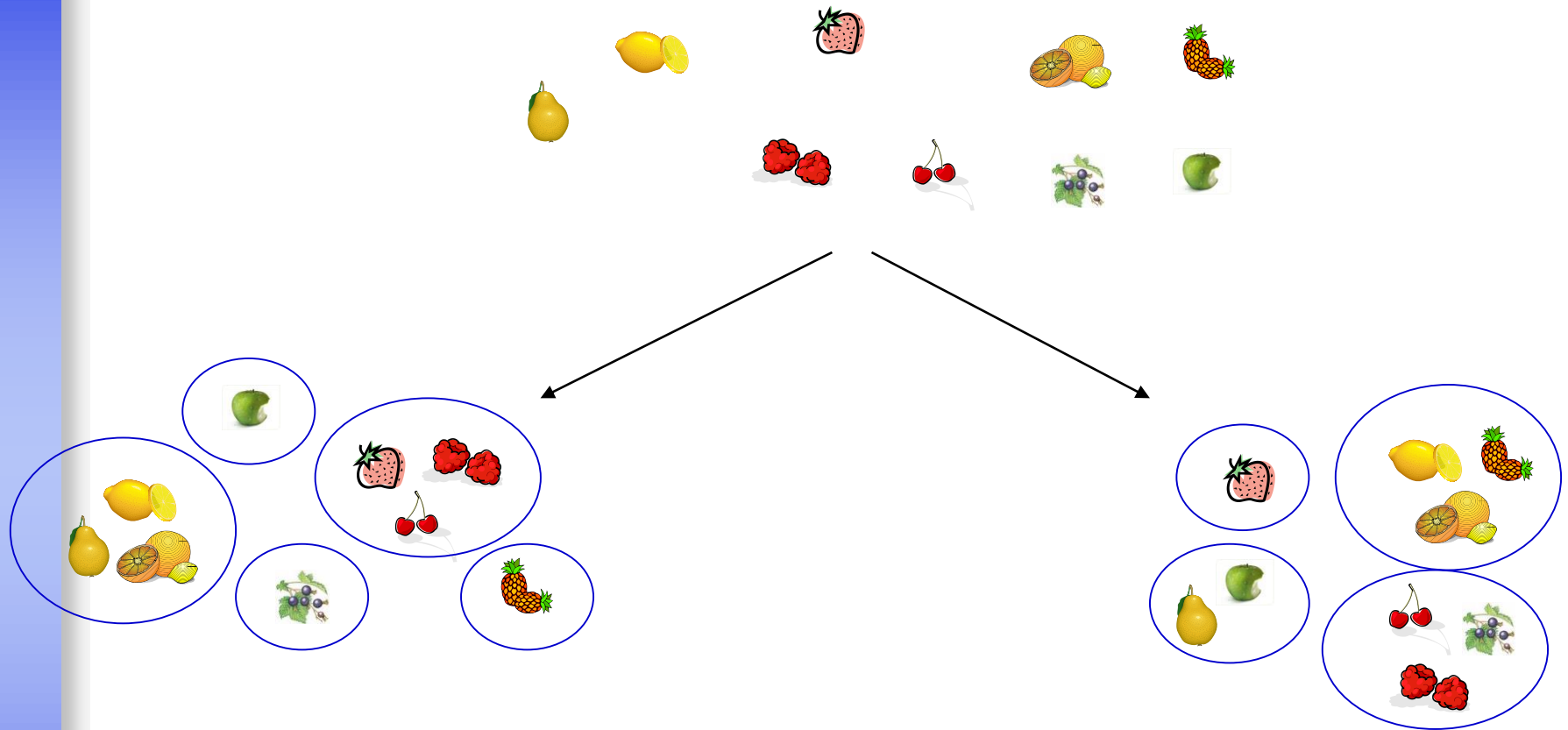


FreeSortR: Un package pour l'analyse de données de tri libre.

