *quis* ‘who’ (inflectable) tagged as PRON, but *quando* ‘when’ (“uninflectable”, but clearly an old ablative form tied to *qui* ‘what, which’, cf. de Vaan, 2008, s. v.) as ADV (both with *PronType=Int* or *PronType=Rel*), which becomes striking when one considers the fact that they both substitute for nominal phrases in relative clauses;<sup>67</sup>
3. these elements appear in exactly the same contexts as adjuncts, sometimes even accompanied by adpositions (case) as any member of NOUN;

<sup>66</sup>It seemed to be the case in Bulgarian for elements such as *kǎde* ‘where’, *koǎa* ‘when’, *kak* ‘what’, as reported by Zeman (2018, p. 40), but they currently (v2.13) all appear tagged with ADV in UD.

<sup>67</sup>Consider for one example 37 in the discussion about “adverbial relativizers” in the context of enhanced dependencies at <https://universaldependencies.org/u/overview/enhanced-syntax.html#relative-clauses>: given the same context, *where* and *the episode* act and should be annotated the same way, i. e. as nominals with relation *obl*. This would be valid for all so-called “relative adverbs”.

4. deixis, which is compatible with copular constructions as in

- (17) *I was there*  
 – English EWT reviews-374000-0004

contrary to the behaviour of core adverbs;<sup>68</sup>

5. the possibility to have adnominal dependents like relative clauses (acl:relcl), as in Greek

- (18) *ekeí ópou den éinai dynatí* [...] [*there where [...] is not possible*]  
 – Greek GDT gdt-20020206-ep-sessions\_174-14

6. their strong if not exclusive preference for co-ordination among themselves (Pinkster, 1972, §7.3.1).

Under the same light, we can subsume under PRON also other elements traditionally considered adverbs like Latin reciprocal *invicem* ‘one another’, or English *yes* and *no*, substituting for whole phrases and clauses. In general, this perspective entails considering UD’s class PRON not limited to personal pronouns and/or inflectable elements, but including pro-forms in general (see Anderson et al., 2003, s. vv. *Pro-Adverb* and *Pro-Form*), and possibly also taking into account other prepositional-like elements which have always been problematic for a dependency analysis (see next Section). What is common to all of them is that they realise arguments inside a predication instead of metapredicating it.

### 5.1.5. Relators and adpositions

In many (at least Indo-European) languages, for some elements traditionally considered adverbs for their “signifying time or place”, an apparent oscillation between fully autosemantic and adpositional (and thus synsemantic) behaviour is observed (for Latin, see Ricca, 2010, §3.2.1, Pinkster, 1972, §9, and Pinkster, 1990, §5). An example is Latin PROCU ‘at a distance’, occurring both totally independently from nominal elements, as in

- (19) *cum [...] id frater [...] procul animadvertisset*  
 ‘when the brother saw this **from afar**’  
 – Latin PROIEL 53360

or with its own arguments, sometimes introduced by “proper” prepositions, as in

---

<sup>68</sup>For example, in English EWT there are no elements besides pronominal or relator (see §5.1.5) ones like *HERE, THERE, HOW, OUT, WHERE* etc. in a copular construction out of 287 occurrences of an ADV with a cop dependent.

- (20) *non est autem **procul** a uero quod [...]*  
 ‘It is not, of course, **far from true** to say that [...]  
 – Latin ITTB train-s20873

Similar behaviour is shown e. g. by ABOVE in English or MÉSA ‘inside’ in Greek, as opposed to non-independently appearing words such as English *to*.

While agreeing with Gerdes and Kahane (2016, §4.1) that a “fracture” in the annotation of these two allegedly different construction types is undesirable, and not completely justified, it seems possible to argue that these kinds of elements, sometimes called relators, are still subordinated to their more autosemantic accompanying terms (if any). On the contrary, it does not appear to be useful to treat them as heads as proposed in Gerdes et al. (2018); Osborne and Gerdes (2019), or also Vincent and Börjars (2020), or at least this appears to run counter to the primacy of content words so fundamental in UD (de Marneffe et al., 2021, §2.2.1). Nonetheless, it is not easy to find a solution to this *vexata quaestio*, but in the framework of this paper we can envision two, possibly controversial, solutions, besides living them in the class ADV as is currently the case.

One could be to acknowledge these elements, too, as pro-forms specialised in particular spatio-temporal relations. Thus, they would be treated not too dissimilarly from deictic elements like THERE (cf. 5.1.4), having in common a context-dependent, so not totally autosemantic, meaning, and would side with PRON. Under this light, this category could be considered (*contra* §5.1.1) to contain also terms like Latin CRAS ‘tomorrow’ or German GESTERN ‘yesterday’.<sup>69</sup> these point neither to a specific day like *Monday*, nor to a generic idea like *day*, but acquire meaning only in relation to the time of utterance. In contrast to PROCUL or ABOVE, though, these words do not seem to take their own arguments. In any case, they differ from a core adverb (ADV) like Latin SAEPE ‘frequently’, which has a “fixed”, context-independent meaning (cf. Ricca, 2010, §3.1.2). Therefore, in a phrase like the previous example *procul a uero*, a PRON PROCUL can be assumed to depend on *uero* (from ADJ UERUS ‘true’), possibly with a case relation, not so distant from pronouns acting as determiners (det); at the same time, it would be perfectly fit for a pronoun to appear independently.

A second, more radical solution (that we actually favour) could start from the identification of a spectrum of grammaticalisation between “adverbial” and “true” prepositions, and could thus advocate the introduction of a new part of speech, say REL, forming a dyad of relators with ADP. This class could even be considered to extend to other connectives, i. e. to conjunctions CCONJ and SCONJ. A kind of unitary analysis

---

<sup>69</sup>Note that similar terms are often derived from semantically bleached noun phrases, such as Latin HODIE ‘today’ from *hoc die* ‘(on) this day’ (de Vaan, 2008, s. v.), exactly equivalent to the corresponding English term.

is already present in UD with the case/mark alternation for adpositions (ADP) introducing respectively noun phrases and clauses.

