Multifactorial Exploratory Approaches: exploratory factor analysis

Guillaume Desagulier

Multifactorial Exploratory Approaches

exploratory factor analysis

Guillaume Desagulier[1]

[1]MoDyCo (UMR 7114)
Paris 8, CNRS, Paris Nanterre
Institut Universitaire de France
gdesagulier@univ-paris8.fr

Corpus Linguistics Summer School 2019
June 25th, 2019
University of Birmingham

outline

EFA

EFA has been made popular in linguistics by Biber's studies on register variation (Biber 1991, 1995).

EFA

- EFA is designed to identify patterns of joint variation in a number of observed variables.
- although close to PCA, EFA differs with respect to the following: the number of relevant components, which are called factors, is not determined automatically. It must be chosen before we run the analysis.

EFA

- EFA looks for variables that are highly correlated with a group of other variables. These intercorrelated variables are assumed to measure one underlying variable. This variable, which is not directly observed, but inferred, is latent. It is known as a factor.
- one added value of EFA is that "an error term is added to the model in order to do justice to the possibility that there is noise in the data" (Baayen 2008, p. 127).

inclusion in French

- the same data set serves as input for EFA: `inclusion_FrWaC.txt`
- In base R, we run EFA with `factanal()`

inclusion in French

- based on PCA, we are tempted to specify 3 factors
- unfortunately, this is not going to work because 3 factors are too many for 5 variables in the kind of EFA that `factanal()` performs.

why?

A χ^2 test reports whether the specified number of factors is sufficient. If the *p*-value is smaller than 0.05, more factors are needed. If it is greater than 0.05, no more factors are needed. The test reports that the χ^2 statistic is 12,667.73 on 1 degree of freedom and that the *p*-value is 0. Although a third factor is required, we have no choice but stick to 2 factors. This means that we should be careful when we interpret the results.

inclusion in French

```
> # clear R's memory
> rm(list=ls(all=TRUE))
> # load the data (inclusion_FrWaC.txt)
> data <- read.table(file=file.choose(), header=TRUE, row.names=1, sep="\t")
```

inclusion in French

```
> fa.object <- factanal(data, factors=2)
> fa.object

Call:
factanal(x = data, factors = 2)

Uniquenesses:
centre  coeur milieu  parmi   sein
 0.655  0.436  0.849  0.005  0.005

Loadings:
       Factor1 Factor2
centre  0.587
coeur   0.750
milieu  0.389
parmi  -0.147   0.987
sein   -0.740  -0.669

               Factor1 Factor2
SS loadings      1.626   1.424
Proportion Var   0.325   0.285
Cumulative Var   0.325   0.610

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 12667.73 on 1 degree of freedom.
The p-value is 0
```

inclusion in French

The output displays:

- uniqueness (unexplained variation)
- factor loadings (the loadings that are too close to zero are not displayed)
- the proportions of variance explained by the factors
- the χ^2 test

inclusion in French

factor loadings:

- the higher the loading the more relevant the variable is in explaining the dimensionality of the factor
- *Au milieu de*, *au centre de*, and *au cœur de* define the first factor
- *Parmi* defines the second factor.
- it seems that *au sein de* defines both.

inclusion in French

The proportions of variance explained by the factors

- = eigenvalues
- a factor is considered worth keeping if the corresponding SS loading (i.e. the sum of squared loadings) is greater than 1
- 2 factors are retained because both have eigenvalues over 1. Factor 1 accounts for 32.5% of the variance. Factor 2 account for 28.5% of the variance. Both factors account for 66.9% of the variance.

inclusion in French

Graphic output:

- rotation a procedure meant to clarify the relationship between variables and factors. As its name indicates, it rotates the factors to align them better with the variables.
- varimax rotation: the factor axes are rotated in such a way that they are still perpendicular to each other
- promax rotation: the factor axes are rotated in an oblique way.
- with promax, the resulting model provides a closer fit to the data than with varimax.

inclusion in French

Plotting the loadings of the prepositions on the two factors with varimax rotation:

```
> loadings <- loadings(fa.object)
> plot(loadings, type="n", xlim=c(-1,1))
> text(loadings, rownames(loadings))
```

For promax rotation, set `rotation` to `promax`:

```
> fa.object2 <- factanal(data, factors=2, rotation="promax")
> loadings2 <- loadings(fa.object2)
> plot(loadings2, type="n", xlim=c(-1,1))
> text(loadings2, rownames(loadings2))
```

inclusion in French

Plotting the loadings of the prepositions on the two factors with varimax rotation:

