Correspondence Analysis

(for categorical data)

Contingency Table (Simple CA)

Features within one category:

	Feature1	Feature2	Feature3	Total in rows
Group1	n_{11}	n_{12}	n_{13}	Σ_1
Group2	n_{21}	n_{22}	n_{23}	Σ_2
Group3	n_{31}	n_{32}	n_{33}	Σ_3
Total in columns	$\Sigma_{.1}$	$\Sigma_{.2}$	$\Sigma_{.3}$	Grand Total (n of observations)

Marginal sums in rows

Marginal sums in columns

(Normalized) corresponcence matrix: example

nxy in each cell is devided by *Grand Total*

	Register				
Colour	spoken	fiction	academic	press	**Total in rows**
black	0,0353	0,0715	0,0467	0,127	**0,2805**
blue	0,0082	0,0384	0,0063	0,0369	**0,0898**
brown	0,0021	0,019	0,0021	0,0201	**0,0433**
gray	0,002	0,0211	0,0022	0,0114	**0,0367**
green	0,0067	0,025	0,0078	0,0466	**0,0861**
orange	0,0016	0,0061	0,0008	0,01	**0,0185**
pink	0,0017	0,0127	0,001	0,011	**0,0264**
purple	0,0011	0,0058	0,0007	0,0059	**0,0135**
red	0,0126	0,0436	0,0098	0,0601	**0,1261**
w	[illegible]	0,0708	0,0458	0,0954	**0,2372**
ye	[illegible]	0,0183	0,0032	0,018	**0,0418**
Total in columns	**0,0988**	**0,3323**	**0,1264**	**0,4424**	**1**

Column mass

Row mass

Row profiles:

n_{xy} in each cell is devided by *row total*

	Register				
Colour	spoken	fiction	academic	press	**Total in rows**
black	0,126	0,2547	0,1666	0,4527	1
blue	0,0909	0,4282	0,0699	0,411	1
brown	0,0477	0,4394	0,0484	0,4646	1
gray	0,0552	0,5738	0,0609	0,31	1
green	0,0779	0,2904	0,0903	0,5414	1
orange	0,0873	0,3277	0,0444	0,5405	1
pink	0,0632	0,4806	0,0384	0,4178	1
purple	0,0785	0,4309	0,0549	0,4357	1
red	0,0996	0,3461	0,0775	0,4768	1
white	0,1061	0,2986	0,193	0,4023	1
yellow	0,0559	0,4372	0,0768	0,4301	1
Average row profile	**0,0987**	**0,3324**	**0,1265**	**0,4425**	1

Row profile table and average row profile

Features within one category:

	Feature1	Feature2	Feature3	Total in rows
Group1	0.11	0.63	0.26	
Group2	0.34	0.28	0.38	
Group3	0.13	0.64	0.23	
Average row profile	$\Sigma_{.1}$ / N	$\Sigma_{.2}$ / N	$\Sigma_{.3}$ / N	N

Rows 1 and 3 are more similar than row 2.

Usually, the distance between rows is calculated using chi-squared distance.

Column profiles are calculated the same way...

X^2 distance

Based on Euclidean distance, but weighted by the mass.

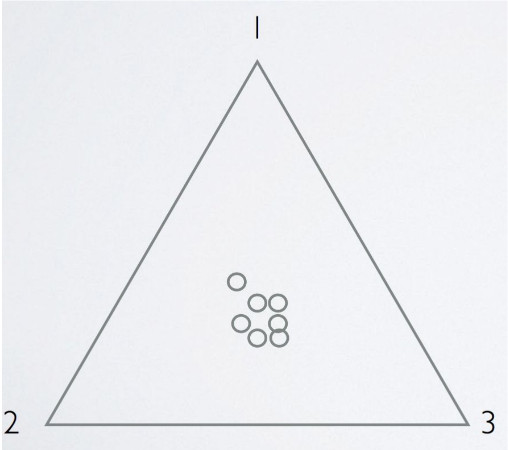
Row mass is an average column profile, **Column mass** is an average row profile

	1	2	3
Row r_1	p_{11}	p_{12}	p_{13}
Row r_2	p_{21}	p_{22}	p_{23}
Column mass	m_1	m_2	m_3

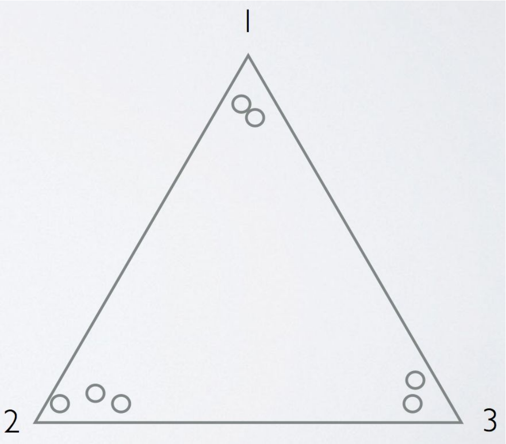
$$d(r_1, r_2) = \sqrt{\sum_j \frac{(p_{r_1 j} - p_{r_2 j})^2}{m_j}}$$

