A tool for detecting words with consensual meaning in verbalization tasks

Belchin Kostov

Universitat Politècnica de Catalunya (Barcelona)

Agrocampus Ouest (Rennes)

Outline

- Objective

- Dataset and data structure

- Analysis method: Multiple Factor Analysis for Contingency Tables (MFACT)
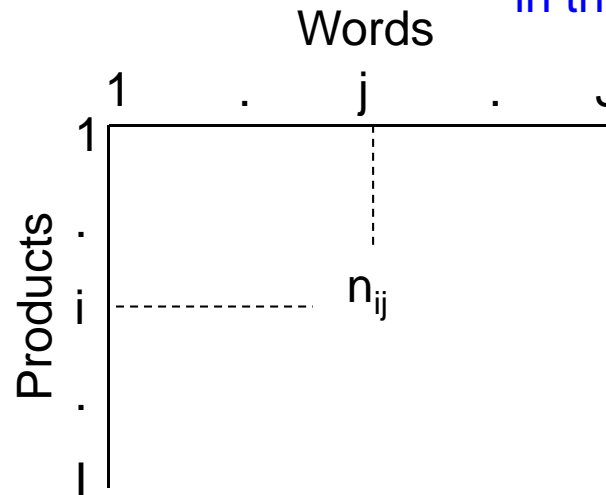
- Consensus measure and validation

- Conclusions

Textual analysis in sensometry

Labelled sorting task

Ultra-flash profiling

Others (check-all-that apply, open-ended questions, etc.)

Verbalization tasks

Assumption: Words are used in the same way by panellists

Words

	1	.	j	.	J

Products

1

.

i n_{ij}

.

I

Objective

Objective: To study similarities between use of the words (consensual words)

Data set

Example: Perfumes

- 12 rows (12 luxury perfumes)

- 98 columns (consumers)

- Each cell corresponds to the words associated with the group to which the product belongs for the consumer

- 198 different words

Sensometrics 2012, Agrocampus
Ouest Rennes

Data structure

	Consumer1	
P1 - Angel	fruity	strong
P2 - Aromatics Elixir	heady	grandmother
P3 - Chanel 5	heady	grandmother
P4 - Cinéma	fruity	middle
P5 - Coco Mademoiselle	fruity	middle
P6 - J adore (EP)	sweet	light
P7 - J adore (ET)	sweet	light
P8 - L instant	fruity	strong
P9 - Lolita Lempicka	fruity	middle
P10 - Pleasures	fruity	strong
P11 - Pure Poison	tangy	deodorant
P12 - Shalimar	tangy	deodorant

	Consumer98	
P1 - Angel		character
P2 - Aromatics Elixir		mature-woman
P3 - Chanel 5		eau-de-cologne
P4 - Cinéma		character
P5 - Coco Mademoiselle	light	natural
P6 - J adore (EP)	fruity	discreet
P7 - J adore (ET)	fruity	discreet
P8 - L instant		mature-woman
P9 - Lolita Lempicka		character
P10 - Pleasures	light	natural
P11 - Pure Poison		eau-de-cologne
P12 - Shalimar		mature-woman

	deodorant	fruity	grandmother	heady	light	middle	strong	sweet	tangy
P1	0	1	0	0	0	0	1	0	0
P2	0	0	1	1	0	0	0	0	0
P3	0	0	1	1	0	0	0	0	0
P4	0	1	0	0	0	1	0	0	0
P5	0	1	0	0	0	1	0	0	0
P6	0	0	0	0	1	0	0	0	0
P7	0	0	0	0	1	0	0	0	0
P8	0	1	0	0	0	0	1	0	0
P9	0	1	0	0	0	1	0	0	0
P10	0	1	0	0	0	0	1	0	0
P11	1	0	0	0	0	0	0	0	1
P12	1	0	0	0	0	0	0	0	1