Admittedly, these phenomena are very complex and a complete treatment of them lies outside of the scope of the present paper, so that the previous suggestions have to be left as tentative and hopefully as a stimulus for further research.

## 6. Conclusions and perspectives

The topic touched upon in this paper is of vast scope: the definition of adverbs in general, and the cases presented in §5 are each dealt with by a rich literature, branching out and intertwining with many other subjects in linguistics. This means that, beyond setting the main points for an improved adverbial annotation schema in UD and beyond (§3 and §4), the discussion in a work like the present one inevitably must be concise when it comes to specific cases (as in §5). This notwithstanding, the author's hope is that the need for "tidying up" a neglected, traditionally poorly defined (§1) part of speech has been convincingly motivated and brought (maybe not for the first time) to the attention of the community. Further, it can be asserted that theoretical and practical problems raised by a more systematic treatment of UD's ADV class can have a positive impact in contributing to an ever more consistent and universal annotation style, also beyond adverbs themselves.

## Acknowledgements

I wish to thank the anonymous reviewers who helped to shape a chaotic heap of ideas into something which, I hope, is now understandable. Also, I thank the community of annotators for the invaluable contributes that the confrontation of ideas brings about every day.

## Bibliography

- Anderson, Susan, Paul C. Jordan, Dwight Day, Eugene Emil Loos, and J. Douglas Wingate, editors. *Glossary of linguistic terms*. LinguaLinks library 5.0 plus. SIL International Digital Resources, Dallas, TX, USA, 2003. URL <https://glossary.sil.org/term>. online resource.
- Andrews, Avery DeLano. The major functions of the noun phrase. In Shopen, Timothy, editor, *Language Typology and Syntactic Description*, volume 1: Clause Structure, chapter 3, pages 132–223. Cambridge University Press, Cambridge, UK, second edition, 2007. ISBN 9780521588577. doi: 10.1017/CBO9780511619427. URL <https://www.cambridge.org/core/books/abs/language-typology-and-syntactic-description/major-functions-of-the-noun-phrase/9D1FECC4F028BADD7F1FF8CC5716D607>.
- Bamman, David and Gregory Crane. The ancient Greek and Latin dependency treebanks. In Sporleder, Caroline, Antal van den Bosch, and Kalliopi Zervanou, editors, *Language Technology for Cultural Heritage, Theory and Applications of Natural Language Processing*, pages 79–98, Berlin/Heidelberg, Germany, 2011. Springer. doi: 10.1007/978-3-642-20227-8\_5. URL <http://nlp.perseus.tufts.edu/docs/latech.pdf>.
- Barz, Irmhild, Cathrine Fabricius-Hansen, Peter Eisenberg, Reinhard Fiehler, Thomas A. Fritz, Peter Gallmann, Damaris Nübling, and Jörg Peters. *Die Grammatik*. Number 4 in *Der Duden in zwölf Bänden*. Dudenverlag, Mannheim, Germany; Vienna, Austria; Zürich, Switzerland, eighth edition, July 2009. ISBN 978-3-411-04048-3. URL <https://de.book-info.com/isbn/3-411-04048-3.htm>. NB: now superseded by a completely revised 10th edition.
- Borges Völker, Emanuel, Maximilian Wendt, Felix Hennig, and Arne Köhn. HDT-UD: A very large Universal Dependencies Treebank for German. In Rademaker, Alexander and Francis Tyers, editors, *Proceedings of the Third Workshop on Universal Dependencies (udw, SyntaxFest 2019)*, pages 46–57, Paris, France, August 2019. Association for Computational Linguistics (ACL). doi: 10.18653/v1/W19-8006. URL <https://aclanthology.org/W19-8006>.
- Bos, Gijssbertha F. L’adverbe en latin. *Bulletin de la Société de Linguistique de Paris*, 62(1):106–122, 1967.
- Busa, Roberto. *Index Thomisticus*. Frommann-Holzboog, Stuttgart - Bad Cannstatt, Germany, 1974–1980. Now browsable at <http://www.corpusthomicum.org/it/index.age>.
- Cecchini, Flavio Massimiliano. *Formae reformandae*: for a reorganisation of verb form annotation in Universal Dependencies illustrated by the specific case of Latin. In Osenova, Petya, Kiril Simov, Myriam de Lhoneux, and Reut Tsarfaty, editors, *Proceedings of the Fifth Workshop on Universal Dependencies (udw, SyntaxFest 2021)*, pages 1–15, Sofia, Bulgaria, December 2021. The Association for Computational Linguistics (ACL). ISBN 978-1-955917-17-9. URL <https://aclanthology.org/2021.udw-1.1/>.
- Cecchini, Flavio Massimiliano, Marco Passarotti, Paola Marongiu, and Daniel Zeman. Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies. In de Marneffe, Marie-Catherine, Teresa Lynn, and Sebastian Schuster, editors, *Proceedings of the Second Workshop on Universal Dependencies (udw 2018)*, pages 27–36, Brussels, Belgium, November 2018. The Association for Computational Linguistics (ACL). ISBN 978-1-948087-78-0. doi: 10.18653/v1/W18-6004. URL <https://www.aclweb.org/anthology/W18-6004>.