Inertia ≈ variance for rows (columns) of the table

Inertia is small: all points in the center



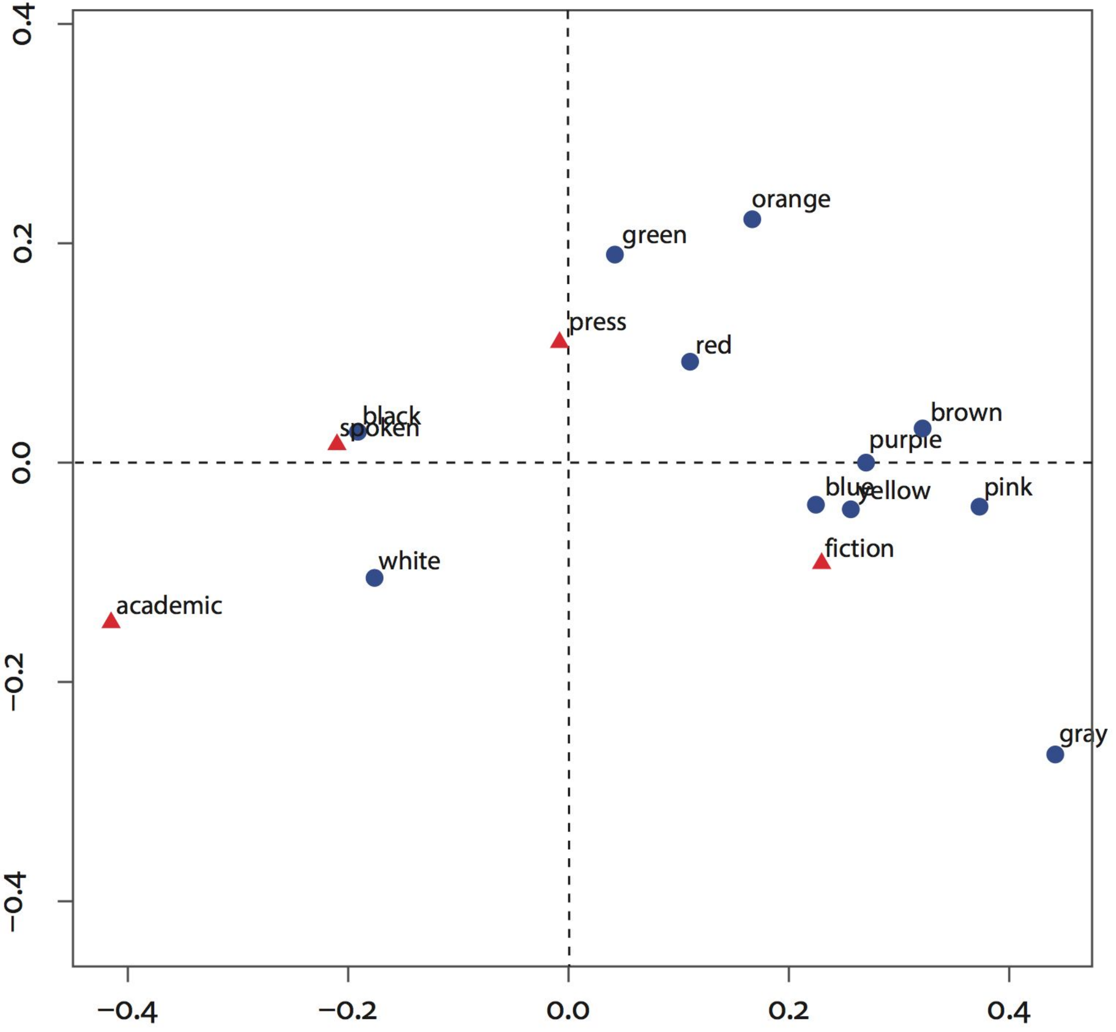
Inertia is large: all points on the poles



Here, the **points** are row profiles and the **poles (vertices)** are column profiles.