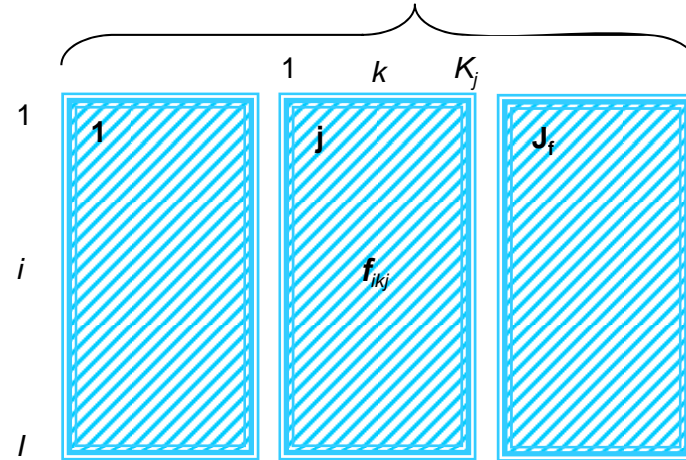|	character	discreet	eau-de-cologne	fruity	light	mature-woman	natural
P1	1	0	0	0	0	0	0
P2	0	0	0	0	0	1	0
P3	0	0	1	0	0	0	0
P4	1	0	0	0	0	0	0
P5	0	0	0	0	1	0	1
P6	0	1	0	1	0	0	0
P7	0	1	0	1	0	0	0
P8	0	0	0	0	0	1	0
P9	1	0	0	0	0	0	0
P10	0	0	0	0	1	0	1
P11	0	0	1	0	0	0	0
P12	0	0	0	0	0	1	0

Analysis = ?

Multiple Factor Analysis for Contingency Tables (MFACT)

- Analysis of multiple frequency and categorical tables (Bécue & Pagès 2004)

- Extended to mixture data with quantitative, categorical and frequency sets (Bécue & Pagès 2008)

- Adopts MFA approach (balance the influence of different sets of variables)

- Available in MFA function of FactoMineR package of *R*

Multiple frequency tables

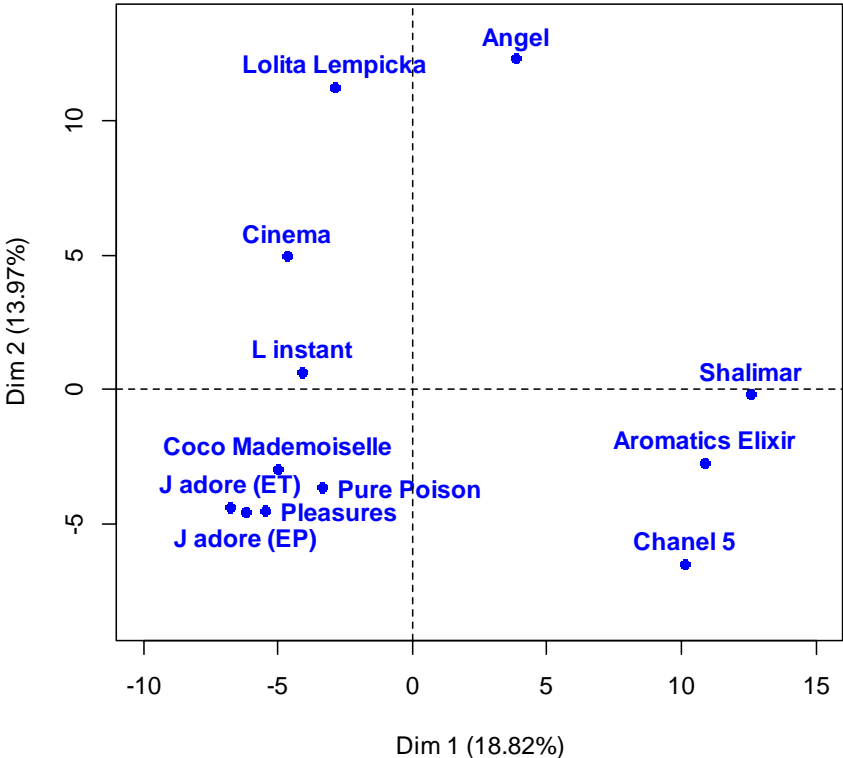Several sets of words (frequency tables)