Philippe Courcoux
Unité de Sensométrie et Chimiométrie
ONIRIS Nantes
philippe.courcoux@oniris-nantes.fr

Epreuve de tri libre

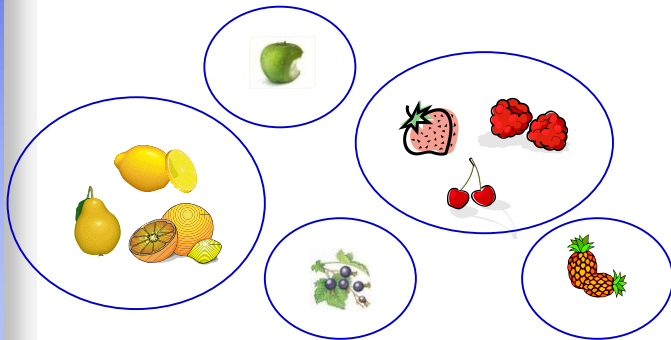


Méthode descriptive et non verbale
qui consiste à évaluer les similarités entre produits

(+ tâche de verbalisation)

Epreuve de tri libre

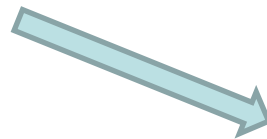
Chaque sujet fournit une partition de l'ensemble des stimuli



Données individuelles

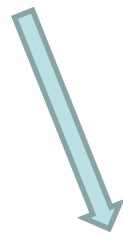
Classes disjointes

{1, 2, 4} {6, 7, 9} {3} {5} {8}

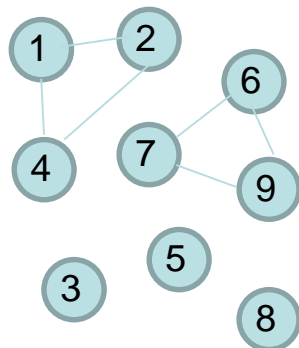


Matrice de dissimilarités

0	0	1	0	1	1	1	1	1
0	0	1	0	1	1	1	1	1
1	1	0	1	1	1	1	1	1
0	0	1	0	1	1	1	1	1
1	1	1	1	0	1	1	1	1
1	1	1	1	1	0	1	1	1
1	1	1	1	1	1	0	1	1
1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	0



Graphe



Analyse d'un ensemble de partitions sur les mêmes objets

✓ Méthodes de classification

- **par arbres**

Classification ascendante hiérarchique
Arbres à distances additives

- **par partitionnement**

Partition centrale ou consensus

✓ Méthodes factorielles

- **Multidimensionnal Scaling (MDS)**

Partition consensus (1)

Soit P un ensemble de partitions $P_1 \dots P_K$

Objectif : construire une partition consensus C
dont l'accord est maximal avec les éléments de P
(i.e. minimisant la distance aux partitions initiales)

*Régnier (1965) ; Marcotorchino & Michaud (1982) ; Barthélémy & Leclerc (1995),
Gordon & Vichi (1998) ; Krieger & Green (1999) ; Guénoche (2011)*

Partition consensus ... ou Partition centrale ou Partition médiane

$$\underset{C}{\text{Arg min}} \sum_{k=1}^K d(C, P_k)$$

Partition consensus (2)

$$\equiv \underset{C}{\text{Arg max}} \frac{1}{K} \sum_{k=1}^K R^a(C, P_k)$$

avec $R^a(,)$ le coefficient du Rand ajusté

- *Pas 1.* Initialisation de la partition C à T classes
- *Pas 2.* Alternativement, affecter chaque stimulus au groupe maximisant ce critère
- *Pas 3.* Stopper lorsque le critère n'est plus amélioré. Sinon, revenir au pas 2.