Fictive example: vertex1 (1, 0, 0) .. then row (0.98, 0.01, 0.01) is close to vertex1
vertex2 (0, 1, 0) row (0.33, 0.33, 0.34) is in the center
vertex3 (0, 0, 1)

The closer the row and column profiles, the more contingent (bound) they are



Correspondence map: color names and speech registers

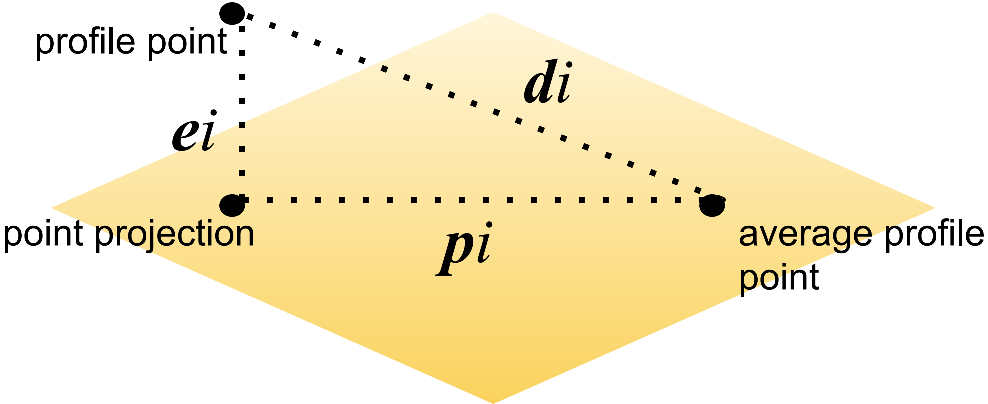
- black-white
- non-primary colors
- politics and recipes/diet

Where is an average row and color profiles?

Correspondence map and moment of inertia

Individual **inertia product** for each profile ***i:***

$$\underbrace{m_i}_{\textit{profile mass}}\ \underbrace{d_i^2}_{\textit{distance to average profile}}$$



Inertia product for the whole table:

$$\sum m_i d_i^2$$

Proportion of information explained by the map (**explained variance**)

residual inertia

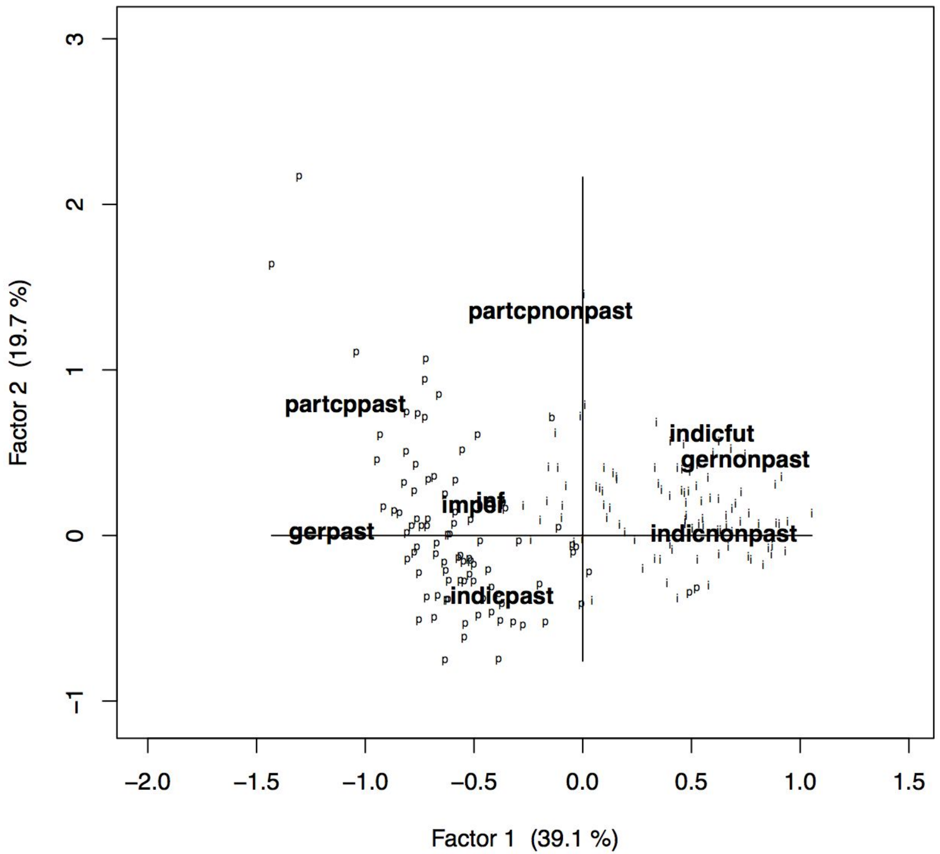
$$\sum_i m_i d_i^2 = \overbrace{\sum_i m_i p_i^2}^{\text{explained variance}} + \overbrace{\sum_i m_i e_i^2}^{\text{residual inertia}}$$

