

Instructions for CoNLL 2017 UD Shared Task

First Author

Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Second Author

Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Abstract

This document contains the instructions for preparing a camera-ready manuscript for the proceedings of *CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies* (*CoNLL 2017 UD Shared Task* for short). The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used for both papers submitted for review and for final versions of accepted papers. Authors are asked to conform to all the directions reported in this document.

Please always include official *CoNLL 2017 UD Shared Task* results, available at the official *CoNLL 2017 UD Shared Task* website.¹ You may also include additional results obtained after the official evaluation period at TIRA ended and the evaluation test sets were released.

There are a few references we would like you to use in the system description papers to give proper credit, especially to data providers: the core Universal Dependencies paper from LREC 2016 (Nivre et al., 2016), the UD version 2.0 datasets (Nivre et al., 2017a), the baseline UDPipe models (Straka et al., 2016), the baseline SyntaxNet models (Weiss et al., 2015) and the evaluation platform TIRA (Potthast et al., 2014). There is also a reference to the overview (intro) paper in the proceedings (Zeman et al., 2017). If you use the CLAS metric in any way, please cite (Nivre and

Fang, 2017). The `udst2017.bib` file available in the latex style package contains all these .bib entries - please use them.

If you do any post-TIRA-evaluation yourself, using the development and gold evaluation data released May 18, 2017, please cite them, too, as (Nivre et al., 2017b). Please note that the reference provide in the .bib file is provisional only, since we are still collecting the full list of authors (we will publish the final bib reference before the camera ready deadline).

Please also don't forget to fill in the structured Google Forms questionnaire about your system.² The organizers will use these data for the overview paper about *CoNLL 2017 UD Shared Task*.

By default, system description papers will be presented as posters. If there are any exceptions, we will contact those concerned individually.

1 Credits

This document has been adapted from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL-2017 by Margaret Mitchell, ACL-2012 by Maggie Li and Michael White, those from ACL-2010 by Jing-Shing Chang and Philipp Koehn, those for ACL-2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, those for ACL-2005 by Hwee Tou Ng and Kemal Oflazer, those for ACL-2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats. Those versions were written by several people, including John Chen, Henry S. Thompson and Donald Walker.

¹<http://universaldependencies.org/conll17/results.html>

²<https://docs.google.com/forms/d/19gqRx7vTi6q33giFvDUTxFhrLAYSEfDPCE1HgZCn90>

Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

2 Introduction

The following instructions are directed to authors of system description papers submitted to the *CoNLL 2017 UD Shared Task*. All authors are required to adhere to these specifications. Authors are required to provide a Portable Document Format (PDF) version of their papers. **The proceedings are designed for printing on A4 paper.**

3 General Instructions

Manuscripts must be in two-column format. Exceptions to the two-column format include the title, authors' names and complete addresses, which must be centered at the top of the first page, and any full-width figures or tables (see the guidelines in Subsection 3.5). **Type single-spaced.** Start all pages directly under the top margin. See the guidelines later regarding formatting the first page. The manuscript should be printed single-sided and its length should not exceed the maximum page limit described in Section 5. **Please do not number the pages in the camera-ready version.**

At the main conference, papers are anonymized by commenting out `\aclfinalcopy` at the top of this document. Since *CoNLL 2017 UD Shared Task* submissions are not anonymous, the `\aclfinalcopy` command is uncommented; please leave it uncommented even for initial submission. However, for the final camera ready copy, fill in the paper number as assigned by START into `\def\aclpaperid{***}` definition at the top.

3.1 Electronically-available resources

The LaTeX `udst2017.*` style, bibtex and sample files are available as a zipped package at <http://universaldependencies.org/conll17/>.

3.2 Format of Electronic Manuscript

For the production of the electronic manuscript you must use Adobe's Portable Document Format (PDF). PDF files are usually produced from L^AT_EX using the `pdflatex` command. If your version of L^AT_EX produces Postscript files, you can convert these into PDF using `ps2pdf` or `dvipdf`. On Win-

dows, you can also use Adobe Distiller to generate PDF.

