



HAL
open science

Méthodologies statistiques pour le recueil et l'analyse de données de caractérisation de grands ensembles de stimuli : Application à la caractérisation sensorielle de produits et à la caractérisation de gestes culinaires

Margot Brard

► To cite this version:

Margot Brard. Méthodologies statistiques pour le recueil et l'analyse de données de caractérisation de grands ensembles de stimuli : Application à la caractérisation sensorielle de produits et à la caractérisation de gestes culinaires. Statistiques [math.ST]. Institut national supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage (AGROCAMPUS OUEST), 2019. Français. NNT: . tel-02124794v1

HAL Id: tel-02124794

<https://hal.science/tel-02124794v1>

Submitted on 9 May 2019 (v1), last revised 18 Jun 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE

AGROCAMPUS OUEST
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 600
Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation
Spécialité : *Statistique/Modélisation en écologie, géosciences, agronomie et alimentation*

Par

Margot BRARD

Méthodologies statistiques pour le recueil et l'analyse de données de caractérisation de grands ensembles de stimuli

Application à la caractérisation sensorielle de produits & à la caractérisation de gestes culinaires

Thèse présentée et soutenue à AGROCAMPUS OUEST (Rennes), le 29/01/2019

Unité de recherche : Institut de Recherche Mathématique de Rennes (IRMAR), UMR 6625 du CNRS
Thèse N° : 2019-1 / G-17

Rapporteurs avant soutenance :

Bernadette GOVAERTS
Hervé ABDI

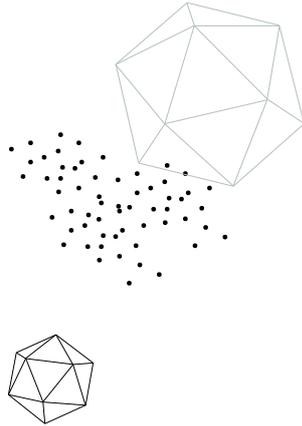
Professeure, Université Catholique de Louvain, Belgique
Professeur, Université du Texas, Dallas, États-Unis

Composition du Jury :

Présidente : Françoise NAU Professeure, AGROCAMPUS OUEST, Rennes, France
Examineur : El Mostafa QANNARI Professeur, ONIRIS, Rennes, France
Dir. de thèse : David CAUSEUR Professeur, AGROCAMPUS OUEST, Rennes, France
Co-encadrant de thèse : Sébastien LÊ Maître de conférences, AGROCAMPUS OUEST, Rennes, France

Invité

Fabrice CLOCHARD Encadrant CIFRE



RÉSUMÉ

MÉTHODOLOGIES STATISTIQUES POUR LE RECUEIL ET L'ANALYSE DE DONNÉES DE CARACTÉRISATION DE GRANDS ENSEMBLES DE STIMULI – APPLICATION À LA CARACTÉRISATION SENSORIELLE DE PRODUITS & À LA CARACTÉRISATION DE GESTES CULINAIRES

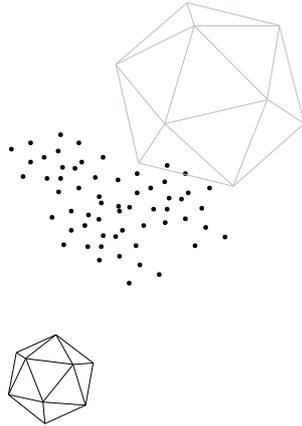
Dans le secteur agroalimentaire, la caractérisation des gestes culinaires (protocoles culinaires, recettes, etc.) - à travers leur observation et leur analyse - est considérée de plus en plus comme un levier d'innovation. En analysant la manière avec laquelle leurs produits sont appréhendés en cuisine par les usagers (consommateurs ou professionnels), les industriels peuvent en effet déceler des pistes d'amélioration ou des idées de nouveaux produits. Dans ce travail de recherche, nous proposons de nouvelles méthodes dédiées à la caractérisation d'un grand nombre de gestes culinaires. Leur caractère innovant repose sur le fait qu'elles relèvent d'une approche quantitative, et non d'une approche qualitative comme cela est usuellement le cas dans le domaine de l'étude des usages. Elles s'inspirent en partie de deux méthodes de caractérisation utilisées en analyse sensorielle : le tri libre et la Q-méthodologie binaire. Nous voyons comment des développements méthodologiques apportés à ces dernières permettent de caractériser des grands ensembles de stimuli. Ces développements sont liés à la fois à la procédure de recueil des données et à la procédure d'analyse statistique des données. En particulier, nous proposons plusieurs procédures statistiques permettant d'aborder des problématiques variées : l'analyse statistique d'un ensemble de partitions contenant des données manquantes, la classification non supervisée de profils d'évaluations binaires basée sur la notion d'accord inter-évaluateurs, etc. Nous voyons ensuite que les deux méthodes de caractérisation quantitatives 'améliorées' sont applicables à des gestes culinaires.

MOTS CLÉS : méthodologies statistiques, grands ensembles de stimuli, caractérisation sensorielle, caractérisation de gestes culinaires

STATISTICAL METHODOLOGIES FOR THE COLLECTION AND ANALYSIS OF
DATA CHARACTERIZING LARGE SETS OF STIMULI – APPLICATION TO THE
SENSORY CHARACTERIZATION OF PRODUCTS & TO THE
CHARACTERIZATION OF CULINARY TECHNIQUES

In the agri-food sector, the characterization of culinary techniques (culinary protocols, recipes, etc.) - through their observation and analysis - is increasingly seen as a lever for innovation. By analyzing how their products are used in the kitchen by the users (consumers or professionals), the manufacturers can detect improvement tracks or ideas for new products. In this research, we propose new methods dedicated to the characterization of a great number of culinary techniques. Their innovative nature is based on the fact that they are based on a quantitative approach, and not on a qualitative approach as is usually the case. They partially draw their inspiration in two methods of characterization used in sensory analysis: the free sorting and the binary Q-methodology. We see how methodological developments brought to the latter make it possible to characterize large sets of stimuli. These developments are both related to the procedure of data collection and to the procedure of statistical analysis of the data. In particular, we propose several statistical procedures to address various issues: the statistical analysis of a set of partitions containing missing data, the unsupervised agreement-based clustering of a set of profiles of binary evaluations, etc. We then see that these two ‘improved’ quantitative methods of characterization can be successfully applicable to culinary techniques.

KEYWORDS : statistical methodologies, large sets of stimuli, sensory characterization, characterization of culinary techniques



REMERCIEMENTS

Ce n'est pas sans une pointe d'émotion que je m'attèle ici à la rédaction de ces quelques lignes. Elles marquent en effet la fin d'une belle expérience, à la fois professionnelle et humaine. Certains voient la thèse de doctorat comme un chemin de croix, une sorte de parcours du combattant dont l'objectif est de sortir indemne. Il n'en est rien me concernant. J'ai en effet toujours pris un plaisir indescriptible à me pencher dans la gestion de ce travail doctoral. Si cela a été rendu possible, c'est notamment parce que j'ai eu la chance d'évoluer au sein d'un environnement idéal, composé de personnes à la fois inspirantes et bienveillantes.

Tout d'abord, je remercie très chaleureusement les rapporteurs de cette thèse, Bernadette GOVAERTS et Hervé ABDI. Merci pour l'intérêt que vous avez porté à mon travail, et pour les conseils précieux que vous m'avez donnés pour améliorer ce manuscrit de thèse. Merci également à Françoise NAU et El Mostafa QANNARI d'avoir accepté de compléter ce jury de thèse.

La thèse de doctorat n'est pas un travail solitaire mais un véritable travail d'équipe. Ces travaux de recherche n'auraient jamais pu voir le jour sans ceux qui m'ont accompagnée au quotidien. C'est pourquoi j'adresse un immense merci à Sébastien LÊ et David CAUSEUR, eux qui m'ont encadrée pendant ces trois années. Vous m'avez donné le goût pour la statistique et la recherche scientifique. Mais plus important encore, vous avez su me *coacher*, m'épauler, et me donner confiance en moi pendant toute cette thèse de doctorat. Je souhaite à tout doctorant de pouvoir être *couvé* par des directeurs de thèse aussi bienveillants que vous. Je vous remercie également de m'avoir offert l'opportunité de faire le tour du monde pour présenter mes travaux de recherche dans des conférences internationales. Enfin, merci de vous être démenés il y a trois ans pour que cette thèse puisse voir le jour.

J'étends ces remerciements à l'ensemble des membres de l'Unité Pédagogique de mathématiques appliquées d'AGROCAMPUS OUEST : Héléna, pour ta bonne humeur et ta réactivité sans pareil, Elisabeth et Karine ; Mathieu, Magalie, François et Marie-Pierre. J'ai pris un grand plaisir à partager avec vous plusieurs tranches importantes de ma vie. D'abord l'une de vos étudiantes, je suis ensuite passée *du côté obscur de la force...* en prenant successivement la casquette d'ingénieure d'études au sein de l'UP puis celle de doctorante. J'ai alors pu vous

voir sous un nouveau jour et découvrir des facettes de votre personnalité que je n'avais pas eu l'occasion de découvrir sur les bancs de l'école. Ce que je retiens de vous, c'est votre amicalité et votre bienveillance. Certains disent que la thèse est aussi une belle aventure humaine au-delà d'une aventure professionnelle. C'est bien vrai. Merci également aux doctorants dont j'ai croisé le chemin à AGROCAMPUS OUEST : Emeline, Vincent et Tâm ; et Florian. Florian, je t'envoie des ondes positives pour ces prochains mois. Ne doute pas de toi : si la qualité de ta thèse est proportionnelle à ta gentillesse, elle ne pourra être qu'excellente.

J'adresse ensuite mes remerciements à Fabrice CLOCHARD et Freddy THIBURCE, pour avoir accepté que le Centre Culinaire Contemporain finance ce travail doctoral dans le cadre d'une convention CIFRE établie avec l'Association Nationale Recherche Technologie (ANRT). Je vous remercie de m'avoir accordé votre confiance dans un contexte économique difficile. Je sais que c'était un pari risqué.

La suite de mes remerciements s'adresse plus globalement aux équipes du Centre Culinaire Contemporain, *actuelles ou passées* : Mélanie, Laure, Delphine, Sabrina, Audrey, Chabada, Loris, Véro, Sophie, Yannick I^{er}, Sylvie, Yannick II^{ème}, Khadija, Ronald, Jean-Marie, et tous les autres. Votre gentillesse, votre bonne humeur et votre écoute ont rendu le quotidien de cette thèse particulièrement agréable et motivant. Enfin, merci à Marie. Au bureau, je te remercie d'avoir toujours porté un regard attentif et curieux sur mes travaux de recherche. En dehors, merci d'avoir été une épaule sur laquelle je pouvais m'appuyer. Si le coup de foudre amical dans un contexte professionnel existe, c'est clairement avec toi que je l'ai eu.

J'ai par ailleurs eu la chance de pouvoir bénéficier d'un comité de thèse *de choc*. Par conséquent, merci à Catherine GUÉRIN, Philippe COURCOUX et Yvonnick NOËL. Je vous remercie pour votre accompagnement pendant ces trois années de thèse. J'ai apprécié nos échanges scientifiques passionnants. Vos encouragements, eux, m'ont toujours permis de repartir regonflée à *bloc* pour une nouvelle année.

Je remercie bien sûr tous les étudiants d'AGROCAMPUS OUEST qui, au cours de ces trois années, m'ont permis de tester mes outils sur des données réelles : Clément, Anas, Mathieu et Victor ; Charlotte, Justine et Antoine ; Anouk, Andréa, Ornella et Thomas ; Morgane, Marie-Anne et Thibaut ; Lisa, Phuc-Loi, Benjamin et Jules.

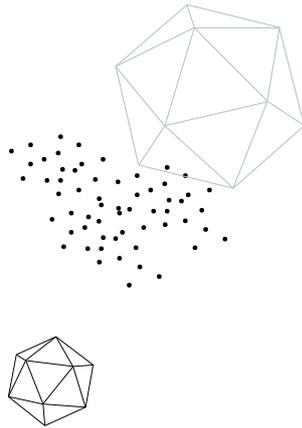
Évidemment, je ne pourrai pas achever ces lignes sans remercier de tout cœur mes proches. Eux qui m'ont fourni et me fournissent chaque jour la force de donner le meilleur de moi-même.

À mes amies (*les sardines, Cécile et les autres*), vous qui êtes toujours présentes pour moi.