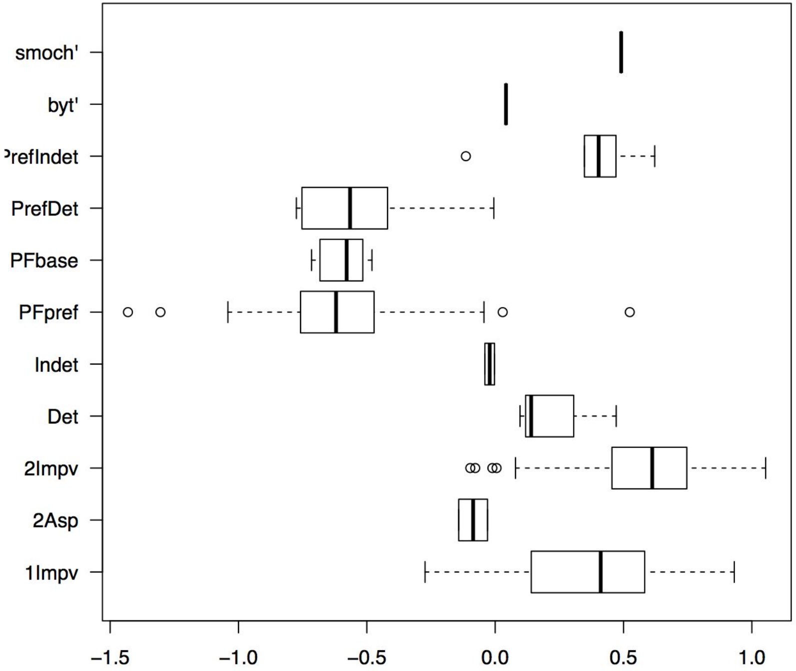
```
> summary(ca.bc)

Principal inertias (eigenvalues):

dim     value      %     cum%   scree plot
1       0.043730   77.9  77.9   ************************
2       0.010787   19.2  97.1   *****
3       0.001650   2.9   100.0
        ----       ---
Total:  0.056167   100.0
```

- Row-to-row distances on the CA map represent the approximate χ^2-distances between the row profiles.
- Column-to-column distances on the CA map represent the approximate χ^2-distances between the column profiles.
- There is no direct interpretation of row-to-column or column-to-row distances. Interpret the dimensions first, and then examine how the profiles are located with regard to the dimensions of variation (Greenacre 2007: 72).

(Levshina 2015)

CA example: grammatical profiles of Russian verbs in Journalistic texts, p(erfective) and i(mperfective) labels overlaid





Factor1 values distribution in some well-known verb classes (post-hoc analysis)

Multiple correspondence analysis

- long format table, binary matrix

	Categorical variables						
Observations	v_1	v_2	v_3	v_4	v_5	v_6	v_7
1	1	1	0	0	1	0	0
2	0	1	0	1	0	0	1
3	0	1	0	1	0	0	1
4	1	0	1	1	0	0	1
5	0	1	0	1	0	0	1
6	0	1	1	0	0	1	1
7	0	0	0	1	0	0	0
8	0	1	0	1	0	0	1
...	...	...	...	...	...	...	...

Multiple correspondence analysis

- Burt matrix

	Male	Female	Slight	Medium	High	<20	20-50	>50
Male	87	0	33	45	9	26	47	14
Female	0	163	27	111	25	43	89	31
Slight	33	27	60	0	0	14	48	7
Medium	45	111	0	111	0	14	107	18
High	9	25	0	0	79	9	30	3
<20	26	43	14	14	9	37	0	0
20-50	47	89	48	107	30	0	185	0
>50	14	31	7	18	3	0	0	28

27 is a number of observations which have both *Female* and *Slightly active*

Multiple correspondence analysis

MCA is CA applied to Burt matrix (or some of its variants)