- Cecchini, Flavio Massimiliano, Timo Korhakangas, and Marco Passarotti. A New Latin Treebank for Universal Dependencies: Charters between Ancient Latin and Romance Languages. In Calzolari, Nicoletta, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, pages 933–942, Marseille, France, May 2020a. European Language Resources Association (ELRA). ISBN 979-10-95546-34-4. doi: 10.5281/zenodo.3862154. URL <https://aclanthology.org/2020.lrec-1.117>.
- Cecchini, Flavio Massimiliano, Rachele Sprugnoli, Giovanni Moretti, and Marco Passarotti. UDante: First Steps Towards the Universal Dependencies Treebank of Dante’s Latin Works. In Monti, Johanna, Felice Dell’Orletta, and Fabio Tamburini, editors, *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020, Bologna, Italy, March 1–3 2021)*, pages 99–105, Turin, Italy, 2020b. Associazione italiana di linguistica computazionale (AILC), aAccademia University Press. ISBN 979-12-80136-28-2. doi: 10.4000/books.aaccademia.8653. URL [http://ceur-ws.org/Vol-2769/paper\\_14.pdf](http://ceur-ws.org/Vol-2769/paper_14.pdf).
- Cinque, Guglielmo. *Adverbs and Functional Heads*. Oxford Studies in Comparative Syntax. Oxford University Press, Oxford, UK, 1999. ISBN 978-0-19-511526-0. doi: 10.1093/oso/9780195115260.001.0001. URL [https://www.academia.edu/58433772/Adverbs\\_and\\_Functional\\_Heads](https://www.academia.edu/58433772/Adverbs_and_Functional_Heads).
- Croft, William. *Syntactic Categories and Grammatical Relations*. University of Chicago Press, Chicago, IL, USA; London, UK, 1991. URL <https://archive.org/details/syntacticcategor0000crof/mode/2up>.
- Croft, William. *Radical Construction Grammar*. Oxford University Press, Oxford, UK, 2001. doi: 10.1093/acprof:oso/9780198299554.001.0001.
- de Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, July 2021. ISSN 0891-2017. doi: 10.1162/coli\_a\_00402. URL <https://direct.mit.edu/coli/article/47/2/255/98516/Universal-Dependencies>.
- de Vaan, Michiel. *Etymological Dictionary of Latin and the other Italic Languages*. Number 7 in Leiden Indo-European Etymological Dictionary Series. Brill, Leiden, Netherlands; Boston, MA, USA, 2008. ISBN 978-90-04-16797-1. URL <https://brill.com/view/title/12612>.
- Eckhoff, Hanne Martine, Kristin Bech, Gerlof Bouma, Kristine Eide, Dag Haug, Odd Einar Haugen, and Marius Jøhndal. The PROIEL treebank family: a standard for early attestations of Indo-European languages. *Language Resources and Evaluation*, 52(1):29–65, 2018. doi: 10.1007/s10579-017-9388-5.
- Gaffiot, Félix. *Dictionnaire Latin-Français*. 2016. URL [gaffiot.fr](http://gaffiot.fr).
- Gamba, Federica and Daniel Zeman. Universalising Latin Universal Dependencies: a harmonisation of Latin treebanks in UD. In Grobol, Loïc and Francis Tyers, editors, *Proceedings of the Sixth Workshop on Universal Dependencies (UDW, GURT/SyntaxFest 2023)*, pages 7–16, Washington, D. C., USA, March 2023. Association for Computational Linguistics (ACL). URL <https://aclanthology.org/2023.udw-1.2>.

- Gerdes, Kim and Sylvain Kahane. Dependency Annotation Choices: Assessing Theoretical and Practical Issues of Universal Dependencies. In Friedrich, Annemarie and Katrin Tomanek, editors, *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-x 2016)*, pages 131–140, Berlin, Germany, August 2016. Association for Computational Linguistics (ACL). doi: 10.18653/v1/W16-1715. URL <https://aclanthology.org/W16-1715>.
- Gerdes, Kim, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. *SUD* or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to *UD*. In de Marneffe, Marie-Catherine, Teresa Lynn, and Sebastian Schuster, editors, *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 66–74, Brussels, Belgium, November 2018. The Association for Computational Linguistics (ACL). ISBN 978-1-948087-78-0. doi: 10.18653/v1/W18-6008. URL <https://aclanthology.org/W18-6008/>.
- Glare, Peter Geoffrey William. *Oxford Latin Dictionary*. Oxford Languages. Oxford University Press, Oxford, UK, 2<sup>nd</sup> edition, 2012. ISBN 978-0-19-958031-6. URL <https://global.oup.com/academic/product/oxford-latin-dictionary-9780199580316?cc=us&lang=en&#>.
- Greenough, James Bradstreet, George Lyman Kittredge, Albert Andrew Howard, and Benjamin Leonard D'Ooge. *New Latin Grammar for Schools and Colleges*. Dickinson College Commentaries, Carlisle, PA, USA, 2014. URL <https://dcc.dickinson.edu/grammar/latin/credits-and-reuse>.
- Hajič, Jan. Building a Syntactically Annotated corpus: The Prague Dependency Treebank. In Hajičová, Eva and Barbora Vidová Hladká, editors, *Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová*, pages 106–132. Karolinum, Prague, Czech Republic, 1998. ISBN 80-7184-601-5. URL <https://ufal.mff.cuni.cz/pdt2.0/publications/Hajic1998.pdf>. Complemented by the guidelines for the analytical layer of PDT 2.0 at <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/index.html>.
- Hajič, Jan, Jarmila Panevová, Eva Buráňová, Zdeňka Uřešová, and Alevtina Bémová. Annotations at analytical level. Instructions for annotators. Technical report, Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, Prague, Cech Republic, 1999. URL <https://ufal.mff.cuni.cz/pdt-c/publications/PDT20-a-man-en.pdf>.
- Hajič, Jan, Eduard Bejček, Alevtina Bémová, Eva Buráňová, Eva Fučíková, Eva Hajičová, Jiří Havelka, Jaroslava Hlaváčová, Petr Homola, Pavel Ircing, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, David Mareček, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Michal Novák, Petr Pajas, Jarmila Panevová, Nino Peterek, Lucie Poláková, Martin Popel, Jan Popelka, Jan Romportl, Magdaléna Rysová, Jiří Semecký, Petr Sgall, Johanka Spoustová, Milan Straka, Pavel Straňák, Pavlína Synková, Magda Ševčíková, Jana Šindlerová, Jan Štěpánek, Barbora Štěpánková, Josef Toman, Zdeňka Uřešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, and Zdeněk Žabokrtský. Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0), 2020. URL <http://hdl.handle.net/11234/1-3185>. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Hajič, Jan, Eva Hajičová, Marie Mikulová, and Jiří Mírovský. Prague Dependency Treebank. In Ide, Nancy and James Pustejovsky, editors, *Handbook of Linguistic Annotation*,