Merci bien sûr à ma formidable famille, mon pilier de toujours. Merci à vous,

mamounette et *papounet*, pour m'avoir toujours soutenue et accompagnée dans la réalisation de tous mes projets, qu'ils soient professionnels ou autres. J'espère que cette thèse vous rendra fiers. Merci à toi, *mon amour de sœur*, pour ton accompagnement au quotidien. Je sais que tu es ma plus fervente supportrice, tout comme moi je suis la tienne. Un jour, peut-être (et pour mon plus grand plaisir !), *le duo des sœurs BRARD* se reformera au boulot pour conquérir le monde de la data science. Enfin, je ne pourrai omettre de te dédier quelques mots, mon ange. Si un jour tu lis ces lignes, sache que tes coups de pieds d'encouragement *in utero* n'ont fait que dupliquer ma motivation et ma détermination lors de ces derniers mois de thèse.

Et pour finir, un merci infini à mon conjoint. Merci, *ma tourte*, pour le soutien psychologique et logistique que tu m'as apporté pendant cette thèse de doctorat. Je te remercie d'avoir réussi à faire semblant de t'intéresser à cette thèse quand j'en avais besoin ; et je ne saurai jamais comment te remercier suffisamment d'avoir accepté d'écouter chacune de mes présentations orales en boucle (*15 fois... avec un accent anglais largement perfectible*) avant mes conférences. Plus globalement, merci de n'avoir jamais douté de moi et de m'avoir appris à ne jamais rien lâcher, à me surpasser et à toujours rebondir face aux difficultés. Certains me demandent comment j'ai pu gérer en parallèle une fin de thèse de doctorat, une procédure de licenciement économique, une grossesse puis une maternité. Avec toi à mes côtés, c'était facile. *Sans toi à mes côtés, c'eût été impossible.*



AVANT PROPOS

Ce travail de thèse a été conduit sous le format d'une convention CIFRE (Convention Industrielle de Formation par la REcherche). Sa mise en place est le fruit d'un partenariat entre deux entités : l'équipe de mathématiques appliquées d'AGROCAMPUS OUEST (rattachée à l'IRMAR - Institut de Recherche Mathématique de Rennes - du CNRS) et les équipes du Centre Culinaire Contemporain.

Étant située au sein d'une école d'ingénieur en agroalimentaire, l'équipe de mathématiques appliquées travaille sur des problématiques statistiques tournées vers l'agroalimentaire, et notamment vers l'analyse sensorielle. Elle est notamment à l'origine du développement de nouvelles méthodes de caractérisation sensorielle telles que le Napping® (Pagès, 2003). Elle a également encadré plusieurs thèses en sensométrie (statistique pour l'analyse sensorielle) portant sur des problématiques variées telles que l'analyse statistique des données de tri libre (Cadoret, 2010), des données de profil idéal (Worch, 2012), des données de commentaires libres (Kostov, 2015), ou des données de digit-tracking (T. Lê, 2015).

De leur côté, les équipes du Centre Culinaire Contemporain accompagnent les industriels agroalimentaires dans leur processus d'innovation. Dans ce cadre, elles sont régulièrement amenées à réaliser des études d'usages culinaires. Ces dernières consistent notamment à observer et à analyser les techniques culinaires des usagers - à la fois consommateurs et professionnels.

Le rapprochement de ces deux entités est dû à leur confrontation respective à une problématique émergente similaire : la nécessité de développer une méthode quantitative d'évaluation permettant la caractérisation de grands ensembles de stimuli, à savoir des grands ensembles de produits industriels dans le cadre de l'analyse sensorielle et des grands ensembles de gestes culinaires dans le cadre des études d'usages culinaires. La vocation de ce travail de recherche pluridisciplinaire est de proposer des solutions méthodologiques - en termes de recueil de données et d'analyse statistique notamment - à ce questionnement. Ce manuscrit de thèse jongle donc entre des propositions de développements méthodologiques adaptés et leurs applications à l'analyse sensorielle ainsi qu'à l'étude des usages culinaires. Ainsi, ces deux domaines d'expertise ont constitué à la fois une source d'inspiration et un

domaine d'application privilégié de ce travail. Les développements méthodologiques proposés peuvent cependant être étendus à d'autres domaines de recherche.

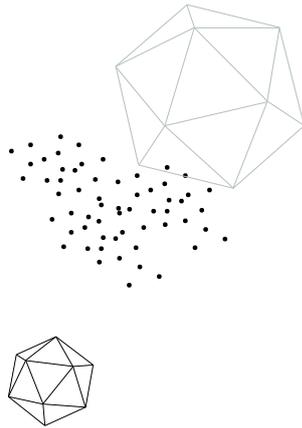


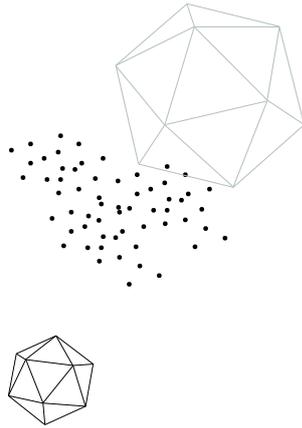
TABLE DES MATIÈRES

Introduction	1
1 De la caractérisation des gestes culinaires à la caractérisation sensorielle des produits : des interrogations similaires	3
1.1 Caractérisation des gestes culinaires	3
1.2 Caractérisation sensorielle de produits	10
1.3 Analogie entre les interrogations rencontrées dans les deux domaines	13
2 Méthodologie pour la caractérisation perceptuelle et conceptuelle de grands ensembles de stimuli par des évaluateurs non-experts	14
2.1 Problématique de recherche	14
2.2 Démarche de recherche privilégiée	15
2.3 Pistes envisagées	17
3 Organisation de la thèse	25
3.1 Caractérisation perceptuelle	25
3.2 Caractérisation conceptuelle	26
I Caractérisation perceptuelle	29
1 Recueil des données de Tri Séquentiel Agglomératif	31
1 Recueil des données	33
1.1 Réalisation de la tâche de caractérisation perceptuelle par l'évaluateur : description du protocole de Tri Séquentiel Agglomératif	34
1.2 Organisation de l'ensemble des stimuli par l'expérimentateur	37
2 Format des données	41
3 Validation du protocole de recueil des données	42
3.1 Mise en place des deux méthodes de caractérisation concurrentes	43
3.2 Comparaison des performances des deux méthodes de caractérisation concurrentes	45
4 Conclusion	54

2	Analyse statistique des données de Tri Séquentiel Agglomératif	57
1	Format des données	60
2	Analyse statistique des données orientée stimuli	63
2.1	Analyse des caractéristiques partagées avec les données de tri libre verbalisé : prise en compte des partitions finales sur les stimuli	63
2.2	Analyse des caractéristiques spécifiques aux données de Tri Séquentiel Agglomératif : prise en compte des partitions intermédiaires sur les stimuli	71
3	Analyse statistique des données orientée évaluateurs	80
3.1	Définition de la distance inter-évaluateurs	80
3.2	Création du dendrogramme et stratégie de coupure	80
3.3	Interprétation de la partition d'évaluateurs	81
4	Illustration sur des données réelles de caractérisation de produits industriels : données 'parfums de gels douche'	85
4.1	Données	86
4.2	Résultats de l'analyse statistique orientée parfums	86
4.3	Résultats de l'analyse statistique orientée consommateurs	89
5	Conclusion	91
3	Application du Tri Séquentiel Agglomératif à la caractérisation de gestes culinaires	95
1	Données 'recettes de tarte au citron meringuée'	97
1.1	Procédure de recueil des données	97
1.2	Format des données	101
2	Résultats de l'analyse statistique	102
2.1	Résultats orientés recettes	102
2.2	Résultats orientés évaluateurs	107
3	Conclusion	110
Annexe : Implémentation dans R de l'analyse statistique des données de Tri Séquentiel Agglomératif		113
1	Fonction <code>AnalyseSAS</code>	115
1.1	Arguments en entrée	115
1.2	Sorties	118
2	Fonction <code>plot.AnalyseSAS</code>	120
2.1	Arguments en entrée	120
2.2	Sorties	121

II	Caractérisation conceptuelle	125
4	Recueil et analyse statistique orientée stimuli des données de Q-méthodologie binaire	127
1	Recueil des données	129
1.1	Réalisation de la tâche de caractérisation conceptuelle par l'évaluateur	129
1.2	Format des données	130
2	Analyse statistique des données orientée stimuli	131
2.1	Définition de la distance inter-stimuli et représentation euclidienne associée	131
2.2	Amélioration de la visualisation de la représentation euclidienne des stimuli	148
3	Illustration sur des données réelles de caractérisation de produits industriels : données 'concept d'innovation et parfums de shampoings'	151
3.1	Données	152
3.2	Résultats de l'analyse statistique orientée parfums	152
4	Conclusion	154
5	Analyse statistique orientée évaluateurs des données de Q-méthodologie binaire	155
1	Format des données	157
2	Analyse statistique des données orientée évaluateurs	158
2.1	Analyse de la structure de désaccord au sein d'un ensemble de profils d'évaluations binaires	159
2.2	Classification non supervisée de profils d'évaluations binaires	167
3	Illustration sur des données réelles de caractérisation de produits industriels : données 'concept d'innovation et parfums de shampoings'	184
3.1	Données	185
3.2	Résultats de l'analyse statistique orientée consommateurs	185
4	Conclusion	189
5	Annexe	191
	Annexe : Implémentation dans R de l'analyse statistique des données de Q-méthodologie binaire	195
1	Fonction <code>AnalyseBinQmet</code>	197
1.1	Arguments en entrée	197
1.2	Sorties	199
2	Fonction <code>plot.AnalyseBinQmet</code>	199
2.1	Arguments en entrée	200
2.2	Sorties	201
	Conclusion	205

Liste des travaux	211
Bibliographie	215



INTRODUCTION

RÉSUMÉ : Dans le secteur agroalimentaire, la caractérisation des usages culinaires - à travers leur observation et leur analyse - est de plus en plus considérée comme un levier d'innovation. En particulier, les gestes culinaires (protocoles culinaires, recettes, etc.) sont devenus des objets d'étude particulièrement intéressants pour les industriels. En effet, en analysant la manière avec laquelle leurs produits sont appréhendés en cuisine par les usagers (consommateurs ou professionnels), ces derniers peuvent déceler des pistes d'amélioration ou des idées de nouveaux produits. Dans ce contexte, la pratique courante consiste à recourir à des méthodes qualitatives d'observation pour caractériser les gestes culinaires. Ces dernières sont réalisées par un évaluateur expert unique qui caractérise les gestes de manière analytique. Étant difficiles et coûteuses à mettre en place, elles reposent généralement sur la caractérisation d'un nombre réduit de gestes culinaires. Dès lors, ces approches qualitatives s'avèrent limitées pour comprendre la variabilité liée à la subjectivité du jugement, obtenir un jugement sur les gestes autre que descriptif, et caractériser un nombre important de gestes. Ces limitations s'apparentent à celles relatives à la caractérisation sensorielle de produits. L'objectif de cette introduction est d'établir un parallèle entre les différents verrous relatifs à la caractérisation des gestes culinaires d'une part, et ceux identifiés en analyse sensorielle d'autre part. L'objectif est également de démontrer qu'il est possible de tirer avantage des méthodes quantitatives sensorielles pour aborder la problématique de la caractérisation des gestes culinaires.

Sommaire

1	De la caractérisation des gestes culinaires à la caractérisation sensorielle des produits : des interrogations similaires	3
1.1	Caractérisation des gestes culinaires	3
1.1.1	Innovation par les usages	3
1.1.2	Méthodes qualitatives de caractérisation des usages	5
1.1.2.1	Principe général	5
1.1.2.2	Du recueil à l'analyse des données	7
1.1.2.3	Verrous et interrogations associées	9
1.2	Caractérisation sensorielle de produits	10
1.2.1	Analyse sensorielle	10
1.2.2	Méthodes quantitatives de caractérisation sensorielle	11
1.2.2.1	Principe général	11
1.2.2.2	Du recueil à l'analyse des données	11
1.2.2.3	Verrous et interrogations associées	12
1.3	Analogie entre les interrogations rencontrées dans les deux domaines	13
2	Méthodologie pour la caractérisation perceptuelle et conceptuelle de grands ensembles de stimuli par des évaluateurs non-experts	14
2.1	Problématique de recherche	14
2.2	Démarche de recherche privilégiée	15
2.3	Pistes envisagées	17
2.3.1	Caractérisation perceptuelle : focus sur la méthodologie du tri libre	18
2.3.2	Caractérisation conceptuelle : focus sur la Q-méthodologie binaire	21
3	Organisation de la thèse	25
3.1	Caractérisation perceptuelle	25
3.2	Caractérisation conceptuelle	26

Pour comprendre l'objectif de ce travail de recherche, nous proposons dans un premier temps de préciser quelques notions de contexte. Nous introduisons des éléments relatifs aux deux grands types de caractérisations abordés ici : d'une part la caractérisation des gestes culinaires (problématique initiale ayant motivé ce travail de recherche), et d'autre part la caractérisation sensorielle de produits. Nous présentons notamment les approches méthodologiques qui permettent d'aborder ces problématiques de caractérisation ainsi que les interrogations qui leur sont associées. L'analogie entre ces deux types de problématiques - *a priori* éloignées - nous amène, dans un second temps, à présenter les hypothèses liées à ce travail de recherche et la démarche privilégiée. Dans un troisième et dernier temps, nous présentons succinctement la manière avec laquelle cette thèse est organisée.