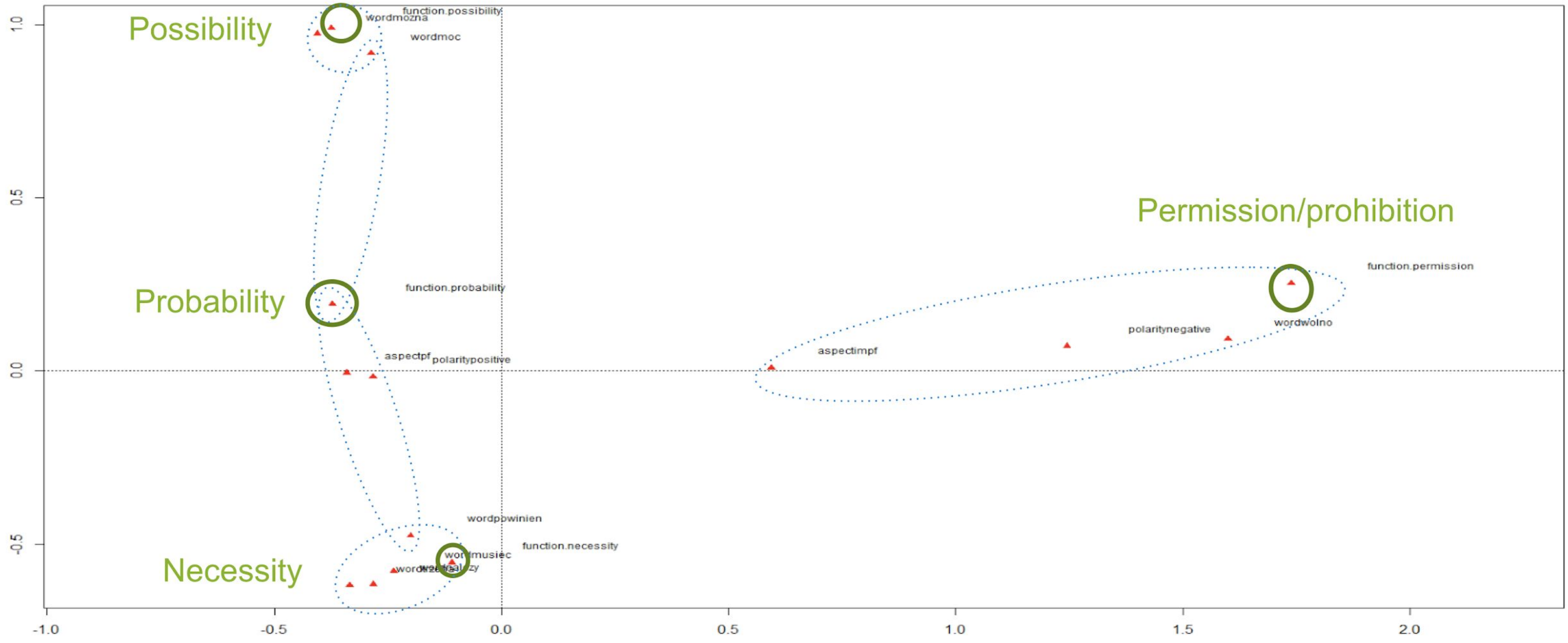
Burt matrix: D x D^T (transposed)

$$
\begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}^T = \begin{pmatrix} 2 & 1 & 1 & 1 & 0 & 2 \\ 1 & 2 & 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 3 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 & 2 \\ 2 & 2 & 1 & 2 & 2 & 4 \end{pmatrix}
$$

n observations x *k* variables

Burt matrix *k* x *k*

Example: Polish modal words (Divjak et al. 2015)



FUNCTION, WORD, ASPECT, POLARITY

Interpreting axes

“Interpreting an axis amounts to finding out what is similar, on the one hand, between all the elements figuring **on the right of the origin** and, on the other hand between al l that is written **on the left**; and expressing with conciseness and precision, the contrast (or opposition) between the two extremes.”

Benzecri (1992, p. 405)

We take into account the **contributions** of points and **deviations**.

Baseline criterion is an **average contribution** = total contrib/(n of variables).
The interpretation of an axis is based on the categories which contributions to axis exceed the criterion.

Interpretation of Dim1

FUNCTION	left	right	deviation
possibility	0.4		2.7 (sum)
probability	0.4		
necessity	0.15		
permission/ prohibition		1.75	