- pages 555–594. Springer, Dordrecht, the Netherlands, June 2017. ISBN 978-94-024-0879-9. doi: 10.1007/978-94-024-0881-2\_21. URL [https://link.springer.com/chapter/10.1007/978-94-024-0881-2\\_21](https://link.springer.com/chapter/10.1007/978-94-024-0881-2_21).
- Hajič, Jan, Eduard Bejček, Jaroslava Hlaváčová, Marie Mikulová, Milan Straka, Jan Štěpánek, and Barbora Štěpánková. Prague Dependency Treebank - Consolidated 1.0. In Calzolari, Nicoletta, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Héléne Mazo, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, pages 5208–5218, Marseille, France, 2020. European Language Resources Association (ELRA). URL <http://www.lrec-conf.org/proceedings/lrec2020/LREC-2020.pdf>.
- Haspelmath, Martin. Word-class-changing inflection and morphological theory. In Booi, Geert Evert and Jaap van Marle, editors, *Yearbook of Morphology 1995*, Yearbook of Morphology, pages 43–66, Dordrecht, the Netherlands, 1996. Kluwer Academic Publishers. ISBN 978-0-7923-3984-7. doi: 10.1007/978-94-017-3716-6\_3. URL [https://www.researchgate.net/publication/240421084\\_Word-class\\_Changing\\_Inflection\\_and\\_Morphological\\_Theory](https://www.researchgate.net/publication/240421084_Word-class_Changing_Inflection_and_Morphological_Theory).
- Haspelmath, Martin. The European linguistic area: Standard Average European. In Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, editors, *Language Typology and Language Universals*, volume 20/2 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter 107, pages 1492–1509. De Gruyter Mouton, Berlin, Germany, 2001. doi: 10.1515/9783110194265-044. URL <https://zenodo.org/records/1236769>.
- Haspelmath, Martin. On s, a, p, t, and r as comparative concepts for alignment typology. *Linguistic Typology*, 15(3):535–567, 2011. doi: 10.1515/LITY.2011.035. URL [https://www.researchgate.net/publication/275733313\\_On\\_S\\_A\\_P\\_T\\_and\\_R\\_as\\_comparative\\_concepts\\_for\\_alignment\\_typology](https://www.researchgate.net/publication/275733313_On_S_A_P_T_and_R_as_comparative_concepts_for_alignment_typology).
- Haspelmath, Martin. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 51(s1000 – Jubilee Issue: 50 Years Folia Linguistica):31–80, 2017. doi: 10.1515/flin-2017-1005. URL <https://www.degruyter.com/document/doi/10.1515/flin-2017-1005/html>.
- Haspelmath, Martin. Four kinds of lexical items: Words, lexemes, inventorial items, and mental items. Draft, February 2024. URL <https://ling.auf.net/lingbuzz/007880>.
- Hengeveld, Kees. *Non-Verbal Predication*. Number 15 in Functional Grammar Series (FGS). De Gruyter Mouton, Berlin, Germany, 1992. ISBN 9783110137132. doi: 10.1515/9783110883282.
- Höhn, Georg F. K. Towards a consistent annotation of nominal person in Universal Dependencies. In Osenova, Petya, Kiril Simov, Myriam de Lhoneux, and Reut Tsarfaty, editors, *Proceedings of the Fifth Workshop on Universal Dependencies (UDW, SyntaxFest 2021)*, pages 75–83, Sofia, Bulgaria, December 2021. The Association for Computational Linguistics (ACL). ISBN 978-1-955917-17-9. URL <https://aclanthology.org/2021.udw-1.6/>.
- Hummel, Martin. Attribution in Romance: Reconstructing the oral and written tradition. *Folia Linguistica Historica*, 47(Historica-vol-34):1–42, 2013. doi: 10.1515/flih.2013.001. URL <https://www.degruyter.com/document/doi/10.1515/flih.2013.001/html>.

