

Author Guidelines for RANLP Proceedings*

Nicolas Nicolov[†]
Umbria Inc.
1655 Walnut Str.
Boulder, CO 80302
nicolas_nicolov@yahoo.com

Abstract

This document provides a L^AT_EX sample (of the formatting) of a paper for the International Conference on Recent Advances in Natural Language Processing (RANLP).

If the abstract is longer than half the textheight perhaps you are trying to say too much.

Keywords

Keywords are useful to the organizers to decide who should review a submission. Also the readers of the proceedings will find these helpful.

1 Introduction

This document is a sample of how to format a submission (and later, the camera ready copy) to the International Conference on Recent Advances in Natural Language Processing (RANLP). It is formatted using a popular scientific typesetting system—L^AT_EX (actually L^AT_EX 2_ε). In the rest of the document we mention some of the features that are available to authors. More importantly, we also provide recommendations about the use of certain formatting features.

We have two goals in writing these guidelines: (i) if you are not a L^AT_EX 2_ε user we would like to show you some of the available functionality and make it easy for you to typeset your document in L^AT_EX; (ii) if you are a proficient L^AT_EX user we'd like the formatting to be as straightforward as possible.

The main reason scientists use L^AT_EX is the quality of the final manuscript. People learning about the system often compare it to WYSIWYG (what you see is what you get) interfaces and find various aspects unintuitive (namely, you don't see how things appear). Actually there is a reason for this—L^AT_EX is a mark-up language and how things appear is dictated by templates. Other reasons for using L^AT_EX are: free availability and cross-platform compatibility.

This document is not a guide on good writing! It will not teach you *what to say* and in what order; it is neither a guide on *how to say* it but on how to *typeset* it. Even worse, this document is not a short guide to the L^AT_EX system but rather a few comments on basic commands related to common structures of a RANLP submission.

2 Anatomy of a paper

Let's start with the basic structure of a paper:

```
\documentclass{ranlp}
\begin{document}
\title{...}
\author{
  Sherlock Homes\footnote{Has worked
    on cases with Dr.\,Watson.}\
  S.H.\ Consulting      \
  221b Baker Street    \
  London, England      \
  \email{sh}{logic.co.uk} \
}
\date{}
\maketitle
\begin{abstract}
In this paper ...
\end{abstract}
\keywords{Statistical MT}

\section{Introduction}
Language processing is ...

\section{Formalism}
...
\section{Experiments}
...
\section{Conclusions}
...
\bibliographystyle{abbrv}
\bibliography{ranlp-biblio}
\end{document}
```

3 ‘Compiling’ the paper

Put this in a file `F.tex`; fill in the sections with your great thoughts (plain text) and you are mostly there. Then you run the commands:

```
latex F
bibtex F
latex F
latex F
dvips -Pcmz -f -t a4 < F.dvi > F.ps
ps2pdf F.ps > F.pdf
```

This will ultimately produce the file `F.pdf`. Alternatively, you can modify the supplied `Makefile` to reflect the name of your submission and run `make`.

You probably would like to use more personal names than `F` or `ranlp`. Why? Because you don't want your

*Footnote on the title—to be avoided at all cost.

[†]Formerly at IBM T.J. Watson Research Center.

submission to be overwritten by the submission of another person who decided to name their file using the same name.

4 The authors

At the beginning of the document there is a description of the authors. It is also good to discuss among the co-authors the order in which the names will appear well in advance. There are many reasons people fall out of friendship. For scientists disagreements about the order of authors ranks fairly high on such a list of reasons.

5 The body of the paper

Typically, the body of a paper is organized into a hierarchical structure, with numbered (or unnumbered) headings for sections, subsections, sub-subsections, and even smaller units. The heading for the current section is produced by the command:

```
\section{The body of the paper}
```

The numbering, font size and bold face are produced automatically. If you want a sub-subsection or smaller part to be unnumbered in your output, simply append an asterisk to the command name (e.g., `\subsection*{...}`). Examples of both numbered and unnumbered headings will appear throughout this document.

Use a blank line to represent a paragraph break.

6 Type changes and *special* characters

You can indicate *italicized* words or phrases in your text with the command `\textit{}` (or `\emph{}`); **boldening** with the command `\textbf{}`; **typewriter-style** (for instance, for computer code) with `\texttt{}`; **sans serif** with `\textsf{}` and **underlining** with `\underline{}`.

6.1 Macros

If you find you are often using a term that you'd like to typeset in a special way it is a good idea to consider a macro.

```
% Support Vector Machine
\newcommand{\SVM}{\textsc{svm}}
```

When you use the macro you may need to pay attention to the next character. If the next character is a punctuation you are ok:

```
‘‘We used an \SVM.’’
```

produces: “We used an SVM.”