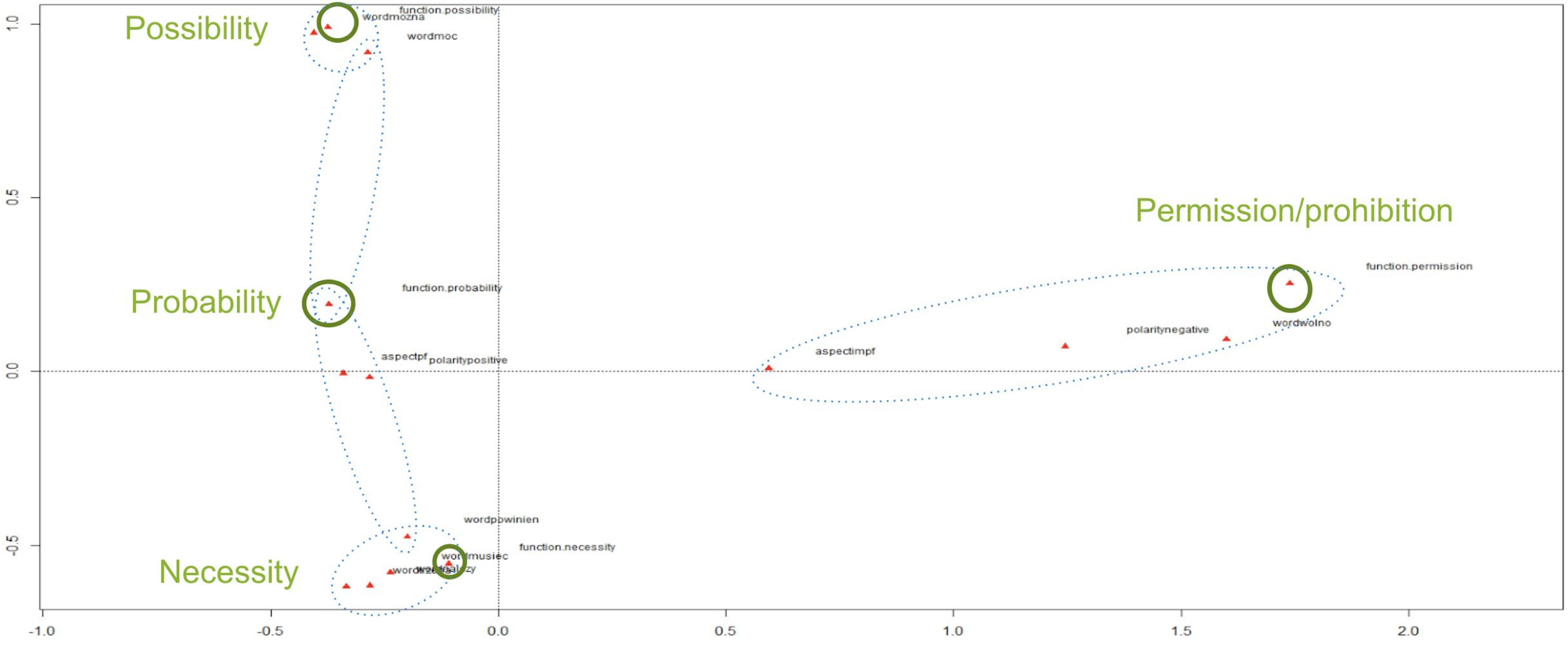
ASPECT	left	right	deviation
pf	0.3		0.9
impf		0.6	

Total contribution = 2.7+0.9+3.6+1.5 = 8.7
Average contribution = 8.7/4 = 2.2

WORD	left	right	deviation
można ‘it is possible’	0.4		3.6
móc ‘can’	0.3		
musieć ‘must’	0.25		
powinien ‘should’	0.3		
należy ‘it is necessary’	0.35		
trzeba ‘it is required’	0.4		
wolno ‘it is allowed’		1.6	

POLARITY	left	right	deviation
positive	0.25		1.5
impf		1.25	

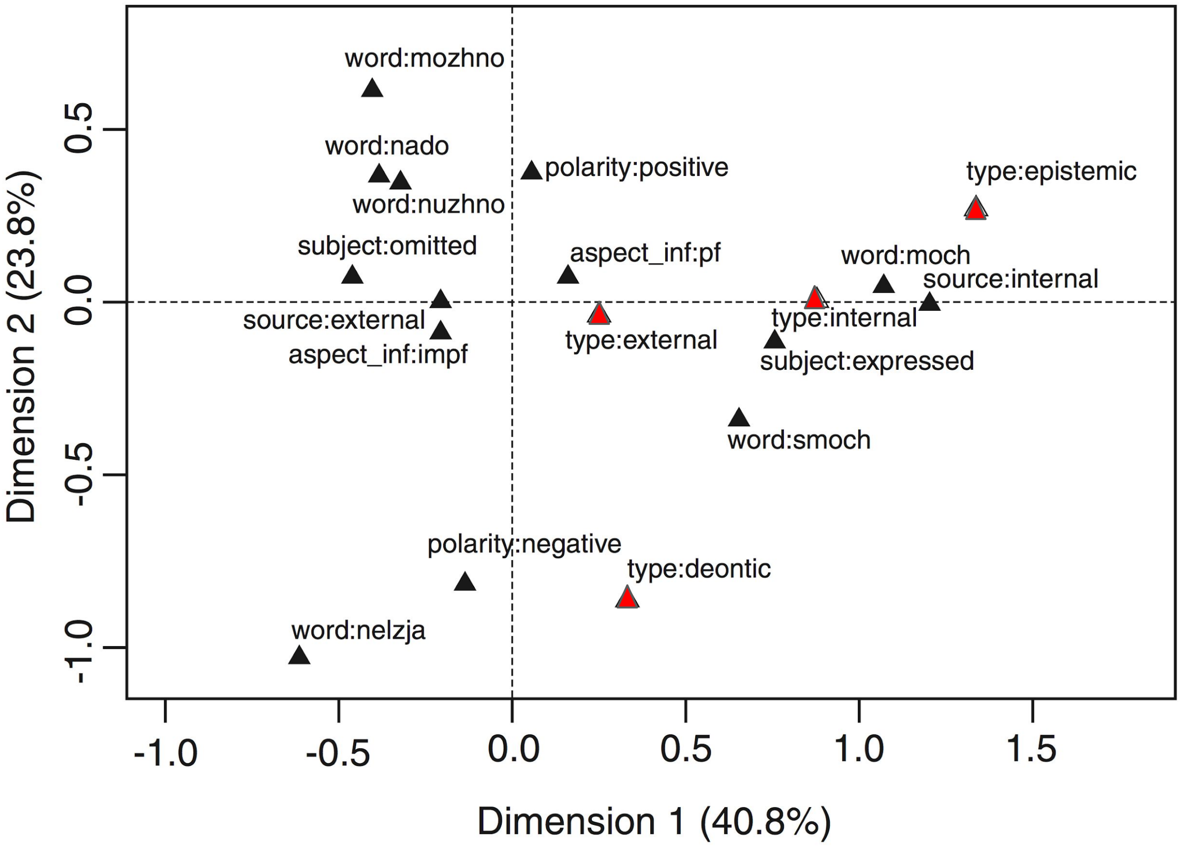
Dim1 opposes permission/prohibition to other functions (FUNCTION) and *wolno* to other modal words (WORD)
ASPECT and POLARITY do not contribute to Dim2 and hardly contribute to Dim1