- Kortmann, Bernd. *Adverbial Subordination*. Number 18 in Empirical Approaches to Language Typology. Mouton de Gruyter, Berlin, Germany, 1997. doi: 10.1515/9783110812428.
- Kroon, Caroline. The relationship between grammar and discourse: evidence from the Latin particle *quidem*. *Journal of Latin Linguistics*, 9(2):577–590, December 2005. doi: 10.1515/joll.2005.9.2.577.
- Law, Vivien. *The History of Linguistics in Europe*. Cambridge University Press, Cambridge, UK, 2003. doi: 10.1017/CBO9781316036464.
- Lehmann, Christian. Univerbation. *Folia Linguistica Historica*, 54(s41-s1 - Historica):205–252, 2020. doi: 10.1515/flih-2020-0007. URL [https://www.researchgate.net/publication/347444414\\_Univerbation](https://www.researchgate.net/publication/347444414_Univerbation).
- Lewis, Charlton Thomas and Charles Short. *A Latin Dictionary*. Clarendon Press, Oxford, UK, 1879. ISBN 978-1-99-985578-9. URL <https://logeion.uchicago.edu>.
- Litta, Eleonora and Marco Passarotti. (When) inflection needs derivation: a word formation lexicon for Latin. In Holmes, Nigel, Marijke Ottink, Josine Schrickx, and Maria Selig, editors, *Lemmata Linguistica Latina*, volume 1: Words and Sounds, pages 224–239. De Gruyter, Berlin, Boston, December 2019. ISBN 978-3-11-064758-7. doi: 10.1515/9783110647587-015. URL <http://www.degruyter.com/view/books/9783110647587/9783110647587-015/9783110647587-015.xml>. Interrogable online at <http://wfl.marginalia.it/>.
- Lynch, John, Malcolm Ross, and Terry Crowley, editors. *The Oceanic Languages*. Routledge language family series. Routledge, London, UK; New York, NY, USA, 2011. ISBN 9780700711284. doi: 10.2307/j.ctv893h2b.12. URL <https://www.routledgehandbooks.com/doi/10.4324/9780203820384>.
- Marcus, Mitch, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330, June 1993. URL <https://aclanthology.org/J93-2004>.
- Marouzeau, Jules. *L'ordre des mots dans la phrase latine*. Number xxiv in Collection d'études latines publiée par la société des études latines - Série scientifique. Les Belles Lettres, Paris, France, 1949. doi: 10.3406/crai.1950.78601. URL <https://archive.org/details/lordredesmotsdan03marouoft/>.
- Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher David Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In Calzolari, Nicoletta, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France, May 2020. European Language Resources Association (ELRA). URL <https://aclanthology.org/2020.lrec-1.497>.
- Nivre, Joakim, Ali Basirat, Luise Dürlich, and Adam Moss. Nucleus Composition in Transition-based Dependency Parsing. *Computational Linguistics*, 48(4):849–886, 12 2022. ISSN 0891-2017. doi: 10.1162/coli\_a\_00450. URL [https://doi.org/10.1162/coli\\_a\\_00450](https://doi.org/10.1162/coli_a_00450).

- Osborne, Timothy. *A Dependency Grammar of English*. John Benjamins, Amsterdam, Netherlands; Philadelphia, PA, USA, 2019. ISBN 9789027203458. doi: 10.1075/z.224. URL <https://benjamins.com/catalog/z.224>.
- Osborne, Timothy and Kim Gerdes. The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: a journal of general linguistics*, 4(1):17, 2019. ISSN 2397-1835. doi: 10.5334/gjgl.537. URL <https://www.glossa-journal.org/article/id/5124/>.
- Osenova, Petya and Kiril Simov. БТБ-ТР05: BulTreeBank Stylebook. Technical report 5, BulTreeBank project, Sofia, Bulgaria, August 2004. URL <http://bultreebank.org/wp-content/uploads/2017/04/BTB-TR05.pdf>.
- Partanen, Niko, Rogier Blokland, KyungTae Lim, Thierry Poibeau, and Michael Rießler. The First Komi-Zyrian Universal Dependencies Treebanks. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 126–132, Brussels, Belgium, November 2018. Association for Computational Linguistics (ACL). doi: 10.18653/v1/W18-6015. URL <https://aclanthology.org/W18-6015>.
- Passarotti, Marco. The Project of the Index Thomisticus Treebank. In Berti, Monica, editor, *Digital Classical Philology*, number 10 in Age of Access? Grundfragen der Informationsgesellschaft, pages 299–320. De Gruyter Saur, Berlin, Germany; Boston, MA, USA, 2019. doi: 10.1515/9783110599572-017. URL <https://www.degruyter.com/view/book/9783110599572/10.1515/9783110599572-017.xml>.
- Pfeifer, Wolfgang et al. Etymologisches Wörterbuch des Deutschen. In *DWDS – Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart*. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, Germany, 1993. URL <https://www.dwds.de/d/wb-etymbw>. Digitised and revised version.
- Pinkster, Harm. *On Latin adverbs*. Number 6 in North-Holland linguistic series. North-Holland (Elsevier), Amsterdam, the Netherlands, 1972. ISBN 978-0-7204-6186-2. doi: 10.1515/9789048503995.
- Pinkster, Harm. *Latin Syntax and Semantics*. Romance linguistics. Routledge, London, UK; New York, NY, USA, 1990.
- Pinkster, Harm. *The Oxford Latin Syntax*, volume 1. Oxford University Press, Oxford, UK, 2015. ISBN 978-0-19-928361-3. doi: 10.1093/acprof:oso/9780199283613.001.0001.
- Polian, Gilles. Morphology. In Aissen, Judith, Nora Clearman England, and Roberto Zavala Maldonado, editors, *The Mayan Languages*, Routledge language family series, chapter 8, pages 201–225. Routledge, London, UK; New York, NY, USA, 2017. doi: 10.4324/9781315192345-8. URL <https://www.routledge.com/The-Mayan-Languages/Aissen-England-ZavalaMaldonado/p/book/9780415738026>.
- Prokopidis, Prokopis and Haris Papageorgiou. Universal Dependencies for Greek. In de Marnette, Marie-Catherine, Joakim Nivre, and Sebastian Schuster, editors, *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies*, number 135 in Linköping Electronic Conference Proceedings, pages 102–106, Gothenburg, Sweden, May 2017. Linköping University Electronic Press. URL <https://aclanthology.org/W17-0413>.

