

Méthodes de traitement statistique des données issues d'une épreuve de tri libre

P.Faye^{1,2}, P.Courcoux^{1,2}, E.M.Qannari^{1,2}, A.Giboreau³

¹ ONIRIS, Unité de Sensométrie et Chimiométrie, Nantes, F-44322, France ;

² Université Nantes Angers Le Mans, France.

pauline.faye@oniris-nantes.fr, philippe.courcoux@oniris-nantes.fr, elmostafa.qannari@oniris-nantes.fr

³ Centre de Recherche de l'Institut Paul Bocuse Château du vivier, BP 25, 69130 Ecully, France

agnes.giboreau@institutpaulbocuse.com

Résumé

En évaluation sensorielle, l'épreuve de catégorisation dite 'tri libre' connaît une popularité grandissante du fait de sa facilité de mise en œuvre et sa capacité de procurer une représentation pertinente des différences perceptives entre des produits. Plusieurs méthodes de traitement statistique des données issues d'une épreuve de catégorisation ont été proposées. L'objectif est de faire une étude bibliographique des méthodes utilisées et de les comparer sur la base d'un exemple afin d'orienter l'utilisateur dans son choix d'une stratégie d'analyse.

Mots clés : méthodes factorielles, classification, tri libre, catégorisation.

Abstract

The interest for free sorting in sensory evaluation is gaining ground as it provides a easy-to-use and reliable means to assess perceptive differences between products. Several statistical procedures have been proposed for the analysis of the data from a free sorting procedure. The aim is to review the proposed statistical methods and to compare their outcomes on the basis of a case study.

Key words: factorial analysis, classification, free sorting task, categorization.

1. Introduction

En agro-alimentaire comme dans d'autres secteurs d'activités tels que la cosmétique ou l'industrie automobile, l'analyse sensorielle est un passage obligé pour la conception de nouveaux produits et le contrôle de la qualité. Plusieurs épreuves peuvent être mises en place selon les objectifs poursuivis. Parmi ces épreuves, nous distinguons l'épreuve du profil sensoriel qui consiste à décrire un ensemble de produits à l'aide d'une liste de descripteurs sensoriels. Un panel de sujets est sélectionné et entraîné pour donner des évaluations sous forme de notes qui reflètent des intensités perçues pour chacun des descripteurs considérés (variables). Très généralement, les différences entre les produits sont testées par analyse de variance puis synthétisées sur une carte factorielle.

Compte tenu de nouvelles exigences opérationnelles en particulier l'augmentation du nombre de produits à comparer et grâce au transfert de connaissances ou d'approches méthodologiques issues de différents domaines scientifiques, de nouvelles méthodes d'investigations ont émergé en évaluation sensorielle, notamment l'épreuve de tri libre, également appelée épreuve de catégorisation. Celle-ci a été, depuis la fin des années 1960, utilisée en psychologie expérimentale, en sociologie, en sciences de l'éducation... Pour une revue relativement exhaustive sur le sujet, nous renvoyons le lecteur au livre de Coxon (1999). Basée sur un processus holistique de catégorisation (évaluation globale), l'épreuve de tri libre s'avère plus naturelle et, de ce fait, permet d'évaluer un plus grand nombre de produits que les approches sensorielles analytiques. Du fait de son caractère intuitif, global, non verbal et comparatif, elle est particulièrement adaptée pour appréhender la perception de produits par des consommateurs (Lawless et al, 1995 ; Faye et al, 2004 ; Cartier et al, 2006), étudier l'organisation des connaissances et des représentations (Picard et al, 2003 ; Dacremont et al, 2006 ; Manetta et al, 2007) ou révéler des différences de comportement entre consommateurs en termes culturel (Chrea et al, 2004 ; Blancher et al, 2007), en termes d'expertise (Giboreau et al, 2001 ; Soufflet et al, 2004 ; Lelièvre et al, 2008 ; Ballester et al, 2008) ou encore en termes d'âge (Ricciuti, 2006).

Lors d'une épreuve de tri libre, le sujet est invité à partitionner les produits en autant de classes qu'il/elle juge nécessaire avec la consigne que les produits d'une même classe sont ceux qu'il/elle perçoit comme étant proches. Ainsi les données collectées se présentent sous forme de

partitions d'un même ensemble de produits ; chaque partition étant produite par un sujet. Notons qu'en plus de partitionner les produits, les sujets sont souvent invités à décrire les groupes constitués à l'aide des mots qui leur semblent pertinents. Généralement, cette description verbale (ou « verbatims ») est utilisée comme élément d'interprétation des résultats.

D'un point de vue statistique, plusieurs démarches d'analyse des données de tri libre ont été proposées. Nous distinguons deux grandes familles de méthodes statistiques permettant d'analyser les données issues d'une épreuve de catégorisation:

- 1- les méthodes factorielles (dont le positionnement multidimensionnel ou multidimensional scaling, MDS),
- 2- les méthodes de classification (dont les arbres additifs).

De manière générale, le recours à telle ou telle autre méthode d'analyse dépend du domaine d'application. L'utilisation des arbres additifs est relativement populaire dans les applications de la catégorisation en psychologie (Dubois, 1991 ; Poitevineau, 2002) alors qu'en psycho-acoustique, ce sont plutôt les méthodes relevant de MDS qui sont le plus souvent utilisées (Mc Adams et al, 1995 ; Gygi et al, 2007). Dans le domaine de l'analyse sensorielle qui nous intéresse plus particulièrement ici, la tendance est d'utiliser des méthodes de type MDS (Lawless et al, 1995 ; King et al, 1998 ; Faye et al, 2004 ; Parr et al, 2007), bien que des méthodes alternatives relevant de l'analyse des correspondances multiples aient été récemment proposées dans ce cadre (Qannari et al, 2009 ; Cadoret et al, 2009).

Notre objectif n'est pas de comparer les performances des différentes méthodes, qui sont le plus souvent plus complémentaires que concurrentes (classification, analyse factorielle...) mais plutôt de dresser un panorama des méthodes statistiques de traitement des partitions issues de l'épreuve de tri libre, et ce, sans prétendre à l'exhaustivité. Aussi, nous avons jugé utile de les présenter et de les illustrer au travers une étude de cas, qui porte sur la catégorisation d'un ensemble de 16 arômes par 31 sujets non experts. *In fine*, nous souhaitons donner au lecteur une vision plus claire des méthodes existantes et l'orienter dans le choix d'une stratégie d'analyse qui correspond à ses objectifs.

2. L'épreuve de tri libre et l'étude de cas

2.1. L'épreuve de tri libre

Comme déjà évoqué, le tri libre est une épreuve simple et naturelle (intuitive) qui permet d'évaluer un grand nombre de produits. L'intérêt peut porter sur les différences perçues entre les produits et, dans ce cas, l'épreuve s'apparente à une méthode de caractérisation sensorielle. Cependant, l'épreuve de tri libre peut également être mise en œuvre pour étudier des différences entre sujets ou groupes de sujets et, de ce point de vue, elle s'inscrit dans une démarche expérimentale de psychologie cognitive.

L'épreuve de tri libre consiste à présenter l'ensemble des produits aux sujets qui participent à l'épreuve et leur demander de les regrouper en fonction des ressemblances et différences perçues globalement. La plupart du temps, la consigne spécifie aux sujets ces deux aspects de la tâche de la manière suivante : « regrouper les produits qui se ressemblent » et « mettre dans des groupes séparés les produits qui semblent différents ». Le nombre de groupes et le nombre de produits par groupe sont laissés à l'appréciation de chaque sujet. Cependant, les sujets sont alertés sur le fait qu'ils doivent éviter de grouper tous les produits dans un même groupe ou mettre chaque produit dans une classe séparée. Une fois les groupes formés, il peut être demandé aux sujets de décrire chacun d'eux avec leurs propres termes.

De manière concrète, l'expérimentateur peut chercher à répondre aux questions suivantes :

- quels sont les produits qui sont proches / distincts d'un point de vue perceptif ? selon quels critères donnés par les sujets ?

- existe-t-il des groupes de sujets présentant des différences dans leur perception et dans leur représentation des produits ?

Compte tenu de ces objectifs, il apparaît qu'en particulier les méthodes factorielles ou les méthodes de classification, utilisées seules ou conjointement, peuvent aider l'utilisateur à trouver des éléments de réponse à ces interrogations.

2.2. L'étude de cas

Seize arômes issus du coffret « Le nez du vin » (Edition Jean Lenoir, 2006) ont été sélectionnés pour l'expérimentation. Ces arômes sont utilisés pour illustrer des sources odorantes familières pour des sujets français et identifiables dans des vins blancs ou des vins rouges. Les indications données par les concepteurs des seize arômes sont consignées dans le tableau 1.

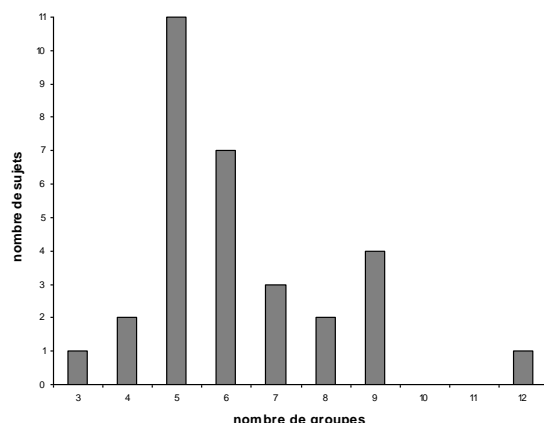
TAB 1. Présentation des arômes en fonction du type de vin et de la nature de la note (les symboles en regard des notes aromatiques seront utilisés dans les représentations graphiques)

Type de vin	Notes	Arômes
blanc	Fruitées □	Citron
		Pamplemousse
		Ananas
		Poire
	Florales ○	Miel
	Animales *	Beurre
	Grillées ◇	Pain grillé
Noisette grillée		
rouge	Fruitées ■	Fraise
		Framboise
		Cassis
		Cerise
	Végétales et épicées ▲	Poivron vert
		Réglisse
		Poivre
	Grillées ◆	Fumé

Un groupe composé de 31 étudiants en agro-alimentaire (ONIRIS) ont participé à l'expérimentation. Les sujets ne sont pas experts en œnologie et n'ont pas suivi de formation à la dégustation. Les arômes codés par un nombre aléatoire à trois chiffres ont été présentés simultanément aux sujets dans des flacons dans un ordre aléatoire. Après avoir senti un à un l'ensemble des arômes, chaque sujet a constitué des groupes en fonction des ressemblances et dissemblances perçues. Par la suite, chaque sujet a été invité à décrire les groupes de produits qu'il/elle a formés.

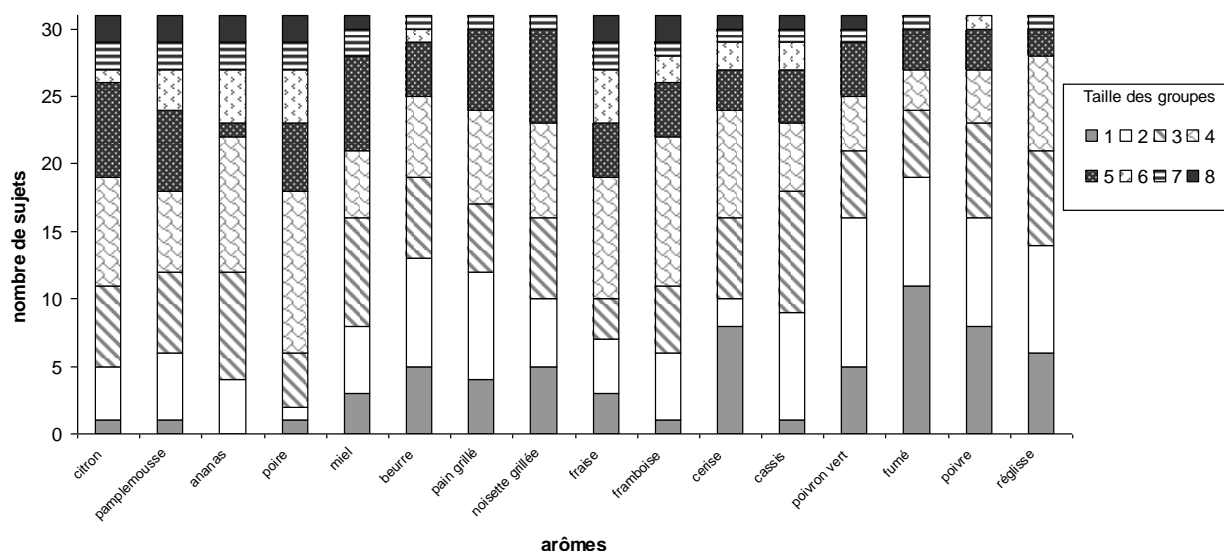
Afin de mieux les appréhender, décrivons sommairement la structure des données collectées à l'issue de cette expérimentation. D'après la figure 1, il apparaît que la majorité des sujets ont classé les arômes dans 5 ou 6 groupes. En moyenne, les sujets ont réalisé 6.2 groupes.