FUNCTION, WORD, ASPECT, POLARITY

Supplementary variables

Russian modal words (Lyashevskaya et al. 2017): WORD, POLARITY, ASPECT, SOURCE. Supplem.: TYPE



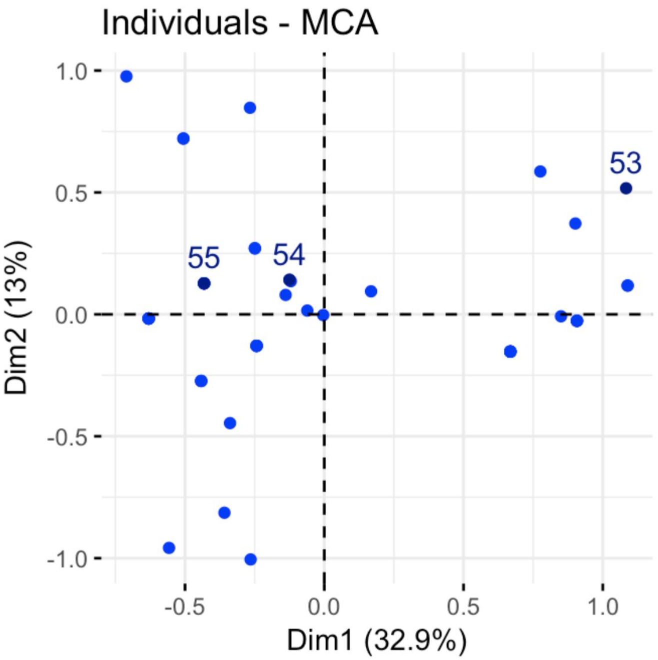
The axes are calculated based on the “active” variables.

Supplementary variables are added for the purposes of interpretation (they do not contribute to the axes).

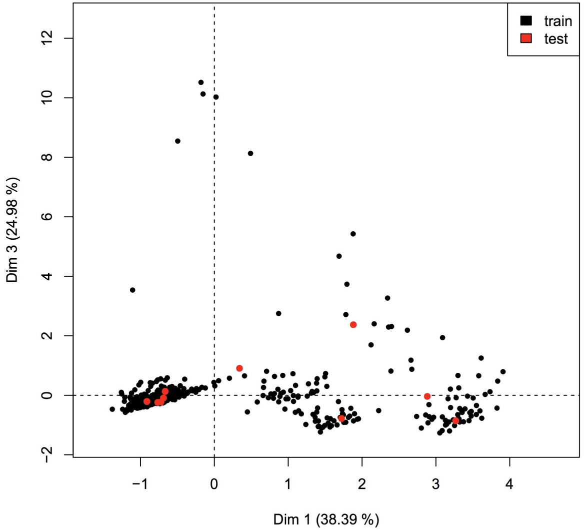
Supplementary variables can be both categorical (qualitative) and numeric (qualitative).