transfert

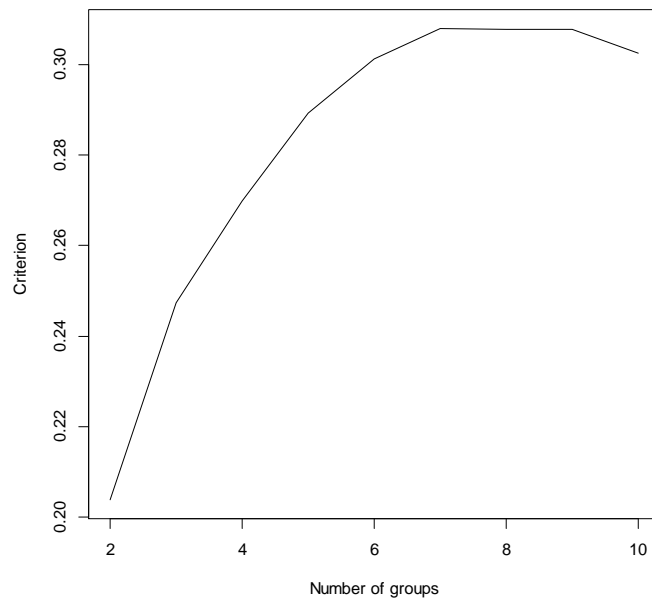
Application

31 sujets; 16 aromes

```
> data(AromaSort)
> AromaSort
```

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29	S30	S31
Lemon	5	1	4	1	1	1	1	2	3	1	1	1	1	4	2	1	2	2	2	3	3	4	3	1	2	6	1	2	8	1	1
Grapefruit	5	3	5	1	2	1	2	2	3	1	1	1	1	4	2	1	1	2	1	2	3	6	3	1	2	6	1	2	2	1	1
Pineapple	4	1	5	1	5	1	2	1	1	1	5	1	1	1	1	1	1	3	1	2	2	4	1	3	7	6	3	3	4	1	2
Pear	5	1	5	1	1	3	2	2	1	2	1	1	1	1	1	1	2	2	1	3	5	4	2	3	6	1	1	3	4	1	1
Honey	4	2	3	2	3	6	3	5	2	3	7	2	1	4	5	5	3	1	5	4	4	3	1	2	2	7	5	4	12	1	1
Butter	1	4	5	2	3	5	3	5	4	4	4	5	2	4	9	5	6	4	3	5	4	2	1	6	4	5	3	5	9	5	3
Grilledbread	1	2	3	5	5	5	3	3	2	3	9	2	3	2	3	2	3	1	4	1	4	2	4	6	4	4	2	1	1	5	3
Grilledhazelnut	5	2	3	2	5	5	3	3	4	3	6	2	3	2	3	2	3	5	5	1	4	2	3	4	4	5	2	7	6	5	6
Strawberry	5	1	5	1	1	2	2	1	1	2	1	1	1	1	6	1	2	3	1	3	6	6	1	3	1	1	3	3	3	1	4
Raspberry	3	3	4	4	1	2	2	1	4	4	1	1	1	1	1	1	2	2	2	3	5	2	2	3	1	1	3	3	4	4	4
Cherry	3	3	5	3	5	6	1	1	1	2	2	1	3	3	7	8	2	4	6	7	7	4	3	2	8	2	3	3	7	1	3
Blackcurrant	3	4	4	3	4	5	2	5	4	3	7	1	2	4	1	1	4	3	6	5	4	1	2	1	4	7	1	3	9	3	2
Greenpepper	4	3	1	4	4	3	3	4	2	5	8	2	1	3	4	7	3	1	3	2	2	2	5	4	7	4	1	6	11	3	2
Smoked	2	2	2	4	3	4	3	4	7	3	3	3	5	2	4	3	5	1	5	9	1	5	6	2	3	3	2	1	10	5	5
Pepper	2	2	2	1	2	3	1	2	6	5	5	4	4	3	8	6	4	4	6	8	3	3	3	5	2	8	4	4	11	2	1
Licorice	2	4	3	4	2	4	3	3	5	4	7	3	5	3	5	4	3	1	4	6	3	5	2	2	5	9	4	4	5	2	3

Aroma data



Partition optimale

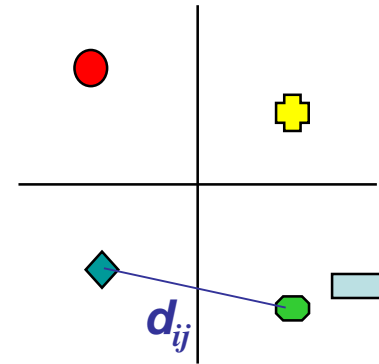
{citron, pamplemousse} {ananas, poire, fraise, framboise}

{miel, pain grillé, noisette}

{beurre, cassis} {fumé, réglisse}

{cerise} {poivre vert} {poivre}

			δ_{ij}		



dissimilarités observées

représentation euclidienne

Indices de stress (Kruskal)