If the next character is a space:

```
We used an \SVM\ to induce a classifier.
```

The safest way is to put braces around it:

```
We used severals {\SVM}s in cascade.
```

Macro definitions are placed before the `\begin{document}` command. If you have many macros you might want to group them into a file to be used by your other papers. In the demo directory you will find a file called `ran1pmacros.sty` (loaded by `ran1p.cls`).

7 A rather long section w-w-w-w title: An example

Short headings might be a bit too cryptic. Very long headings slow down a reader who is skimming the paper (not too mention that esthetically they don't look so great).

7.1 Capitalization

The title of your paper should have content words capitalized (the first letter is capital; the rest are lower-case unless the word is an acronym). Well, if one of the words is iPod we will make an exception for your paper. From there onwards the less you capitalize the better. There are some terms that are commonly used by researchers in the field and you will need to mention them. Resist the temptation to capitalize them. The main reasons are again slowing readers and putting emphasis on less important aspects. Terms like ‘information technology’, ‘the web’, etc. are good candidates to keep lowercased. Our recommendation is not to capitalize non-initial content words in section headings. If you have a subheading following a column (:) you may capitalize the next word (see Section 7).

7.2 Indentation

No indentation of:

- the paragraph after section headings (this happens automatically if you use `\section` commands);
- first paragraph on new pages;
- top paragraph of second column;
- paragraph after lists environments;
- paragraph after figures/tables/equations.

Use `\noindent`.

7.3 Tables

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial mention. To ensure this proper “floating” placement of tables, use the environment `table` to enclose the table's contents and the table caption. The contents of the table itself must go in the `tabular` environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules.

Language	Users millions	Comments
English	327	Lots of folks use English
Chinese	153	Chinese is catching up
Spanish	87	All of Latin America
Japanese	86	Supplements phone chats
German	58	They even dub movies
Portuguese	34	All of Brazil is on orkut

Table 1: *Top seven languages on the web*

Note (1) the sparing use of horizontal and vertical lines; (2) table captions are *under* the table.

Referring to figures/tables inline: “In Table 1 we show the popular languages in the weblog dataset.” First, we associate a label with the table:

```
\begin{table}[htb] % here/top/bottom
\centering
\begin{tabular}{rrl}
...
\end{tabular}
\caption{Top seven languages on the web}
\label{tab:weblog-data}
\end{table}
```

Then we use a reference to the table:

```
‘‘In Table~\ref{tab:web-users} we show
```

The ~ encodes a non-breakable space.

If the figure/table is wide you may consider placing it so that it spans the width of the two columns:

```
\begin{table*}[...] ... \end{table*}
```

As with a single-column table, this wide table will “float” to a location deemed more desirable.

7.4 Figures

Like tables, figures cannot be split across pages; the best placement for them is typically the top or the bottom of the page nearest their initial cite. To ensure this proper “floating” placement of figures, use the environment `figure` to enclose the figure and its caption. This sample document contains examples of `.eps` and `.ps` files to be displayable with L^AT_EX.



Fig. 1: *A sample black and white graphic (.eps format)*

You can associate a label with a figure and refer to the figure using that label as is done for tables.

As was the case with tables, you may want a figure that spans two columns. To do this, and still to ensure proper “floating” placement of the table/figure, instead of enclosing the figure and its caption in a `figure` environment use the environment `figure*`:

```
\begin{figure*}[...] ... \end{figure*}
```

Note that either `.eps` or `.ps` formats are used; use the `\epsfig` or `\psfig` commands as appropriate for the different file types.

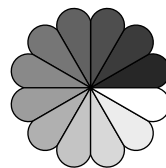


Fig. 2: *Black and white graphic (.ps format) that has been resized*

It is possible that a large figure can “float” onto the next page while the next figure can still fit on the current page and thus causing non-consecutive ordering of figures. One way to have a figure appear physically before subsequent figures is to place it earlier in the L^AT_EX source and let it float onto the desired page. These kind of fine tuning changes should be done when you are finished with the content of the paper.

8 Captions

Do not end the captions of tables/figures with periods. One line captions look nice and are easy to read. You may wish not have parentheses italicized in captions by toggling the italics around the parentheses: `\emph{()}` and `\emph{()}`. Compare the parentheses in the captions of Fig. 1 and Fig. 2.

9 Citations

Let us cite a few colleagues and see what happens: [1].

