

# Phonetic fieldwork and experiments with the phonfieldwork package for R

George Moroz

Linguistic Convergence Laboratory, NRU HSE

08.08.2020, Grupo de estadística para el estudio del lenguaje

Presentation is available here: [tinyurl.com/y6lf5ch4](https://tinyurl.com/y6lf5ch4)



# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

## Most phonetic research consists of the following steps:

- 1 Formulate a research question. Think of what kind of data is necessary to answer this question, what is the appropriate amount of data, what kind of annotation you will do, what kind of statistical models and visualizations you will use, etc.
- 2 Create a list of stimuli.
- 3 Elicite list of stimuli from speakers who signed an Informed Consent statement, agreeing to participate in the experiment to be recorded on audio and/or video.
- 4 Annotate the collected data.
- 5 Extract the information from annotated data.
- 6 Create visualizations.
- 7 Evaluate your statistical models.
- 8 Report your results.
- 9 Publish your data.

## Most phonetic research consists of the following steps:

- 1 Formulate a research question. Think of what kind of data is necessary to answer this question, what is the appropriate amount of data, what kind of annotation you will do, what kind of statistical models and visualizations you will use, etc.
- 2 Create a list of stimuli.
- 3 Elicite list of stimuli from speakers who signed an Informed Consent statement, agreeing to participate in the experiment to be recorded on audio and/or video.
- 4 Annotate the collected data.
- 5 Extract the information from annotated data.
- 6 Create visualizations.
- 7 Evaluate your statistical models.
- 8 Report your results.
- 9 Publish your data.

The `phonfieldwork` package is created for helping with items 3, 4 (partially), 5, 6, and 9.

## Why/when do you need the phonfieldwork package?

These ideal plan hides some technical subtasks:

- creating a presentation for elicitation task
- renaming and concatenating multiple sound files
- automatic annotation in Praat TextGrids [[Boersma and Weenink 2019](#)]
- creating a searchable .html table with annotations, spectrograms and ability to hear sound
- converting multiple formats (Praat, ELAN [[Brugman et al. 2004](#)] and EXMARaLDA [[Schmidt and Wörner 2009](#)])

# Why/when do you need the phonfieldwork package?

These ideal plan hides some technical subtasks:

- creating a presentation for elicitation task
- renaming and concatenating multiple sound files
- automatic annotation in Praat TextGrids [[Boersma and Weenink 2019](#)]
- creating a searchable .html table with annotations, spectrograms and ability to hear sound
- converting multiple formats (Praat, ELAN [[Brugman et al. 2004](#)] and EXMARaLDA [[Schmidt and Wörner 2009](#)])

All of these tasks can be solved by a mixture of different tools:

- any programming language can handle automatic file renaming
- Praat contains scripts for concatenating and renaming files

# Why/when do you need the phonfieldwork package?

These ideal plan hides some technical subtasks:

- creating a presentation for elicitation task
- renaming and concatenating multiple sound files
- automatic annotation in Praat TextGrids [[Boersma and Weenink 2019](#)]
- creating a searchable .html table with annotations, spectrograms and ability to hear sound
- converting multiple formats (Praat, ELAN [[Brugman et al. 2004](#)] and EXMARaLDA [[Schmidt and Wörner 2009](#)])

All of these tasks can be solved by a mixture of different tools:

- any programming language can handle automatic file renaming
- Praat contains scripts for concatenating and renaming files

phonfieldwork provides a functionality that will make it easier to solve those tasks independently of any additional tools. You can also compare the functionality with other packages: rPraat [[Bořil and Skarnitzl 2016](#)], textgRid [[Reidy 2016](#)], pypmi [[Lubbers and Torreira 2013](#)]

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite



## Install phonfieldwork

phonfieldwork is an R package, so you need to install [R](#), [RStudio](#) (optional) or use [rstudio.cloud](#). There are two possibilities for installing package in R:

- official version from CRAN

```
install.packages("phonfieldwork")
```

- development version from GitHub

```
devtools::install_github("agricolamz/phonfieldwork")
```

## Install phonfieldwork

phonfieldwork is an R package, so you need to install [R](#), [RStudio](#) (optional) or use [rstudio.cloud](#). There are two possibilities for installing package in R:

- official version from CRAN

```
install.packages("phonfieldwork")
```

- development version from GitHub

```
devtools::install_github("agricolamz/phonfieldwork")
```

Since this package is under [rOpenSci review](#) there is a chance that in couple months the address could be changed to [ropensci/phonfieldwork](#), and documentation page will be moved from [agricolamz.github.io/phonfieldwork](#) to [docs.ropensci.org/phonfieldwork](#).

```
library("phonfieldwork")  
packageVersion("phonfieldwork") # Unreleased version
```

```
## [1] '0.0.8'
```

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

**Data renaming**

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

**Data merging**

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

**Data annotation**

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

**Data extraction**

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

**Data visualization**

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite



# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Outline of the talk

Introduction

Installation of the package

Creating your presentation

Data renaming

Data merging

Data annotation

Data extraction

Data visualization

Reading data from different linguistic sources

Creating a data viewer – template for the data sharing

Cite

# Cite the package

```
citation("phonfieldwork")
```

```
##  
## Moroz G (2020). _Phonetic fieldwork and experiments with phonfieldwork  
## package_. <URL: https://CRAN.R-project.org/package=phonfieldwork>.  
##  
## A BibTeX entry for LaTeX users is  
##  
## @Manual{,  
##   title = {Phonetic fieldwork and experiments with phonfieldwork package},  
##   author = {George Moroz},  
##   year = {2020},  
##   url = {https://CRAN.R-project.org/package=phonfieldwork},  
## }
```

- Boersma, P. and Weenink, D. (2019). Praat: doing phonetics by computer (version 5.3.51)[computer program] version 6.0.25, retrieved 15 november 2019 from <http://www.praat.org/>.
- Bořil, T. and Skarnitzl, R. (2016). Tools rpraat and mpraat. In Sojka, P., Horák, A., Kopeček, I., and Pala, K., editors, *Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings*, pages 367–374, Cham. Springer International Publishing.
- Brugman, H., Russel, A., and Nijmegen, X. (2004). Annotating Multi-media/Multi-modal Resources with ELAN. In *LREC*.
- Lubbers, M. and Torreira, F. (2013). Pympl-ling: a Python module for processing ELANs EAF and Praats TextGrid annotation files.

- Reidy, P. (2016). *textgRid: Praat TextGrid Objects in R*. R package version 1.0.1.
- Schmidt, T. and Wörner, K. (2009). Exmaralda—creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 19(4):565–582.