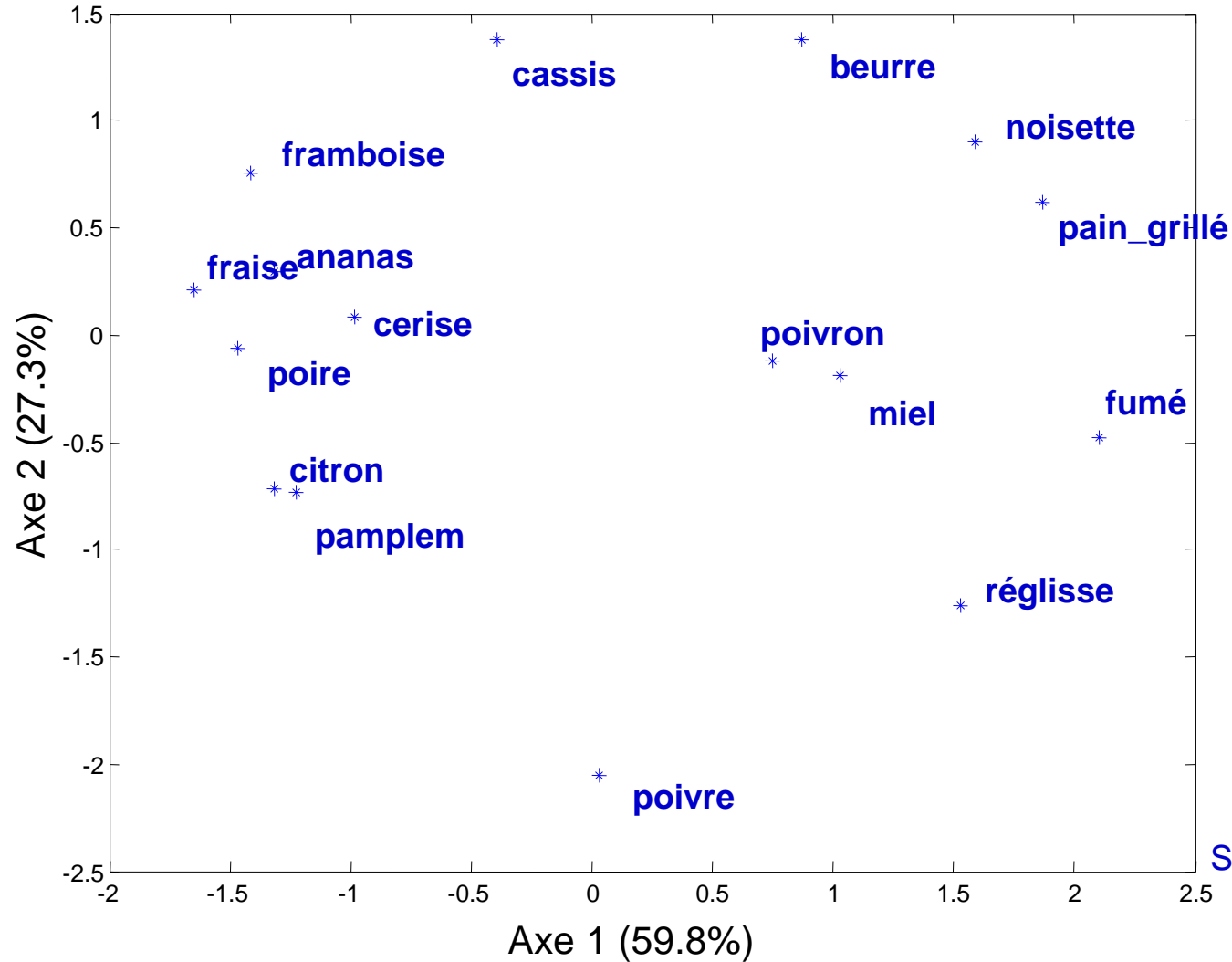
$$\frac{\sum_{i,j} (\delta_{ij} - d_{ij})^2}{\sum_{i,j} d_{ij}^2}$$