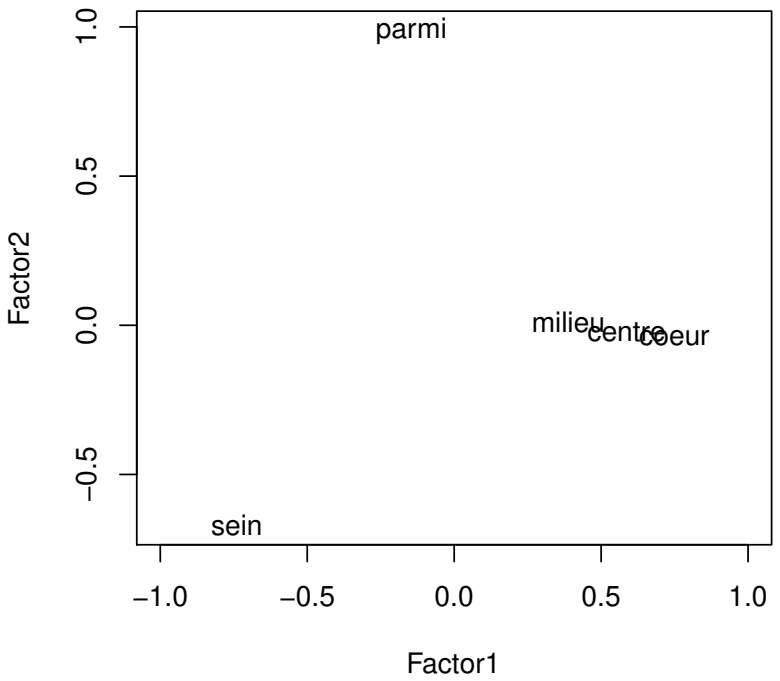

Figure 1: loadings with varimax rotation

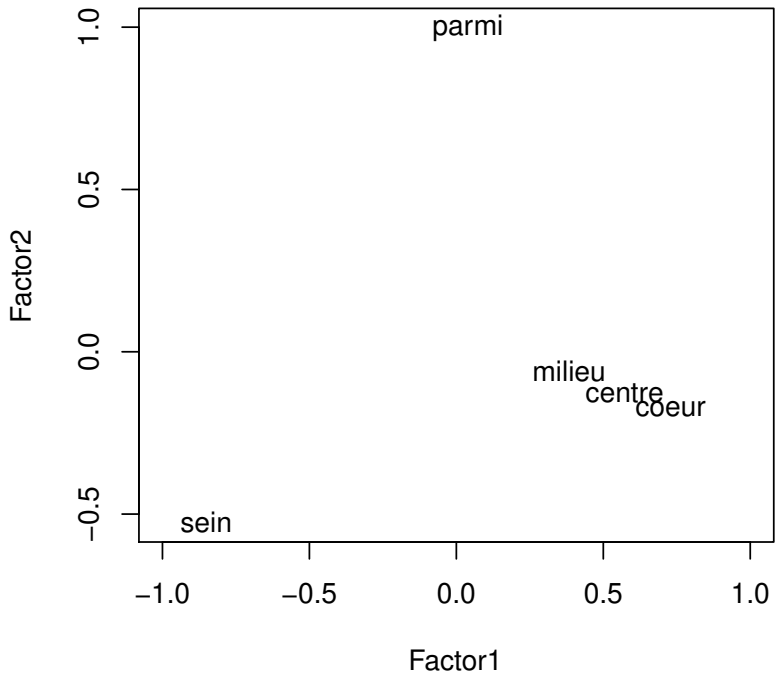

Figure 2: with promax rotation

inclusion in French

summary

The distinctive profiles we obtain with EFA are similar to those we obtained with PCA. The only major difference is the proximity of *au milieu de* with *au centre de* and *au cœur de*. This may be due to the fact that only two factors are retained in the analysis. As far as this data set is concerned, PCA is clearly a better alternative, all the more so as individuals are not taken into account in the graphic output of this kind of EFA.

Practical Handbook of Corpus Linguistics

Guillaume Desagulier (to appear). "Multifactorial exploratory approaches." In: *Practical Handbook of Corpus Linguistics*. Ed. by Magali Paquot and Stefan Thomas Gries. New York: Springer

Bibliography I

- Baayen, R Harald (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Biber, Douglas (1991). *Variation across speech and writing*. Cambridge University Press.
- – (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.
- Desagulier, Guillaume (to appear). "Multifactorial exploratory approaches." In: *Practical Handbook of Corpus Linguistics*. Ed. by Magali Paquot and Stefan Thomas Gries. New York: Springer.