Objectives
- Study the link between the sets of words
- Balance the influence of different sets of words in the global analysis
  - Standardizing the inertia of every cloud on the first principal axis to 1
  - Dividing the weight of the words by the first eigenvalue of separate analysis (each consumers)
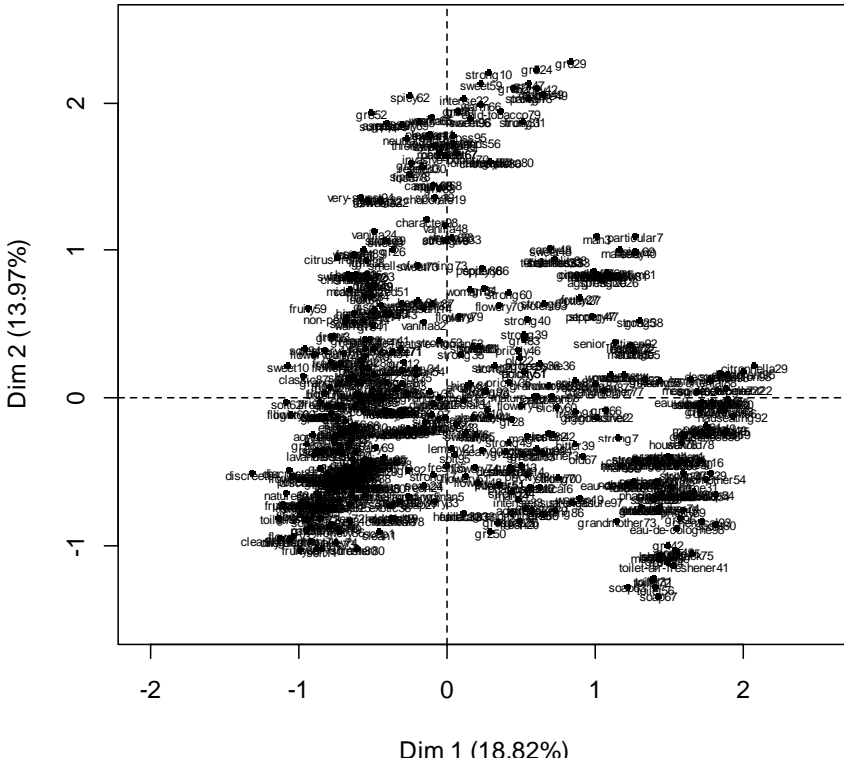
Outputs
- Configuration of products (perfumes)
- Representation of words
- Representation of consumers
- Configuration of products seen by each consumers

