

Durational measurements on the Dutch prefix *ge-*

Using old data from Harald Baayen’s `languageR` package

Felix Golcher* and Alter Ego

2022-05-04 (this version)

It is said that your life flashes before your eyes just before you die.
That is true, it’s called Life.”

– Terry Pratchett, *The Last Continent*

Abstract: This paper is about Dutch prefixes. It uses an example data set from Pluymaekers et al. (2005), cited by Baayen (2008, pp. 126, 338).

1 Introduction

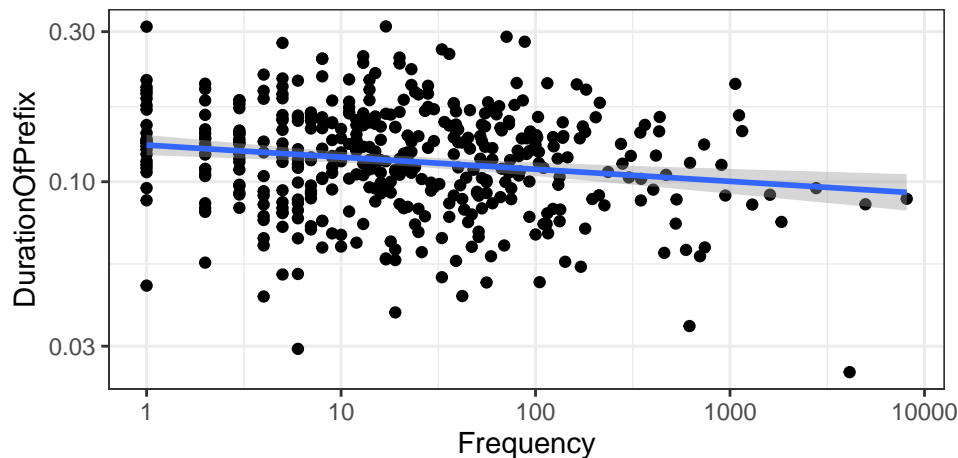


Figure 1: A very interesting graphical description of the data.

We tried to figure out if the frequency of a word has an influence on the duration of Dutch prefixes. Intuitively one could assume that with the duration drops with frequency. For this we recruited **132** Dutch native speakers. They produced **428** different words. Figure 1 gives an overview of the data. We clearly see a downward trajectory.

*Corresponding author, felix.golcher@hu-berlin.de

Table 1: Scientists of the 20th century.

Scientist	Nobel prices	Birthyear	Death
Emmy Noether	0	1882	1935
Rosalind Franklin	0	1920	1958
Vera Rubin	0	1928	2016

Table 1 lists three women who helped to shape the world as it is.

Table 2: This is how our data look like.

	Word	Frequency	Speaker	Sex	YearOfBirth
1	geprikt	13	N01159	male	1,944
2	gepresteerd	25	N01077	male	1,980
3	gevolgd	309	N01032	female	1,939
4	geprikkeld	16	N01128	female	1,979
5	gestaakt	40	N01204	female	1,963
6	geselecteerd	42	N01151	female	1,956

Table 2 gives the first 6 rows and first 5 columns of our data. It clearly shows the problems with automatically created tables: `YearOfBirth` needs reformatting, and `Word` and `Frequency` should not be centered, I think. Of course, `stargazer` can take care of that, provided you figure out all the relevant options.

2 Methods and Materials

We applied `methods` to `materials`.

It can quite generally be said that

$$a + b = b + a$$

where a and b are some real numbers and $a + b$ is the sum of a and b .

There are

1. things that are not dashes
2. things that are dashes
 - short dashes: -
 - a. they are basically not dashes at all, but minus signs.
 - b. This fact is rarely relevant.
 - longer dashes: –
 - really long dashes: —

2.1 Methods

They were great

Here are the [materials](#).

Back to [methods and materials](#)

2.2 Materials

Subjects were told what to do.

Back to [methods and materials](#)

2.3 Results

We computed two linear models with the `lm` function of R and the model formulas

- $\text{DurationOfPrefix} \sim \log(\text{Frequency})$ and
- $\text{DurationOfPrefix} \sim \log(\text{Frequency}) + \text{NumberSegmentsOnset}$.

The table compares both models.

Table 3: our models

	<i>Dependent variable:</i>	
	DurationOfPrefix	
	(1)	(2)
$\log(\text{Frequency})$	-0.004^{***} ($-0.007, -0.002$)	-0.004^{***} ($-0.007, -0.002$)
$\text{NumberSegmentsOnset}$		-0.005 ($-0.013, 0.002$)
Adjusted R^2	0.022	0.024
<i>Note:</i> * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$		

Template to be reproduced ends here

What shall we do?

Update your Rmd file so the output matches this document.

- Add what you find in the “Results” section.
- The function used for the models is `lm()` on the data set `durationsGe`.
- model formulas can be stringified with `Reduce(paste, deparse(formula(yourmodel)))`.
- You might want to look at the argument `omit` for the `stargazer` function.

If you don't remember the syntax, the cheatsheet or google will surely help. If not, I'm there to help.

References

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Frequency and acoustic length: The case of derivational affixes in dutch. *Journal of the Acoustical Society of America*, 118, 2561--2569.