FIG. 1 - Répartition des sujets en fonction du nombre de groupes formés



La figure 2 qui représente la répartition des sujets en fonction de l'effectif des groupes associés à chaque arôme, permet d'observer que les arômes présentés ne constituent pas une catégorie homogène. En effet, certains arômes sont difficiles à regrouper avec les autres. A l'inverse, plusieurs arômes, fruités, ont été regroupés ensemble.

FIG. 2 - Répartition des sujets en fonction de la taille des groupes associés à chaque arôme



Comme pour la plupart des notes fruitées, la majorité des sujets a associé l'arôme « citron » à 2, 3 ou 4 autres arômes, formant ainsi des groupes de 3, 4 ou 5 arômes. Il apparaît que les notes fruitées ont été assez peu isolées à l'exception de l'arôme « cerise » qui a constitué un groupe à lui seul pour 8 sujets. Les arômes des autres notes ont été isolés par un plus grand nombre de sujets et, en particulier, les arômes « fumé » (11 sujets) et « poivre » (8 sujets). Notons également que les notes grillées (« pain grillé », « noisettes grillées », « fumé »), végétales (« poivron vert », « poivre », « réglisse ») et les arômes « beurre » et « cassis » sont souvent associés à un seul autre arôme.

3. Généralités sur le traitement statistique des données

3.1. le codage des données

En préambule, rappelons la « théorie des données » développée par Combs (in Coxon, 1999) qui distingue les observations, matériel brut issu de l'expérimentation, des données issues d'une interprétation et d'une sélection des observations par l'expérimentateur. En effet, l'objectif de l'expérimentation, les hypothèses de travail et les outils de traitement disponibles vont orienter le codage et la transformation des données. De manière plus concrète, considérons que N sujets

réalisent l'épreuve de tri de P Produits. Pour un sujet h donné, h variant de 1 à N , $K^{(h)}$ groupes de produits sont formés. Les différentes approches de codage de ce matériel expérimental sont les suivantes :

- Les partitions : Les groupes de produits réalisés par chaque sujet peuvent être considérés comme une partition des P produits. Ainsi, N sujets ayant réalisé l'expérimentation, nous disposons de N partitions des P produits.
- Les matrices de dissimilarités individuelles : Cette approche, la plus fréquemment utilisée, consiste à consigner les données de chaque sujet dans une matrice carrée de dimension $(P \times P)$ indiquant les dissimilarités entre les produits. Cette matrice est notée ici $\Delta^{(h)} = (\delta_{ij}^{(h)})$. Pour chaque sujet h et chaque paire de produits (i, j) , la cellule $\delta_{ij}^{(h)}$ est égale à 0 si les produits ont été regroupés dans la même classe et 1 sinon. De fait, la diagonale de cette matrice $\Delta^{(h)}$ est composée de 0.
- La matrice de dissimilarités agrégées : La matrice de dissimilarités agrégées, notée $\Delta = (\delta_{ij})$, est déterminée en considérant la somme des matrices de dissimilarités individuelles. Ainsi, la dissimilarité δ_{ij} entre les produits i et j ($i, j = 1, 2 \dots P$) est évaluée par le nombre de sujets qui n'ont pas mis ces deux produits dans le même groupe, soit $\delta_{ij} = \sum_h \delta_{ij}^{(h)}$.
- Le tableau disjonctif complet : Ce tableau est une manière différente et moins fréquente de coder l'information initiale. En effet, la partition de chaque sujet h peut être représentée sous forme d'un tableau disjonctif complet $Z^{(h)}$ de dimensions $(P \times K^{(h)})$ dont les lignes correspondent aux produits et les colonnes aux différents groupes de la partition de produits. Le tableau $Z^{(h)}$ du sujet h comportera donc autant de colonnes que de groupes réalisés. Pour une colonne donnée, chaque cellule est composée de 1 ou 0, selon que le produit est présent ou absent du groupe considéré. Les tableaux disjonctifs complets individuels sont concaténés pour former le tableau disjonctif complet global, noté ici Z .

Le tableau 2 reprend les principales caractéristiques des différents codages possibles des observations issues d'une épreuve de tri libre.

TAB 2. Synthèse des codages des observations issues de l'épreuve de tri libre

Type de codage	Construction	Caractéristiques
Partitions	Groupes de produits par sujet	
Matrices de dissimilarités individuelles	Pour chaque sujet : matrice carrée symétrique $P \times P$	0 si 2 produits dans le même groupe, 1 sinon Diagonale de 0
Matrice de dissimilarités agrégées	Somme des matrices individuelles $P \times P$	Diagonale de 0 Dissimilarité maximum = nombre total de sujets
Tableaux disjonctifs complets individuels	Tableau de P lignes et de $K^{(h)}$ colonnes	1 quand le produit i a été associé au groupe, 0 sinon
Tableau disjonctif complet	Concaténation des tableaux individuels	

Notons que d'autres dissimilarités entre produits ont été proposées. Elles ne seront pas développées ici. Soulignons néanmoins l'approche de Burton (Burton, 1975) qui propose des mesures permettant de pondérer, dans le calcul de la matrice agrégée, les matrices individuelles en tenant compte de la taille des groupes de produits et des paires de produits séparés par les sujets. En tout état de cause, toutes ces mesures de dissimilarités sont basées sur les co-occurrences de paires de produits dans les groupes formés par les sujets. Cependant, il faut noter que Daws (1996) a souligné les limitations d'une telle approche et proposé une stratégie d'analyse basée sur les co-occurrences de triplets de produits. Cette stratégie d'analyse n'est pas détaillée ici. Pour approfondir ce sujet, nous renvoyons le lecteur au livre de Coxon (1999) et à l'article de Daws (1996).

3.2. Les familles de méthodes

Comme mentionné ci-dessus, les méthodes factorielles et de classification sont très utilisées pour le traitement de données de tri libre. Le tableau 3, sans prétendre à l'exhaustivité, donne un panorama de ces méthodes, classées selon leur famille d'appartenance et le type de données de base auxquelles elles s'appliquent. Le tableau indique également une liste de références aux techniques statistiques utilisées et à leur application dans le cadre de l'épreuve de tri libre.

TAB 3. Panorama des méthodes d'analyse de données de tri libre classées par catégories (les méthodes indexées par un astérisque seront illustrées sur la base des données aromatiques)

Famille	Sous-famille	Données de base	Méthodes	Références bibliographiques
Méthodes factorielles	Méthodes apparentées à MultiDimensional Scaling	matrice de dissimilarités agrégées	MDS métrique*	Kruskal, 1964 Gygi et al, 2007
			MDS non métrique*	Schiffman, 1981 Lawless, 1989 Falahee et al, 1997 Faye et al, 2004
		matrice de dissimilarités individuelles	INDSCAL métrique DISTATIS	Caroll et Chang, 1970 Abdi et al, 2007
			INDSCAL non métrique*	Takane et al, 1977
	Méthodes apparentées à l'Analyse des Correspondances Multiples.	tableau disjonctif complet	ACM* MDSORT FAST	Van der Kloot et Van Herk, 1991 Takane, 1981 Cadoret et al, 2009
			pondérations individuelles IDSORT CCSort	Takane, 1982 Qannari et al, 2009
Méthodes de classification		matrice de dissimilarités agrégées	Classification hiérarchique*	Lebart et al, 2006 Coxon, 1999 Lawless, 1989 Giboreau et al, 2001
			Arbres additifs*	Sattah et Tversky, 1977 Barthélémy in Dubois, 1991 Poitevineau, 2002
		partitions individuelles	Partitions centrales*	Marcotorchino et Michaud, 1982 Brenac in Poitevineau, 2002
			Partitions latentes	Wiley, 1967

Le principe général des différentes méthodes (factorielles ou classification) est abordé dans les paragraphes suivants.

4. Les méthodes factorielles

Il est bien connu que les méthodes factorielles permettent de représenter les proximités entre produits dans un espace multidimensionnel de faible dimension. Pour le traitement des données de tri libre, nous avons choisi de séparer les méthodes factorielles en deux sous-familles : les méthodes s'apparentant à Multidimensional Scaling (MDS) et les méthodes s'apparentant à l'Analyse des Correspondances Multiples (ACM).

D'un point de vue conceptuel, les axes factoriels peuvent être interprétés comme des variables latentes qui structurent l'espace perceptif. Ces variables sont supposées révéler les critères sous-jacents aux catégorisations des sujets. L'interprétation de ces variables pourra être facilitée par le recours à des variables de formulation des produits, des mesures instrumentales ou les descriptions des groupes procurées par les sujets à l'issue du tri libre, le plus souvent utilisées comme variables supplémentaires, dites aussi illustratives.

4.1. Méthodes apparentées à MultiDimensional Scaling (MDS)

4.1.1. MDS sur matrices de dissimilarités agrégées

Considérons la matrice $\Delta = (\delta_{ij})$ des dissimilarités agrégées entre produits, mesurées, pour deux produits i et j ($i, j=1, \dots, P$), par le nombre de sujets qui n'ont pas placé ces produits dans le même groupe.

Les méthodes MDS (Kruskal, 1964 ; Torgerson, 1958) permettent de déterminer un espace de représentation de faible dimension dans lequel les distances entre produits reflètent au mieux les dissimilarités originales. Soit X , l'espace de représentation de dimension fixée par l'utilisateur. Nous désignons par $d_{ij}(X)$ la distance entre les produits i et j dans X . L'espace de représentation optimal est obtenu par minimisation d'un critère appelé Stress (Kruskal, 1964) :

$$Stress = \left[\sum_i \sum_j (\delta_{ij} - d_{ij}(X))^2 / \sum_i \sum_j d_{ij}(X)^2 \right]^{1/2}$$

Ce problème de minimisation est résolu de manière itérative. Un algorithme de résolution couramment utilisé est SMACOF (De Leeuw, 1988 ; Heiser, 1990 ; Borg et Groenen, 1997). Naturellement, plus le stress est proche de 0, meilleure est la représentation des dissimilarités initiales. Kruskal (1964) note qu'une valeur du critère Stress inférieure à 0.05 révèle une bonne qualité de représentation c'est-à-dire un bon ajustement entre les dissimilarités originales et les distances euclidiennes dans l'espace de représentation. A contrario, des valeurs supérieures à 0.20 indiquent un mauvais ajustement. Outre la qualité de la représentation, le calcul du stress permet de déterminer le nombre d'axes qui réalise un bon compromis entre la qualité de restitution des dissimilarités originales et la complexité de la représentation. Pour cela, il est recommandé d'examiner la courbe qui présente l'évolution du stress en fonction du nombre d'axes considérés. La stabilisation ou une décroissance moins marquée du stress à partir d'une certaine dimension signifie que l'ajout d'un axe supplémentaire dans la configuration n'apporte pas un gain significatif en termes de qualité de représentation des dissimilarités entre produits. Il s'agit donc d'identifier une rupture de pente dans la courbe d'évolution du stress et de retenir la configuration correspondante à cette rupture, dans la mesure où la qualité de représentation est acceptable. Alternativement, nous pouvons utiliser le ρ de Spearman, indice de corrélation de rang entre les dissimilarités originales et les distances calculées sur la base de l'espace de représentation. Ce coefficient augmentant avec la dimension de l'espace, il s'agit plus ici d'observer son amplitude de variation en fonction de la dimension que de le comparer à un seuil.