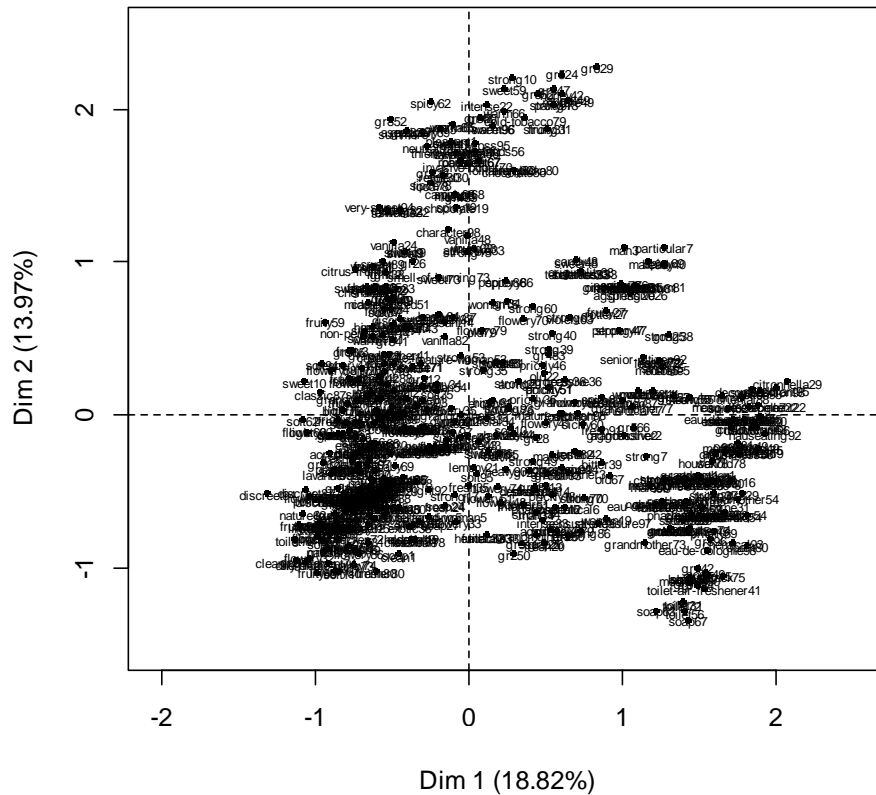# Global representation of perfumes and words

Individual factor map
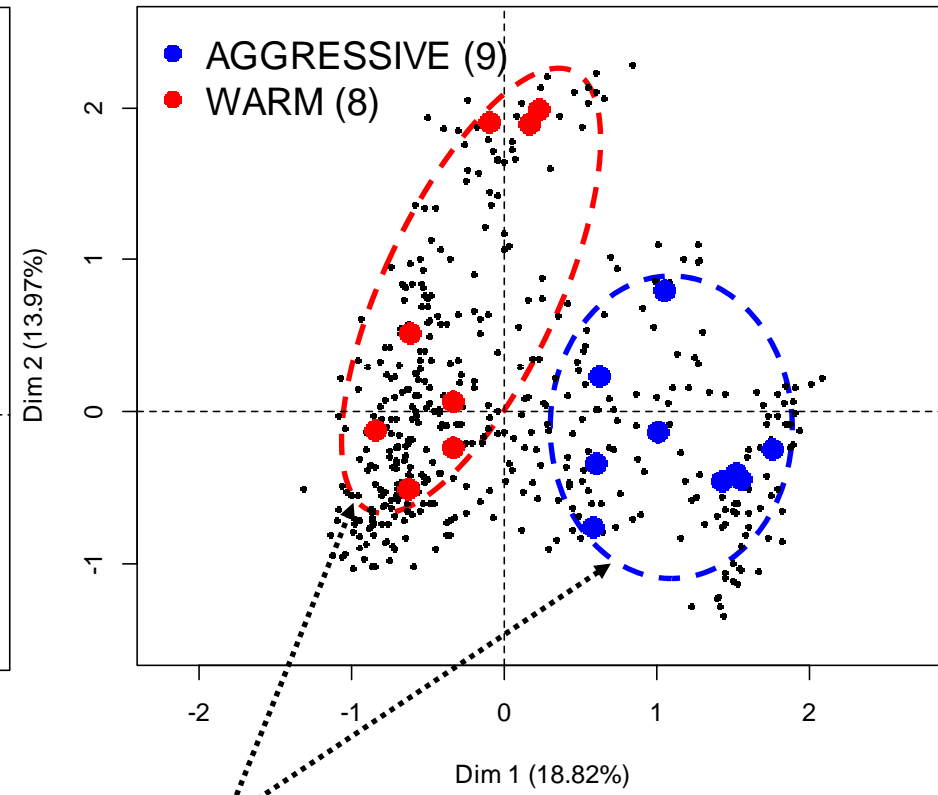
Factor map for the contingency table(s)

How to decide consensus?

