

Sample Paper

Sample Student

This is a sample abstract for the sample paper just to show how you can add an abstract. It's not necessary, but you could if you wanted.

Table of contents

1	Introduction	2
2	Intermediate section	3
3	Discussion of results	5
4	Conclusion	5
	References	6

List of Figures

1	Scatterplot of document scores with the first two factors as dimensions and relevant registers annotated.	4
---	---	---

List of Tables

1	Numerical variables and their definitions.	2
2	Loadings of the four factors.	4

1 Introduction

In this sample paper a few tables, plots and pieces of text are reproduced from the register analysis case study. The goal is to illustrate some things to be expected from a final paper for this course.

Your paper should have at least an introduction and a conclusion; the specific organization of the sections in between is not important, as long as it makes sense given your content and argumentation. The introduction should at least contain:

- a presentation of the research question(s);
- a description of the corpus and/or dataset used in the study;
- and a summary of the structure of the paper.

The description of the corpus should cover its source, its size in number of tokens and types (if you had access to the full corpus) and any relevant internal composition (e.g. regional variation, authors, topic categories, etc.). If you didn't work on a corpus directly but downloaded a dataset, you should also refer to its source and describe what the rows and columns represent. A full description of the dataset may also be included in an appendix or the README if you feel it is not so relevant for the argumentation of the paper. An example would be:

The corpus used for this study is a TEI-compliant version of the Brown corpus: it has 1 161 028 tokens and 381 244 types distributed in 500 and 15. The files were processed to obtain a number of automatically extractable variables for factor analysis, resulting in a dataframe with 500 rows and 14 columns. The meaning of the variables is described in Table 1 — this study is admittedly very modest in comparison to Biber (1988).

Table 1 shows that tables can be written either in R chunks or manually (for example, to easily add markdown formatting and footnotes), but in any case they should have a caption and be cross-referenceable.

Table 1: Numerical variables and their definitions.

Name	Value
ttr	Type token ratio (number of types divided by number of tokens) ¹
word_len	Average word length, in characters
p_mw	Proportion of combined tags; typically clitics as in <i>she's</i>
p_c	Proportion of punctuation characters
p_ppss	Proportion of personal pronouns nominative, besides third person singular

¹The type token ratio of different texts can only be compared if they have similar lengths, which is the case with the Brown corpus.

Name	Value
p_noun	Proportion of nouns
p_bigr	Number of unique word bigrams, relative to document size
p_nomin	Proportion of nominalisations (nouns ending in <i>-tion</i> , <i>-ment</i> , <i>-ness</i> , or <i>-ity</i>)
p_pobi	Number of unique pos tag bigrams, relative to document size
p_adj	Proportion of adjectives
p_neg	Number of negations, relative to document size
p_adv	Proportion of adverbs
p_qual	Number of qualifiers, relative to document size

Somewhere in your article you should refer to the main packages used for the analysis: *at least* R itself (R Core Team 2022) and the package(s) used for the main statistical technique, it can include `{tidyverse}` and `{mclm}`, assuming you would be using them (Wickham et al. 2019; Speelman & Montes 2022; Wickham, Hester & Ooms 2021). This can be done either in the introduction or in the section where you describe the technique of choice.

2 Intermediate section

There should be a section dedicated to the methodology, in which you introduce the technique(s) used and its/their relationship to the research question(s).

Normally you wouldn't show code in your paper, so you can add `echo: false` to your top YAML. But for specific code that is key to the paper, it might be interesting to show it without evaluating, i.e. indicating `eval: false` in the YAML of the chunk:

```
dataset <- read_tsv(here("register-analysis.tsv"), show_col_types = FALSE)
fa <- dataset %>% data.frame(row.names = "filename") %>%
  as.matrix() %>% factanal(factors = 4, scores = "regression")
```

But, given that the source code is included along with the paper, it is not strictly necessary.

It is most likely relevant to add some literature review as well — similar papers that you have found, references regarding either the technique or the research question. This will require citations, either just pointing to relevant literature or with quotes.

Remember that long quotes (around > 40 words) should be included as block quotes.

Of course, you should use cross references to refer to useful figures and tables such as Figure 1 and Table 2. If your figures or tables are not useful enough to be referenced, then maybe they shouldn't be in the paper in the first place.

Table 2: Loadings of the four factors.

