

Implementing reproducibility in phonetic research: a computational workflow

Stefano Coretta

26/03/2017

Reproducible research

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

[figure reproducible]

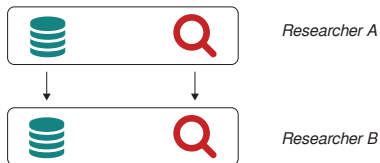
Reproducible research

Implementing
reproducibility
in phonetic
research: a
computational
workflow

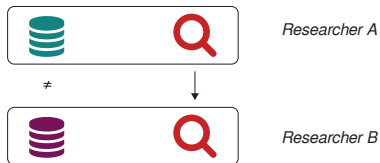
Stefano
Coretta

References

REPRODUCIBILITY



REPLICABILITY



Why should we care?

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

The problem (Sandve et al. 2013):

- difficulty of reproduction
- difficulty of replication
- retracted papers

The solution:

- Reproducible Research (RR)

A case from linguistics

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

The “Yokuts vowels” case (Weigel 2002):

- about **75%** of the data is contrived (Weigel 2005:149)
- some of the generalisations are **wrong** (Blevins 2004)

Reproducible Research in linguistics

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

- linked data (Bird & Simons 2003, Thieberger 2004)
- computational grammar (Maxwell & Amith 2005)
- glossbox (McDonnell & Hall 2017)
- RR in the Speech Sciences (Abari 2012)

RR in phonetics: the workflow

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References



Phase A: source code and documentation

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

Praat scripting:

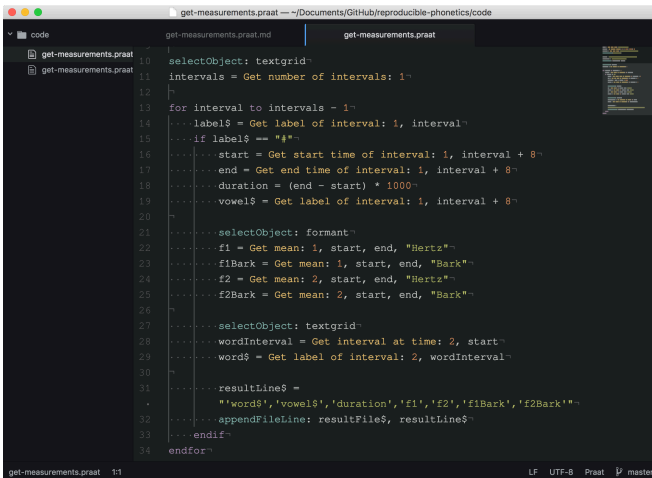
- Atom editor (for syntax highlighting and snippets)
- Literate Markdown
 - tangle: `lmt`
 - weaving: `pandoc`

Atom

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References



The screenshot shows the Atom text editor with a file named 'get-measurements.praat' open. The file is located at '~/Documents/GitHub/reproducible-phonetics/code'. The editor displays a Praat script with the following content:

```
10 selectObject: textgrid~
11 intervals = Get number of intervals: 1~
12 ~
13 for interval to intervals - 1~
14   ....label$ = Get label of interval: 1, interval~
15   ....if label$ == "#~
16     ....start = Get start time of interval: 1, interval + 8~
17     ....end = Get end time of interval: 1, interval + 8~
18     ....duration = (end - start) * 1000~
19     ....vowel$ = Get label of interval: 1, interval + 8~
20     ~
21     ....selectObject: formant~
22     ....f1 = Get mean: 1, start, end, "Hertz"~
23     ....f1Bark = Get mean: 1, start, end, "Bark"~
24     ....f2 = Get mean: 2, start, end, "Hertz"~
25     ....f2Bark = Get mean: 2, start, end, "Bark"~
26     ~
27     ....selectObject: textgrid~
28     ....wordInterval = Get interval at time: 2, start~
29     ....word$ = Get label of interval: 2, wordInterval~
30     ~
31     ....resultLine$ =
32       . "word$', 'vowel$', 'duration', 'f1', 'f2', 'f1Bark', 'f2Bark'"~
33     ....appendFileLine: resultFile$, resultLine$~
34   ....endif~
35 endfor~
```

The status bar at the bottom of the editor shows 'get-measurements.praat 1:1' on the left and 'LF UTF-8 Praat' on the right.

lmt

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

[figure]

pandoc

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

[figure]

Phase B: the speakr package

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

speakr is an R package to aid Praat users:

- aim: tangle and run Praat scripts from within R
- two main functions
 - `lmt()`: tangle a Praat script
 - `praatRun()`: run a Praat script

Phase B: the speakr package

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

```
# Tangle a Praat script
```

```
lmt("code/get-measurements.praat.md")
```

```
# Run the script
```

```
praatRun("code/get-measurements.praat")
```

```
# Read the results of the script
```

```
vowels <- read_csv("results/vowels.csv") %>%
```

```
  mutate_if(is.character, as.factor) %>%
```

```
  mutate(vowel = factor(vowel, c("i", "e", "a",  
                                "o", "u")))
```

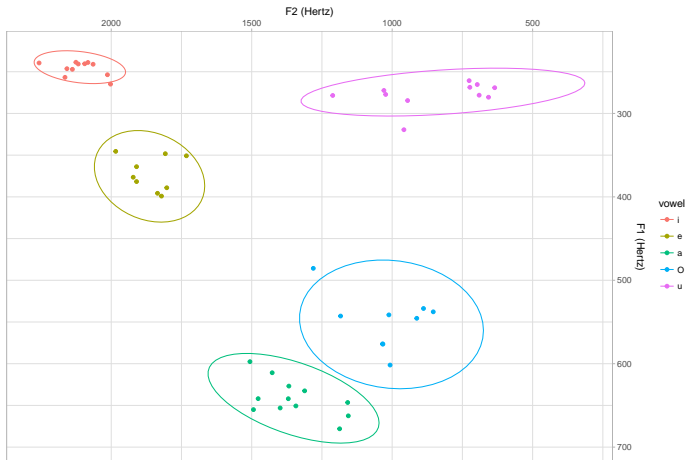
Phase B: the speakr package

Implementing
reproducibility
in phonetic
research: a
computational
workflow

Stefano
Coretta

References

Vowel plot of one speaker of Italian



- Abari, Kálmán. 2012. Reproducible research in speech sciences. *International Journal of Computer Science Issues* 9(6). 43–52.
- Bird, Steven & Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 557–582.
- Blevins, Juliette. 2004. A reconsideration of Yokuts vowels. *International Journal of American Linguistics* 70(1). 33–51.
- Maxwell, Michael & Jonathan D. Amith. 2005. Language documentation: the Nahuatl grammar. In A. Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing*, 474–485. Berlin Heidelberg: Springer-Verlag.
- McDonnell, Bradley & Patrick Hall. 2017. Developing methods for reproducible research in linguistics: a first step.

- Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor & Eivind Hovig. 2013. Ten simple rules for reproducible computational research. *PLoS Computational Biology* 9(10). 1–4.
- Thieberger, Nicholas. 2004. Documentation in practice: Developing a linked media corpus of South Efate. In Peter K. Austin (ed.), *Language documenta and description*, vol. 2, Hans Rausing Endangered Languages Project, School of Oriental and African Studies, University of London.
- Weigel, William. 2005. *Yowlumne in the Twentieth century*: University of California, Berkley dissertation.
- Weigel, William F. 2002. The Yokuts canon: A case study in the interaction of theory and description. Paper presented at the annual meeting of the Linguistics Society of America, January 2002, San Francisco.