MDS métrique

$$\frac{\sum_{i,j} (f(\delta_{ij}) - d_{ij})^2}{\sum_{i,j} d_{ij}^2}$$

MDS non métrique

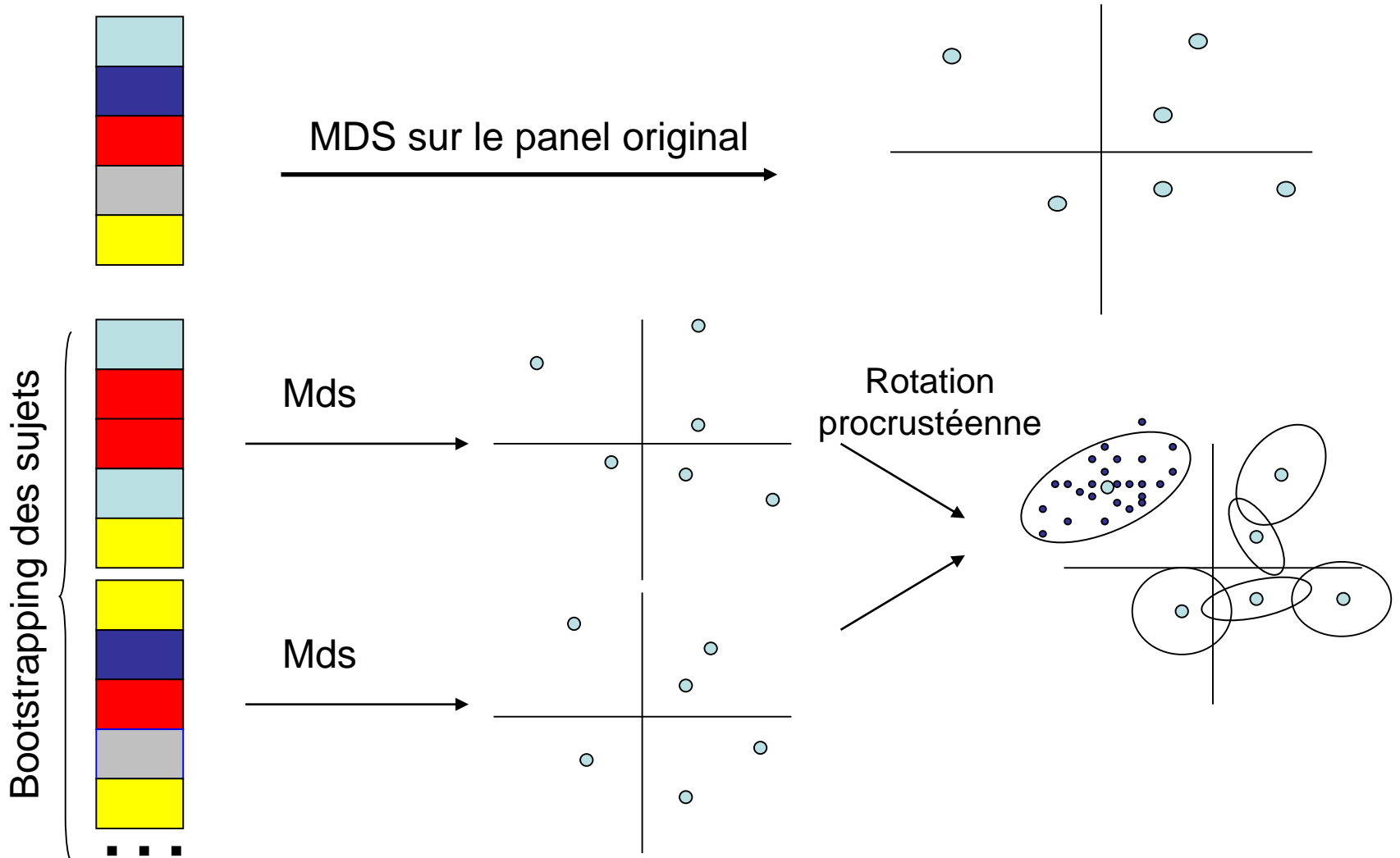
MDS non métrique: configuration