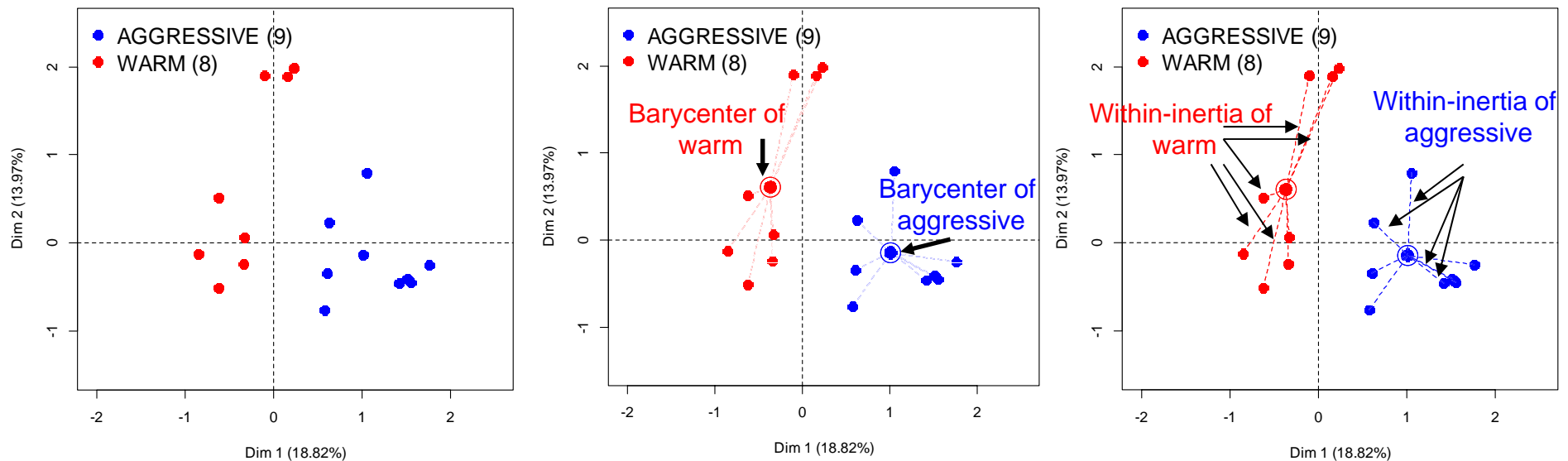**Is there a consensus for these words?**

Consensual words

Consensual words: *words associated to similar perfumes by different consumers*

→ Find the consensual words
  – Measure of consensus
  – Validation to define a consensus criterion
→ Representation with only consensual words