Notons que certains algorithmes se basent sur un critère appelé 'S-Stress' qui est similaire au critère Stress à ceci près que les carrés des distances et des dissimilarités sont considérées. Ceci conduit à un problème de résolution plus simple. Cependant, le principal inconvénient est qu'il tend à favoriser les distances les plus grandes dans la solution retenue au détriment des distances les plus faibles.

Il est important de souligner que plusieurs techniques de MDS peuvent être utilisées en fonction de la nature des dissimilarités considérées. A la différence des algorithmes de MDS dits métriques qui sont basés sur les valeurs des dissimilarités, la méthode MDS avec transformation non métrique (Schiffman, 1981) suppose que ce ne sont pas tant les dissimilarités (entre produits) qui sont importantes mais plutôt leurs ordres (relation d'ordre sous jacente aux valeurs données par les dissimilarités). Pour l'établissement de la carte factorielle, les dissimilarités peuvent être transformées pour autant que cela ne remette pas en cause la structure d'ordre entre ces dissimilarités. De manière formelle, la méthode MDS non métrique est basée sur la minimisation du critère : $Stress = \left[\sum_i \sum_j (f(\delta_{ij}) - d_{ij}(X))^2 / \sum_i \sum_j d_{ij}(X)^2 \right]^{1/2}$ où f est une fonction monotone croissante (à déterminer par l'algorithme de résolution) des dissimilarités δ_{ij} .

Ainsi, il est également possible de mettre en œuvre un algorithme de MDS non métrique, qui selon nous, est plus adapté aux données issues de l'épreuve de tri libre (Lawless et al, 1995 ; Faye

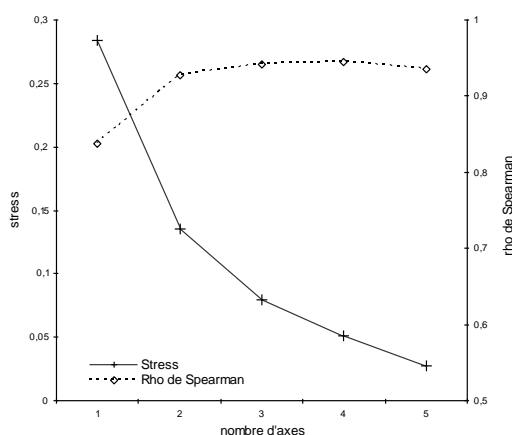
et al, 2004). En effet, les sujets forment des groupes produits en fonction des proximités perçues mais n'évaluent ni les distances entre produits ni les distances entre les groupes de produits sur une échelle de notation. D'un point de vue conceptuel, cette analyse suppose que seul l'ordre des dissimilarités reflète réellement la structure de données (Falahee et Mac Rae, 1997). En pratique, les résultats obtenus entre les deux approches (métrique et non métrique) sont souvent comparables.

Il est à noter que des logiciels standards (SAS, SPSS, Excel Stat) permettent de réaliser aussi bien les méthodes MDS métrique que non métrique.

Application de la MDS non métrique à l'étude de cas

Nous avons appliqué la méthode MDS non métrique sur les dissimilarités agrégées obtenues à partir des données de tri libre. La figure 3 présente l'évolution du stress et du rho de Spearman en fonction de la dimension de la configuration. Il apparaît qu'à partir de deux dimensions, la courbe d'évolution du stress marque une inflexion et le rho de Spearman semble atteindre un palier d'une valeur supérieure à 0.9. Cependant, la solution à deux dimensions présente un stress nettement supérieur à 10%. Avec un stress de 8%, la configuration de dimension 3 est retenue pour garantir un bon ajustement entre les dissimilarités initiales et les proximités entre produits.

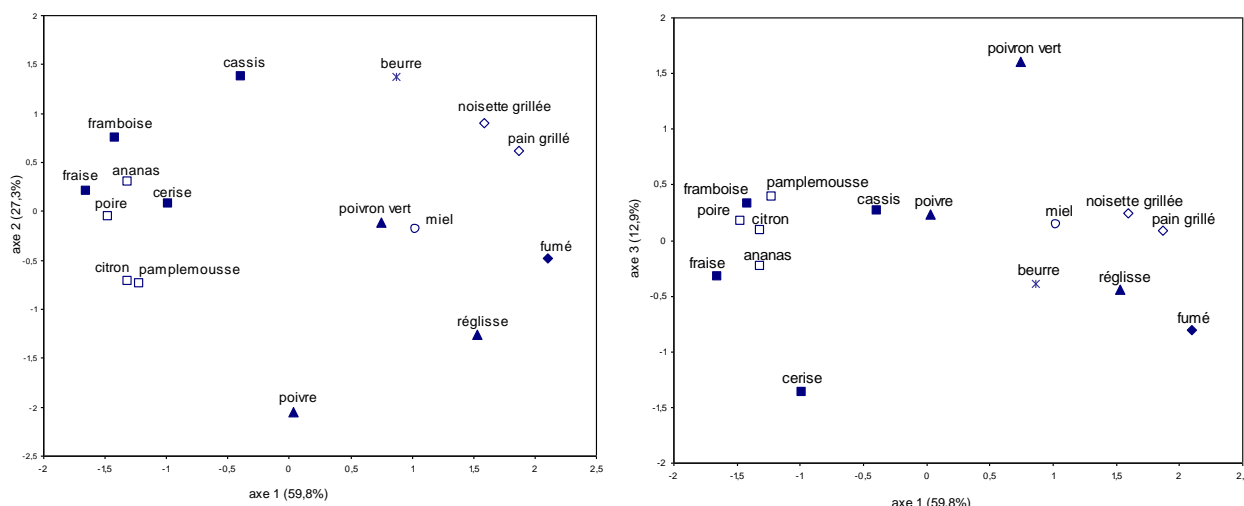
FIG. 3- Evolution du stress et du rho de Spearman en fonction de la dimension de l'espace de représentation (modèle MDS non métrique sur matrice de dissimilarités agrégées appliqué à l'étude de cas).



La figure 4 présente la configuration des produits sur la base des trois premiers axes de l'analyse MDS non métrique. Nous indiquons également sur cette figure l'importance relative des dispersions des produits sur les trois axes : l'inertie totale considérée, ici, est calculée sur la base des trois axes retenus.

Les arômes se répartissent en groupes assez nets sur le premier plan de la configuration. Le premier axe sépare les notes fruitées des notes grillées, florales, animales, végétales et épicées. Au niveau des notes fruitées, le deuxième axe permet d'isoler le « cassis » et les deux arômes agrumes, « citron » et « pamplemousse » des autres arômes fruités : « cerise », « framboise », « fraise », « ananas » et « poire ». Au niveau des autres notes, les arômes sont positionnés en quatre pôles. Le premier pôle est constitué de deux arômes de notes grillées, « noisette grillée » et « pain grillé », et du « beurre » qui se détache légèrement de ces derniers. Par ailleurs, les trois arômes des notes végétales, « poivre », « poivron vert » et « réglisse », sont répartis sur la partie inférieure de l'axe 2, le « réglisse » se trouvant à proximité du « fumé », le « poivron vert » du « miel » et le « poivre » étant isolé. L'axe 3 permet de dissocier spécifiquement la « cerise » des arômes « fraise », « framboise », « poire » et « ananas », le « poivron vert » du « miel ». Bien que proches sur le premier plan de la configuration, ces arômes sont séparés sur le troisième axe, ce qui indique qu'une partie des sujets les a positionnés dans des groupes différents (un regroupement consensuel de ces deux produits induirait une proximité graphique sur tous les axes de la configuration). Ceci confirme ce que nous avons constaté dans la figure 2.

FIG. 4 - Configuration des produits sur la base des axes 1-2 et 1-3 (modèle MDS non métrique sur matrices agrégées appliqué à l'étude de cas)



La méthode MDS métrique a également été appliquée à ce jeu de données (graphiques reportés en annexe). Il ressort que, pour un choix de dimension donné, la méthode métrique fournit des valeurs de stress plus élevées que l'algorithme non métrique (respectivement 25%, 17%, 11% et 8% pour deux, trois, quatre et cinq axes). Pour la solution de dimension 3, la structure de la configuration des arômes se recoupe dans une large mesure avec celle obtenue à l'aide de la MDS non métrique. Cependant, nous avons noté que les produits sont plus dispersés avec la MDS métrique et, de ce fait, il est plus difficile d'identifier des regroupements de produits aussi clairement que dans le cas de la MDS non métrique.

4.1.2. MDS sur matrices de dissimilarités individuelles (INDSCAL/ DISTATIS)

- INDSCAL

Une des limites de la démarche précédente basée sur MDS sur dissimilarités agrégées est de ne pas tenir compte des différences individuelles entre sujets (Lawless et al, 1995). La méthode MDS sur dissimilarités individuelles basée sur le modèle INDSCAL (Carrol et Chang, 1970), modèle à pondération individuelle, permet de pallier cette limite. Cette méthode permet de positionner les produits dans un espace multidimensionnel de faible dimension et de déterminer le poids que chaque sujet accorde à chacun des axes. Ce poids est directement lié à la dispersion des produits sur chaque axe. Si, comme cela est indiqué ci-dessus, les axes factoriels reflètent des variables latentes qui structurent l'espace perceptif, l'introduction de poids permet de tenir compte du fait que les sujets ne perçoivent pas nécessairement ces variables latentes avec la même intensité.

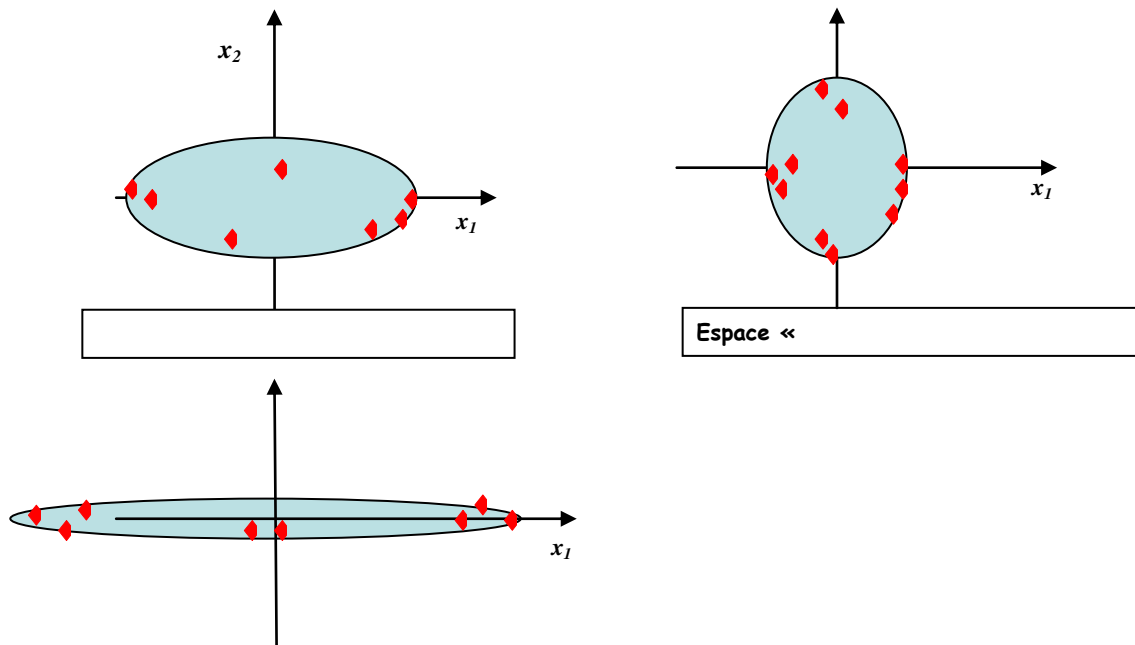
Soit X l'espace de représentation, nous désignons par $d_{ij}^{(h)}(X)$ la distance entre les produits i et j pour le sujet h dans cet espace. Soit x_{ia} la coordonnée du produit i sur l'axe a de X et $w_a^{(h)}$ le poids du sujet h sur cet axe. L'espace de représentation optimal est obtenu par minimisation du Stress défini par:

$$\text{Stress} = \left[\sum_h \sum_{ij} (\delta_{ij}^{(h)} - d_{ij}^{(h)}(X))^2 / \sum_h \sum_{ij} d_{ij}^{(h)}(X)^2 \right]^{1/2} \text{ avec } d_{ij}^{(h)}(X) = \sqrt{\sum_a w_a^{(h)} (x_{ia} - x_{ja})^2}$$

Le problème de minimisation associé au modèle INDSCAL est résolu de manière itérative (Carrol et Chang, 1970).

