

Combining local and global word statistics

Ayman Farahat

Department of Computer Science

Nonesuch State University

Utopia, NS 12345

jane.doe@cs.nsu.edu John Smith

Department of Linguistics

Another State University

Collegetown, AS 98765

jsmith@ling.asu.edu

Abstract

This document contains the instructions for preparing a submission to the main track in ACL-2004. The instructions for submitting a camera-ready manuscript for the proceedings of ACL-2004 will be published later, but the format will be identical to that described here. This document conforms to its own specifications, and is therefore an example of what your manuscript should look like. Authors are asked to conform to all the directions reported in this document.

A major *format difference* between the submitted version and the final, camera-ready manuscript (and probably the only one) is that **authors' names and affiliation should not be included in the version submitted for review.**

1 Introduction

The following instructions are directed to authors of papers submitted to the main track in ACL-2004. The submission is electronic. Precise instructions for submission of camera-ready copies of the accepted papers will follow; however, the formatting of the final version will be similar to the submission formatting. All authors are required to adhere to these specifications for the submission. (Note that the proceedings will be printed on DIN A4 paper, and the final version *must* be in A4 format.)

Authors from countries in which access to word-processing systems is limited should contact the PC Walter Daelemans (walter.daelemans@ua.ac.be) as soon as possible.

2 Using the provided style files

The easiest way to correctly format your paper is to use one of the style files provided on the conference web page. For instance, \LaTeX users can use the **acl04.sty** style file, and Microsoft Word users can use the **acl04.dot** document template.

Figure 1 shows a skeleton document for \LaTeX users. To use it, simply copy the text in the figure to a file and then replace the parts flagged by **TODO** as appropriate for your document. Figure 2 shows the simple Bib \TeX file used by the skeleton document.

3 General Instructions

Manuscripts must be in two-column format. Exceptions to the two-column format include the title, which must be centered at the top of the first page, and any full-width figures or tables (see the guidelines in Subsection 4.2). **Use single spacing.** See the guidelines later regarding formatting the first page.

The maximum length of a manuscript is eight (8) pages for the main conference (see Subsection 4.1 for exact page layout and Section 5 for additional information on the maximum number of pages).

4 Format of Electronic Manuscript

For the production of the submission you must use Adobe's Portable Document Format (PDF). This format can be generated from postscript files: on Unix systems, you can use `ps2pdf` for this purpose; under Microsoft Windows, Adobe's Distiller can be used. Note that some word processing programs generate PDF which may not include all the necessary fonts (esp. tree diagrams, symbols). 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 processor may generate very large postscript/PDF files, where each page is rendered as an image. Such images may reproduce poorly. In this case, try alternative ways of obtaining the postscript and/or 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.

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

```

\documentclass[11pt]{article}
\usepackage{acl04}
\usepackage{times}
\usepackage{latexsym}

\title{TODO: put title here}

\author{}

\date{}

\begin{document}
\maketitle

\begin{abstract}
TODO: put the abstract here
\end{abstract}

\section{Introduction}
TODO: add introduction; and,
perhaps cite WordNet \cite{Miller-90}

\section{TODO: add body of paper}

\section{Conclusion}
TODO: add conclusion

\section*{Acknowledgements}
TODO: add acknowledgements

\bibliographystyle{acl}
%% TODO: use base name of your .bib file
\bibliography{my-citations}

\end{document}

```

Figure 1: Skeleton L^AT_EX document illustrating use of **acl04.sty** and **acl.bst**

```

@Article{Miller-90,
  author = {G. Miller},
  title = {Special Issue on {WordNet}},
  journal = {International Journal of
            Lexicography},
  year = {1990},
  volume = {3(4)},
}

```

Figure 2: Bib \TeX file **my-citations.bib** used in skeleton document

```

\usepackage{times}
\usepackage{latexsym}

```

in the preamble.

Additionally, it is of utmost importance to specify the **A4 format** when formatting the paper. When working with `dvips`, for instance, one should specify `-t a4`.

4.1 Layout

Format the manuscript two columns to a page, in the manner these instructions are formatted. In no case should the font for the body of the paper be smaller than 11pt, and the dimensions of the text area proper must be 230mm x 160mm (9" by 6.3") or less.

Papers should not be submitted formatted for any other paper size. Exceptionally, authors for whom it is *impossible* to format on A4 paper may use *US-letter* paper. In this case, they should keep the text area as given above, and modify the bottom and right margins as necessary.

4.2 The First Page

Center the title across both columns. Do not include author names or affiliations. 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. Long titles should be typed on two lines without a blank line intervening. Approximately, put the title at 1in from the top of the page.

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 0.25in 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.

Text: Begin typing the main body of the text immediately after the abstract, observing the two-column format as shown in the present document. Type single spaced.

Indent when starting a new paragraph. For reasons of uniformity, use Adobe's **Times Roman** fonts, with 11 points for text and subsection headings, 12 points for section headings and 15 points for the title. If Times Roman is unavailable, use **Computer Modern Roman** (L^AT_EX2e's default; see section 4 above). Note that the latter is about 10% less dense than Adobe's Times Roman font.

4.3 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 with the section number and the subsection number separated by a dot, in Arabic numerals. Do not number subsections.

Citations: Follow the "Guidelines for Formatting Submissions" to *Computational Linguistics* that appears in the first issue of each volume, if possible. That is, citations within the text appear in parentheses as [Gusfield1997] or, if the author's name appears in the text itself, as Gusfield Gusfield:97. Append lowercase letters to the year in cases of ambiguities. Treat double authors as in [Aho and Ullman1972], but write as in [Chandra et al.1981] when more than two authors are involved. Collapse multiple citations as in [Gusfield1997, Aho and Ullman1972].

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 Association1983]. 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* [Association for Computing Machinery1983].

The provided L^AT_EX and BibT_EX style files 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.

The **Acknowledgements** section should go as the last section immediately before the references. Do not number the acknowledgements section.

4.4 Footnotes

Footnotes: Put footnotes at the bottom of the page. They may be numbered or referred to by asterisks or other symbols.¹ Footnotes should be separated from the text by a line.²

4.5 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. Do not use color illustrations as they may reproduce poorly.

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.

5 Length of Submission

Eight pages (8) is the maximum length of papers for submissions to the ACL-2004 main conference. All illustrations, references, and appendices must be accommodated within these page limits, observing the formatting instructions given in the present document. Papers that do not conform to the specified length and formatting requirements are subject to be rejected without review.

Acknowledgements

This document has been adapted from the instructions for ACL 2003. These were adapted by Tom O’Hara and Dan Roth from the instructions for the ACL-2002 proceedings by Dekang Lin and Eugene Charniak. It was adapted in turn from the ACL-01 proceedings by Norbert Reithinger, Giorgio Satta, and Roberto Zamparelli. The ACL-01 instructions were elaborated from similar documents used for previous editions of the ACL and EACL annual meetings. Those versions were written by several people, including John Chen, Henry S. Thompson and Donald

¹This is how a footnote should appear.

²Note the line separating the footnotes from the text.

Walker. Additional elements were taken from the formatting instructions of the Xth *International Joint Conference on Artificial Intelligence*.

References

- [Aho and Ullman1972] Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.
- [American Psychological Association1983] American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.
- [Association for Computing Machinery1983] Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.
- [Chandra et al.1981] Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.
- [Gusfield1997] Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.