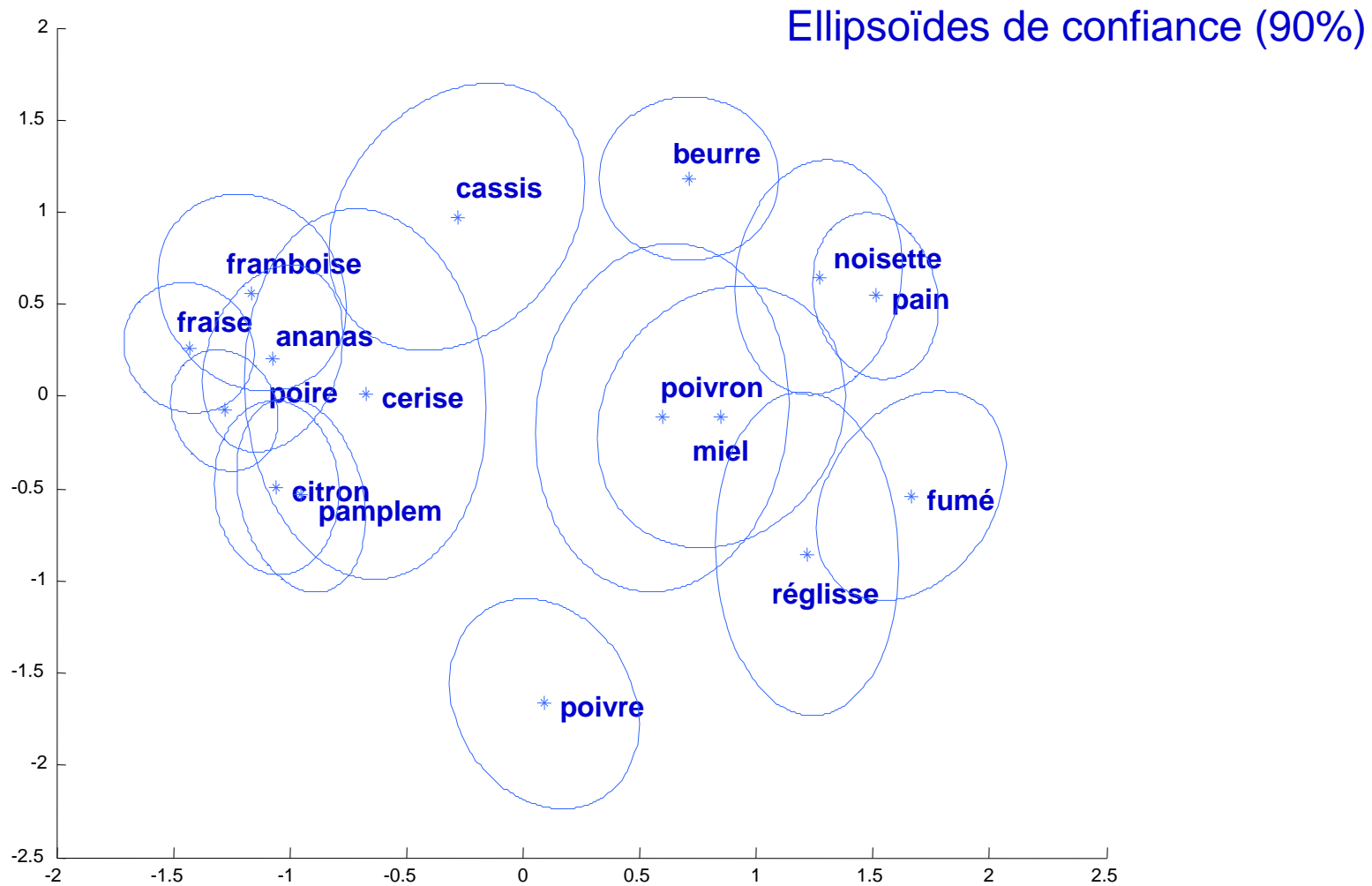


Solution en dimension 3

Stress=0.051
 $\rho_s(\delta_{ij}, d_{ij}) = 0,94$
 $R(\delta_{ij}, d_{ij}) = 0,92$