1 De la caractérisation des gestes culinaires à la caractérisation sensorielle des produits : des interrogations similaires

1.1 Caractérisation des gestes culinaires

1.1.1 Innovation par les usages

En donnant naissance à de nouvelles technologies ou à de nouveaux produits, l'innovation est un moteur majeur de la croissance économique. Depuis quelques années, le débat sur l'innovation a beaucoup porté sur l'‘innovation orientée usager’ ou ‘innovation par les usages’. L'efficacité de cette démarche d'innovation justifie le fait que l'intérêt qui lui est porté reste constant depuis des années (voir Figure 1).

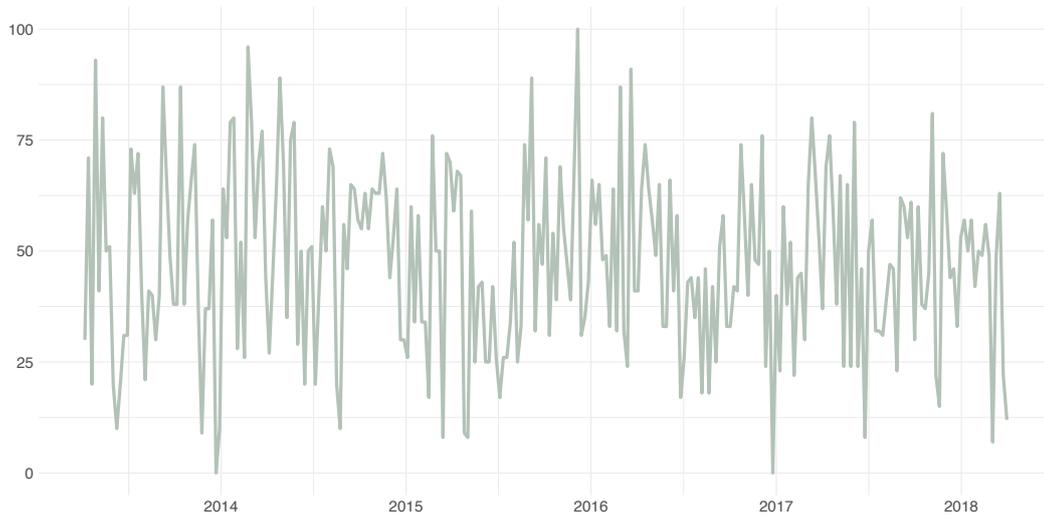


Figure 1: Évolution de l'intérêt pour la recherche associée au thème ‘user innovation’ sur google entre avril 2013 et avril 2018. Les résultats reflètent la proportion de recherches portant sur ce thème pour une période spécifique, par rapport au taux d'utilisation le plus élevé (valeur de 100). Ainsi, une valeur de 50 signifie que le thème a été recherché moitié moins souvent que sur la période où il a atteint son maximum.

L'innovation par les usages consiste à prendre en compte les usagers dans le processus de Recherche & Développement (R&D) industriel. Ceux-ci sont en effet les plus à même d'identifier ce qu'ils souhaitent et ce dont ils ont besoin. La vocation finale d'une telle approche est d'optimiser les chances de réussite commerciale des nouveaux concepts développés (Douthwaite, Keatinge, & Park, 2001; von Hippel, 2005).

Cette démarche d'innovation a été décrite pour la première fois dans les années 1970 par von Hippel (von Hippel, 1976, 1978). Sa description s'est basée sur l'observation du phénomène suivant : beaucoup d'innovations de produits ont été

initiées, inventées ou développées grâce à des usagers (de Jong & von Hippel, 1978). À l'aube de l'innovation par les usages, ces exemples concernaient principalement des secteurs industriels à forte dominante technologique de type 'business to business' (B2B). Dans ce cadre, nous pouvons notamment citer des exemples d'instruments scientifiques ou de circuits imprimés (von Hippel, 1976, 1988). D'autres références existent également dans des secteurs industriels de type 'business to consumer' (B2C). Certains considèrent par exemple que le développement initial du Vélo Tout Terrain (VTT) est imputable à de jeunes cyclistes. Frustrés par leurs caractéristiques techniques, ces usagers ont modifiés eux-mêmes leurs vélos classiques pour les adapter aux chemins plus accidentés (Luthje, Herstatt, & von Hippel, 2005).

Depuis les prémices de l'innovation par les usages, la définition donnée à ce terme a été considérablement étendue. Ainsi, elle ne fait plus seulement référence aux situations dans lesquelles l'innovation a été entièrement pensée et initiée par les usagers. Plus précisément, elle décrit désormais également toutes les situations dans lesquelles les usagers ont été impliqués lors d'une ou plusieurs phases du processus de développement (voir les phases en question sur la Figure 2). Selon cette définition étendue, l'innovation par les usages prend aujourd'hui différentes formes. Elle peut, par exemple, faire référence à des situations dans lesquelles l'industriel a initialement pensé l'innovation et a ensuite impliqué les usagers dans le processus de R&D (Grunert et al., 2008).

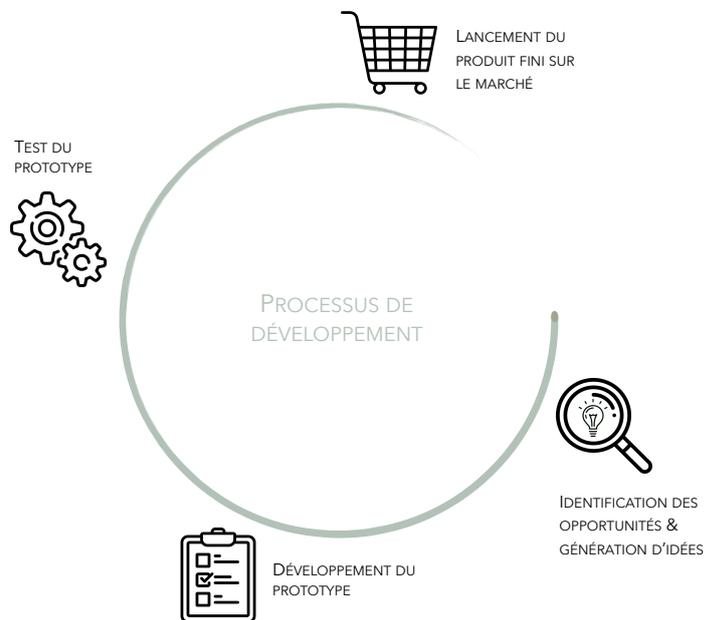


Figure 2: Phases du processus de développement d'un nouveau produit.

Bien que largement plébiscitée dans d'autres secteurs industriels, l'innovation par les usages est peu utilisée en agroalimentaire (Busse & Siebert, 2017; Grunert et al., 2008). Le faible niveau de développement technologique et la faible intensité de recherche dans ce secteur traditionnel peuvent expliquer ce déficit (Bayona-Saez, Cruz-Cázares, García-Marco, & García, 2017; Bibliardi & Galati, 2016; Galanakis, 2016; Schoen, 2017). De plus, ce secteur a longtemps privilégié l'augmentation de la capacité de production et la réduction des coûts inhérents à celle-ci (Costa & Jongen, 2006). Dès lors, les lourds investissements (humains et financiers) réservés à la production ont été réalisés au détriment de ceux dédiés à la R&D. Ensuite, l'agroalimentaire est un secteur industriel mature dans lequel les produits développés sont majoritairement de type incrémental - ces derniers ne modifient pas de manière radicale le secteur et leur développement ne nécessite que peu d'opérations complexes (Costa & Jongen, 2006; Grunert et al., 2008). Ceci est notamment dû au fait que les usagers sont perçus par les industriels du secteur comme conservateurs (Sarkar & Costa, 2008). De plus, ce secteur industriel présente la particularité de développer majoritairement des produits dédiés à un marché de masse. Cette particularité du secteur complique la mise en place d'une interaction directe et personnalisée avec la communauté d'usagers. Pour finir, certains aspects fragilisent l'implémentation pratique d'une démarche d'innovation orientée usager au sein des entreprises du secteur. Parmi ceux-ci, nous pouvons par exemple citer le manque de recommandations concrètes pour la mise en œuvre efficace d'une telle démarche, ainsi que le manque d'interaction entre les équipes de R&D et marketing (Fortuin & Omta, 2009; van Kleef, van Trijp, & Luning, 2005).

Malgré cela, les spécialistes de l'agroalimentaire s'entendent depuis longtemps sur le fait qu'une implication systématique de l'utilisateur dans le processus de R&D est primordiale. Une telle implication de l'utilisateur, et en particulier du consommateur, est en effet reconnue pour augmenter les chances de succès une fois le produit mis sur le marché (Grunert et al., 2008; Stewart-Knox & Mitchell, 2003; Trijp & van Kleef, 2008; van Kleef et al., 2005).

1.1.2 Méthodes qualitatives de caractérisation des usages

1.1.2.1 Principe général

L'implication de l'utilisateur peut être réalisée durant chaque phase du processus de développement (voir les phases en question sur la Figure 2). Cette implication a pourtant généralement lieu en aval du processus lors des phases de développement, de test, et de lancement sur le marché. Cependant, il est de plus en plus reconnu que le succès d'un nouveau produit dépend nettement de la qualité de l'amont du processus, lors de la phase d'identification des opportunités (van Kleef & van Trijp, 2007). L'objectif de cette phase est de faire émerger de nouvelles idées d'innovation qui répondent aux attentes des usagers. En pratique, cette démarche est souvent difficile à mettre en place. En effet, les usagers ne sont pas toujours capables d'exprimer eux-mêmes leurs besoins ou attentes. On parle alors de 'besoins latents'

(Grunert et al., 2008; van Kleef et al., 2005).

L'étude de ces besoins latents peut se faire grâce à une large palette de méthodes principalement issues du domaine des sciences sociales. Ces approches relèvent des sciences qualitatives dans le sens où, comme nous le verrons par la suite, elles mettent l'accent sur le recueil de données non-numériques qui sont ensuite analysées via une procédure interprétative. Un état de l'art de leur utilisation pour étudier les besoins latents est proposé dans van Kleef et al. (2005). Les *méthodes d'observation* sont des outils utilisés dans ce cadre. En pratique, elles consistent à observer les usagers en situation réelle d'usage d'un produit. À travers cette démarche, l'objectif est de mettre en évidence des signaux faibles reflétant différents aspects : des besoins latents, des frustrations entraînées par le produit ou la situation d'usage, ou des usages inattendus du produit (van Kleef & van Trijp, 2007). Dans un contexte de développement de produit, les méthodes d'observation peuvent être utilisées pour répondre à des problématiques opérationnelles très variées. En effet, elles peuvent être utilisées lors de la phase d'identification d'opportunités mais aussi lors de la phase de test de produit pour identifier comment les usagers utilisent le produit développé. Par nature, ces méthodes d'observation peuvent être opposées aux méthodes déclaratives qui consistent à demander aux usagers d'auto-décrire leurs comportements. Contrairement à ces dernières, les méthodes d'observation présentent l'avantage d'éviter certains biais tels que les erreurs, les oublis ou l'autocensure (Buck, 2007; Lacour, 2004).

Dans le cas particulier du secteur agroalimentaire, les méthodes d'observation sont sous-utilisées (Buck, 2007). Et pourtant, comme dans les autres secteurs industriels, elles représentent une source d'informations intéressantes dans un contexte de R&D. Leonard and Rayport (1997) présente un bon exemple de leur intérêt en agroalimentaire. Celui-ci décrit la mise en place d'une approche d'observation par Cheerios, entreprise initialement spécialisée dans les céréales pour le petit-déjeuner. L'objectif était d'observer et d'analyser comment les consommateurs utilisaient leurs produits. Les résultats de cette étude ont été surprenants. Ils ont montré que la dégustation des céréales dans le bol du petit-déjeuner n'était pas leur seule utilisation. Plus précisément, Cheerios a observé que certains parents mettaient les céréales dans un sachet. De cette façon, leur jeunes enfants pouvaient partir avec leurs céréales et les utiliser à n'importe quel moment de la journée comme collations. Ces résultats ont poussé l'entreprise à développer un nouveau produit snacking à base de céréales pour les enfants.