Supplementary individuals

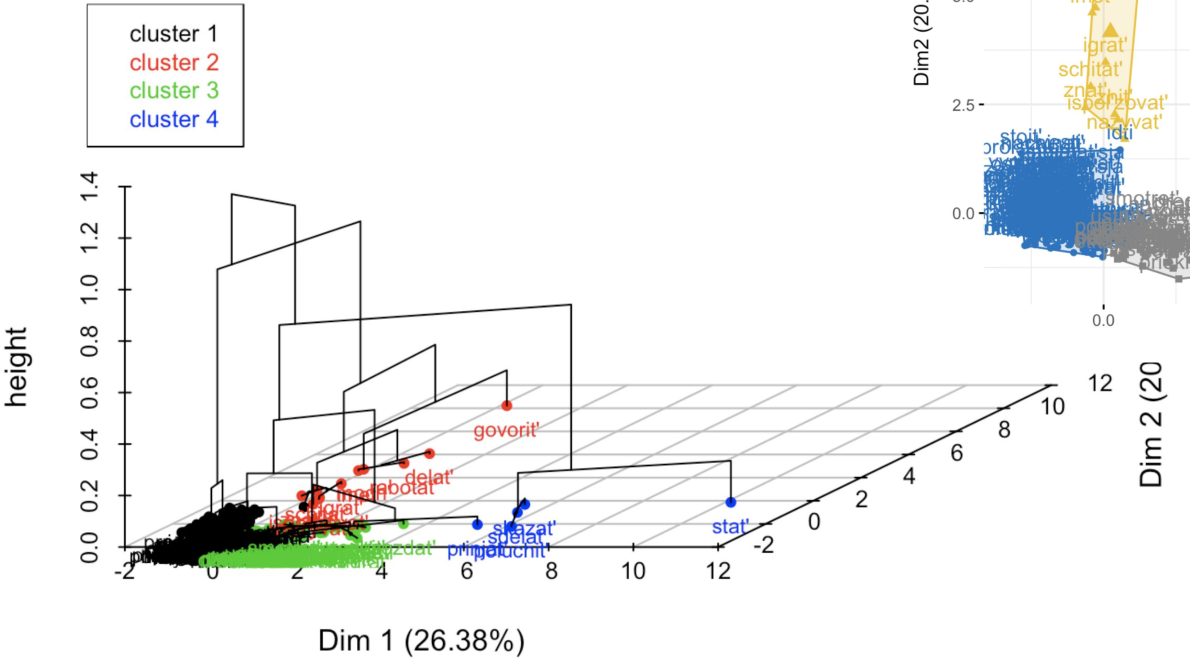
Example: train and test data. The row distances are calculated based on train data. The coordinates of supplementary individual observations are predicted using only the information of (M)CA performed on “active” rows and columns.

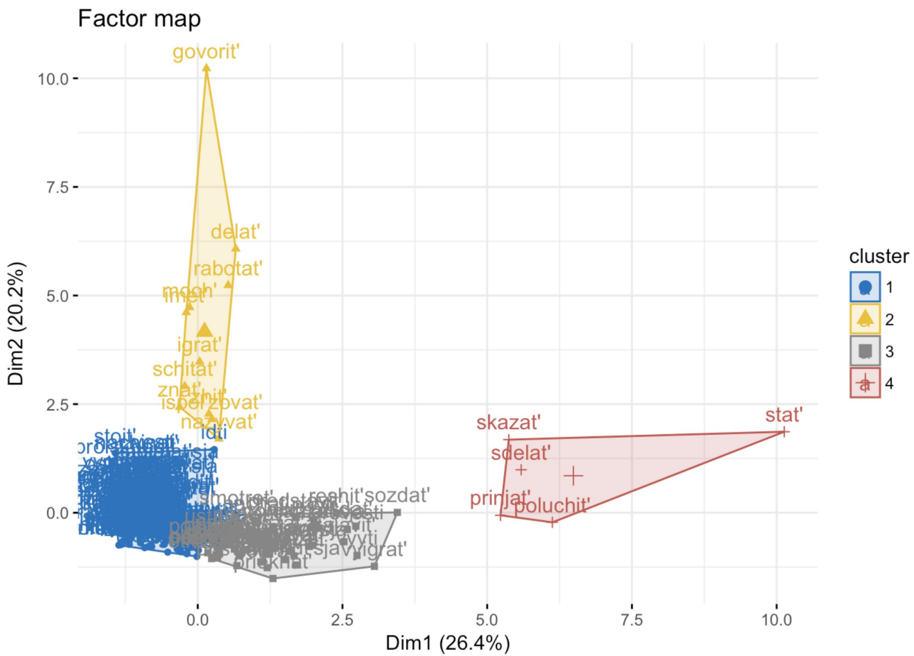


Will supplementary individuals be plotted within the same clusters or not?



Further steps: clusterisation and classification based on new dimensions





Example:
HCPC(),
Package:
FactoMineR