Comme cela est schématisé sur la figure 5, le modèle INDSCAL procure une représentation des produits qui est, selon les sujets, allongée ou contractée le long de chaque axe. Ce modèle procure également une représentation des sujets en fonction des poids qu'ils affectent aux différents axes. A titre d'illustration, cette figure représente les configurations des produits de trois sujets typiques ainsi que les poids de ces sujets pour les deux axes.

FIG. 5 – Illustration de la représentation des produits pour trois sujets typiques 1, 2 et 3 et des poids des trois sujets sur les axes



S3

Il est clair que le modèle INDSCAL donne des informations utiles sur le positionnement des produits les uns par rapport aux autres et permet également de mettre en évidence des différences entre les sujets ce qui, comme nous l'avons souligné ci-dessus, peut être un objectif en soi pour une épreuve de catégorisation.

Cependant, il faut souligner que son aspect métrique nous semble peu adapté aux données issues du tri libre comme cela a été discuté ci-dessus. C'est pourquoi nous préconisons d'adopter une version non métrique de l'approche INDSCAL. A notre connaissance, cette approche tout à fait adaptée aux données issues d'une épreuve de tri libre n'a jamais été utilisée en évaluation sensorielle.

La méthode INDSCAL avec transformation non métrique développée dans le contexte de la psychométrie par Takane et al (1977) est basée sur la minimisation du critère suivant :

$$Stress = \left[\sum_h \sum_{ij} (f^{(h)}(\delta_{ij}^{(h)}) - d_{ij}^{(h)}(X))^2 / \sum_h \sum_{ij} d_{ij}^{(h)}(X)^2 \right]^{1/2} \text{ avec } d_{ij}^{(h)}(X) = \sqrt{\sum_a w_a^{(h)} (X_{ia} - X_{ja})^2}$$

où $f^{(h)}$ ($h=1, \dots, H$) sont des fonctions (à déterminer) monotones et croissantes.

MDS INDSCAL métrique et non métrique sont disponibles dans les logiciels SAS et SPSS.

- DISTATIS

Dans cette partie concernant le traitement des dissimilarités individuelles par une approche s'apparentant à MDS, nous pouvons citer la démarche DISTATIS proposée par Abdi et al (2007). Cette méthode a été conçue comme « une généralisation de la méthode MDS au cas de dissimilarités organisées en trois voies », c'est-à-dire en multi-tableaux (soit ici N matrices $P \times P$). Elle a été appliquée au cas spécifique des données issues du tri libre (Abdi et al, 2007). DISTATIS consiste à calculer les matrices de produits scalaires (formes de Torgerson) associées aux dissimilarités individuelles des sujets. Par la suite, ces matrices sont soumises à la méthode STATIS (Lavit, 1988). Ceci passe par le calcul des coefficients RV qui reflètent l'accord mutuel des sujets sur la configuration des produits. Une configuration compromis est obtenue tenant compte de l'accord de chaque sujet avec l'ensemble du panel en ce sens qu'un sujet a d'autant plus de contribution dans la définition du compromis qu'il est en accord avec le panel. Malgré son caractère

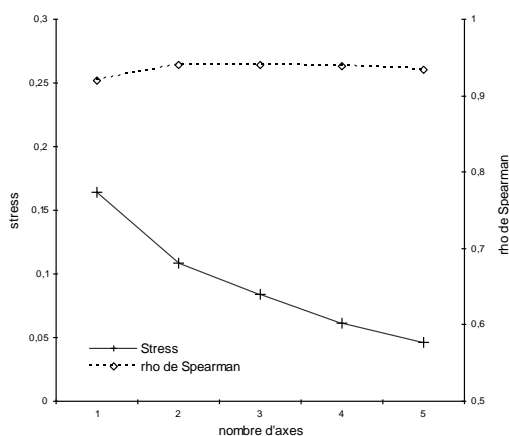
métrique peu adapté aux données de tri libre, la méthode DISTATIS présente l'avantage d'apporter un éclairage sur l'accord des sujets (coefficients RV).

En ce qui concerne la détermination du compromis par pondération des sujets en fonction de leur accord global, nous pouvons souligner que le modèle INDSCAL permet davantage de flexibilité. En effet, INDSCAL tient compte dans la détermination de l'espace de représentation de l'accord et des différences entre les sujets sur chaque axe de représentation alors que, DISTATIS tient compte de l'accord global (pour toutes les dimensions) entre les sujets.

Application de la MDS INDSCAL pour l'étude de cas

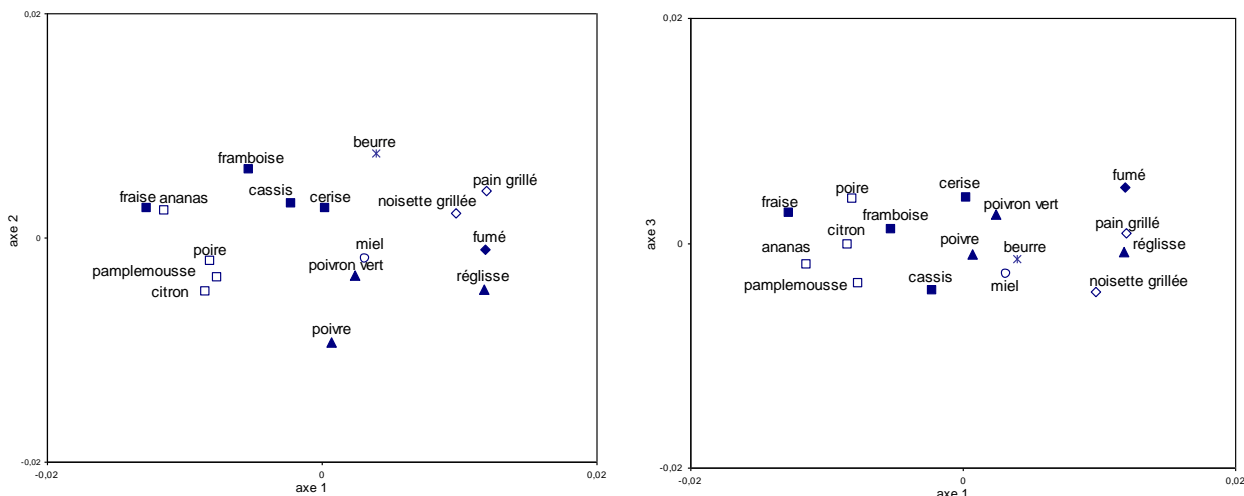
Nous avons appliqué la méthode MDS non métrique sur les dissimilarités individuelles obtenues à partir des données de tri libre des seize arômes de vin. D'après la figure 6, l'évolution du stress marque une décroissance moins importante et le rho de Spearman présente des valeurs stables proches de 0.92 dès la deuxième dimension. Cependant, bien que ne réalisant qu'un faible gain, le troisième axe permet d'obtenir un stress inférieur à 10%, seuil considéré comme acceptable en termes d'ajustement du modèle.

FIG. 6 - Evolution du stress et du rho de Spearman en fonction de la dimension de l'espace de représentation. (modèle INDSCAL appliqué à l'étude de cas).



La figure 7 présente la configuration des produits sur la base des trois premiers axes de l'analyse MDS non métrique sur matrices de dissimilarités individuelles. Sur chaque axe, nous avons choisi de représenter les produits de manière à ce que la dispersion corresponde à la moyenne des dispersions données par les sujets (moyenne des poids $w_a^{(h)}$, $h=1...H$, pour l'axe a). Pour chaque sujet, cette configuration sera allongée ou contractée horizontalement ou verticalement en fonction des poids spécifiques associés au sujet considéré.

FIG. 7 - Configuration des produits sur les axes 1-2 et 1-3 (modèle INDSCAL appliqué à l'étude de cas).

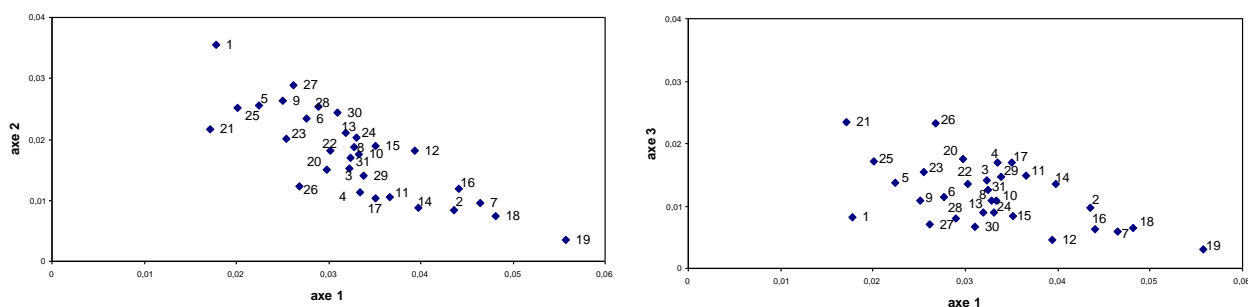


La structure de la configuration et le positionnement relatif des arômes sont globalement similaires à ce que nous avons obtenu à l'issue de la méthode MDS non métrique sur dissimilarités agrégées, notamment sur le premier plan. En effet, le premier axe sépare les notes fruitées des notes grillées, florales, animales, végétales et épicées. La principale différence réside dans la dispersion des arômes des notes fruitées sur le premier plan et de tous les arômes sur le

deuxième plan de la configuration. Ainsi, quatre pôles peuvent être identifiés : (« poire », « pamplemousse », « citron »), (« fraise », « ananas »), (« framboise »), (« cassis », « cerise »). Le troisième axe permet de dissocier la « cerise » du « cassis » et la « poire » du « citron » et du « pamplemousse ». Le positionnement relatif des autres notes sur le premier plan est assez similaire à celui obtenu sur dissimilarités agrégées : le « poivron » est isolé, le « poivron vert » et le « miel » sont proches ainsi que le « réglisse » et le « fumé ». Notons que le « beurre » est isolé des deux notes grillées, « pain grillé » et « noisettes grillées ». Le troisième axe permet de dissocier le « poivron » du « miel ».

La figure 8 représente les poids des sujets associés aux trois axes et reflète les différences entre sujets quant à la « perception » des variables latentes sous jacentes aux axes. A titre d'exemple, il apparaît clairement que les sujets 1, 19, 21 et 26 se détachent des autres sujets. En effet, la configuration du sujet 19 semble être principalement basée sur le premier axe alors que le sujet 1 semble disposer les produits sur le deuxième axe et les sujets 21 et 26, sur le troisième axe.

FIG. 8 - Poids des sujets sur les axes 1-2 et 1-3



4.1.3. Synthèse sur les méthodes apparentées à la MDS

En ne considérant que l'ordre des proximités entre produits, la transformation non métrique permet d'apporter davantage de flexibilité dans l'interprétation des occurrences collectées et de se prémunir du codage plus ou moins arbitraire des dissimilarités par 0 et 1. Ainsi, il apparaît que, d'un point de vue conceptuel, cette hypothèse est cohérente avec la procédure de tri libre et la tâche de catégorisation. Ce point a été souligné de manière claire par Poitevineau (in Dubois, 2009) qui considère que la prise en compte de l'ordre et non des distances est « *une des raisons du succès de ces dernières méthodes (i.e. MDS non métriques) dans des domaines où la connaissance des propriétés formelles des objets est encore fragmentaire ou d'une grande complexité (en particulier les sciences humaines et sociales).* ».