L'exemple précédent relève du développement d'un produit alimentaire ayant vocation à être ingéré. Cependant, les méthodes d'observation peuvent également être utilisées dans d'autres contextes de R&D dans le secteur agroalimentaire. Par exemple, elles peuvent être mises en œuvre pour observer comment les usagers réalisent une technique culinaire donnée (p. ex. le foisonnement d'une crème fraîche). Cette observation peut permettre d'identifier d'éventuelles frustrations

pouvant être sources de nouvelles idées de développement d'ustensiles culinaires. Dans un autre registre, elles peuvent également être utilisées pour déterminer de quelle manière les usagers mettent en œuvre un produit alimentaire dans une préparation culinaire (p. ex. une huile végétale dans une mayonnaise). L'observation de ces usages peut permettre au fabricant d'identifier le protocole culinaire majoritairement utilisé par les usagers. Il peut ainsi l'utiliser comme protocole d'évaluation pour comparer les performances culinaires de son produit par rapport à la concurrence, et ce dans les conditions réelles d'usage. En effet, si les protocoles utilisés par les fabricants sont éloignés des conditions réelles d'usage, le risque est qu'un produit jugé performant par le fabricant s'avère décevant une fois mis en œuvre par les usagers. La spécificité de ces exemples d'application repose sur le fait que l'accent est porté sur le *geste culinaire* réalisé par l'usager. À ce jour, aucun travail scientifique n'a porté sur ce type de problématiques, et ce malgré le souhait important des industriels du secteur de les aborder. Nous verrons par la suite que la caractérisation de ces gestes culinaires, à travers leur observation et leur analyse, est au cœur de ce projet de recherche.

1.1.2.2 Du recueil à l'analyse des données

Les méthodes d'observation des usages consistent à recueillir des données d'observation des usages et à les analyser. Ces deux étapes sont généralement réalisées par une seule et même personne. Dans un contexte de R&D, cette personne est le plus souvent un expert. En effet, un expert présente la faculté de traduire les besoins latents des usagers en actionnables opérationnels. Il peut par exemple s'agir d'un ingénieur R&D, d'un designer produit ou d'un ergonome (Leonard & Rayport, 1997).

Dans un premier temps, l'unique évaluateur expert a la charge de recueillir les données d'observation. Chacune de ces données correspond en réalité à la description d'un usage observé. En pratique, l'observation des usages à proprement parler peut prendre différentes formes. Elle dépend par exemple du rôle - actif ou passif - de l'évaluateur. Dans le cas d'une observation passive, l'évaluateur observe seulement les usages en condition réelle sans interagir directement avec l'usager. Au contraire, dans le cas d'une observation active, l'évaluateur peut interpeller l'usager et le questionner sur son comportement au cours de l'observation (Buck, 2007). L'observation des usages peut également être réalisée de manière directe ou indirecte. Lorsqu'elle est réalisée de manière indirecte, l'évaluateur n'est pas physiquement présent avec l'usager et les usages sont alors filmés par le biais de caméras.

Dans le cadre de ce projet de recherche, nous nous focalisons sur les méthodes d'observation réalisées de manière indirecte et passive. En évitant toute interaction humaine, elles présentent l'avantage de limiter les éventuels impacts de l'évaluateur sur le comportement de l'usager. Par la suite, chaque séquence audiovisuelle

constitue un stimulus qui est caractérisé par l'évaluateur. Lors de cette étape de caractérisation, ce dernier est amené à retranscrire ce qu'il observe sur les vidéos. La procédure classique consiste à effectuer une transcription littérale. L'évaluateur traduit alors la vidéo en texte (Bailey, 2008; Garcez, Duarte, & Eisenberg, 2011). Cette caractérisation est 'perceptuelle' dans le sens où l'évaluateur décrit simplement de manière analytique ce qu'il observe sur les vidéos. C'est cette étape de caractérisation perceptuelle qui permet d'obtenir les données d'observation. Dans un second et dernier temps, l'évaluateur a la charge d'analyser ces données. Étant présentées sous la forme d'un corpus de textes, il les analyse grâce à des procédures dédiées au traitement de matériel écrit (Garcez et al., 2011).

De manière générale, les procédures de recueil et d'analyse des données d'observation sont coûteuses, difficiles à mettre en place et chronophages (Buck, 2007; von Hippel & Katz, 2002). En conséquence, le nombre d'utilisateurs observés en situation réelle d'usage est souvent limité. Plus précisément, les méthodes d'observation impliquent généralement une quinzaine d'utilisateurs au maximum.

La Figure 3 représente de manière schématique le protocole de mise en œuvre des méthodes d'observation abordé dans cette section.

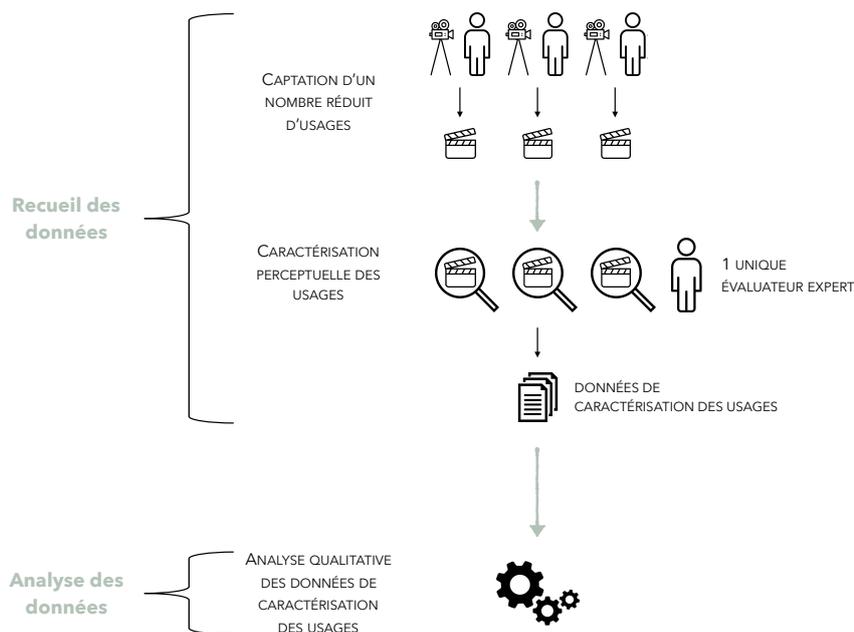


Figure 3: Protocole correspondant à une méthode d'observation qualitative permettant la caractérisation des usages.

1.1.2.3 Verrous et interrogations associées

La méthodologie décrite précédemment présente différents verrous techniques qui sont détaillés un à un dans cette section.

Tout d'abord, les données de caractérisation des stimuli peuvent refléter une certaine interprétation subjective de la part de l'évaluateur expert (van Kleef et al., 2005). La tâche de caractérisation doit normalement être réalisée de manière la plus objective possible. Il arrive cependant que la grille de lecture des vidéos soit conditionnée par des aspects relatifs à l'évaluateur tels que son approche théorique (Bonnin, 1998). Ce conditionnement peut, de la même façon, impacter les résultats de l'analyse qualitative des données de caractérisation. Ceci s'avère d'autant plus problématique que ces tâches sont réalisées par un seul et même évaluateur. Il est alors impossible d'étudier la variabilité liée à la subjectivité de la caractérisation des usages. Une manière d'aborder cette difficulté serait de diversifier les sources dont sont issues les données de caractérisation, en remplaçant l'unique évaluateur expert - potentiellement conditionné par son approche théorique - par plusieurs évaluateurs.

Ensuite, la caractérisation des usages se fait de manière uniquement perceptuelle. Et pourtant, comme tout stimulus, un usage peut à la fois être exploré à travers sa représentation perceptuelle et sa représentation conceptuelle. La représentation perceptuelle fait notamment référence aux caractéristiques intrinsèques du stimulus (p. ex. les différents comportements observés sur la vidéo correspondant à l'usage). La représentation conceptuelle définit, elle, à quel point le stimulus représente un concept donné selon l'évaluateur. Dans un contexte de R&D, il pourrait pourtant être intéressant d'étudier la représentation conceptuelle des usages. En distinguant ceux qui relèvent du concept de 'bon usage' de ceux qui n'en relèvent pas, il serait possible de déterminer les caractéristiques de l'usage qui permettent de véhiculer cette perception de 'bon usage'. Une façon de dépasser cette limite serait de réaliser également la caractérisation des usages de manière conceptuelle.

Enfin, comme énoncé précédemment, le nombre d'usages qui sont caractérisés est souvent limité. Le risque est alors de ne pas capter toute la variabilité existante. Il est également impossible d'évaluer la prédominance de certains usages par rapport à d'autres au sein de la population. Cette caractéristique est partagée avec un certain nombre de méthodologies qualitatives. Bien qu'elles fournissent des informations particulièrement intéressantes, elles ne sont pas statistiquement représentatives (Buck, 2007). Ce problème de non-représentativité est particulièrement important dans le cadre de la caractérisation des gestes culinaires. En effet, il existe une multitude de façons de réaliser un même geste culinaire. Le protocole de fabrication d'une mayonnaise peut ainsi être différent d'une personne à l'autre : chacune utilise des ustensiles culinaires différents ou effectue des gestes différents. Cette variété des gestes culinaires est si importante qu'un musée audiovisuel leur a récemment été dédié. Lancé par Hervé This dans le cadre du Centre de Gas-

tronomie Moléculaire, ce musée a vocation à répertorier différents gestes culinaires réalisés par des cuisiniers ou des amateurs. Une manière d’aborder cette difficulté serait de capter et de caractériser un nombre plus conséquent d’usages.

1.2 Caractérisation sensorielle de produits

1.2.1 Analyse sensorielle

L’analyse sensorielle (ou évaluation sensorielle) consiste à étudier les réponses engendrées par un produit sur les cinq sens de l’être humain (c.-à-d. la vue, l’odorat, le toucher, le goût et l’ouïe) (Stone & Sidel, 2004). Elle permet de comprendre et de prédire la réponse d’un consommateur à un produit de manière rigoureuse et précise. La vocation finale d’une telle approche est d’optimiser les chances de réussite commerciale des nouveaux produits développés.

Dans un contexte de développement de produit, l’analyse sensorielle peut être utilisée pour répondre à des problématiques opérationnelles très variées. Elle peut ainsi être utilisée pour comparer des produits avec des produits de la concurrence mais aussi pour déterminer si des changements de recettes modifient la perception des consommateurs. Une description détaillée des possibilités d’utilisation de cette approche dans un contexte industriel est présentée dans Talavera and Chambers (2017).

L’analyse sensorielle est utilisée dans des secteurs variés tels que l’industrie de la cosmétique ou de la parfumerie fine (Boinbaser, Parente, Castura, & Ares, 2015; Brard & Lê, 2016; Moussour, Lavarde, Pensé-Lhéritier, & Bouton, 2016; Worch, Crine, Gruel, & Lê, 2014), l’industrie textile (Zhu et al., 2010), ou l’industrie automobile (Blumenthal & Herberth, 2015; Giboreau, Navarro, Faye, & Dumortier, 2001; Kergoat et al., 2012). C’est cependant dans le secteur agroalimentaire que l’analyse sensorielle est la plus utilisée. L’intérêt précoce de ce secteur industriel pour cette approche peut expliquer cet attrait. L’analyse sensorielle a été utilisée pour la première fois dans les années 1940 dans l’objectif de prédire l’acceptabilité des rations alimentaires fournies aux forces armées américaines (Peryam, Pilgrim, & Peterson, 1954). Dès lors, le secteur agroalimentaire a compris l’intérêt de cette discipline dans un contexte de R&D et l’a intégrée au sein de ses entreprises. Bien qu’initialement dédiée à la caractérisation de produits ayant vocation à être ingérés, l’analyse sensorielle est désormais de plus en plus utilisée pour étudier d’autres types de stimuli. Elle est ainsi fréquemment utilisée pour caractériser des packagings (Heide & Olsen, 2017; Lefebvre et al., 2010; Murray & Delahunty, 2000; Tijssen, Zandstra, de Graaf, & Jager, 2017) ou des ustensiles culinaires (Carvalho & Spence, 2018; Cavazzana, Larsson, Hoffmann, Hummel, & Haeh, 2017; Faye, Courcoux, Giboreau, & Qannari, 2013; Mirabito, Oliphant, Doorn, Watson, & Spence, 2017).

1.2.2 Méthodes quantitatives de caractérisation sensorielle

1.2.2.1 Principe général

Les réponses humaines que l'analyse sensorielle permet d'étudier sont de deux types : les réponses dites 'perceptuelles' (c.-à.d. analytiques) et les réponses dites 'hédoniques'. Dans le premier cas, il s'agit de déterminer si des produits sont perçus différemment par les sens humains. Si tel est le cas, l'objectif est alors de déterminer sur quelles caractéristiques reposent ces différences. Dans le second cas, il s'agit de déterminer de quelle manière les produits sont appréciés (ou préférés les uns par rapport aux autres) par les consommateurs (Lawless & Heymann, 2010).

