

Visualising Typological Relations: Using Heat Maps with WALS

Richard Littauer

University of Saarland
Computational Linguistics Department
66041 Saarbrücken, Germany
richard.littauer@gmail.com

Abstract

This paper presents a novel way of visualising linguistic typological data. Computational methods have only recently been applied to large typological databases. Many studies have focused on discovering relations between languages using typology, often using sophisticated statistical techniques. However, few papers have provided new ways of visually presenting the resulting data. Here, we show that one can use the data from the World Atlas of Language Structures to develop both heat maps and admixture graphs that visually show the interconnected relationships between languages and language families. We hope that the images will bring a new perspective to the data, resulting in interesting findings and illuminating areas of research.

1 Credits

This document has been adapted from the instructions for earlier ACL proceedings, including 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. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence*.

2 Introduction

The following instructions are directed to authors of papers submitted to EACL-2012 or accepted

for publication in its proceedings. 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 will be printed on A4 paper.** Authors from countries in which access to word-processing systems is limited should contact the publication chairs Fabrice Lefèvre (fabrice.lefevre@univ-avignon.fr) and Adrià de Gispert (ad465@eng.cam.ac.uk) as soon as possible.

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 ??). **Type single-spaced.** Start all pages directly under the top margin. See the guide-lines later regarding formatting the first page. Do not number the pages.

3.1 Electronically-available resources

EACL-2012 provides this description in $\text{\LaTeX}2\text{e}$ (eac12012.tex) and PDF format (eac12012.pdf), along with the $\text{\LaTeX}2\text{e}$ style file used to format it (eac12012.sty) and an ACL bibliography style (acl.bst). These files are all available at <http://www.eacl2012.org>. A Microsoft Word template file (eac12012.dot) is also available at the same URL. We strongly recommend the use of these style files, which have been appropriately tailored for the EACL-2012 proceedings. If you have an option, we recommend that you use the $\text{\LaTeX}2\text{e}$ version. **If you will be using the**

Microsoft Word template, we suggest that you anonymize your source file so that the pdf produced does not retain your identity. This can be done by removing any personal information from your source document properties.

3.2 Format of Electronic Manuscript

For the production of the electronic manuscript you must use Adobe’s Portable Document Format (PDF). This format can be generated from postscript files: on Linux/Unix systems, you can use `ps2pdf` for this purpose; under Microsoft Windows, you can use Adobe’s Distiller, or if you have `cygwin` installed, you can use `dvipdf` or `ps2pdf`. 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 to obtain 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.

Additionally, it is of utmost importance to specify the **A4 format** (8.27in × 11.69in) when formatting the paper. When working with `dvips`, for instance, one should specify `-t a4`.

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 chair above 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: 1in
- Top margin: 1in
- Bottom margin: 1in

- Column width: 3in
- Column height: 9.6in
- Gap between columns: 0.2in

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 chair 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 ??.) Long title should be typed on two lines without a blank line intervening. Approximately, put the title at 1in 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., "Zhou," not "ZHOU"). The affiliation should contain the author's complete address, and if possible an electronic mail address. Leave about 0.75in between the affiliation and the body of the first 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.

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. 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. 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 ?? above). Note that the latter is about 10% less dense than Adobe's Times Roman font.

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

Citations: Citations within the text appear in parentheses as (?) or, if the author's name appears in the text itself, as Gusfield (?). Append lowercase letters to the year in cases of ambiguities. Treat double authors as in (?), but write as in (?) when more than two authors are involved. Collapse multiple citations as in (?: ?). Also refrain

from using full citations as sentence constituents. We suggest that instead of

"(?) showed that ..."

you use

"Gusfield (?) showed that ..."

If you are using the provided L^AT_EX and BibT_EX style files, you can use the command `\newcite` to get "author (year)" citations.

As reviewing will be double-blind, the submitted version of the papers should not include the authors' names and affiliations. Furthermore, self-references that reveal the author's identity, e.g.,

"We previously showed (?) ..."

should be avoided. Instead, use citations such as

"Gusfield (?) previously showed ..."

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* (?). 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* (?).

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**.

Acknowledgment sections should go as a last section immediately before the references. Do not number the acknowledgement section.

3.7 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 Footnotes should be in 9 point font.

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 and should be placed at the top of a page. 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 10 point text.

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

Papers may consist of up to nine (9) pages of content and an unlimited number of reference pages. Papers that do not conform to the specified length and formatting requirements are subject to re-submission.

6 Other Issues

Those papers that had software and/or dataset submitted for the review process, should also submit it with the camera-ready paper. Besides, the software and/or dataset should not be anonymous.

Please note that the publications of EACL 2012 will be publicly available at ACL Anthology (<http://aclweb.org/anthology-new/>) on April 19th, 2012, the week before the start of the conference.

Since some of the authors may have plans to file patents related to their papers in the conference, we are sending this reminder that April 19th, 2012 may be considered to be the official publication date, instead of the opening day of the conference.

Acknowledgments

Do not number the acknowledgment section. Do not include this section when submitting your paper for review.

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.
- Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.
- Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

¹This is how a footnote should appear.

²Note the line separating the footnotes from the text.