Please make sure that your PDF file includes all the necessary fonts (especially tree diagrams, symbols, and fonts with Asian characters). When you print or create the PDF file, there is usually an option in your printer setup to include none, all or just non-standard fonts. Please make sure that you select the option of including ALL the fonts. **Before sending it, test your PDF by printing it from a computer different from the one where it was created.** Moreover, some word processors may generate very large PDF files, where each page is rendered as an image. Such images may reproduce poorly. In this case, try alternative ways to obtain the PDF. One way on some systems is to install a driver for a postscript printer, send your document to the printer specifying "Output to a file", then convert the file to PDF.

It is of utmost importance to specify the **A4 format** (21 cm x 29.7 cm) when formatting the paper. When working with `dvips`, for instance, one should specify `-t a4`. Or using the command `\special{papersize=210mm,297mm}` in the latex preamble (directly below the `\usepackage` commands). Then using `dvipdf` and/or `pdflatex` which would make it easier for some.

Print-outs of the PDF file on A4 paper should be identical to the hardcopy version. If you cannot meet the above requirements about the production of your electronic submission, please contact the publication chairs as soon as possible.

3.3 Layout

Format manuscripts two columns to a page, in the manner these instructions are formatted. The exact dimensions for a page on A4 paper are:

- Left and right margins: 2.5 cm
- Top margin: 2.5 cm
- Bottom margin: 2.5 cm
- Column width: 7.7 cm
- Column height: 24.7 cm
- Gap between columns: 0.6 cm

Papers should not be submitted on any other paper size. If you cannot meet the above requirements

about the production of your electronic submission, please contact the publication chairs above as soon as possible.

3.4 Fonts

For reasons of uniformity, Adobe's **Times Roman** font should be used. In $\text{\LaTeX}2\text{e}$ this is accomplished by putting

```
\usepackage{times}
\usepackage{latexsym}
```

in the preamble. If Times Roman is unavailable, use **Computer Modern Roman** ($\text{\LaTeX}2\text{e}$'s default). Note that the latter is about 10% less dense than Adobe's Times Roman font.

Type of Text	Font Size	Style
paper title	15 pt	bold
author names	12 pt	bold
author affiliation	12 pt	
the word "Abstract"	12 pt	bold
section titles	12 pt	bold
document text	11 pt	
captions	11 pt	
abstract text	10 pt	
bibliography	10 pt	
footnotes	9 pt	

Table 1: Font guide.

3.5 The First Page

Center the title, author's name(s) and affiliation(s) across both columns. Do not use footnotes for affiliations. Do not include the paper ID number assigned during the submission process. Use the two-column format only when you begin the abstract.

Title: Place the title centered at the top of the first page, in a 15-point bold font. (For a complete guide to font sizes and styles, see Table 1) Long titles should be typed on two lines without a blank line intervening. Approximately, put the title at 2.5 cm from the top of the page, followed by a blank line, then the author's names(s), and the affiliation on the following line. Do not use only initials for given names (middle initials are allowed). Do not format surnames in all capitals (e.g., use "Mitchell" not "MITCHELL"). Do not format title and section headings in all capitals as well except for proper names (such as "BLEU")

Command	Output	Command	Output
$\{\backslash"a\}$	ä	$\{\backslash" c\}$	ç
$\{\backslash^e\}$	ê	$\{\backslash u\ g\}$	ğ
$\{\backslash"i\}$	ì	$\{\backslash l\}$	ł
$\{\backslash.I\}$	İ	$\{\backslash~n\}$	ñ
$\{\backslash o\}$	ø	$\{\backslash H\ o\}$	ö
$\{\backslash' u\}$	ú	$\{\backslash v\ r\}$	ř
$\{\backslash aa\}$	å	$\{\backslash ss\}$	ß

Table 2: Example commands for accented characters, to be used in, e.g., $\text{BIB}\text{\TeX}$ names.

that are conventionally in all capitals. The affiliation should contain the author's complete address, and if possible, an electronic mail address. Start the body of the first page 7.5 cm from the top of the page.

The title, author names and addresses should be completely identical to those entered to the electronic paper submission website in order to maintain the consistency of author information among all publications of the conference. If they are different, the publication chairs may resolve the difference without consulting with you; so it is in your own interest to double-check that the information is consistent.