Les méthodes de caractérisation sensorielle consistent à mesurer ces réponses et à les analyser. Ces approches relèvent des sciences quantitatives dans le sens où, comme nous le verrons par la suite, elles mettent l'accent sur un recueil de données qui sont ensuite analysées via une procédure statistique.

1.2.2.2 Du recueil à l'analyse des données

La méthode classique pour le recueil et l'analyse des données sensorielles est la méthode du profil sensoriel conventionnel, qui prend notamment ses origines dans la *Quantitative Descriptive Analysis* (QDA) (Stone, Sidel, Oliver, Woolsey, & Singleton, 1974). Cette méthode est la référence dans le cadre de l'évaluation sensorielle puisque recommandée par les normes AFNOR (NF ISO 13299, Mai 2010). C'est donc sur celle-ci que nous nous focalisons ici pour introduire les méthodes quantitatives de caractérisation sensorielle.

Cette méthode a été développée dans l'objectif de fournir une description sensorielle des produits la plus fiable possible. Elle repose sur une décomposition du produit en une liste d'attributs sensoriels qui sont ensuite évalués sur une échelle graduée reflétant leur intensité. Plus précisément, une liste de 5 à 20 attributs est établie. Cette liste est sensée couvrir toutes les caractéristiques sensorielles de la gamme de produits considérée. Elle ne doit comporter que des attributs relatifs aux caractéristiques perceptuelles des produits. Tous les termes relatifs à un jugement hédoniques sont ainsi évincés. Les produits à caractériser sont ensuite présentés un à un à l'évaluateur. Pour chacun des produits, il lui est alors demandé de quantifier l'intensité de chaque attribut à l'aide d'une échelle graduée. Cette tâche de quantification est réalisée par tous les évaluateurs du panel. Les données numériques ainsi obtenues sont analysées par le biais de techniques statistiques telles que l'analyse de la variance ou l'Analyse en Composantes Principales.

Pour garantir sa précision et son objectivité, cette méthode est fondée sur l'utilisation d'un panel d'une dizaine d'évaluateurs qui présentent de bonnes performances. Un évaluateur est jugé performant dès lors qu'il est capable de discriminer les produits, qu'il est répétable dans le temps, et qu'il est en accord

avec les autres membres du panel. Cette performance n'est obtenue qu'au prix d'un entraînement intensif et régulier. Cet entraînement permet aux évaluateurs de devenir de véritables experts de l'évaluation sensorielle.

Au-delà de cet aspect, la performance de cette méthode dépend du nombre de produits évalués durant l'expérience. Il est en effet raisonnable de penser que, si l'expérience est trop longue, les évaluateurs peuvent être sujets à une certaine fatigue sensorielle. Celle-ci s'exprime notamment par une saturation du goût ou de l'odorat. L'apparition d'une telle fatigue chez les évaluateurs est problématique dès lors qu'elle peut altérer la qualité des données sensorielles collectées. Le nombre de produits évalués lors d'une tâche de profil sensoriel conventionnel varie ainsi de 5 à 15 produits (Pagès & Périnel, 2004). Dans les situations impliquant un plus grand nombre de produits, des solutions telles que l'organisation de plusieurs sessions d'évaluation ou l'utilisation de plans de présentation incomplets des produits peuvent être utilisées.

La Figure 4 représente de manière schématique le protocole de mise en œuvre des méthodes de caractérisation sensorielle abordé dans cette section.

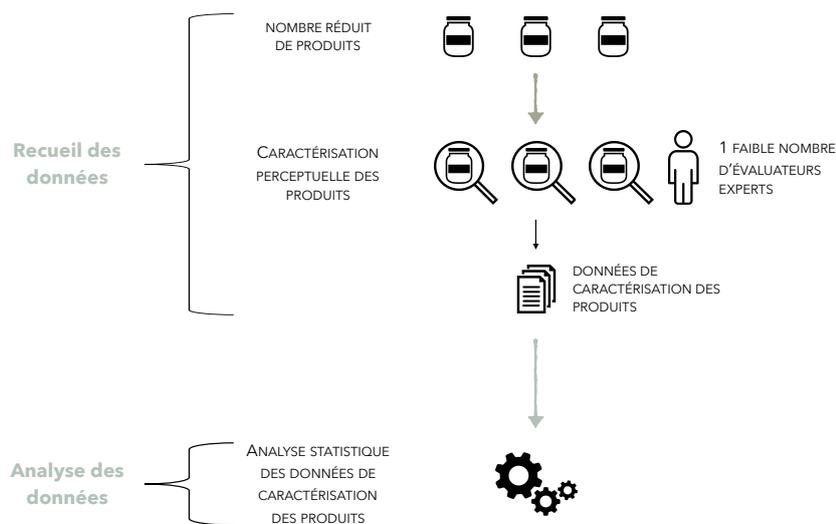


Figure 4: Protocole correspondant à une méthode sensorielle quantitative permettant la caractérisation de produits.

1.2.2.3 Verrous et interrogations associées

La méthodologie décrite précédemment présente différents verrous techniques qui sont détaillés un à un dans cette section.

Tout d'abord, l'entraînement intensif des évaluateurs sensoriels s'avère limitant. Il

est ainsi reconnu pour engendrer des coûts de temps et financiers non négligeables que les acteurs du secteur agroalimentaire cherchent à réduire (Delarue, 2015). Au delà de cet aspect, il fait des évaluateurs de véritables outils de mesure calibrés. À cause de ce conditionnement théorique des évaluateurs, il est impossible d'étudier la variabilité liée à la subjectivité de la caractérisation sensorielle. Une manière d'aborder cette difficulté serait de diversifier les sources dont sont issues les données sensorielles, en substituant le panel d'évaluateurs experts par un panel d'évaluateurs non-experts tels que des consommateurs. Nous verrons par la suite comment la résolution de cette interrogation a donné naissance à des développements méthodologiques spécifiques dans le domaine de l'analyse sensorielle.

Ensuite, la tâche de caractérisation sensorielle se fait de manière uniquement perceptuelle. Et pourtant, comme tout stimulus, un produit peut être à la fois exploré à travers sa représentation perceptuelle et sa représentation conceptuelle (Thomson, 2015). La représentation perceptuelle fait notamment référence aux caractéristiques intrinsèques du stimulus (p. ex. les différentes caractéristiques sensorielles du produit perçues par l'évaluateur). La représentation conceptuelle définit, elle, à quel point le stimulus représente un concept donné selon l'évaluateur. Dans un contexte de R&D, il pourrait cependant être intéressant d'étudier la représentation conceptuelle des produits. En distinguant ceux qui relèvent d'un concept ciblé par les développeurs (p. ex. le concept de fraîcheur) de ceux qui n'en relèvent pas, il serait possible d'identifier les caractéristiques sensorielles qui permettent de véhiculer ce concept. Ces caractéristiques constitueraient des leviers opérationnels de développement. Une façon de dépasser cette limite serait de réaliser également la caractérisation des produits de manière conceptuelle. Nous verrons par la suite comment la résolution de cette interrogation a, là encore, donné naissance à des méthodes de caractérisation sensorielle spécifiques.

Enfin, comme énoncé précédemment, le nombre de produits qui sont caractérisés est souvent limité. Le risque est alors de ne pas capter toute la variabilité existante. Cet aspect s'avère d'autant plus problématique dans le contexte actuel de l'hyper choix auquel est confronté le consommateur. Les situations impliquant un nombre important de produits à évaluer sont en effet de plus en plus fréquentes (Heymann, 2017). C'est par exemple le cas si le but de l'expérience est de caractériser toute une gamme de produits sans en sélectionner arbitrairement et subjectivement un sous-ensemble. Une manière d'aborder cette difficulté serait de développer des méthodes spécifiquement dédiées à la caractérisation sensorielle de grands ensembles de produits. Nous verrons par la suite que cette interrogation n'a été que peu étudiée en analyse sensorielle.

1.3 Analogie entre les interrogations rencontrées dans les deux domaines

Un parallèle peut être établi entre les différentes interrogations relatives à la caractérisation des gestes culinaires d'une part, et celles identifiées en analyse sensorielle

d'autre part. Cette comparaison est résumée dans le Tableau 1.

Table 1: Comparaison entre les interrogations rencontrées dans le domaine de la caractérisation des gestes culinaires d'une part, et dans le domaine de la caractérisation sensorielle de produits d'autre part. Il est spécifié si des démarches ont déjà été mises en œuvre dans le domaine pour résoudre ces interrogations.

Caractérisation de gestes culinaires		Caractérisation sensorielle de produits	
Interrogation	Résolution déjà entreprise	Interrogation	Résolution déjà entreprise
Substitution de l'évaluateur expert par un panel d'évaluateurs non-experts	Non	Substitution du panel d'évaluateurs experts par un panel d'évaluateurs non-experts	Oui
Ajout d'une dimension conceptuelle pour la caractérisation des gestes	Non	Ajout d'une dimension conceptuelle pour la caractérisation des produits	Oui
Caractérisation d'un nombre important de gestes	Non	Caractérisation d'un nombre important de produits	Non

2 Méthodologie pour la caractérisation perceptuelle et conceptuelle de grands ensembles de stimuli par des évaluateurs non-experts

2.1 Problématique de recherche

Dans la section précédente, la caractérisation sensorielle de produits a été décrite comme une approche plébiscitée par les industriels du secteur agroalimentaire. En permettant de répondre à de nombreuses problématiques relatives au développement de produit, elle s'inscrit pleinement dans leur stratégie d'innovation. Bien que très peu abordée dans la littérature, la caractérisation des gestes culinaires semble également s'inscrire comme un nouvel outil pour l'innovation agroalimentaire.

Cependant, nous avons vu que ces deux approches présentent des verrous qui limitent leur utilisation. Ces derniers sont similaires d'une approche à l'autre. Nous retrouvons parmi ceux-ci la possibilité de caractériser un nombre important de stimuli, d'ajouter une dimension conceptuelle à la tâche de caractérisation, et de substituer le panel d'évaluateurs experts par des non-experts. Même si l'analogie entre les interrogations rencontrées dans ces deux domaines est directe, la possibilité d'utiliser des outils similaires pour y répondre ne l'est pas. Aussi,

nous avons rappelé que, à ce jour, aucun outil ne permet de résoudre totalement ces interrogations dans l'un ou l'autre des deux domaines d'application. Nous avons cependant précisé que le domaine de l'analyse sensorielle a déjà mis en place des réflexions scientifiques pour y remédier.

L'objectif de ce travail de recherche est de développer deux méthodologies permettant de dépasser ces limites. Plus précisément, elles doivent reposer sur la caractérisation d'un grand ensemble de stimuli par un panel d'évaluateurs non-experts. La tâche de caractérisation doit être réalisée de manière perceptuelle dans le cadre de la première méthodologie, et de manière conceptuelle dans le cadre de la seconde. Elles doivent toutes les deux présenter la flexibilité d'être à la fois applicables sur des stimuli correspondant à des vidéos représentant des gestes culinaires, et sur des stimuli correspondant à des produits industriels. De cette manière, ce travail de recherche est dédié à l'apport d'une contribution conjointe aux deux domaines de caractérisation dont il est question ici.

2.2 Démarche de recherche privilégiée

Une manière d'aborder cette problématique est de tirer avantage des méthodes quantitatives sensorielles pour développer les deux méthodologies, et d'étudier ensuite leur applicabilité sur des vidéos de gestes culinaires. C'est cette démarche de recherche que nous avons privilégiée dans le cadre de ce travail.

Le domaine de l'analyse sensorielle est en effet celui qui présente une plus grande expérience en ce qui concerne la résolution des interrogations abordées ici (voir Table 1). Malgré tout, aucune méthodologie sensorielle ne permet de les régler totalement. La première phase de la démarche de recherche consiste alors à améliorer les méthodes sensorielles existantes pour répondre à ces interrogations. Si ce développement s'avère concluant, se pose ensuite la question de la possibilité d'utiliser ces méthodes 'améliorées' dans le cadre de la caractérisation des gestes culinaires. Rappelons que c'est cette problématique a initialement motivé la mise en place de ce projet. La deuxième phase de la démarche de recherche consiste ainsi à changer la nature des stimuli considérés, en utilisant dans ce cadre des vidéos de gestes culinaires réalisés par des usagers. Une représentation schématique de cette démarche de recherche est exposée sur la Figure 5.