Variable	Factor1	Factor2	Factor3	Factor4
ttr			0.98	
word_len	0.88	-0.35		
p_bigr			0.92	
p_pobi	-0.35		0.42	0.62
p_mw		0.95		
p_c	-0.40	0.54		
p_nomin	0.79			
p_noun	0.61	-0.39		-0.47
p_ppss	-0.44	0.53		0.31
p_adj	0.68	-0.30		
p_neg		0.87		
p_adv	-0.38	0.33		0.58
p_qual				0.67

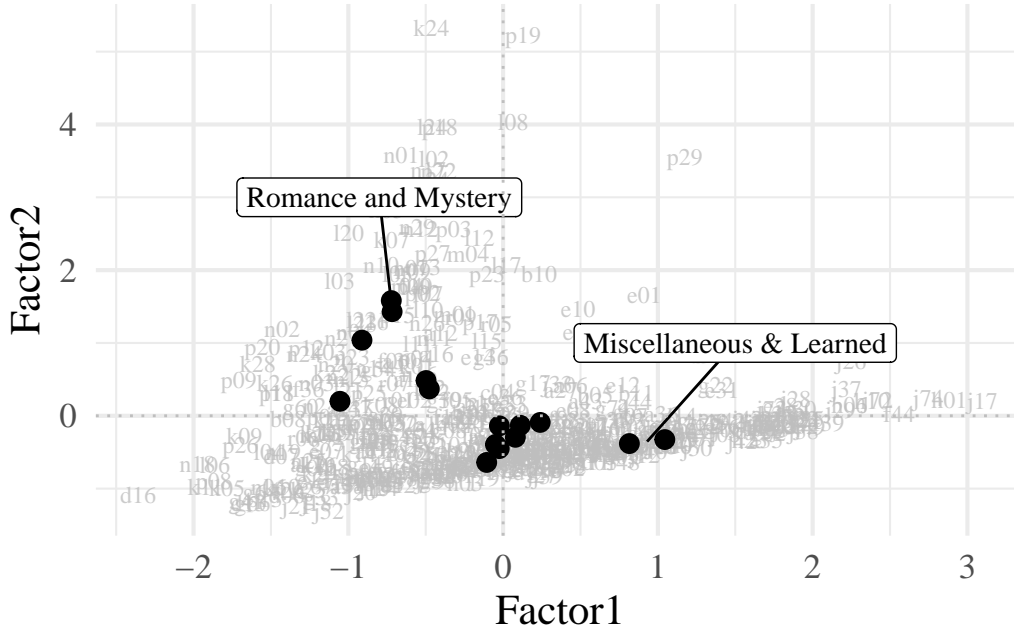


Figure 1: Scatterplot of document scores with the first two factors as dimensions and relevant registers annotated.

3 Discussion of results

It might also be useful to add linguistic examples, such as (1), to illustrate concrete examples and enrich your argumentation.

- (1) Nevertheless , the theory that the determining influence of the hypothalamic balance has a profound influence on the clinical behavior of neuropsychiatric patients has not yet been tested on an adequate number of patients .

If your examples are not in English, you should add a translation. You can use `{glossr}` to aid with pairing an original text and a translation, even if you don't need aligned lines, as shown in (1). Note that in PDF output the two types of examples cannot be combined because they don't continue the numbering.

- (1) *Example-ID*

Este ejemplo está en español.

“This example is in Spanish.”

4 Conclusion

The conclusion should recap the paper, summarizing what was said in each section and how everything ties together. It might feel redundant, in particular in relation to the introduction, but that's precisely the point: for someone who has not heard your ideas, redundancy is key to understand. Be clear about how each section contributes to your main point and what the take-home message is.

Of course, the language of the paper should be formal and academic (I appreciate puns and jokes, but don't lower the register too much). Coherence in the ideas, cohesion between the sentences and appropriate use of the technical vocabulary and of connectors (e.g. *however*, *in contrast*, *while...*) are important and **will be evaluated**. Not because of “language” evaluation but because these aspects are crucial for understanding, and if the reader needs to read a sentence many times and/or have previous knowledge of your process in order to understand the text, it's not well written. I also recommend checking out the `{spelling}` package to run some spelling checks on your files!

! Important

Enjoy your holidays and best of luck!

References

- Biber, Douglas. 1988. *Variation across Speech and Writing*. First. Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>.
- R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Speelman, Dirk & Mariana Montes. 2022. *Mclm: Mastering corpus linguistics methods*. <https://CRAN.R-project.org/package=mclm>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester & Jeroen Ooms. 2021. *xml2: Parse XML*. <https://CRAN.R-project.org/package=xml2>.