Abstract: Type the abstract at the beginning of the first column. The width of the abstract text should be smaller than the width of the columns for the text in the body of the paper by about 0.6 cm on each side. Center the word **Abstract** in a 12 point bold font above the body of the abstract. The abstract should be a concise summary of the general thesis and conclusions of the paper. It should be no longer than 200 words. The abstract text should be in 10 point font.

Text: Begin typing the main body of the text immediately after the abstract, observing the two-column format as shown in the present document. Do not include page numbers.

Indent: When starting a new paragraph. Use 11 points for text and subsection headings, 12 points for section headings and 15 points for the title.

3.6 Sections

Headings: Type and label section and subsection headings in the style shown on the present document. Use numbered sections (Arabic numerals) in order to facilitate cross references. Number subsections tablwth the section number and the subsection number separated by a dot, in Arabic numerals. Do not number subsections.

Citations: Citations within the text appear in

output	natbib	previous ACL style files
(Gusfield, 1997)	\citep	\cite
Gusfield (1997)	\citet	\newcite
(1997)	\citeyearpar	\shortcite

Table 3: Citation commands supported by the style file. The citation style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.

parentheses as (Gusfield, 1997) or, if the author’s name appears in the text itself, as Gusfield (1997). Using the provided L^AT_EX style, the former is accomplished using \cite and the latter with \shortcite or \newcite. Collapse multiple citations as in (Gusfield, 1997; Aho and Ullman, 1972); this is accomplished with the provided style using commas within the \cite command, e.g., \cite{Gusfield:97,Aho:72}. Append lowercase letters to the year in cases of ambiguities. Treat double authors as in (Aho and Ullman, 1972), but write as in (Chandra et al., 1981) when more than two authors are involved. Collapse multiple citations as in (Gusfield, 1997; Aho and Ullman, 1972). Also refrain from using full citations as sentence constituents.

We suggest that instead of

“(Gusfield, 1997) showed that ...”

you use

“Gusfield (1997) showed that ...”

If you are using the provided L^AT_EX and BibT_EX style files, you can use the command \citet (cite in text) to get “author (year)” citations.

If the BibT_EX file contains DOI fields, the paper title in the references section will appear as a hyperlink to the DOI, using the hyperref L^AT_EX package. To disable the hyperref package, load the style file with the nohyperref option: \usepackage[nohyperref]{acl2017}

Digital Object Identifiers: As part of our work to make ACL materials more widely used and cited outside of our discipline, ACL has registered as a CrossRef member, as a registrant of Digital Object Identifiers (DOIs), the standard for registering permanent URNs for referencing scholarly materials. As of 2017, we are requiring all camera-ready references to contain the appropriate DOIs (or as a second resort, the hyperlinked ACL Anthology Identifier) to all cited works. Thus, please ensure that you use BibT_EX records that

contain DOI or URLs for any of the ACL materials that you reference. Appropriate records should be found for most materials in the current ACL Anthology at <http://aclanthology.info/>.

As examples, we cite (Goodman et al., 2016) to show you how papers with a DOI will appear in the bibliography. We cite (Harper, 2014) to show how papers without a DOI but with an ACL Anthology Identifier will appear in the bibliography.

Please do not use anonymous citations and do not include acknowledgements when submitting your papers. Papers that do not conform to these requirements may be rejected without review.

References: Gather the full set of references together under the heading **References**; place the section before any Appendices, unless they contain references. Arrange the references alphabetically by first author, rather than by order of occurrence in the text. Provide as complete a citation as possible, using a consistent format, such as the one for *Computational Linguistics* or the one in the *Publication Manual of the American Psychological Association* (American Psychological Association, 1983). Use of full names for authors rather than initials is preferred. A list of abbreviations for common computer science journals can be found in the *ACM Computing Reviews* (for *Computing Machinery*, 1983).

The L^AT_EX and BibT_EX style files provided roughly fit the American Psychological Association format, allowing regular citations, short citations and multiple citations as described above.

Appendices: Appendices, if any, directly follow the text and the references (but see above). Letter them in sequence and provide an informative title: **Appendix A. Title of Appendix**.