- Ramat, Paolo. Les adverbess latins de point de vue de l'indo-européenn. In Van Laer, Sophie and Michèle Fruyt, editors, *Adverbess et évolution linguistique en latin*, Kubaba – Série Grammaire et Linguistique, pages 13–24, Paris, France, January 2009. L'Harmattan. ISBN 978-2-296-07442-2. URL [https://www.academia.edu/7263906/Les\\_adverbess\\_latins\\_du\\_point\\_de\\_vue\\_de\\_l\\_indo\\_européenn](https://www.academia.edu/7263906/Les_adverbess_latins_du_point_de_vue_de_l_indo_européenn).
- Ramat, Paolo and Davide Ricca. Prototypical adverbs: on the scalarity/radiality of the notion of ADVERB. *Rivista di linguistica (now: Italian Journal of Linguistics)*, 6(2):289–326, 1994. URL [https://www.italian-journal-linguistics.com/app/uploads/2021/06/8\\_Ramat-Ricca.pdf](https://www.italian-journal-linguistics.com/app/uploads/2021/06/8_Ramat-Ricca.pdf).
- Ricca, Davide. Adverbs. In Baldi, Philip and Pierluigi Cuzzolin, editors, *New Perspectives on Historical Latin Syntax*, volume 2 of *Trends in Linguistics. Studies and Monographs (TiLSM)*, pages 109–191. De Gruyter Mouton, Berlin, Germany, 2010. ISBN 9783110190823. doi: 10.1515/9783110215458.109. Constituent Syntax: Adverbial Phrases, Adverbs, Mood, Tense.
- Rosén, Hannah. Coherence, sentence modification, and sentence-part modification – the contribution of particles. In Baldi, Philip and Pierluigi Cuzzolin, editors, *New Perspectives on Historical Latin Syntax*, volume 1 of *Trends in Linguistics. Studies and Monographs (TiLSM)*, pages 317–441. De Gruyter Mouton, Berlin, Germany, 2010. ISBN 9783110190823. doi: 10.1515/9783110205626.317. Syntax of the Sentence.
- Savary, Agata, Manfred Sailer, Yannick Parmentier, Michael Rosner, Victoria Rosén, Adam Przepiórkowski, Cvetana Krstev, Veronika Vincze, Beata Wójtowicz, Gyri Smørdal Losnegaard, Carla Parra Escartín, Jakub Waszczuk, Matthieu Constant, Petya Osenova, and Federico Sangati. PARSEME – PARSing and Multiword Expressions within a European multilingual network. In *7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2015)*, Poznań, Poland, November 2015. URL <http://typo.uni-konstanz.de/parseme/images/papers/parseme-ltc-2015.pdf>.
- Savary, Agata, Sara Stymne, Virginica Barbu Mititelu, Nathan Schneider, Carlos Ramisch, and Joakim Nivre. PARSEME Meets Universal Dependencies: Getting on the Same Page in Representing Multiword Expressions. *The Northern European Journal of Language Technology (NEJLT)*, 9(1), February 2023. doi: 10.3384/nejlt.2000-1533.2023.4453. URL <https://nejlt.ep.liu.se/article/view/4453>.
- Scatton, Ernest A. Bulgarian. In Comrie, Bernard and Greville G. Corbett, editors, *The Slavonic Languages*, Routledge language family series, chapter 5, pages 188–248. Routledge, London, UK; New York, NY, USA, 1993. ISBN 9780415280785. URL <https://www.routledge.com/The-Slavonic-Languages/Corbett-Comrie/p/book/9780415280785>.
- Schachter, Paul and Timothy Shopen. Parts-of-speech systems. In Shopen, Timothy, editor, *Language Typology and Syntactic Description*, volume 1: Clause Structure, chapter 1, pages 1–60. Cambridge University Press, Cambridge, UK, second edition, 2007. ISBN 9780521588577. doi: 10.1017/CBO9780511619427.001. URL <https://www.cambridge.org/core/books/abs/language-typology-and-syntactic-description/partsofspeech-systems/C298EDBE67E9C2D4DBE5E43E212584E1>.
- Seyoum, Binyam Ephrem, Yusuke Miyao, and Baye Yimam Mekonnen. Universal Dependencies for Amharic. In McCrae, John P., Christian Chiarcos, Thierry Declerck, Jorge Gracia, and Bettina Klimek, editors, *Proceedings of the Eleventh International Conference on Language*

- Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). ISBN 979-10-95546-19-1. URL <https://aclanthology.org/L18-1350/>.
- Silveira, Natalia, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Christopher David Manning. A Gold Standard Dependency Corpus for English. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2897–2904, Reykjavík, Iceland, May 2014. European Language Resources Association (ELRA). URL <https://aclanthology.org/L14-1067/>.
- Spencer, Andrew. *Lexical Relatedness*. Oxford University Press, Oxford, UK, November 2013. ISBN 9780199679928. doi: 10.1093/acprof:oso/9780199679928.001.0001.
- Spencer, Andrews and Ana Ribeiro Luís. *Clitics*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge, UK, 2012. ISBN 9780521682923. doi: 10.1017/CBO9781139033763. URL [www.cambridge.org/9780521682923](http://www.cambridge.org/9780521682923).
- Stassen, Leon. *Intransitive Predication*. Oxford Studies in Typology and Linguistic Theory. Oxford University Press, Oxford, UK, 2003. ISBN 9780199258932. doi: 10.1093/oso/9780198236931.001.0001. URL <https://global.oup.com/academic/product/intransitive-predication-9780198236931?cc=it&lang=en&>.
- Tesnière, Lucien. *Éléments de Syntaxe Structurale*. Klincksieck, Paris, France, 1959. URL <https://archive.org/details/LucienTesniereElementsDeSyntaxeStructurale>.
- Tesnière, Lucien. *Elements of Structural Syntax*. John Benjamins, Amsterdam, Netherlands; Philadelphia, PA, USA, 2015. ISBN 9789027269997. doi: 10.1075/z.185. URL <https://www.jbe-platform.com/content/books/9789027269997>. Translation of: Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Klincksieck, Paris, France.
- Tyers, Francis M. and Robert Henderson. A corpus of K'iche' annotated for morphosyntactic structure. In Mager, Manuel, Arturo Oncevay, Annette Rios, Ivan Vladimir Meza Ruiz, Alexis Palmer, Graham Neubig, and Katharina Kann, editors, *Proceedings of the First Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)*, pages 10–20, Online, June 2021. Association for Computational Linguistics (ACL). ISBN 78-1-954085-44-2. doi: 10.18653/v1/2021.americasnlp-1.2. URL <https://aclanthology.org/2021.americasnlp-1.2/>.
- van der Auwera, Johan. Introduction. In van der Auwera, Johan and Dónall P. Ó Baoill, editors, *Adverbial Constructions in the Languages of Europe*, volume 20/3 (EUROTYPE) of *Empirical Approaches to Language Typology*, chapter 1, pages 1–23. De Gruyter Mouton, Berlin, Germany, 1998. doi: 10.1515/9783110802610.1. URL <https://www.degruyter.com/document/doi/10.1515/9783110802610.1/html>.
- van der Auwera, Johan and Dónall P. Ó Baoill, editors. *Adverbial Constructions in the Languages of Europe*, volume 20/3 (EUROTYPE) of *Empirical Approaches to Language Typology*. De Gruyter Mouton, Berlin, Germany, 1998. doi: 10.1515/9783110802610. URL <https://www.degruyter.com/document/doi/10.1515/9783110802610/html>.