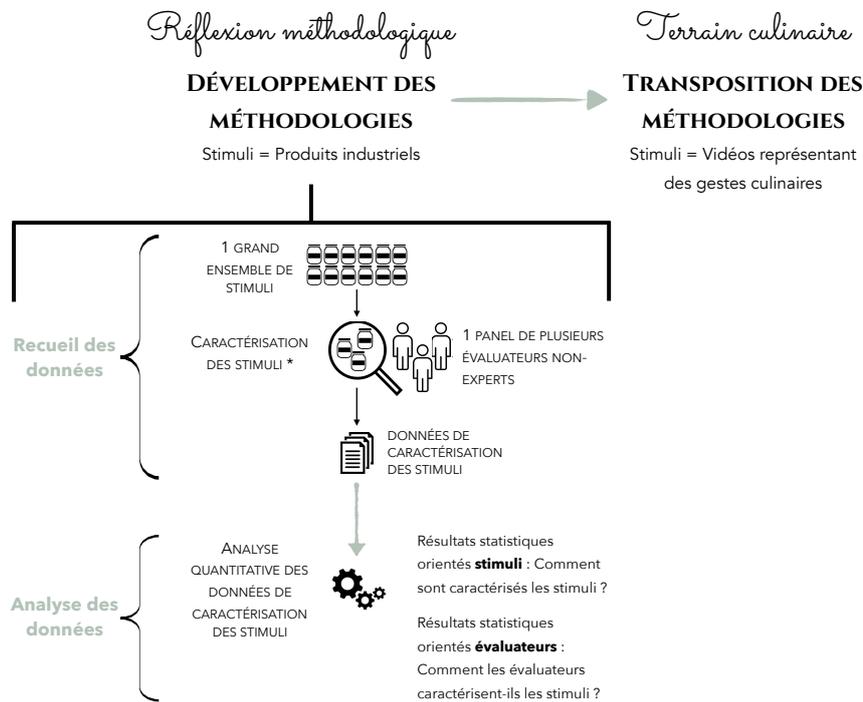


Figure 5: Démarche de recherche mise en place dans le cadre de ce projet.

* La caractérisation est réalisée de manière perceptuelle pour l'une des méthodologies, et de manière conceptuelle pour l'autre.

Ce travail de recherche s'avère être pluridisciplinaire. Il se situe à l'interface entre deux grandes familles de disciplines scientifiques souvent jugées antagonistes : d'une part la recherche qualitative (à travers les méthodes d'observation des usages), et d'autre part la recherche quantitative (à travers les méthodes de caractérisation sensorielle). L'originalité de ce projet peut donc être abordée au travers de plusieurs points de vue.

Observée sous l'angle de la recherche qualitative, elle repose sur le développement d'une méthodologie permettant d'appréhender des vidéos d'usages de manière quantitative. Ceux-ci étaient jusqu'alors abordés de manière qualitative. En tirant avantage d'une démarche quantitative, il est possible d'obtenir une caractérisation d'un nombre important d'usages grâce à plusieurs évaluateurs non-experts. Grâce à une analyse statistique des données de caractérisation collectées, il est ainsi possible d'étudier la variabilité des usages étudiés. De la même façon, il est possible d'étudier la variabilité liée à la subjectivité du jugement fourni par les évaluateurs.

Observée sous l'angle de la recherche quantitative, le caractère innovant de ce projet repose principalement sur le format inhabituel des stimuli étudiés. L'utilisation de vidéos représentant des usages est en effet peu courante, si ce n'est jamais vue. Ce développement répond à l'une des attentes actuelles du domaine de l'analyse

sensorielle et des sciences du consommateur. Celui-ci désire augmenter le nombre d'études réalisées dans des conditions réelles de vie pour obtenir une meilleure validité écologique (Jaeger et al., 2017).

De manière plus globale, l'originalité de ce projet de recherche relève de l'objet d'étude dont il est principalement question ici : le geste culinaire. À notre connaissance, l'étude de cet objet n'a jamais donné lieu à la réalisation de travaux scientifiques dédiés.

2.3 Pistes envisagées

Nous explorons ici les solutions identifiées en analyse sensorielle pour répondre aux interrogations soulevées dans le cadre de cette thèse. Celles-ci sont des amorces de pistes méthodologiques que nous privilégions pour aborder la problématique de recherche. En particulier, nous nous attachons à présenter les approches mises en place pour substituer les évaluateurs experts par des non-experts, pour caractériser les stimuli (c.-à-d. les produits) de manière à la fois perceptuelle et conceptuelle, et pour gérer la caractérisation de grands ensembles de stimuli.

Tout d'abord, la problématique de la substitution du panel d'évaluateurs experts par des non-experts a été résolue il y a quelques années. Dans un premier temps, il a été démontré que ces derniers sont capables d'établir une caractérisation sensorielle de qualité. Plus précisément, ils sont en mesure de fournir des données sensorielles comparables à celles des évaluateurs experts (Bruzzone, Ares, & Giménez, 2012; Husson, Le Dien, & Pagès, 2001; Moskowitz, 1996; Worch, Lê, & Punter, 2010). Au-delà de leur capacité à caractériser les produits, il a été prouvé dans un second temps que les évaluateurs non-experts (généralement les consommateurs) apportent un point de vue intéressant dans un contexte de R&D. En identifiant la manière avec laquelle ils perçoivent les produits, il est possible de dégager des leviers de développement ou d'amélioration (MacFie, 2007). Cette intégration de l'évaluateur non-expert dans le processus d'analyse sensorielle a orienté la recherche vers le développement de nouvelles méthodes de recueil et d'analyse de données sensorielles. Les évaluateurs n'étant plus entraînés à la caractérisation sensorielle, l'utilisation de la méthode du profil sensoriel conventionnel était en effet compromise. La première direction prise à travers ces orientations a consisté à augmenter la taille du panel. En augmentant le nombre d'évaluateurs, il est en effet possible d'obtenir une meilleure discrimination statistique des produits (Delarue, 2015). Ceci permet en partie de contrebalancer le manque d'entraînement et d'expertise du panel (Worch, 2012). Alors qu'une dizaine d'évaluateurs experts sont nécessaires en profil sensoriel conventionnel, les panels de non-experts sont généralement composés d'une cinquantaine d'évaluateurs (Delarue, 2015). La seconde direction prise a consisté à développer des méthodologies de recueil de données sensorielles spécifiques. Celles-ci sont moins analytiques et plus faciles à utiliser par des évaluateurs non-entraînés que la méthode conventionnelle.

Par la suite, nous présentons deux méthodes sensorielles développées spécifiquement dans ce cadre. Celles-ci présentent des propriétés liées aux autres interrogations identifiées dans le cadre de ce travail. C'est pour cette raison que nous avons choisi de présenter ces deux approches en particulier parmi toutes celles dédiées aux évaluateurs non-experts. Nous verrons dans quelle mesure elles peuvent être utilisées pour caractériser - de manière perceptuelle ou conceptuelle - un nombre important de produits.

2.3.1 Caractérisation perceptuelle : focus sur la méthodologie du tri libre

La quasi-totalité des méthodologies sensorielles dédiées aux évaluateurs non-experts permet d'appréhender la représentation perceptuelle des produits. L'enthousiasme historique des spécialistes de l'analyse sensorielle pour l'étude de cette représentation peut expliquer cet essor.

Ces méthodologies peuvent être divisées en deux grandes familles : les méthodologies d'élicitation directe d'une part, et les méthodologies d'élicitation indirecte d'autre part. Les premières supposent qu'il existe une relation étroite entre une sensation sensorielle et sa description verbale (Bech & Zacharov, 2006). Le produit est alors évalué à travers sa décomposition en une liste d'attributs sensoriels. Le Check-All-That-Apply est l'une de ces approches (Adams, Williams, Lancaster, & Foley, 2007). Elle consiste à demander à l'évaluateur de cocher, au sein d'une liste d'attributs sensoriels prédéfinis, tous ceux qui représentent le produit. Dès lors, la représentation perceptuelle des produits est établie conditionnellement aux attributs prédéfinis. C'est l'une des critiques pouvant être faites à ce type de procédures. Au contraire, les méthodologies d'élicitation indirecte supposent qu'une sensation sensorielle peut être séparée de sa description verbale (Bech & Zacharov, 2006). Le produit est alors évalué dans sa globalité, et non à travers sa décomposition en une liste d'attributs sensoriels. Ces approches consistent à demander à l'évaluateur d'étudier les similarités et les différences sensorielles globales qu'il perçoit au sein de l'ensemble des produits. L'évaluateur doit utiliser son propre prisme de perception pour réaliser cette tâche. Plus précisément, il utilise ses propres critères pour évaluer les produits et n'est pas obligé de spécifier ces critères. Dès lors, la représentation perceptuelle des produits est établie indépendamment de tout attribut prédéfini. Leur capacité à capturer la perception sensorielle brute des évaluateurs, et en particulier des consommateurs, fait de ces méthodologies d'élicitation indirecte des approches populaires en analyse sensorielle.

La tâche de tri libre, également appelée tâche de catégorisation, est l'une de ces approches. Elle a été extrapolée à partir du domaine de la psychologie (Lawless, 1989). Elle consiste à demander à l'évaluateur de former des groupes homogènes de produits (voir Figure 6). Deux produits doivent être dans le même groupe (resp. dans des groupes différents) s'ils présentent des caractéristiques perceptuelles similaires (resp. différentes). Comme pour toute méthodologie d'élicitation

indirecte, cette tâche doit être réalisée selon les propres critères de l'évaluateur.

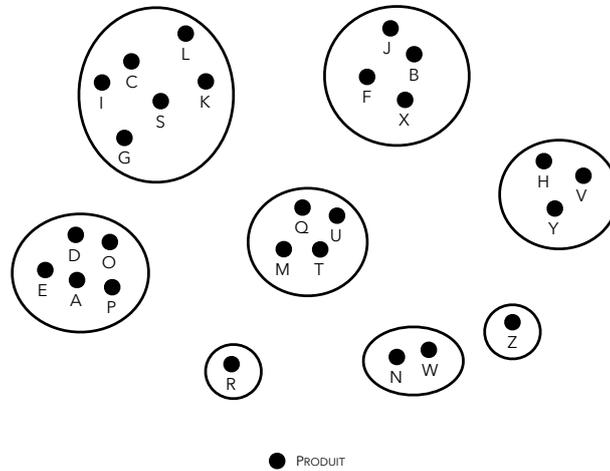


Figure 6: Exemple d'une tâche de tri libre réalisée par un évaluateur sur 26 produits. Cet évaluateur a créé 8 groupes de produits.

Les données recueillies à l'échelle du panel sont ensuite analysées au travers de deux types de procédures statistiques. La première est orientée 'produits'. Elle a pour objectif d'étudier la variabilité perceptuelle observée au sein de l'ensemble de produits. Grâce à cette procédure, il est possible d'identifier les principales similitudes et différences sensorielles observées parmi les produits. La seconde est, elle, orientée 'évaluateurs'. Elle a pour objectif d'étudier la variabilité de perception observée au sein du panel.

Malgré ses nombreux avantages, la tâche de tri libre est limitée par le nombre de produits qu'elle permet d'évaluer (Chollet, Valentin, & Abdi, 2014; Courcoux, Qannari, & Faye, 2015; Derndorfer & Baiertl, 2014; Fleming, Ziegler, & Hayes, 2015; Rogeaux, Lawlor, Punter, & Delarue, 2011). Cette méthode est en effet basée sur une procédure de comparaison des produits. Lorsque le nombre de produits devient trop important, une saturation de la mémoire et une fatigue sensorielle peuvent apparaître chez les évaluateurs. Celles-ci peuvent entraîner des problèmes de confusion entre les produits et alors altérer la qualité des données sensorielles recueillies (Chollet et al., 2014). Il existe peu de références dans la littérature concernant le nombre de produits optimal à considérer lors d'une tâche de tri libre. Le fait que ce nombre optimal dépende grandement de la nature du produit peut expliquer cet aspect. Entraînant moins de saturation de la mémoire et de fatigue, certains types de produits sont en effet plus difficiles à évaluer que d'autres. Certaines tendances générales peuvent pourtant être trouvées dans la littérature. Le nombre de produits varie généralement de dix à vingt produits (Lawless & Heymann, 2010). Certains auteurs font par ailleurs référence à un

nombre optimal de douze produits (Chollet et al., 2014).

Cependant, les expérimentateurs sont de plus en plus confrontés à des ensembles de produits dont la taille dépasse ces recommandations. Dans le cadre de certaines procédures sensorielles, il est important que l'espace produit considéré présente une variabilité perceptuelle importante. C'est par exemple le cas de la cartographie externe des préférences (*preference mapping* en anglais). Sa vocation étant d'identifier les caractéristiques perceptuelles qui impactent la préférence des consommateurs, il est primordial que les produits évalués soient très différents sur le plan sensoriel. L'expérimentateur peut alors être amené, en amont, à soumettre un grand nombre de produits à une tâche de tri libre. Il peut ainsi en extraire un sous-ensemble aussi variable que possible. Comment procéder dans ces situations impliquant un grand ensemble de produits en tri libre ? La littérature met en évidence deux solutions principales à ce problème.