3.7 Footnotes

Footnotes: Put footnotes at the bottom of the page and use 9 points text. They may be numbered or referred to by asterisks or other symbols.³ Foot-

³This is how a footnote should appear.

notes should be separated from the text by a line.⁴

3.8 Graphics

Illustrations: Place figures, tables, and photographs in the paper near where they are first discussed, rather than at the end, if possible. Wide illustrations may run across both columns. Color illustrations are discouraged, unless you have verified that they will be understandable when printed in black ink.

Captions: Provide a caption for every illustration; number each one sequentially in the form: “Figure 1. Caption of the Figure.” “Table 1. Caption of the Table.” Type the captions of the figures and tables below the body, using 11 point text.

3.9 Accessibility

In an effort to accommodate the color-blind (as well as those printing to paper), grayscale readability for all accepted papers will be encouraged. Color is not forbidden, but authors should ensure that tables and figures do not rely solely on color to convey critical distinctions. Here we give a simple criterion on your colored figures, if your paper has to be printed in black and white, then you must assure that every curves or points in your figures can be still clearly distinguished.

4 Translation of non-English Terms

It is also advised to supplement non-English characters and terms with appropriate transliterations and/or translations since not all readers understand all such characters and terms. Inline transliteration or translation can be represented in the order of: original-form transliteration “translation”.

5 Length of Submission

The *CoNLL 2017 UD Shared Task* Workshop accepts submissions of system papers, consisting of eight (8) pages of content plus unlimited pages for references.

All illustrations and tables that are part of the main text must be accommodated within these page limits, observing the formatting instructions given in the present document.

Acknowledgments

The acknowledgments should go immediately before the references. Do not number the acknowledgments section.

⁴Note the line separating the footnotes from the text.

References

- Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.
- American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.
- Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. [Alternation](#). *Journal of the Association for Computing Machinery* 28(1):114–133. <https://doi.org/10.1145/322234.322243>.
- Association for Computing Machinery. 1983. *Computing Reviews* 24(11):503–512.
- James Goodman, Andreas Vlachos, and Jason Naradowsky. 2016. [Noise reduction and targeted exploration in imitation learning for abstract meaning representation parsing](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 1–11. <https://doi.org/10.18653/v1/P16-1001>.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.
- Mary Harper. 2014. [Learning from 26 languages: Program management and science in the babel program](#). In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, page 1. <http://aclweb.org/anthology/C14-1001>.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association, Portoro, Slovenia, pages 1659–1666.
- Joakim Nivre and Chiao-Ting Fang. 2017. Universal dependency evaluation. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*. pages 86–95.
- Joakim Nivre et al. 2017a. [Universal Dependencies 2.0](#). LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University, Prague, <http://hdl.handle.net/11234/1-1983>. <http://hdl.handle.net/11234/1-1983>.
- Joakim Nivre et al. 2017b. [Universal Dependencies 2.0 CoNLL 2017 shared task development and test data](#). LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University, Prague, <http://hdl.handle.net/>

11234/1-2184. <http://hdl.handle.net/11234/1-2184>.

Martin Potthast, Tim Gollub, Francisco Rangel, Paolo Rosso, Efstathios Stamatatos, and Benno Stein. 2014. Improving the reproducibility of PAN's shared tasks: Plagiarism detection, author identification, and author profiling. In Evangelos Kanoulas, Mihai Lupu, Paul Clough, Mark Sanderson, Mark Hall, Allan Hanbury, and Elaine Toms, editors, *Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. 5th International Conference of the CLEF Initiative (CLEF 14)*. Springer, Berlin Heidelberg New York, pages 268–299. https://doi.org/10.1007/978-3-319-11382-1_22.

Milan Straka, Jan Hajič, and Jana Straková. 2016. UD-Pipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association, Portoro, Slovenia.

David Weiss, Chris Alberti, Michael Collins, and Slav Petrov. 2015. Structured training for neural network transition-based parsing. *CoRR* abs/1506.06158. <http://arxiv.org/abs/1506.06158>.

Daniel Zeman, Filip Ginter, Jan Hajič, Joakim Nivre, Martin Popel, Milan Straka, and et al. 2017. CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Association for Computational Linguistics, pages 1–20.