10 Colors

No doubt about it—colors are catchy! **RED**, **BLUE**, **GREEN**, etc. However, you should consider the following:

1. Ultimately the proceedings will be printed in black and white; this will make the colors come out as shades of grey and may not be easily distinguishable from one another.
2. Many reviewers print your submission and read a hardcopy and many printers are set by default to black and white printing; reviewers may not necessarily browse through the whole submission to decide they should print it in color or on a color printer.

11 Oh, my paper is too long

After the reviewing process a common request from reviewers is to ask you elaborate on aspects they didn’t quite understand. If you already are at the maximum

length of a submission you have a problem. Obviously some bits will have to be omitted. But you have spent so much time on these experiments! And this paragraph/table/picture looks so good! Everybody has the same problem. The organizers have provided a template which is very compact and already allows you to put a lot of information. Think about the main idea you want to convey. Maybe actually you are trying to say more than one thing. Shouldn't the rest of the important ideas be put in another paper? Maybe for the next RANLP conference? Then again, you can have a nice, well structured paper albeit a short one which will interest people and they will contact you for the longer report on which this paper is based. Maybe you can mention the URL for that longer report or cite it.

11.1 Cheating with the length

Here are some obvious ideas to condense your paper, not all of them make the paper look very nice but ...

- using `small` font for list environments and maybe some tables too; *like here—see the source*
- put this great but secondary idea in a footnote;
- put a table/figure/equation in one column only;
- redraw figures to take less space;
- control vertical spacing between elements `\vspace*{-.}` or `\[-1.5em]`; *like here—see the source*
- to ensure uniformity of the proceedings please don't modify the margins or the text height!

A section that has only a single subsection looks very suspicious. You may want to consider using the unnumbered second level heading produced by the command:

```
\subsection*{...}
```

12 Spellcheck the source!

If you are a non-native speaker why don't you ask a friend of yours who is a native speaker (who went to an exchange in the UK / Ireland / US / Canada / Australia) to have a quick look at your paper. The spellchecker can't tell you if this phrase sounds rather awkward in English. And it would be nice if the reader flies over your text and not stumble on words and sentences trying to find out what it is you were trying to say.

The 'basket of words', 'forecasting values', 'nerve-network', 'web-captured' are usually referred to as the 'bag of words', 'predicting values', 'neural networks', 'web-crawled', etc.

13 Page numbering

Do not use page numbering for the camera ready copies (because you wouldn't know the final pages in the proceedings where your paper will be placed). This also means you wouldn't be able to use `\pageref{}` commands. For your submission it is helpful to have page numbers.

Add the command `\pagenumbering{arabic}` above `\begin{document}` if you would like to have the page numbers appear.

14 Bibliography

Two commands at the end of your \LaTeX source will do the magic:

```
\bibliographystyle{abbrv}
\bibliography{ranlp-biblio}
```

The first command (`\bibliographystyle{abbrv}`) specifies how the bibliography will be formatted. The second (`\bibliography{ranlp-biblio}`) simply states that your bibliography database with declarative specification of the fields of each bibliographical entry is located in file `ranlp-biblio.bib`. By varying `\bibliographystyle{...}` you can have different formatting for the bibliographical entries (and the inline citations) without having to change anything else in your document. Normally you would have a single bibliographical database file which will be referred to all of your papers, reports, thesis, etc.

15 Available resources

Where can you get this great \LaTeX typesetting system?

- If you are using UNIX/LINUX chances are \LaTeX is already installed on your network. You can create `*.eps` figures using `XFig`.
- For PCs here are some (among many) options:
 - `cygwin` includes \LaTeX . You need to ensure you have selected it when you installed `cygwin`. You can always add to an existing installation (www.cygwin.com). `XFig` is also available on `cygwin`.
 - `MikTeX` is another good implementation for Windows (www.miktex.org);
- \LaTeX is also available for Apple machines (OSX):
 - `iTeXMac` (<http://itexmac.sourceforge.net>);
 - `TeXShop`—it won an Apple Design Award in 2002 (<http://darkwing.uoregon.edu/koch/texshop>);
 - `BibDesk`—a graphical bibliography manager (<http://bibdesk.sourceforge.net>).

Acknowledgments

This section is optional; it is a location for you to acknowledge grants, funding, editing assistance and what have you. In the present case, for example, the RANLP organizers would like to thank ACM, Gerald Murray in particular, for allowing us to adapt the ACM stylesheet for our conference. The heading is not numbered because we use:

```
\section*{Acknowledgments}
```

Alternatively if you are short on space you can use:

```
{\small
  \noindent
  \textbf{Acknowledgments.} We thank
  ...
}
```

16 Conclusions

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the L^AT_EX book, the citations in this paper are to articles which are used as examples only.

References

- [1] D. S. Jurafsky and J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, New Jersey, 2000.