La première solution repose sur une présentation des produits selon un plan d'expérience incomplet (Chollet et al., 2014; Courcoux et al., 2015). Chaque évaluateur n'évalue alors qu'un sous-ensemble composé d'un nombre restreint de produits. Cependant, l'évaluateur ne fournit qu'une évaluation perceptuelle partielle de l'ensemble de produits. Bien que l'intégralité des produits soit évaluée à l'échelle du panel, les dimensions de variabilité perceptuelle identifiées par l'analyse statistique des données restent approximées.

La seconde solution n'a, pour le moment, donné lieu à aucune publication scientifique (Chollet et al., 2014). Elle consisterait à utiliser comme prototypes certains produits du grand ensemble (Chollet, Lelièvre, Abdi, & Valentin, 2011). Ceux-ci seraient choisis en amont par l'expérimentateur. Le grand ensemble de produits serait ensuite divisé en plusieurs sous-ensembles. Chaque évaluateur devrait alors comparer sensoriellement chaque sous-ensemble de produits avec les prototypes. Les évaluations obtenues sur les différents sous-ensembles seraient par la suite agrégées. Cette procédure fournirait, à l'échelle individuelle, une évaluation perceptuelle de l'intégralité des produits. Contrairement à l'utilisation d'un plan de présentation incomplet, cette solution présenterait ainsi l'avantage d'identifier des dimensions de variabilité perceptuelle qui ne seraient pas approximées. Cependant, elle serait limitée par le fait que, contrairement à la tâche de tri libre classique, l'évaluateur n'aborderait pas l'espace produit à travers son propre prisme de perception. Plus précisément, la perception de l'évaluateur serait conditionnée par les prototypes prédéfinis. Il est en effet raisonnable de penser que le choix des prototypes pourrait affecter la manière avec laquelle il perçoit les produits. Cet impact potentiel des prototypes a d'ailleurs été prouvé dans le cadre d'autres méthodologies similaires comme le Polarized Sensory Positioning (Teillet, 2015) : de Saldamando, Delgado, Herencia, Giménez, and Ares (2013); Teillet, Schlich, Urbano, Cordelle, and Guichard (2010).

En conséquence, les techniques traditionnellement utilisées en tri libre pour gérer des grands ensembles de produits ne sont pas optimales. Cette méthode n'est donc pas directement utilisable pour caractériser de manière perceptuelle un grand nombre de produits dans le cadre de ce travail de recherche.

2.3.2 Caractérisation conceptuelle : focus sur la Q-méthodologie binaire

Pendant longtemps, les spécialistes de l'analyse sensorielle n'ont porté que peu d'attention à la représentation conceptuelle des produits. Alors que de nombreuses méthodes sont disponibles pour étudier la représentation perceptuelle d'un produit, peu d'entre elles sont spécifiquement dédiées à l'étude de sa représentation conceptuelle.

Bien que des processus de classement soient également rapportés dans la littérature (Bucher, Collins, Diem, & Siegrist, 2016; Lee & O'Mahony, 2005), les méthodologies classiquement dédiées à l'étude de la représentation conceptuelle des produits sont principalement basées sur un processus de notation. Ce dernier peut impliquer une échelle non structurée (Ballester, Dacremont, Fur, & Etiévant, 2005) ou structurée (Carr, Craig-Petsinger, & Hadlich, 2001; Picard, Tempere, de Revel, & Marchand, 2015). Chaque évaluateur indique sur cette échelle dans quelle mesure le produit représente le concept donné. En raison de la nature intrinsèque de la tâche, les concepts qui peuvent être étudiés grâce à ces approches sont ceux qui peuvent être mesurés à l'aide d'un gradient (c.-à-d. dont l'intensité peut être facilement appréhendée par l'évaluateur).

Cependant, certains concepts peuvent difficilement être étudiés via ce type de mesure quantitative. C'est le cas des concepts multidimensionnels. C'est de l'étude de ces concepts dont il est question dans le cadre de cette thèse, ceux-ci étant décrits par la suite comme des concepts 'complexes'. L'évaluation de ces derniers repose *a priori* sur un ensemble de plusieurs critères. C'est par exemple le cas du concept d'innovation. Ces concepts complexes s'opposent ainsi aux concepts unidimensionnels (p. ex. le concept de fraîcheur), dont l'évaluation repose *a priori* sur un unique critère. Le nombre de critères impliqués dans l'évaluation étant plus important, les concepts multidimensionnels sont plus complexes à appréhender par un évaluateur que les concepts unidimensionnels. L'utilisation d'une échelle de notation pour les évaluer peut alors s'avérer compromise.

Certains auteurs ont proposé une méthode spécifiquement dédiée à l'étude de ces concepts complexes. Cette dernière implique une tâche de tri basée sur deux catégories prédéfinies : l'une représente le concept et l'autre représente sa négation (voir Figure 7). L'évaluateur est chargé de trier les produits en utilisant ces deux catégories. La première utilisation de ce type de protocole a été mise en place dans le cadre des travaux de Jaffré, Valentin, Dacremont, and Peyron (2009). L'objectif était d'étudier le concept de potentiel de vieillissement à travers un ensemble d'échantillons de vins rouges.

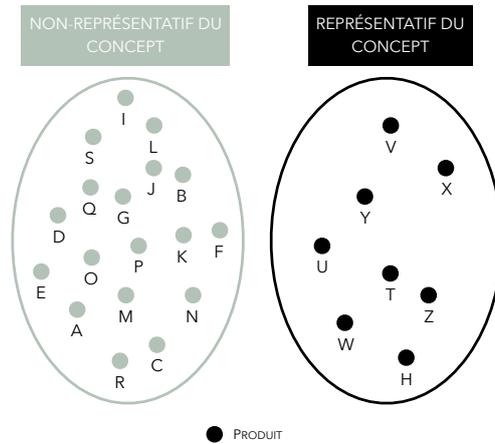


Figure 7: Exemple d'un tri basé sur deux catégories prédéfinies (l'une représentant le concept et l'autre représentant sa négation) réalisée par un évaluateur sur 26 produits.

Le choix de l'ensemble de produits utilisés lors de l'expérience doit être fait avec soin. Il doit inclure à la fois des produits représentatifs du concept et des produits non-représentatifs du concept. Les exemples et contre-exemples jouent en effet un rôle important dans la définition d'un concept (Vezin, 1972). Plus le concept est subjectif, plus ce choix des produits est difficile. Comment identifier des produits représentatifs et d'autres non-représentatifs si l'on s'attend à ce que la perception du concept varie sensiblement d'une personne à l'autre ? Dans ces situations, il est raisonnable de choisir un grand nombre de produits. Ceci permet d'augmenter la probabilité d'avoir à la fois des exemples du concept et des contre-exemples dans l'ensemble de produits.

Ce type de protocole de recueil de données présente de fortes similitudes avec le Q-sort, une des étapes de la Q-méthodologie. Cette approche est utilisée en psychologie pour étudier la subjectivité humaine (c.-à-d. la diversité d'opinion à propos d'un sujet d'étude) (Stephenson, 1953). Elle peut être décomposée en trois étapes principales. Dans un premier temps, l'expérimentateur choisit des assertions qui sont intentionnellement en relation avec le sujet d'étude. Celles-ci ayant vocation à comprendre en finesse le sujet d'étude, l'expérimentateur en choisit généralement un grand nombre. Une trentaine d'assertions est considérée comme une quantité courante (Davis & Michelle, 2011). Dans un second temps, les évaluateurs du panel trient individuellement les assertions dans des catégories prédéfinies qui correspondent à un continuum d'accord (voir Figure 8). C'est pourquoi cette tâche, dénommée 'Q-sort', est souvent considérée comme un cas particulier de tâche de tri fixée et graduée (Coxon, 1999). Dans un troisième et dernier temps, l'expérimentateur analyse les données correspondant aux Q-sorts fournis par les évaluateurs. Les catégories utilisées dans les Q-sorts étant ordinales, ces données sont également de nature ordinale. Les techniques statistiques traditionnellement utilisées pour

analyser ces données relèvent principalement des techniques d'analyse en facteurs (Brown, 1993; Watts & Stenner, 2005). Celles-ci permettent d'extraire des variables latentes (ou facteurs) qui maximisent la part de variance commune entre les variables initiales (c.-à-d. entre les Q-sorts des évaluateurs) (Kim & Mueller, 1978). Ces variables latentes peuvent être considérées comme des types d'opinion à propos du sujet d'étude. L'analyse en facteurs fournit à l'expérimentateur des *factor loadings*. Ils représentent la corrélation entre les variables initiales (c.-à-d. les Q-sorts des évaluateurs) et chaque variable latente extraite (c.-à-d. les types d'opinion). Ceux-ci permettent ainsi de déterminer à quel type d'opinion est relié chaque évaluateur du panel. L'expérimentateur a également à sa disposition des *factor scores*. Ils représentent la moyenne des scores donnés à chaque individu statistique (c.-à-d. à chaque assertion) pour chacune des variables initiales (c.-à-d. par chacun des évaluateurs) liées au facteur extrait (c.-à-d. au type d'opinion). Ils sont, eux, utilisés pour définir un Q-sort moyen pour chaque type d'opinion. Cette analyse statistique permet ainsi d'obtenir des groupes d'évaluateurs, chacun exprimant un type d'opinion du sujet étudié. Pour plus de détails sur la Q-méthodologie, le lecteur peut se référer aux travaux de Brown (1993) et Brown (2000) ainsi qu'à quelques exemples de mise en œuvre de la Q-méthodologie en sciences du consommateur (ten Klooster, Visser, & de Jong, 2008; Zhang, Lusk, Miroso, & Oey, 2016).

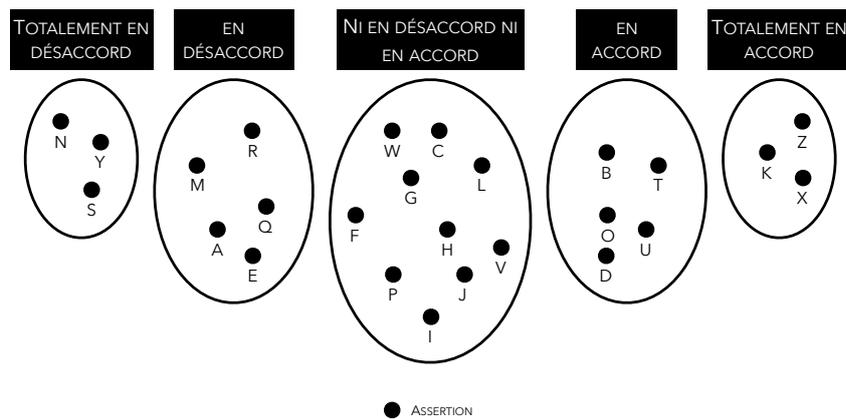


Figure 8: Exemple d'un Q-sort réalisé par un évaluateur avec 26 assertions.

Certaines similitudes existent entre la tâche de tri basée sur deux catégories prédéfinies représentant le concept et sa négation et le Q-sort. Il s'agit effectivement de deux outils dédiés à la mesure de la subjectivité qui reposent sur une tâche de tri impliquant des catégories prédéfinies par l'expérimentateur. Ces similitudes ont conduit certains auteurs à clairement définir la tâche de tri basée sur deux catégories prédéfinies représentant le concept et sa négation comme un Q-sort (Courcoux et al., 2015). Cependant, certaines différences majeures entre les deux protocoles doivent être soulignées.

La première de ces différences découle directement de la nature des catégories

prédéfinies par l'expérimentateur. Alors que des catégories ordinales sont utilisées dans un Q-sort, ce sont des catégories catégorielles - et plus précisément binaires - qui sont utilisées lors d'une tâche de tri basée sur deux catégories représentant le concept et sa négation. De cette différence de nature des catégories résulte naturellement une différence entre les types de données qu'elles permettent de recueillir. Il s'agit de données ordinales dans un cas, et de données catégorielles - et plus précisément binaires - dans l'autre cas.