Measure of consensus

→ Within-inertia for the words pronounced by each consumer

   – Coordinates of words on the first dimensions of MFACT
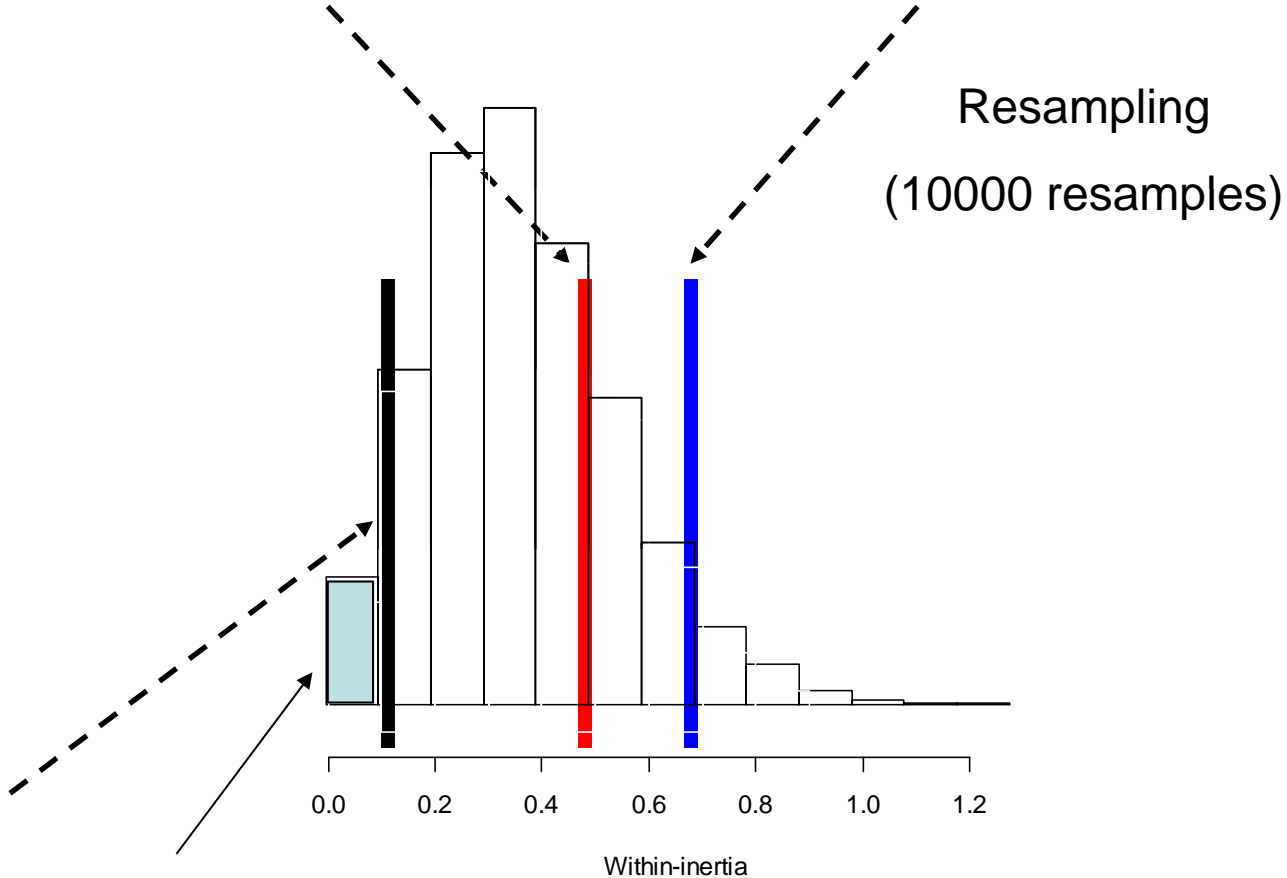
   – Weights of words (column weights MFACT)

How we can decide if the value of within-inertia is small or not?

Validation criterion

	Dim.1	Dim.2
acid49	1.772	-0.204
acid69	-0.605	0.034
acidity51	0.525	0.178
acrid10	-0.905	-0.243
acrid21	0.331	0.223
aggressive2	1.012	-0.131
aggressive26	1.054	0.793
aggressive28	1.422	-0.455
aggressive30	1.553	-0.444
aggressive36	0.628	0.228
aggressive56	1.765	-0.242
aggressive71	1.515	-0.406
aggressive79	0.579	-0.753
aggressive94	0.605	-0.341
alcohol17	0.714	0.011
alcohol33	1.592	-0.422
alcohol81	1.086	0.843
....	....	....
....	....	....
....	....	....
winter64	1.019	0.840
winter80	-0.338	-0.234
woman15	0.155	0.738
woman19	0.894	-0.672
woman32	-0.979	-0.710
wooded11	1.206	0.156
wooded58	1.104	0.157
young11	-0.620	0.975
young32	-0.448	1.339
young33	0.034	1.093
young5	-0.554	0.342
youth2	-0.092	-0.036