Robustesse de la configuration : rééchantillonnage des sujets

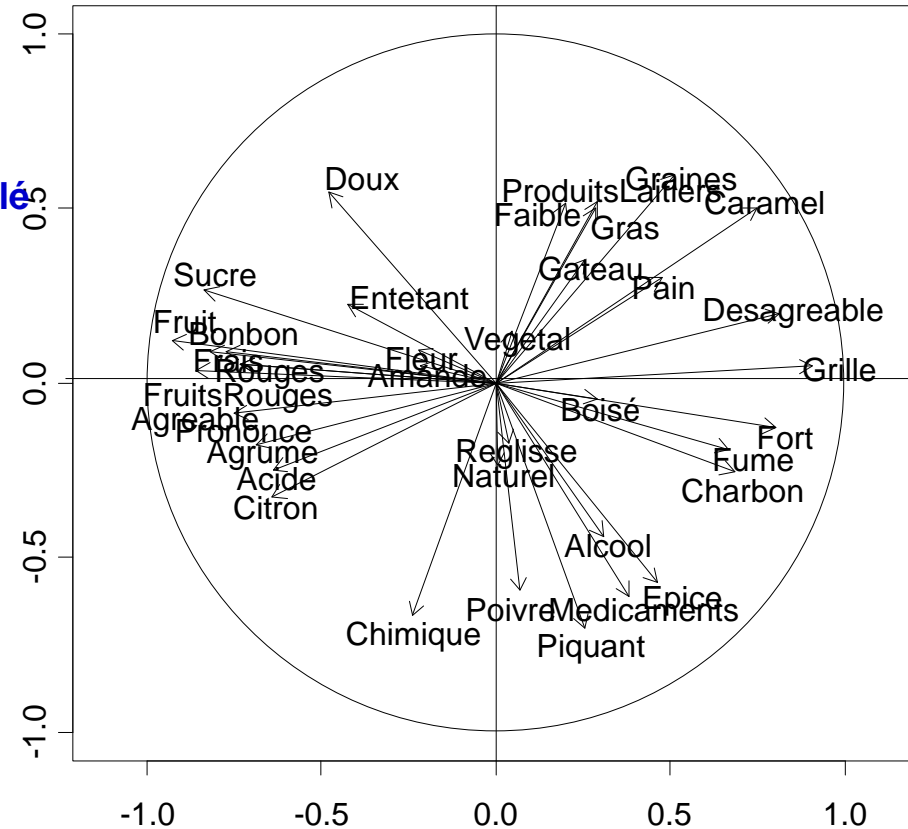
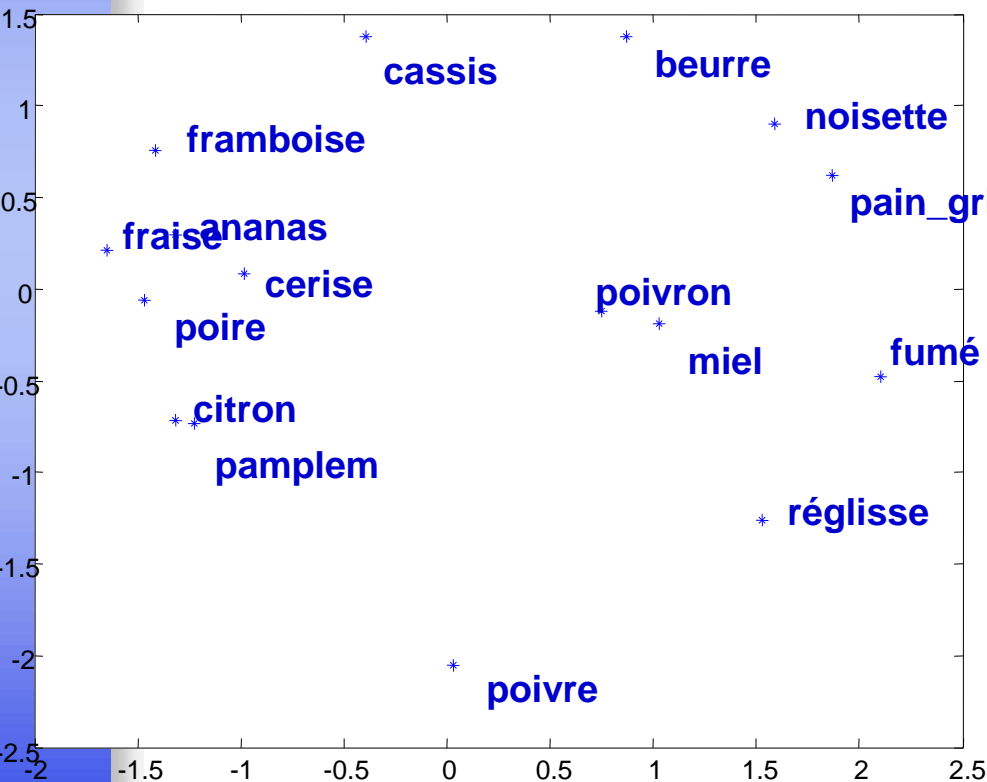




Robustesse de la configuration/ Qualité du consensus entre sujets

Interprétation de la configuration

Corrélations entre occurrences des termes (phase de verbalisation) et coordonnées des stimuli sur les axes de la MDS



Techniques utilisables pour comparer des résultats de classifications

Quelques extensions:

- *Classification en utilisant les arbres à distances additives,*
- *Gestion des tris incomplets (sorting en Blocs Incomplets),*
- *Adaptation pour les tris multiples et les tris hiérarchiques*