Par ailleurs, les modèles MDS-INDSCAL présentent l'avantage de considérer les données individuelles et de prendre en compte la diversité des sujets.

4.2. Les méthodes apparentées à l'Analyse des Correspondances Multiples (ACM)

4.2.1. L'Analyse des Correspondances Multiples

L'analyse des correspondances multiples (ACM) est une méthode descriptive d'analyse factorielle multivariée. Généralisation de l'analyse des correspondances simples développée par Benzécri (in Lebart et al, 1997), l'ACM représente pour les variables qualitatives ce que l'analyse en composantes principales représente pour les variables quantitatives (Greenacre, 1993). Elle consiste à réaliser une analyse des correspondances simples sur le tableau disjonctif complet Z , concaténation des tableaux disjonctifs individuels $Z^{(h)}$ de dimension $(P \times K^{(h)})$ (cf § 3.1). Il en résulte une configuration X , constituée d'axes mutuellement orthogonaux. Pour l'interprétation des données, il est usuel de ne retenir que les premiers axes de la configuration, le nombre d'axes retenu étant déterminé en considérant les deux objectifs suivants : synthétiser l'information initiale (nombre d'axes restreint) et restituer suffisamment d'information (nombre d'axes pouvant être élevé). Une des caractéristiques fondamentales de l'ACM est l'utilisation de la distance de Chi² pour calculer les distances entre modalités (groupes formés par les sujets) et individus (produits). Ainsi, la distance entre deux produits dépend non seulement de leur co-occurrence dans les groupes formés par les sujets mais également de la taille de ces groupes. L'idée de base étant que deux

produits sont réunis dans des groupes de petites tailles sont plus proches entre eux que s'ils étaient réunis dans des groupes de plus grandes tailles. Ils seront donc d'autant plus éloignés des produits des autres groupes. Cela signifie, qu'en mettant des produits dans des groupes isolés (de petites tailles) les sujets les ont identifiés comme spécifiques et très nettement différents des autres.

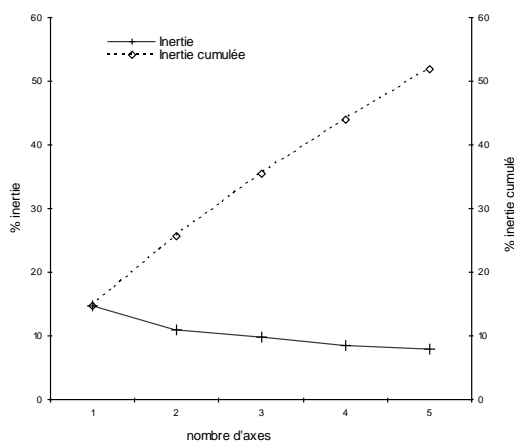
L'utilisation de l'ACM dans le cadre d'une épreuve de catégorisation a été préconisée par Van der Kloot et Van Herk (1991). Cependant, il faut souligner que Takane (1981) a développé une procédure pour le traitement statistique des données de tri libre, appelée MDSORT. Celle-ci consiste à représenter, pour chaque sujet h , les produits dans un espace de dimension fixée a priori à partir du tableau disjonctif complet $Z^{(h)}$ associé à ce sujet. Ainsi, un même produit est représenté autant de fois qu'il y a de sujets. Le principe de MDSORT consiste à déterminer un espace de représentation de manière à ce que l'inertie intra-produits soit la plus faible possible ou, ce qui revient au même, l'inertie inter-produits soit la plus grande possible. En examinant la solution donnée par Takane (1981), il est facile de constater, qu'en définitive, ceci conduit à l'ACM, comme cela a été souligné par plusieurs auteurs (Van der Kloot et Van Herk, 1991 ; Qannari et al, 2009).

Nous pouvons également citer FAST (Factoriel Approach for Sorting Task data) développée par Cadoret et al (2009) comme méthode d'analyse de données de tri libre apparentée à l'ACM. Outre la configuration « produits » issue de l'ACM, cette méthode permet de construire des ellipses de confiance autour des produits par la technique de rééchantillonnage (Boostraping). De plus, les proximités entre les sujets sont représentées dans un espace multidimensionnel, issu de l'Analyse Factorielle Multiple (AFM) des tableaux disjonctifs complets individuels $Z^{(h)}$. Le lien entre les deux représentations (produits et sujets) est assuré par le fait que, dans le cas de tableaux disjonctifs complets, les axes obtenus par l'AFM et l'ACM sont identiques à un facteur près.

Application de l'ACM à l'étude de cas

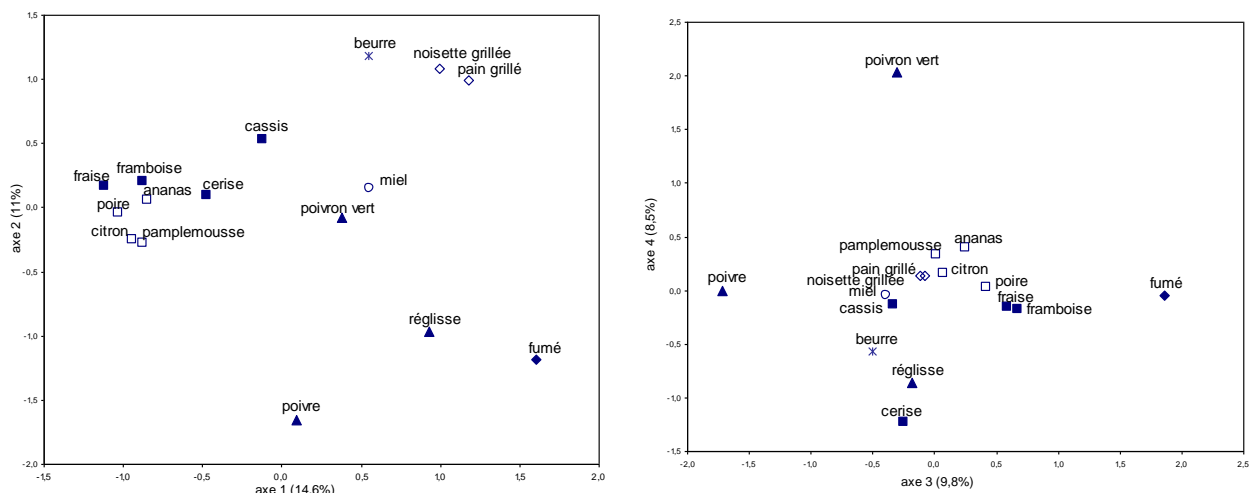
Nous avons appliqué l'ACM sur les dissimilarités agrégées obtenues à partir des données de tri libre des seize arômes de vin. La figure 9 présente l'inertie restituée par chacun des axes ainsi que l'inertie cumulée. On constate que l'inertie restituée par les premiers axes est assez faible. Ce phénomène s'observe fréquemment dans le cadre de l'ACM. Cependant, un palier semble apparaître entre la quatrième et la cinquième dimension. Avec 44% d'inertie restituée et une qualité de représentation considérée comme acceptable, la configuration de dimension 4 a été retenue.

FIG 9. Evolution de l'inertie et de l'inertie cumulée en fonction du nombre d'axes factoriels (ACM appliquée à l'étude de cas)



La figure 10 présente la configuration des produits sur la base des axes factoriels 1 et 2 puis 3 et 4. La configuration des produits sur le premier plan factoriel est assez proche de celle obtenue avec MDS sur matrice de dissimilarités agrégées. Cependant, les arômes sont regroupés en pôles moins nombreux que sur les configurations MDS. On retrouve la séparation entre notes fruitées et autres notes sur le premier axe. Sur le premier plan factoriel, les notes fruitées ne semblent former qu'une seule catégorie, mis à part le cassis qui se détache. En effet, à la différence des configurations MDS, les arômes agrumes, « citron » et « pamplemousse », sont associés aux autres notes fruitées. Par ailleurs, le « poivre » est isolé sur le deuxième axe, et on retrouve les proximités entre « poivron » et « miel », « réglisse » et « fumé » et « beurre », « noisette grillée » et « pain grillé ». Le troisième axe factoriel isole également le « poivre » mais sépare aussi le « fumé » du « réglisse ». Le quatrième axe permet de dissocier la « cerise » des autres notes fruitées ainsi que le « poivron vert » du « miel ». Ce dernier axe peut être rapproché du troisième axe procuré par MDS sur matrice de dissimilarités agrégées.

FIG. 10 - Configuration des produits sur les axes 1-2 et 3-4 (ACM appliquée à l'étude de cas)



Notons, en particulier, les positions extrêmes du « fumé » et du « poivre » sur l'axe 3. Comme le montrait la figure 2, le « fumé » et le « poivre » sont les arômes les plus isolés par les sujets. De ce fait, leur distance aux autres arômes, calculée à l'aide de la distance de χ^2 , s'en trouve d'autant plus importante.

4.2.2. Analyse du tableau disjonctif complet avec pondérations individuelles

Takane, après avoir proposé en 1981 la méthode MDSORT qui comme, nous l'avons vu, a des liens de parenté très forts avec l'ACM, a également proposé la méthode IDSORT (Takane, 1982). L'idée consiste ici à considérer, pour chaque sujet, la matrice de dissimilarités calculées sur la base d'une distance de χ^2 , à partir des tableaux disjonctifs complets individuels. Par la suite, ces matrices de dissimilarités sont soumises au programme INDSCAL. L'intérêt par rapport à MDSORT (et donc l'ACM) est de tenir compte des différences individuelles des sujets et d'exhiber des poids qui reflètent ces différences.

Dans la même veine, Qannari et al (2009) ont proposé une méthode appelée CCSort qui s'insère dans le cadre de l'analyse multi-tableaux. CCSort consiste, à partir des tableaux disjonctifs individuels $Z^{(h)}$, d'opérer une standardisation de chaque colonne par division par la racine carrée de son effectif marginal, ce qui induit la métrique de χ^2 entre les produits. La méthode CCSort préconise également d'effectuer une pondération à l'échelle de chaque tableau afin de tenir compte du nombre de groupes réalisés par chaque sujet. Par la suite, les tableaux ainsi obtenus sont soumis à l'Analyse en Composantes Communes et Poids Spécifiques (Hanafi et Qannari, 2008). Tout comme l'analyse par pondérations individuelles MDS-INDSCAL, cette méthode permet de déterminer les composantes sous-jacentes à l'espace de représentation ainsi que les poids associés pour chaque sujet à chaque dimension. Cependant, comme la méthode CCSort procède de manière séquentielle, les solutions correspondant à des dimensions différentes sont emboîtées, contrairement à la méthode MDS-INDSCAL. Ceci présente un avantage en termes de facilité d'interprétation mais présente également un inconvénient en termes de qualité de représentation.

4.2.3. Synthèse sur les méthodes apparentées à l'ACM

Il est important de rappeler que les méthodes apparentées à l'ACM sont basées sur l'utilisation de la distance du χ^2 qui prend en compte l'effectif des groupes dans le calcul des proximités. S'il est effectivement prouvé que cette métrique présente des propriétés mathématiques intéressantes, il n'en demeure pas moins vrai qu'elle peut poser des problèmes dans le cas de faibles effectifs. C'est le cas pour l'épreuve du tri libre lorsque les sujets considèrent qu'un ou deux produits, par exemple, forment une catégorie à part. Ceci est relativement courant dans ce type d'épreuves et peut avoir une incidence importante sur la détermination des cartes factorielles et des distances entre produits.