	Dim.1	Dim.2
acrid21	0.331	0.223
aggressive79	0.579	-0.753
winter80	-0.338	-0.234
woman32	-0.979	-0.710

	Dim.1	Dim.2
acid49	1.772	-0.204
aggressive36	0.628	0.228
alcohol17	0.714	0.011
winter64	1.019	0.840

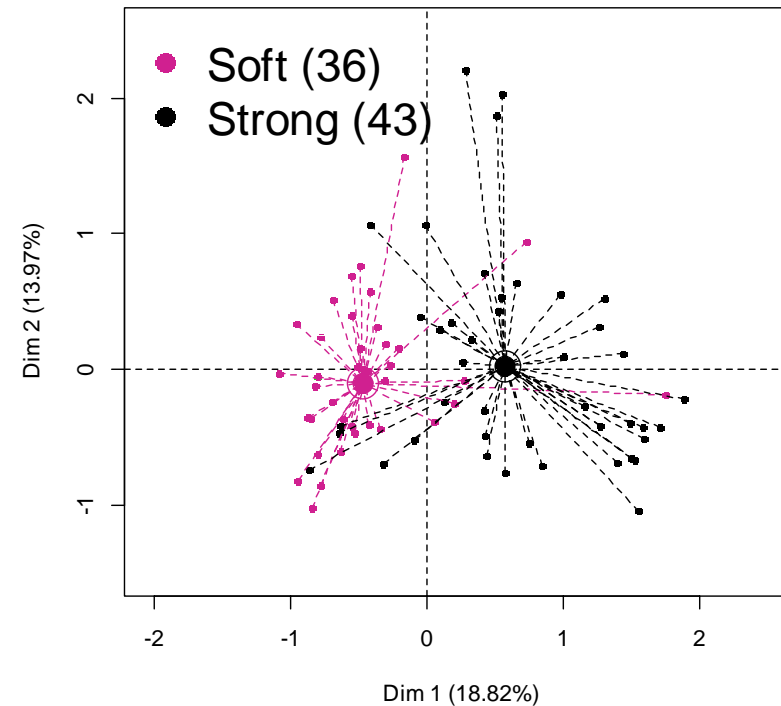Resampling

(10000 resamples)

Within-inertia

Probability to have a within-inertia smaller than the observed one
=251/10000=0.0251

Consensual words for perfumes

	pvalues	words.sel.panel	words.sel.freq
flowery	0.0000	46	200
soft	0.0000	36	146
fruity	0.0003	35	105
eau-de-cologne	0.0005	5	6
light	0.0007	19	80
discreet	0.0029	8	29
fresh	0.0070	16	48
aggressive	0.0107	9	25
toilet	0.0217	11	19
medicine	0.0237	4	6
young	0.0251	4	12
grandmother	0.0260	12	39
old	0.0305	9	26
soap	0.0501	11	22
sweet	0.0629	32	106
male	0.0693	5	13
nature	0.0967	6	16
vanilla	0.1255	7	19
heady	0.1505	7	13
spicy	0.2021	12	30
prickly	0.2101	11	31
strong	0.2147	43	140
rose	0.2157	4	15
gr6	0.3027	5	5
heavy	0.3342	7	22
warm	0.5003	8	25
chemical	0.6172	7	18
summer	0.6350	4	17
peppery	0.7058	6	17
intense	0.7086	4	12
gr4	0.9157	13	26
gr5	0.9302	13	24
gr2	0.9381	13	34
gr3	0.9994	17	42
gr1	0.9995	11	31

Consensual words

Soft (36)
Strong (43)

Dim 2 (13.97%)

Dim 1 (18.82%)

Representation of the consensual words

Individual factor map

Consensual words

Conclusions

- We have proposed an original method that gives a clear description of the products
- Words are selected from a statistical criterion
- This method runs with any number of dimensions

For the future

- Use of contextual information on consumers (age, sex)
- Multilingual consumer panels

Thanks for your attention!