- Van Laer, Sophie and Michèle Fruyt, editors. *Adverbes et évolution linguistique en latin*. Kubaba – Série Grammaire et Linguistique. L'Harmattan, Paris, France, January 2009. ISBN 978-2-296-07442-2. URL [https://www.editions-harmattan.fr/livre-adverbes\\_et\\_evolution\\_linguistique\\_en\\_latin\\_sophie\\_van\\_laer\\_michele\\_fruyt-9782296074422-27699.html](https://www.editions-harmattan.fr/livre-adverbes_et_evolution_linguistique_en_latin_sophie_van_laer_michele_fruyt-9782296074422-27699.html).
- Vincent, Nigel and Kersti Börjars. Heads and history. In Bárány, András, Theresa Biberauer, Jamie Douglas, and Sten Vikner, editors, *Syntactic architecture and its consequences I: Syntax inside the grammar*, number 9 in Open Generative Syntax, pages 133–157, Berlin, Germany, 2020. Language Science Press. doi: 10.5281/zenodo.3972840.
- Wackernagel, Jacob. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen*, 1(1):333–436, 1892. doi: 10.1515/9783110242430.333.
- Wackernagel, Jacob. *Lectures on Syntax*. Oxford University Press, Oxford, UK, April 2009. ISBN 9780198153023. doi: 10.1093/acprof:oso/9780198153023.001.0001. URL <https://academic.oup.com/book/306>.
- Yavrumyan, Marat M. Universal Dependencies for Armenian. In *International Conference on Digital Armenian – Abstracts*, Paris, France, October 2019. URL <http://www.inalco.fr/actualite/conference-internationale-digital-armenian-3-5-octobre>.
- Zeman, Daniel. *The world of tokens, tags and trees*. Number 19 in Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky (ÚFAL), Prague, Czech Republic, 2018. ISBN 9788088132097. URL <https://mttalks.ufal.ms.mff.cuni.cz/books/2018-zeman>.
- Zeman, Daniel, Martin Popel, Milan Straka, Jan Hajic, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinkova, Jan Hajic jr., Jaroslava Hlavacova, Václava Kettnerová, Zdenka Uresova, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria dePaiva, Kira Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonca, Tatiana Lando, Rattima Nitisaroj, and Josie Li. CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada, August 2017. Association for Computational Linguistics (ACL). doi: 10.18653/v1/K17-3001. URL <http://www.aclweb.org/anthology/K/K17/K17-3001.pdf>.
- Zeman, Daniel, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Salih Furkan Akkurt, Gabrielé Aleksandravičiūtė, Ika Alfina, Avner Algom, Khalid Alnajjar, Chiara Alzetta, Erik Andersen, Lene Antonsen, Tatsuya Aoyama, Katya Aplonova, Angelina Aquino, Carolina Aragon, Glyd Aranes, Maria Jesus Aranzabe, Bilge Nas Arıcan, Þórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Katla Ásgeirsdóttir,

Deniz Baran Aslan, Cengiz Asmazoğlu, Luma Ateyah, Furkan Atmaca, Mohammed Atia, Aitziber Atutxa, Liesbeth Augustinus, Mariana Avelãs, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Starkaður Barkarson, Rodolfo Basile, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Shabnam Behzad, Juan Belieni, Kepa Bengoetxea, İbrahim Benli, Yifat Ben Moshe, Gözde Berk, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnè Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, António Branco, Kristina Brokaitė, Aljoscha Burchardt, Marisa Campos, Marie Candito, Bernard Caron, Gauthier Caron, Catarina Carvalho, Rita Carvalho, Lauren Cassidy, Maria Clara Castro, Sérgio Castro, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Liyanage Chamila, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Jinho Choi, Jayeol Chun, Juyeon Chung, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Claudia Corbetta, Daniela Corbetta, Francisco Costa, Marine Courtin, Benoît Crabbé, Mihaela Cristescu, Vladimir Cvetkoski, Ingerid Løyning Dale, Philemon Daniel, Elizabeth Davidson, Leonel Figueiredo de Alencar, Mathieu Dehouck, Martina de Laurentiis, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Adrian Doyle, Timothy Dozat, Kira Droganova, Magali Sanches Duran, Puneet Dwivedi, Christian Ebert, Hanne Eckhoff, Masaki Eguchi, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Farah Essaidi, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Federica Favero, Jannatul Ferdaousi, Marília Fernanda, Hector Fernandez Alcalde, Amal Fethi, Jennifer Foster, Theodorus Fransen, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Federica Gamba, Marcos Garcia, Moa Gärdenfors, Fabrício Ferraz Gerardi, Kim Gerdes, Luke Gessler, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Kirian Guiller, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Takahiro Harada, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Yidi Huang, Marivel Huerta Mendez, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Olájidé Ishola, Artan Islamaj, Kaoru Ito, Sandra Jagodzińska, Siratun Jannat, Tomáš Jelínek, Apoorva Jha, Katharine Jiang, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Ritván Karahóga, Andre Kåsen, Tolga Kayadelen, Sarveswaran Kengatharaiyer, Václava Kettnerová, Lilit Kharatyan, Jesse Kirchner, Elena Klementieva, Elena Klyachko, Petr Kocharov, Arne Köhn, Abdulatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Mehmet Köse, Alexey Koshevoy, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sandra Kübler, Adrian Kuqi, Oğuzhan Kuyrukçu, Aslı Kuzgun, Sookyoung Kwak, Kris Kyle, Kâbi Laan, Veronika Laippala, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei

Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Lauren Levine, Cheuk Ying Li, Josie Li, Keying Li, Yixuan Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani, Yi-Ju Jessica Lin, Krister Lindén, Yang Janet Liu, Nikola Ljubešić, Irina Lobzhanidze, Olga Loginova, Lucelene Lopes, Stefano Lusito, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Menel Mahamdi, Jean Maillard, Ilya Makarchuk, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Stella Markantonatou, Héctor Martínez Alonso, Lorena Martín Rodríguez, André Martins, Cláudia Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Tatiana Merzhovich, Niko Miekka, Aaron Miller, Karina Mischenkova, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amir-saeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskiy, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinky Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñaicek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaraj, Alireza Nourian, Maria das Graças Volpe Nunes, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Hulda Óladóttir, Adédáyò Olúòkun, Mai Omura, Emeka Onwuegbuzia, Noam Ordan, Petya Osenova, Robert Östling, Lilja Øvre-lid, Şaziye Betül Özateş, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Teresa Paccosi, Alessio Palmero Aprosio, Anastasia Panova, Thiago Alexandre Salgueiro Pardo, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Giulia Pedonese, Angelika Peljak-Łapińska, Siyao Peng, Siyao Logan Peng, Rita Pereira, Sílvia Pereira, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Andrea Peverelli, Jason Phelan, Claudel Pierre-Louis, Jussi Piitulainen, Yuval Pinter, Clara Pinto, Rodrigo Pintucci, Tommi A Pirinen, Emily Pitler, Magdalena Plamada, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Robert Pugh, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andreia Querido, Andriela Rääbis, Alexandre Rademaker, Mizanur Rahoman, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Joana Ramos, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Mathilde Regnault, Georg Rehm, Arij Riabi, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Putri Rizqiyah, Luisa Rocha, Eiríkur Rögnvaldsson, Ivan Roksandic, Mykhailo Romanenko, Rudolf Rosa, Valentin Roşca, Davide Rovati, Ben Rozonoyer, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Aleksí Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Marta Sartor, Mitsuya Sasaki, Baiba Saulīte, Agata Savary, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Emmanuel Schang, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Syeda Shahzadi, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Maria Shvedova, Janine Siewert, Einar Freyr Sigurðsson, João Silva, Aline Silveira, Natalia Silveira, Sara Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Haukur Barri Símónarson, Kiril Simov, Dmitri Sitchinava, Ted Sither, Maria Skachedubova, Aaron Smith, Isabela



Soares-Bastos, Per Erik Solberg, Barbara Sonnenhauser, Shafi Sourov, Rachele Sprugnoli, Vivian Stamou, Steinþór Steingrímsson, Antonio Stella, Abishek Stephen, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Daniel Swanson, Zsolt Szántó, Chihiro Taguchi, Dima Taji, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Dipta Tanaya, Mirko Tavoni, Samson Tella, Isabelle Teller, Marinella Testori, Guillaume Thomas, Sara Tonelli, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sveinbjörn Þórðarson, Vilhjálmur Þorsteinsson, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Elena Vagnoni, Sowmya Vajjala, Socrates Vak, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Uliana Vedenina, Giulia Venturi, Eric Villemonde de la Clergerie, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Shira Wigderson, Sri Hartati Wijono, Vanessa Berwanger Wille, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Qishen Wu, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Arlisa Yuliawati, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, He Zhou, Hanzhi Zhu, Yilun Zhu, Anna Zhuravleva, and Rayan Ziane. Universal Dependencies 2.13, 2023. URL <http://hdl.handle.net/11234/1-5287>. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

**Address for correspondence:**

Flavio Massimiliano Cecchini  
flaviomassimiliano.cecchini@kuleuven.be  
Erasmushuis  
Blijde-Inkomststraat 21  
3000 Leuven  
Belgium



---

**The Prague Bulletin of Mathematical Linguistics**  
**NUMBER 121 JUNE 2024**

---

## **INSTRUCTIONS FOR AUTHORS**

Manuscripts are welcome provided that they have not yet been published elsewhere and that they bring some interesting and new insights contributing to the broad field of computational linguistics in any of its aspects, or of linguistic theory. The submitted articles may be:

- long articles with completed, wide-impact research results both theoretical and practical, and/or new formalisms for linguistic analysis and their implementation and application on linguistic data sets, or
- short or long articles that are abstracts or extracts of Master's and PhD thesis, with the most interesting and/or promising results described. Also
- short or long articles looking forward that base their views on proper and deep analysis of the current situation in various subjects within the field are invited, as well as
- short articles about current advanced research of both theoretical and applied nature, with very specific (and perhaps narrow, but well-defined) target goal in all areas of language and speech processing, to give the opportunity to junior researchers to publish as soon as possible;
- short articles that contain contraversing, polemic or otherwise unusual views, supported by some experimental evidence but not necessarily evaluated in the usual sense are also welcome.

The recommended length of long article is 12–30 pages and of short paper is 6–15 pages.

The copyright of papers accepted for publication remains with the author. The editors reserve the right to make editorial revisions but these revisions and changes have to be approved by the author(s). Book reviews and short book notices are also appreciated.

The manuscripts are reviewed by 2 independent reviewers, at least one of them being a member of the international Editorial Board.

Authors receive a printed copy of the relevant issue of the PBML together with the original pdf files.

The guidelines for the technical shape of the contributions are found on the web site <https://ufal.mff.cuni.cz/pbml>. If there are any technical problems, please contact the editorial staff at [pbml@ufal.mff.cuni.cz](mailto:pbml@ufal.mff.cuni.cz).