4.3. Synthèse sur les méthodes factorielles

Outre le recours à la distance de Chi^2 pour les méthodes apparentées à l'ACM, la différence entre les deux familles de méthodes factorielles concerne le caractère séquentiel et le caractère global dans la détermination des solutions. De manière plus précise, nous avons vu que dans l'ACM, tout comme dans CCSort, les solutions correspondant à des dimensions croissantes sont emboîtées (une solution de dimension donnée contient toutes les solutions de dimensions inférieures). Il est clair que les méthodes conduisant à des solutions emboîtées procurent une plus grande facilité d'interprétation des résultats. Cependant, elles peuvent conduire à un choix d'un espace de représentation de dimension relativement grande ; chaque axe supplémentaire donnant des éclairages complémentaires par rapport aux autres axes ou, séparant des produits qui n'étaient pas séparés précédemment. A contrario, les méthodes apparentées à MDS procurent la meilleure solution pour un choix de dimension donné, un peu comme si cette solution devait se suffire à elle-même. Comme conséquence, nous pouvons raisonnablement penser que, pour une dimension fixée, les éléments les plus probants ressortent davantage dans une démarche de type MDS que dans une démarche de type ACM.

Par ailleurs, nous pouvons souligner que les méthodes à pondérations individuelles de type INDSCAL (incluant la méthode CCSort), permettent de déterminer des poids spécifiques pour les sujets qui reflètent l'importance que ces sujets attachent à chacun des axes factoriels. Il est vrai que des poids qui reflètent la même idée peuvent être calculés, a posteriori, à partir de l'ACM. Cependant, l'avantage des méthodes apparentées à INDSCAL est de les intégrer explicitement dans le modèle et d'en tenir compte pour la détermination des paramètres du modèle.

5. Les méthodes de classification

Les méthodes de classification permettent de représenter les proximités entre les produits sous forme d'arbres additifs ou, pour la classification hiérarchique, de dendrogrammes (arbres hiérarchiques). Les méthodes de partitionnement permettent de procurer directement des groupes de produits. Dans le cas des données de catégorisation, l'hypothèse de base consiste à considérer qu'il existe une partition latente que chaque sujet essaie de reconstituer. Il s'agit, par conséquent, de déterminer cette partition latente également appelée partition compromis. Nous verrons que la représentation sous forme d'arbres additifs est motivée par l'idée qu'il existe des classes de produits agrégées autour de produits typiques. Comme pour les méthodes factorielles, les méthodes de classification peuvent être appliquées sur les matrices de dissimilarités individuelles ou sur la matrice de dissimilarités agrégées.

5.1. Classification sur tableaux de données agrégées

5.1.1. Classification ascendante hiérarchique

Les méthodes de classification ont pour objectif de regrouper des objets en un nombre restreint de classes homogènes (Lebart et al, 2006). Pour cela, la classification ascendante hiérarchique (CAH) procède à des regroupements par étapes successives. A chaque étape, les deux produits, ou les deux classes de produits, les plus proches, selon un critère d'agrégation à préciser, sont combinés pour former une nouvelle classe. Ainsi au niveau le plus bas, les P produits définissent chacun une classe. Au niveau le plus élevé, il n'existe plus qu'une classe regroupant tous les produits. Les regroupements successifs sont schématisés sous forme d'un arbre hiérarchique ou dendrogramme, de sorte que la hauteur d'un palier reflète une mesure de dissemblance entre les classes. La coupure de l'arbre à un niveau donné permet d'obtenir des partitions d'objets plus ou moins fines selon que le nombre de classes est élevé ou pas.

Le critère le plus couramment utilisé en classification ascendante hiérarchique est le critère de Ward. A chaque étape, nous cherchons à minimiser l'inertie intra-classes (ou, de manière équivalente, maximiser l'inertie inter-classes). Le critère d'agrégation est, par conséquent, égal à la perte d'inertie inter-classes résultant du regroupement de deux classes. La méthode de Ward a

tendance à fournir des classes sphériques et équilibrées en effectifs ce qui n'est pas toujours conforme à la structure des données analysées.

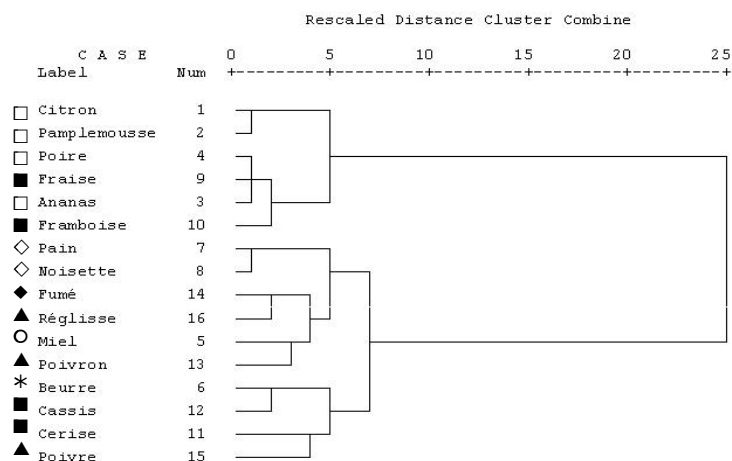
En réalisant une classification ascendante hiérarchique sur la matrice de dissimilarités agrégées déterminées à partir des données issues d'un tri libre, il est clair que nous conférons aux dissimilarités mesurées un caractère métrique, ce qui, comme nous l'avons déjà souligné, présente quelques limitations. Afin de contourner cette difficulté, Lawless (1989) a suggéré de réaliser une classification ascendante hiérarchique sur les coordonnées factorielles obtenues à partir d'une MDS non métrique. Par ailleurs, Coxon (1999) s'interroge sur le critère d'agrégation le plus adapté pour des données de tri. Il suggère à l'expérimentateur de générer plusieurs arbres sur le même jeu de données en faisant varier les critères (critère de Ward, saut minimum, saut maximum). Ceci permet, d'une part, d'identifier les similitudes et les différences entre les solutions et, d'autre part, de retenir le résultat qui semble le mieux interprétable. Il propose également de recourir à une solution intermédiaire consistant à se baser sur le critère d'agrégation de la médiane. D'autres auteurs ont suggéré que le critère du saut minimum était intéressant du fait de la propriété stipulant que l'arbre hiérarchique formé est invariant par toute transformation monotone qui laisserait invariant l'ordre des dissimilarités (Anderberg, 1973 ; page 138). Cette propriété confère à la stratégie de classification sur la base du saut minimum un caractère non métrique.

Application de la Classification Ascendante Hiérarchique à l'étude de cas

Nous avons appliqué la classification ascendante hiérarchique sur la matrice de dissimilarités agrégées obtenues à partir des données de tri libre des seize arômes de vin. La figure 11 présente le dendrogramme obtenu en utilisant le critère d'agrégation de Ward et la distance euclidienne usuelle.

La coupure en deux classes dissocie clairement une partie des arômes fruités (hors « cerise » et « cassis ») des autres notes. Cette dissociation correspond globalement au positionnement des arômes sur le premier axe des configurations obtenues grâce aux méthodes factorielles présentées précédemment. La coupure en trois classes permet de séparer le « beurre », le « poivre » et les deux notes fruitées, « cassis » et « cerise », des notes grillées, du « fumé », du « réglisse », du « poivron » et du « miel ». Notons que cette coupure correspond au positionnement des produits sur le deuxième axe factoriel des configurations présentées précédemment, les arômes « beurre », « cerise », « cassis » et « poivre » étant positionnés sur les deux parties extrêmes de cet axe, les autres arômes étant plus centraux.

FIG. 11 - Dendrogramme avec le critère de Ward (CAH appliquée à l'étude de cas)



La coupure en six classes permet d'isoler au sein des notes fruitées, les deux arômes agrumes des arômes « poire », « fraise », « ananas » et « framboise », les deux arômes grillés, « pain grillé » et « noisette grillée », des arômes « fumé », « réglisse », « miel » et « poivron » et enfin, le « beurre » et le « cassis » de la « cerise » et du « poivre ». Ces résultats sont cohérents avec les proximités observées entre produits dans la configuration issue de la MDS non métrique.

5.1.2. Arbres additifs

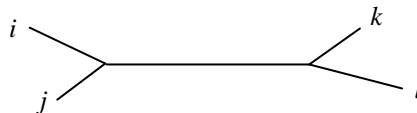
Cette méthode permet de représenter les proximités entre objets sous forme d'un X-arbre valué à distance additive composé de sommets correspondant aux objets (i.e. produits) à classer et de noeuds internes (ou virtuels), reliés par des arêtes. La longueur du chemin entre deux objets

(somme des longueurs des arêtes) correspond à la dissimilarité entre ces deux objets (Sattah et Tversky, 1977 ; Barthélémy in Dubois, 1991 ; Poitevineau, 2002). Un arbre à distance additive n'est donc pas une hiérarchie mais un graphe (Coxon, 1999) ou une topologie (David et Rouby in Poitevineau, 2002), la dénomination « X »-arbre référant à la forme de sa structure. L'algorithme le plus répandu pour réaliser les arbres est Addtree, proposé par Sattah et Tversky (1977) sur la base du modèle de contraste de Tversky.

Afin d'évaluer la qualité de la représentation, il est usuel d'utiliser un critère métrique tel que le stress de Kruskal (Poitevineau, 2002). La spécificité de cette méthode est l'optimisation d'un critère topologique « condition des quatre points », visant à minimiser le nombre de quadruplets mal représentés. Coxon (1999) précise qu'il s'agit d'une généralisation de l'inégalité ultramétrique. Cette condition signifie que deux objets rattachés au même nœud interne sont plus proches que deux objets rattachés à des nœuds internes différents. Elle est représentée dans la figure 12 et peut être énoncée comme suit (Barthélémy, 1991) : si E est l'ensemble des objets à classer, d est une distance arborée sur E si et seulement si pour tout $i, j, k, l \in E$, nous avons :

$$d(i, j) + d(k, l) \leq \max(d(i, k) + d(j, l), d(i, l) + d(j, k))$$

FIG. 12 - Condition des 4 points



Barthélémy (1991) précise que « le modèle de contraste (de Tversky), pour autant qu'on y adhère, permet d'interpréter ces nœuds virtuels comme des entités qui partagent avec un groupe d'objets les caractères communs aux membres de ce groupe ». Rappelons que Rosch (1975) stipule que la catégorisation d'objets naturels ou artificiels se base sur un critère de typicalité c'est-à-dire sur un critère de ressemblance des exemplaires de la catégorie à un prototype ou un air de famille. La structure de l'arbre rend compte des deux notions essentielles sous-jacentes à cette théorie, à savoir :

- l'air de famille qui se situe au niveau des nœuds,
- la ressemblance qui se traduit par la métrique de l'arbre.

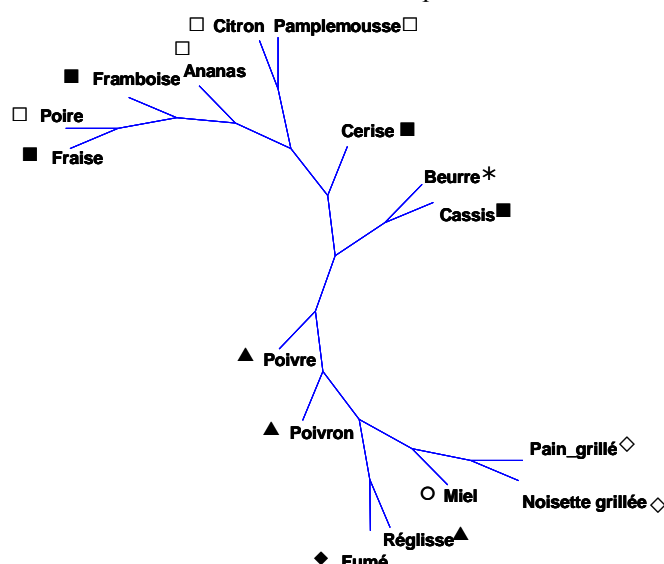
Le critère topologique, évaluant la qualité de la structure du graphe, le plus souvent utilisé sur la base de la condition des quatre points est le taux de quadruplets bien représentés. D'autres critères de ce type peuvent être calculés. Pour plus de détails, le lecteur peut se référer à David et Rouby (in Poitevineau, 2002) et à Poitevineau (2009).

