Implementing reproducibility in phonetic research: a computational workflow

Stefano Coretta

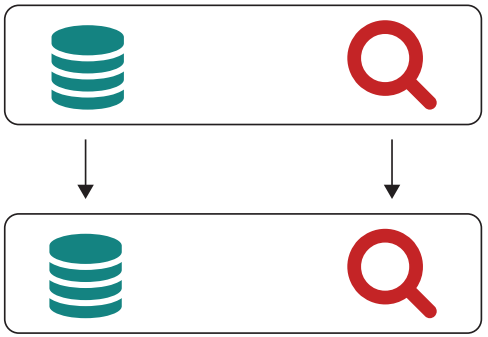
26/03/2017

Reproducible research

[figure reproducible]

Reproducible research

REPRODUCIBILITY



Researcher A

Researcher B

REPLICABILITY

Researcher A

Researcher B

Why should we care?

The problem (Sandve et al. 2013):

- difficulty of reproduction
- difficulty of replication
- retracted papers

The solution:

- Reproducible Research (RR)

A case from linguistics

The "Yokuts vowels" case (Weigel 2002):

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

Reproducible Research in linguistics

- 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

Phase A: source code and documentation

Praat scripting:

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

Phase B: the `speakr` package

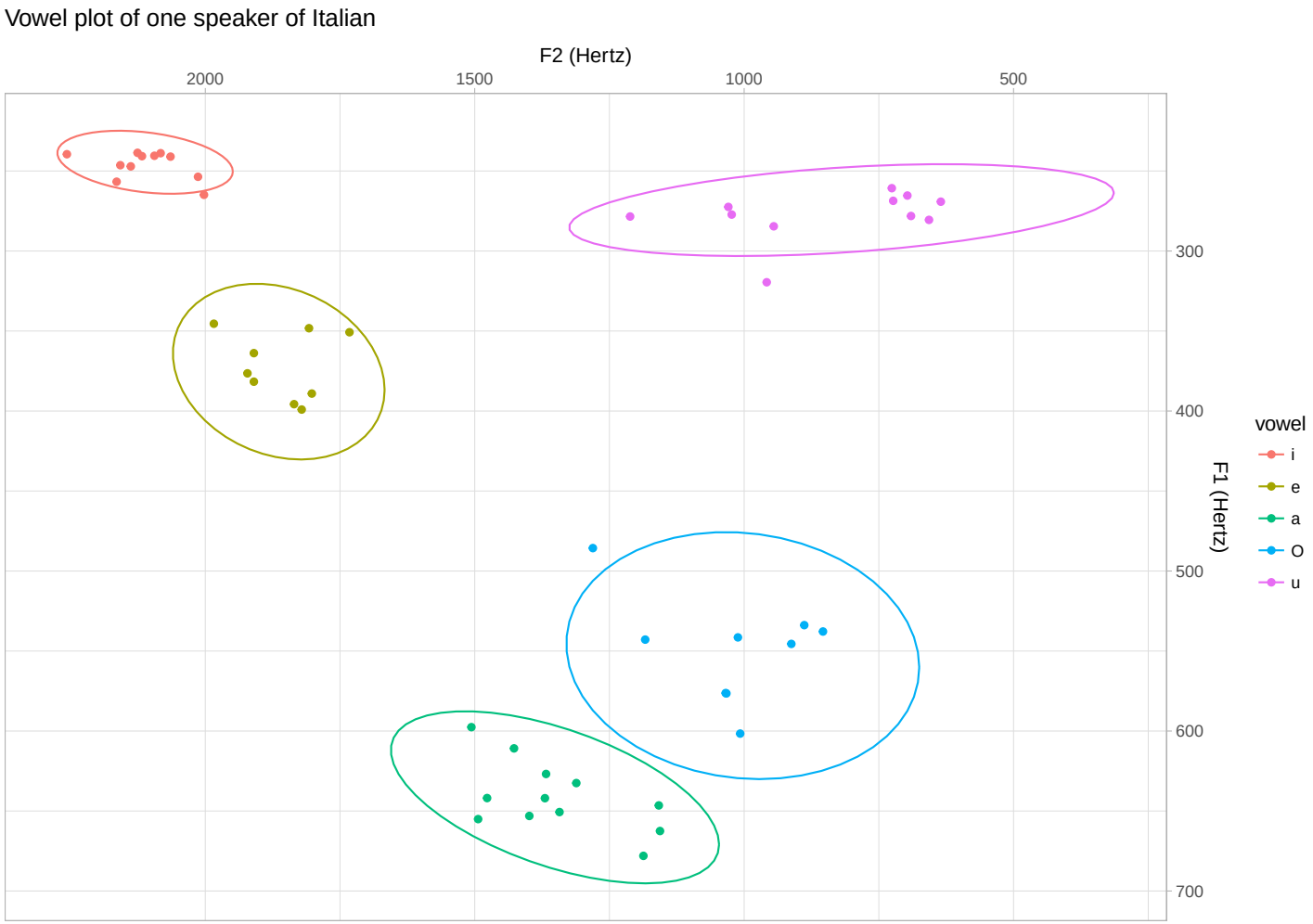
An `R` package to aid phoneticians who use Praat:

- tangle and run Praat scripts from within `R`

Phase B: the speakr package


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



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.