L1ML: Identification of First Language Given English Text

Yuzhe Chen, Jessica Huynh, Ryan Nicoll

Brandeis University

Waltham, Massachusetts USA

{yzchen,jhuynh,rnicoll}@brandeis.edu

Abstract

Each article must include an abstract of 150 to 200 words in Times 9 pt with interlinear spacing of 10 pt. The heading Abstract should be centred, font Times 10 bold. This short abstract will also be used for printing a Booklet of Abstracts containing the abstracts of all papers presented at the Conference.

Keywords: keyword A, keyword B, keyword C

1. Paper

Each manuscript should be submitted on white **A4 paper.** The fully justified text should be formatted in two parallel columns, each 8.25 cm wide, and separated by a space of 0.63 cm. Left, right, and bottom margins should be 1.9 cm. and the top margin 2.5 cm. The font for the main body of the text should be Times 10 pt with interlinear spacing of 12 pt. **Articles must be between 4 and 8 pages in length**, regardless of the mode of presentation (oral or poster).

2. General Instructions

Each paper is allocated between <u>a minimum of four and a maximum of eight pages</u> including figures.

The unprotected PDF files will appear in the on-line proceedings directly as received. Do not print the page number.

3. Page Numbering

Please do not include page numbers in your article. The definitive page numbering of articles published in the proceedings will be decided by the organizing committee.

4. Headings / Level 1 Headings

Headings should be capitalised in the same way as the main title, and centred within the column. The font used is Times 12 bold. There should also be a space of 12 pt between the title and the preceding section, and a space of 3 pt between the title and the text following it.

4.1. Level 2 Headings

The format for level 2 headings is the same as for level 1 Headings, with the font Times 11, and the heading is justified to the left of the column.

4.1.1. Level 3 Headings

The format for level 3 headings is the same as for level 2 headings, except that the font is Times 10, and there should be no space left between the heading and the text.

5. Citing References in the Text

5.1. Bibliographical References

All bibliographical references within the text should be placed in parentheses containing the author's surname followed by a comma before the date of publication (Strötgen and Gertz, 2012). If the sentence already includes the author's name, then it is only necessary to put the date in parentheses: Strötgen and Gertz (2012). When several authors are cited, those references should be separated with a semicolon: (Strötgen and Gertz, 2012; Castor and Pollux, 1992). When the reference has more than three authors, only cite the name of the first author followed by et al. (e.g. (Superman et al., 2000)).

5.2. Language Resource References

5.2.1. When Citing Language Resources

When mentioning language resources, we recommend that they are cited in a similar way to bibliographical references, except that, in order to make them appear in a separate section, you need to use the `citelanguageresource` tag. Thus, a language resource should be cited as (Speecon Consortium, 2014).

5.2.2. When Not Citing Any Language Resource

When no language resource needs to be cited in the paper, you need to comment out a few lines in the `.tex` file:

```
% \usepackage{multibib}
% \newcites{languageresource}{}
% \section{Language Resource References}
% \bibliographystylelanguageresource
%    {lrec2016}
% \bibliographylanguageresource{xample}
```

6. Figures & Tables

6.1. Figures

All figures should be centred and clearly distinguishable. They should never be drawn by hand, and the lines must be very dark in order to ensure a high-quality printed version. Figures should be numbered in the text, and have a caption in Times 10 pt underneath. A space must be left between each figure and its respective caption.

Example of a figure enclosed in a box:

Figure and caption should always appear together on the same page. Large figures can be centred, using a full page.

6.2. Tables

The instructions for tables are the same as for figures.

Figure 1: The caption of the figure.

Level	Tools
Morphology	Pitrat Analyser
Syntax	LFG Analyser (C-Structure)

Table 1: The caption of the table

7. Footnotes

Footnotes are indicated within the text by a number in superscript[1].

8. Copyrights

The The Language Resource and Evaluation Conference (LREC) proceedings are published by the European Language Resources Association (ELRA). They are available online from the conference website.

ELRA's policy is to acquire copyright for all LREC contributions. In assigning your copyright, you are not forfeiting your right to use your contribution elsewhere. This you may do without seeking permission and is subject only to normal acknowledgement to the LREC proceedings. The LREC 2016 Proceedings are licensed under CC-BY-NC, the Creative Commons Attribution-NonCommercial 4.0 International License.

9. Conclusion

Your submission of a finalized contribution for inclusion in the LREC proceedings automatically assigns the above-mentioned copyright to ELRA.

10. Acknowledgements

Place all acknowledgements (including those concerning research grants and funding) in a separate section at the end of the article.

11. Providing References

11.1. Bibliographical References

Bibliographical references should be listed in alphabetical order at the end of the article. The title of the section, "Bibliographical References", should be a level 1 heading. The first line of each bibliographical reference should be justified to the left of the column, and the rest of the entry should be indented by 0.35 cm.

The examples provided in Section 12 (some of which are fictitious references) illustrate the basic format required for articles in conference proceedings, books, journals articles, Ph.D. theses, and chapters of books.

11.2. Language Resource References

Language resource references should be listed in alphabetical order at the end of the article, in the "Language Resource References" section, placed after the "Bibliographical References" section. The title of the "Language Resource References" section, should be a level 1 heading. The first line of each language resource reference should be justified to the left of the column, and the rest of the entry should be indented by 0.35 cm. The example in Section 13 illustrates the basic format required for language resources. In order to be able to cite a language resource, it must be added to the `.bib` file first, as a `@LanguageResource` item type, which contains the following fields:

- `author`: the builder of the resource

- `title`: the name of the resource

- `publisher`: the publisher of the resource (project, organization etc)

- `year`: year of the resource release

- `series`: more general resource set this language resource belongs to

- `edition`: version of the resource

- `islrn`: the International Standard Language Resource Number (ISLRN) of the resource[2]

If you want the full resource author name to appear in the citation, the language resource author name should be protected by enclosing it between `{...}`, as shown in the model `.bib` file.

Appendix: How to Produce the `.pdf` Version

In order to generate a PDF file out of the LaTeX file herein, when citing language resources, the following steps need to be performed:

- Compile the `.tex` file once

- Invoke `bibtex` on the eponymous `.aux` file

- Invoke `bibtex` on the `languageresources.aux` file

- Compile the `.tex` file twice

---

[1]They should be in Times 9, and appear at the bottom of the same page as their corresponding number. Footnotes should also be separated from the rest of the text by a horizontal line 5 cm long.

[2]The ISLRN number is available from http://islrn.org

12. Bibliographical References

Castor, A. and Pollux, L. E. (1992). The use of user modelling to guide inference and learning. *Applied Intelligence*, 2(1):37–53.

Strötgen, J. and Gertz, M. (2012). Temporal tagging on different domains: Challenges, strategies, and gold standards. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3746–3753, Istanbul, Turkey, may. European Language Resource Association (ELRA).

Superman, S., Batman, B., Catwoman, C., and Spiderman, S. (2000). *Superheroes experiences with books*. The Phantom Editors Associates, Gotham City, 20th edition.

13. Language Resource References

Speecon Consortium. (2014). *Dutch Speecon Database*. Speecon Project, distributed via ELRA, Speecon resources, 1.0, ISLRN 613-489-674-355-0.