Le logiciel T-Rex® (<http://www.labunix.uqam.ca/~makarenv/trex.html>) en diffusion libre permet de construire les arbres additifs.

Application des arbres additifs à l'étude de cas

Nous avons appliqué la méthode des arbres additifs sur les dissimilarités agrégées obtenues à partir des données de tri libre des seize arômes de vin. La figure 13 présente l'arbre obtenu à l'aide du logiciel T-Rex®.

FIG. 13 - Arbre à distance additive pour l'étude de cas



Comme précédemment, la dissociation entre les notes fruitées, dans la partie supérieure de l'arbre, et les autres notes, dans la partie inférieure, est bien évidente. Au niveau des notes fruitées, nous retrouvons les quatre pôles : (« poire », « fraise », « framboise », « ananas »), (« citron », « pamplemousse »), (« cerise ») et le « cassis » ici associé à l'arôme « beurre », résultat similaire à celui obtenu par classification ascendante hiérarchique. La distinction entre « poivre », « poivron », « cerise » et « cassis », correspondant au deuxième axe des configurations obtenues à l'aide des méthodes factorielles est visible au centre de l'arbre. Chacun de ces arômes est isolé de manière assez spécifique. Dans la partie inférieure, notons la proximité entre « fumé » et « réglisse » mais également entre « pain grillée » et « noisettes grillées », ces arômes ayant un nœud commun qui reflète la présence d'un air de famille. Notons également le positionnement du « miel », qui se démarque ici du « poivron vert » et semble plus proche des deux notes grillées, résultat que nous n'avons pas trouvé à l'issue de la classification ascendante hiérarchique. Cependant, il est difficile de définir un nombre de catégories sur la base de cette représentation. En effet, aucun critère, même qualitatif, ne permettent d'en juger. Il n'en demeure pas moins vrai que la structure de l'arbre aboutit à une représentation cohérente qui permet d'interpréter les proximités entre produits.

Les résultats des arbres sont cohérents avec ceux de la classification ascendante hiérarchique mais également avec les proximités observées dans les configurations factorielles. Cependant, l'arbre semble prendre plus en compte toute la structure des données dans la représentation et pas seulement les distances correspondant aux plus « grandes dispersions » (cas du miel et poivron vert). Les deux méthodes se basant sur la matrice de dissimilarités agrégées et étant métriques, les différences observées peuvent s'expliquer par la prise en compte des critères topologiques dans les arbres additifs.

5.2. Les méthodes de partitionnement

La méthode de partition centrale consiste à identifier une partition compromis à partir d'un ensemble de partitions (i.e. les partitions réalisées par les sujets). L'avantage de cette méthode est d'obtenir des résultats sous une forme proche des observations (ensemble de classes d'objets ou de produits).

Soit $\Delta^{(h)}$, la matrice de dissimilarités associée à l'individu h et $C=(c_{ij})$ la matrice de dissimilarités de la partition centrale. C est déterminée de manière à minimiser le critère suivant :

$$\sum_h \sum_{i,j} (\delta_{ij}^{(h)} - c_{ij})^2$$
 avec la contrainte que pour tout triplet d'objets (produits) (i, j, k) , C représente

une partition, soit :

- $c_{ij}=0$ ou 1
- $c_{ii}=1$ pour tout i (réflexivité)
- $c_{ij}=c_{ji}$ pour tout i,j (symétrie)
- $c_{ij} + c_{jk} - c_{ik} \leq 1$ (transitivité)

Le critère à minimiser est, de fait, la somme des désaccords des partitions individuelles avec la partition centrale (Brenac in Poitevineau, 2002). Pour un nombre de groupes donné, un algorithme d'optimisation (Gordon et Vichi, 1998) permet de déterminer itérativement la partition centrale.

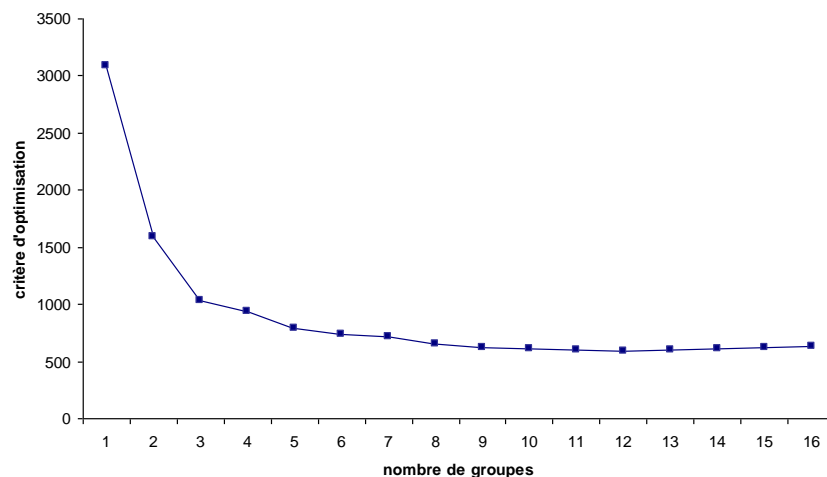
L'évolution du critère en fonction du nombre de groupes permet d'aider l'utilisateur dans le choix d'une partition appropriée.

En pratique, toutes choses étant égales par ailleurs, le nombre de groupes de la partition compromis augmente avec le désaccord entre les sujets. Ainsi, la pertinence d'une partition compromise unique est limitée en cas de fort désaccord entre les sujets. Cependant, nous pouvons retenir que lorsque le nombre de groupes de la partition compromise est bien supérieur à la moyenne des nombres de groupes réalisés par les sujets, cela peut être considéré comme un bon indicateur d'un grand désaccord entre les sujets.

Application du partitionnement à l'étude de cas

Sur les données de tri libre des seize arômes de vin, nous avons déterminé la partition centrale à partir des partitions individuelles des sujets. La figure 14 représente l'évolution du critère d'optimisation en fonction du nombre de groupes de la partition centrale. Le critère est minimum pour huit groupes. Ce nombre est supérieur au nombre moyen de groupes, qui pour mémoire est de 6.2, ce qui semble indiquer un certain désaccord entre les sujets sur les regroupements des produits qu'ils ont formés.

FIG. 14 - Evolution du critère d'optimisation en fonction du nombre de groupes de la partition centrale (application du partitionnement à l'étude de cas)



La partition centrale retenue se compose des groupes suivants : (« citron », « pamplemousse ») (« ananas », « poire », « fraise », « framboise ») (« miel », « pain grillé », « noisette grillée ») (« beurre », « cassis ») (« fumé », « réglisse ») (« cerise ») (« poivron vert ») (« poivre »). Cette partition est cohérente avec les proximités observées sur l'arbre additif (figure 13). Elle présente quelques différences avec la coupure en huit classes de la classification ascendante hiérarchique : le « miel » et le « poivron vert » préalablement associés dans le même groupe, sont ici séparés, l'arôme « miel » étant associé aux deux notes grillées, « pain grillé » et « noisettes grillées ».

Notons qu'il existe une méthode alternative appelée « analyse en partitions latentes » (Wiley, 1967) que nous ne développerons pas ici. L'objectif est de fournir une partition en k classes d'objets, le nombre k étant fixé a priori. Cette démarche suppose l'existence d'une partition latente sous-jacente aux partitions réalisées par les sujets lors de l'expérimentation. La technique consiste à calculer de manière itérative la matrice de probabilités qui lie les partitions formées par les sujets à la partition latente. Pour plus de détails sur cette méthode, nous suggérons de consulter l'article de Wiley (1967).

5.3. Synthèse des méthodes de classification

Les arbres additifs semblent bien adaptés aux données de tri libre car ils permettent de représenter à la fois les distances entre les objets (i.e. les produits) et la structure des liens entre eux. Bien que les distances entre les objets soient calculées sur la base de leurs cooccurrences, les critères topologiques qui permettent de réaliser l'arbre additif intègrent une condition non métrique. Cependant, nous pouvons noter qu'à notre connaissance il n'existe pas de critère décisionnel qui permet de déterminer le nombre de classes à retenir.

Les méthodes de partitionnement présentent l'avantage de fournir des partitions d'objets, résultats similaires à l'information collectée à l'issue de la tâche de tri libre. Cependant, elles ne permettent pas d'avoir une vision globale sur la structure de proximité entre les classes de la partition obtenue. La nécessité de fixer le nombre de groupes a priori présente également un inconvénient bien qu'il existe des procédures pour aider l'utilisateur dans ce choix, telle que celles proposées par Cariou et al (2009).

Notons, qu'il est également possible de réaliser une classification des sujets, permettant de mettre en évidence des groupes présentant des perceptions différentes. Cette dernière démarche, classiquement réalisée dans le cadre de tests consommateurs, dépasse le cadre du présent papier.

6. Conclusion

Dans cet article, les méthodes de traitement des données de catégorisation ont été divisées en deux familles : méthodes factorielles et méthodes de classification. Comme le précise Coxon (1999), ces deux familles de méthodes répondent à des objectifs différents et, de ce fait, sont complémentaires:

- dans une approche plus analytique des produits, les méthodes factorielles permettent d'identifier les distances entre les produits ainsi que les caractéristiques qui les différencient,
- dans une approche plus holistique, les partitions permettent de délimiter les catégories et les arbres d'identifier l'organisation en catégories de produits et d'évaluer la typicité de chacun d'eux au sein de ces catégories.

Les méthodes présentées ici permettent d'analyser les données issues d'une épreuve de tri libre et visent à établir une typologie des produits représentant au mieux les perceptions des sujets. Il est utile de rappeler que chaque méthode correspond en réalité à un modèle mathématique qu'il convient de rapprocher de l'idée que l'expérimentateur se fait du processus cognitif en jeu dans la réalisation de l'épreuve de catégorisation et, en particulier, la représentation des produits dans l'espace perceptif.

Les méthodes factorielles reflètent une idée de représentation des produits sur la base de variables latentes qui structurent l'espace perceptif.

Pour le cas des arbres additifs, le principe sous jacent est l'existence d'une structure d'objets polarisés autour de prototypes représentés par les nœuds de l'arbre.

Les méthodes de partitionnement supposent la représentation des produits sous forme de catégories distinctes que les sujets cherchent à reconstituer. D'après Brenac (in Poitevineau, 2002), cette dernière représentation peut être considérée comme le niveau de base de la théorie de Rosch (1975). Cette théorie stipule que les catégories ou « représentations sémantiques du monde réel » sont organisées selon des structures hiérarchiques et qu'il existe trois niveaux d'abstraction dans la structure catégorielle : le niveau surordonné, le plus général, le niveau de base, le niveau dit privilégié et le niveau subordonné, le plus spécifique (Dubois, 1991).

Des investigations plus poussées concernant ces points seraient utiles et pourraient aboutir non seulement à une meilleure interprétation des résultats mais également à l'élaboration de démarches d'analyse encore plus adaptées à l'épreuve de catégorisation.

Quelle que soit la méthode adoptée, nous avons, en particulier, souligné l'intérêt des méthodes qui confèrent un caractère non métrique aux dissimilarités calculées. C'est le cas de la méthode MDS non métrique et de l'approche INDSCAL non métrique mais c'est aussi, en partie, le cas des arbres additifs. Nous avons également souligné l'intérêt des méthodes qui permettent de tenir

compte des différences individuelles entre sujets dans la détermination de l'espace de représentation. De plus, ces méthodes permettent de jeter un éclairage sur l'accord (et le désaccord) entre sujets.

Un prolongement naturel du travail qui a été réalisé ici est justement d'explorer davantage les différences entre les sujets. Nous pourrions, par exemple, établir une classification des sujets et déterminer, pour chaque classe, un espace de représentation propre, un arbre ou une partition centrale spécifique... Une autre direction de recherche est l'étude de la stabilité de la représentation des produits ou objets (sous forme de carte factorielle, arbre additif, arbre hiérarchique ou partition centrale) en fonction du nombre de sujets et du nombre d'objets. La comparaison de représentations issues de conditions expérimentales différentes est également un sujet de recherche de grand intérêt aussi bien sur le plan pratique que sur le plan de la recherche en statistique.

Listes des abréviations et notations

Abréviations	Désignation	Notations	Désignation
ACCPS	Analyse en Composantes Communes et Poids Spécifiques	N	nombre total de sujets
ACM	Analyse des Correspondances Multiples	P	nombre total de Produits
CAH	Classification Ascendante Hiérarchique	h	indice pour les sujets, variant de 1 à N
CCSORT	Common Components and specific weights analysis for SORTing data	$K^{(h)}$	nombre de groupes formés par le sujet h
DISTATIS	Analysis of Multiple Distance Matrices	$\Delta^{(h)}$	matrice de dissimilarités du sujet h
FAST	Factorial Approach for Sorting Task data	$\delta_{ij}^{(h)}$	pour la paire de produits (i,j) , cellule de la de la matrice $\Delta^{(h)}$
MDS	MultiDimensional Scaling	Δ	matrice de dissimilarités agrégées
MDSORT	a special purpose MultiDimensional scaling for SORTing data	δ_{ij}	pour la paire de produits (i,j) , cellule de la de la matrice Δ
IDSORT	Individual Difference multidimensional scaling program for SORTing data	$Z^{(h)}$	tableau disjonctif complet pour le sujet h , de dimensions $(PxK^{(h)})$
INDSCAL	Individual Difference Scaling ou weighted MDS	Z	concaténation pour tous les sujets des tableaux $Z^{(h)}$
		X	espace de représentation de dimension fixée
		$d_{ij}(X)$	distance entre les produits i et j dans X
		f	fonction monotone croissante des δ_{ij}
		$d_{ij}^{(h)}(X)$	distance entre les produits i et j pour le sujet h dans X
		x_{ia}	coordonnée du produit i sur l'axe a de X
		$w_a^{(h)}$	poids du sujet h sur cet l'axe a de X
		C	matrice de dissimilarité de la partition centrale
		c_{ij}	pour la paire de produits (i,j) , cellule de la de la matrice C

Références

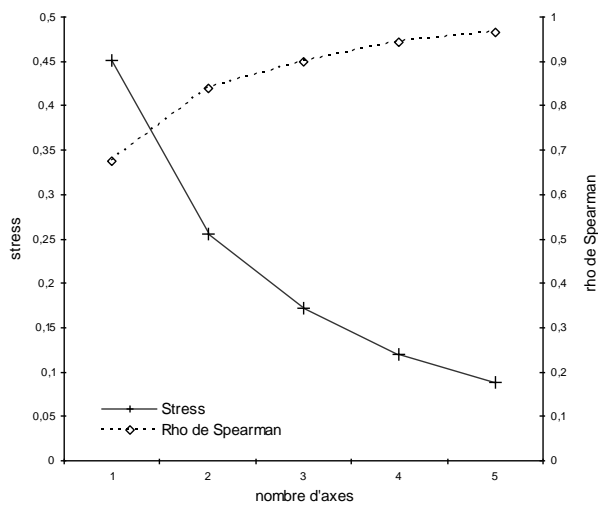
- Abdi H., Valentin D., Chollet S., Chrea C.: Analyzing assessors and products in sorting tasks : DISTATIS, theory and applications. *Food Quality and Preference* **18** (2007) 627-640.
- Anderberg M.K.: *Cluster Analysis for Applications*. Academic Press, New York (1973).
- Barthélémy J.P., Guénoche A.: *Les arbres et les représentations des proximités*. Edition Masson (1988).
- Blancher G., Chollet S., Kesteloot R., Nguyen Hoang D., Cuvelier G., Siefferman J.-M.: French and Vietnamese : How do they describe texture characteristics of the same food ? A case study with jellies. *Food Quality and Preference* **18** (2007) 560 -575

- Ballester J., Patris B., Symoneaux R., Valentin D.: Conceptual vs perceptual wine spaces : Does expertise matter? *Food Quality and Preference* **19** (2008) 267-276
- Borg I., Groenen P.: *Modern Multidimensional Scaling. Theory and Applications* (1997) Springer
- Burton M.L.: Dissimilarity measures for unconstrained sorting data. *Multivariate Behavioral Research* **10** (1975) 409-424
- Cadoret M., Lê S., Pagès J.: A Factorial Approach for Sorting Task data (FAST). *Food Quality and Preference* **20** (2009) 410-417
- Cariou V., Verdun S., Diaz E., Qannari E.M, Vigneau E. : Comparison of three hypothesis testing approaches for the selection of the appropriate number of clusters of variables. *Adv. Data Analysis and Classification* **3** (2009) 227-241
- Carroll J.D., Chang J.J.: Analysis of individual differences in multidimensional scaling via an N-way generalisation of 'Eckart-Young' decomposition. *Psychometrika* **35** (1970) 283-319
- Cartier R., Rytz A., Lecomte A., Poblete F., Krystlik J., Belin E., Martin N.: Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference* **17** (2006) 562-571
- Chrea C., Valentin D., Sulmont-Rossé C., Ly Mai H., Hoang Nguyen D., Abdi H.: Culture and odor categorization : agreement between cultures depends upon the odors. *Food Quality and Preference* **15** (2004) 669-679
- Coxon A.M.: *Sorting Data : Collection and Analysis*, Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-127. Thousand Oaks, CA : Sage (1999)
- Dacremont C., Soufflet I.: Impact of fabric end-use knowledge on handle perception. *Revue européenne de psychologie appliquée* **56** (2006) 273-277
- Daws J.T.: The analysis of free sorting data : Beyond pairwise cooccurrences. *Journal of Classification* **13** (1996) 57-80
- De Leeuw J.: Convergence of the majorization method for multidimensional scaling. *Journal of Classification* **5** (1988) 163-180
- Dubois D. (éditeur): *Sémantique et cognition. Catégories, prototypes, typicalité*. Editions du CNRS, Collection Sciences du langage, Paris (1991).
- Dubois D. (éditeur): *Le Sentir et le Dire. Concepts et méthodes en psychologie et linguistique cognitives*. Editions L'Harmattan, Paris (2009).
- Falahee M., MacRae A.W.: Perceptual variation among drinking waters : the reliability of sorting and ranking data for multidimensional scaling. *Food Quality and Preference* **8** (1997) 389-394
- Faye P., Brémaud D., Teillet E., Courcoux P., Giboreau A., Nicod H.: Perceptive free sorting and verbalization tasks with naive subjects : An alternative to descriptive mapping. *Food Quality and Preference* **15** (2004) 781-791
- Giboreau A., Navarro S., Faye P., Dumortier J.: Sensory evaluation of automotive fabrics : the contribution of categorization tasks and non verbal information to set-up a descriptive method of tactile properties. *Food Quality and Preference* **12** (2001) 311-322
- Gordon A.D., Vichi M.: Partitions of partitions. *Journal of Classification* **15** (1998) 265-285.
- Greenacre, M.J.: *Correspondence analysis in practice*. Academic Press, London (1993)
- Gygi B., Kidd G.R., Watson C.S.: Similarity and categorization of environmental sounds. *Perception & Psychophysics* **69** (6) (2007) 839-855
- Hanafi M., Qannari E.M.: Nouvelles propriétés de l'Analyse en Composantes Communes et Poids Spécifiques. *Revue de Statistiques Appliquées* **149** (2) (2008) 75-97
- Heiser W.J.: A generalized majorization method for least squares multi-dimensional scaling of pseudodistances that may be negative. *Psychometrika* **56** (1990) 7-27
- King M.C., Cliff M.A., Wall J.W.: Comparison of projective mapping and sorting data collection and multivariate methodologies for identification of similarity-of-use of snack bars. *Journal of Sensory Studies* **13** (1998) 347-358
- Kruskal J.B.: Nonmetric multidimensionnal scaling : a numerical method. *Psychometrika* **29** (1964) 115-129.
- Lavit Ch.: *Analyse conjointe de tableaux quantitatifs*. Edition Masson, Paris (1988).

- Lawless H.: Exploration of fragrance categories and ambiguous odors using multidimensional scaling and cluster analysis. *Chemical Senses* **14** (3) (1989) 349-360
- Lawless H., Sheng N., Knoops Stan S.C.P.: Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference* **6** (1995) 91-98
- Lebart L., Morineau A., Piron M.: *Statistiques exploratoires multidimensionnelles*, 2^{ème} édition, Dunod (1997).
- Lebart L., Morineau A., Piron M.: *Statistiques exploratoires multidimensionnelles*, 4^{ème} édition, Dunod (2006).
- Lelièvre M., Chollet S., Abdi H., Valentin D.: What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference* **19** (2008) 697-703
- Manetta C, Santarpia A., Sander E., Montet A., Urdapilleta I.: Categorization of literal and figurative olfactory descriptions in the perception of complex perfumes. *Psychologie française* **52** (2007) 479-497
- MacAdams S., Winsberg S., Donnadieu S., De Soete G., Krimphoff J.: Perceptual scaling of synthesized musical timbres : Common dimensions, specificities and latent subject classes. *Psychological Research* **58** (1995) 177-192
- Marcotorchino J.F., Michaud P.: Agrégation de similarité en classification automatique. *Revue de Statistique Appliquée* **30** (2) (1982) 21-44
- Parr W.V., Green J.A., White K.G., Sherlock R.R.: The distinctive flavour of New Zealand Sauvignon blanc : Sensory characterisation by wine professionals. *Food Quality and Preference* **18** (2007) 849-861
- Picard D., Dacremont C., Valentin D., Giboreau A.: Perceptual dimensions of tactile textures. *Acta Psychologica* **114** (2003) 165-184
- Poitevineau J. (éditeur): Arbres, classes, distances. *Cahiers du LCPE*, n°6 (2002).
- Qannari E.M., Cariou V., Teillet E., Schlich P.: SORT-CC : A procedure for the statistical treatment of free sorting. *Food Quality and Preference* (2009) doi:10.1016/j.foodqual.2009.01.003
- Ricciuti H.N., Marney T., Ricciuti A.E.: Availability and spontaneous use of verbal labels in sorting categorization by 16-23 months olds. *Early Childhood Research Quarterly* **21** (2006) 360-373
- Rosch E.: Cognitive representation of semantic categories. *Journal of experimental psychology, General*. **Vol 104 -3** (1975) 192-233
- Sattah S., Tversky A.: Additive similarity trees. *Psychometrika* **42** (3) (1977) 319-345
- Schiffman S.S., Reynolds M.L., Young F.W.: *Introduction to multidimensional scaling. Theory, methods and applications*. Academic Press, Orlando (1981).
- Soufflet I., Calonnier M., Dacremont C.: A comparison between industrial experts' and novices' haptic perceptual organization: a tool to identify descriptors of the handle fabrics. *Food Quality and Preference* **15** (2004) 689-699
- Takane Y., Young F.W., De Leeuw J.: Nonmetric Individual differences multidimensional scaling : An alternating least squares method with optimal scaling features. *Psychometrika* **42** (1977) 8-67
- Takane Y.: MDSORT: A special purpose multidimensional scaling program for sorting data. *Journal of Marketing Research* **18** (1981) 480-481
- Takane Y.: IDSORT: An individual differences multidimensional scaling for sorting data. *Behavior Research Methods and Instrumentation* **14** (1982) 546
- Torgerson W.S.: *Theory and methods of scaling*. New York, Wiley (1958)
- Van der Kloot W.A., Van Herk H.: Multidimensional scaling of sorting data : A comparison of three procedures. *Multivariate Behavioral Research* **26** (4) (1991) 563-581
- Wiley D.E. : Latent partition analysis. *Psychometrika* **32** (1967) 183-193

Annexe : modèle MDS métrique sur matrice de dissimilarités agrégées appliqué à l'étude de cas

Evolution du stress et du rho de Spearman en fonction de la dimension de l'espace de représentation



Configuration des produits sur la base des axes 1-2 et 1-3