La deuxième de ces différences repose sur les types d'informations que l'expérimentateur va chercher à extraire des données recueillies. Dans le cadre d'un Q-sort, la préoccupation principale de l'expérimentateur est d'identifier des groupes homogènes d'évaluateurs. Dans ce contexte, il n'accorde que peu d'importance aux assertions ayant été utilisées lors du Q-sort. Elles sont seulement utilisées pour comprendre avec plus de finesse les groupes d'évaluateurs identifiés. Dans le cadre d'une tâche de tri basée sur deux catégories représentant le concept et sa négation, la préoccupation principale de l'expérimentateur est de comprendre de quelle manière les différents produits représentent le concept. Cependant, ces représentations conceptuelles peuvent être sujettes à de la subjectivité individuelle. Dès lors, l'expérimentateur fait également de l'évaluateur un objet d'étude. Plus précisément, il cherche là aussi à identifier des groupes homogènes d'évaluateurs ayant une perception similaire du concept à travers l'ensemble de produits. Ces différences d'objectifs des expérimentateurs ne sont pas propres à ce cadre. Elles sont en effet courantes lorsqu'il s'agit de confronter un psychologue à un spécialiste de l'analyse sensorielle. Alors que le psychologue s'intéresse essentiellement à l'individu, le spécialiste de l'analyse sensorielle s'intéresse à la fois au produit et à l'interaction qu'il existe entre ce dernier et l'individu (Lawless & Heymann, 2010).

La méthodologie basée sur deux catégories prédéfinies représentant le concept et sa négation apparaît alors comme une version binaire et adaptée à la caractérisation conceptuelle de grands ensembles de produits de la Q-méthodologie classique. Pour cette raison, et pour éviter toute confusion, nous nous référons par la suite au terme de 'Q-méthodologie binaire' pour décrire cette méthodologie.

Il apparaît que la Q-méthodologie classique et la Q-méthodologie binaire diffèrent sensiblement du point de vue de l'analyse statistique des données qu'elles permettent de recueillir. Ces différences sont résumées dans la Table 2. Dès lors, les techniques statistiques utilisées pour analyser les données de Q-méthodologie ne sont pas appropriées pour analyser les données de Q-méthodologie binaire. C'est une procédure statistique spécifique qui doit être utilisée. Malheureusement, un nombre restreint de réflexions scientifiques ont été menées dans ce cadre. Aussi, le peu de travaux entrepris, notamment par Jaffré et al. (2009), ne permettent pas de tirer partie de toute la richesse des données de Q-méthodologie binaire.

Table 2: Processus d'analyse statistique des données de Q-méthodologie et des données de Q-méthodologie binaire : *inputs* et *outputs*.

Méthodologie	<i>Inputs</i> →	Procédure statistique →	<i>Outputs</i>
Q-méthodologie	Données ordinales	Analyse en facteurs	Orientés évaluateur
Q-méthodologie binaire	Données catégorielles (binaires)	?	Orientés produit & orientés évaluateur

En conséquence, la procédure d'analyse statistique des données de Q-méthodologie binaire n'est pas accomplie. Cette méthode n'est donc pas directement utilisable pour caractériser de manière conceptuelle un grand nombre de produits dans le cadre de ce travail de recherche.

3 Organisation de la thèse

Cette thèse est articulée autour de deux parties principales. La Partie 1 présente la méthodologie dédiée à la caractérisation perceptuelle d'un grand nombre de gestes culinaires. La Partie 2 concerne la méthodologie dédiée à la caractérisation conceptuelle d'un grand nombre de gestes culinaires.

3.1 Caractérisation perceptuelle

L'incapacité du tri libre à être appliqué sur un nombre important de stimuli constitue l'obstacle principal à son utilisation dans le cadre de ce projet de recherche ; et ce malgré le développement de solutions spécifiques (ntmt. l'utilisation de plans de présentation incomplets et l'utilisation de prototypes prédéfinis). Pour dépasser cette limite, nous proposons dans cette thèse une amélioration à cette méthodologie du tri libre : le Tri Séquentiel Agglomératif (TSA). L'applicabilité de cette nouvelle approche pour la caractérisation de gestes culinaires a ensuite été étudiée. Cet axe de travail est développé dans la Partie 1, partie qui comprend les Chapitres numérotés de 1 à 3 ainsi qu'une Annexe.

Le Chapitre 1 présente la procédure de recueil des données de TSA. Celle-ci constitue le cœur innovant de cette nouvelle méthode de caractérisation sensorielle. Nous verrons dans ce cadre comment le TSA peut être vu comme une version particulière du tri libre verbalisé permettant de gérer des grands ensembles de stimuli. Afin d'éprouver ce nouveau protocole de recueil de données, un processus de validation basé sur la réalisation de plusieurs études a été défini. Ce Chapitre 1 présente les résultats de la première étude menée dans ce cadre. Comme nous le verrons, les résultats obtenus sont encourageants.

Le Chapitre 2 est dédié à la procédure d'analyse statistique des données de TSA. Dans ce chapitre, nous verrons que les données de TSA présentent les caractéristiques usuelles des données de tri libre verbalisé. Une étude plus

détaillée de leur format nous amènera à voir qu'elles présentent également des caractéristiques qui leur sont propres. Ceci nous amène à proposer une réflexion méthodologique spécifique pour leur analyse. Nous verrons que la procédure statistique proposée permet de comprendre comment les stimuli ont été caractérisés (approche orientée stimuli). Elle permet également de comprendre comment les évaluateurs ont effectué cette tâche de caractérisation (approche orientée évaluateurs). L'intégralité de ce Chapitre 2 est motivé par une étude dont l'objectif était d'obtenir la caractérisation sensorielle d'un ensemble de 40 parfums de gels douche, cette caractérisation étant réalisée par 60 consommateurs. Les différentes notions abordées dans ces deux premiers chapitres (le Chapitre 1 et le Chapitre 2) ont fait l'objet d'un article actuellement en révision.

Le Chapitre 3 présente le terrain mené au Centre Culinaire Contemporain dans l'objectif de tester l'applicabilité du TSA sur un grand ensemble de gestes culinaires. Son objectif était de caractériser 27 recettes de tarte au citron meringuée, la tâche de caractérisation étant réalisée par 66 évaluateurs. Ce terrain a permis de démontrer les bonnes performances de la méthodologie du TSA pour la caractérisation perceptuelle de grands ensembles de gestes culinaires.

Enfin, l'Annexe de cette partie présente l'implémentation dans R de la procédure statistique dédiée à l'analyse des données de TSA présentée dans le Chapitre 2. Cette implémentation est réalisée sous la forme du développement d'un package R.

3.2 Caractérisation conceptuelle

À notre connaissance, il n'existe pas de démarche statistique optimale des données que la Q-méthodologie binaire permet de recueillir. Ceci constitue l'obstacle principal à l'utilisation de cette méthodologie dans le cadre de ce projet de recherche. Nous proposons dans cette thèse une procédure d'analyse statistique *ad hoc* qui permet de prendre en considération la nature des données. Associée à la procédure de recueil de données, elle a donné naissance à une version améliorée de la Q-méthodologie binaire. Son applicabilité pour la caractérisation de gestes culinaires a ensuite été étudiée. Cet axe de travail est développé dans la Partie 2 de cette thèse, partie qui comprend les Chapitres numérotés de 4 à 6 ainsi qu'une Annexe.

Le Chapitre 4 est dédié à la présentation de la procédure d'analyse statistique orientée stimuli des données. Elle a pour objectif de déterminer comment les stimuli sont associés au concept étudié. Le cœur de cette analyse repose sur la définition d'une distance inter-stimuli. Celle-ci reflète de manière spécifique une différence de représentation du concept. Nous nous attachons à démontrer que les techniques statistiques classiquement utilisées pour obtenir une telle distance ne sont pas adaptées à la nature des données collectées. Plus précisément, nous illustrons comment elles occultent une partie de l'information véhiculée par celles-ci. Ceci nous amène par la suite à proposer une procédure *ad hoc*.

L'intégralité de ce Chapitre 4 est motivée par une étude dont l'objectif était d'obtenir la caractérisation d'un ensemble de 39 parfums de shampoings selon le concept d'innovation, cette caractérisation étant réalisée par 120 consommateurs. Les différentes notions abordées dans ce Chapitre 4 ont fait l'objet d'un article publié : Brard & Lê (2017).

Le Chapitre 5 présente la procédure d'analyse statistique orientée évaluateurs des données. Elle a pour objectif de déterminer comment les évaluateurs perçoivent le concept au sein de l'ensemble de stimuli. Le cœur de cette analyse repose sur la définition d'une distance inter-évaluateurs. Celle-ci reflète de manière spécifique une différence de perception du concept. Cette problématique n'a jamais été abordée dans un contexte de Q-méthodologie binaire. Et pourtant, certaines techniques statistiques utilisées en psychologie permettent d'analyser des données binaires similaires selon une approche orientée évaluateurs. Nous verrons cependant que celles-ci ne permettent pas d'analyser la structure de désaccord au sein du panel de manière optimale. Ceci nous amène, là encore, à proposer une procédure permettant de dépasser cette limite. L'intégralité de ce Chapitre 5 est motivée par l'étude déjà introduite dans le Chapitre 4. Les différentes notions abordées dans ce Chapitre 5 ont fait l'objet d'un article actuellement en révision.

Le Chapitre 6 présente le terrain mené au Centre Culinaire Contemporain dans l'objectif de tester l'applicabilité de la Q-méthodologie binaire sur un grand ensemble de gestes culinaires. Son objectif était de caractériser 39 gestes d'ouverture d'huître selon le concept de 'bon geste', cette tâche de caractérisation étant réalisée par 72 évaluateurs. Ce terrain a permis de s'assurer des bonnes performances de la Q-méthodologie binaire pour la caractérisation conceptuelle de grands ensembles de gestes culinaires.

Enfin, l'Annexe de cette partie présente l'implémentation dans R de la procédure statistique dédiée à l'analyse des données de Q-méthodologie binaire présentée dans les Chapitres 4 et 5. Cette implémentation est réalisée sous la forme du développement d'un package R.

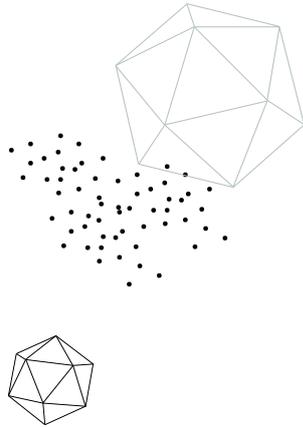
Cette thèse s'achève par une conclusion générale dans laquelle sont rappelés les principaux points évoqués dans ce manuscrit et dans laquelle sont abordées différentes perspectives.

Partie I :

Caractérisation perceptuelle

La première partie de cette thèse présente notre proposition méthodologique dédiée à la caractérisation perceptuelle d'un grand nombre de stimuli : le Tri Séquentiel Agglomératif (TSA). Elle présente également son application à un grand nombre de gestes culinaires.

Dans le Chapitre 1, le protocole de recueil des données de Tri Séquentiel Agglomératif est présenté. Comme nous le verrons, le Tri Séquentiel Agglomératif s'inspire en partie des solutions mises en place pour gérer des grands ensembles de stimuli en tri libre. Ce chapitre présente également les résultats d'un premier terrain de validation du Tri Séquentiel Agglomératif. Dans le Chapitre 2, une procédure d'analyse statistique des données de Tri Séquentiel Agglomératif est proposée. Elle permet d'obtenir des résultats orientés stimuli et des résultats orientés évaluateurs. Une étude du format des données permet de révéler leur originalité : elles présentent les caractéristiques usuelles des données de tri libre verbalisé ainsi que des caractéristiques spécifiques. Nous verrons comment les caractéristiques usuelles sont prises en compte dans la procédure d'analyse. Pour cela, un point actualisé des approches statistiques utilisées pour analyser les données de tri libre verbalisé est proposé. La question de savoir dans quelle mesure il est possible de valoriser les caractéristiques 'spécifiques' des données est également abordée. Dans le Chapitre 3, l'étude de l'applicabilité du Tri Séquentiel Agglomératif à des vidéos de gestes culinaires est abordée. Pour cela, les résultats d'un terrain réalisé au Centre Culinaire Contemporain sont exposés. Enfin, cette partie se conclut par une Annexe dans laquelle est présentée l'implémentation dans R de la procédure d'analyse statistique des données de Tri Séquentiel Agglomératif.



Chapitre 1

RECUEIL DES DONNÉES DE TRI SÉQUENTIEL AGGLOMÉRATIF

RÉSUMÉ : Utilisé depuis de nombreuses années en analyse sensorielle, le tri libre a démontré ses performances pour étudier la représentation perceptuelle de stimuli. Cependant, son utilisation est compromise dès lors que le nombre de stimuli à caractériser devient trop important. Des solutions ont pourtant été mises en place pour gérer ces situations : l'utilisation de plans de présentation incomplets et l'utilisation de prototypes prédéfinis. Celles-ci ne sont malheureusement pas optimales : la première fait émerger des dimensions perceptuelles qui sont approximées à l'échelle du panel d'évaluateurs ; celles révélées par la seconde sont conditionnées par des prototypes prédéfinis. Pour dépasser cette limite, une nouvelle méthodologie est ici proposée : le Tri Séquentiel Agglomératif (TSA). Elle repose en partie sur une combinaison des deux solutions précédentes. L'intégralité des stimuli est présentée à l'évaluateur sous la forme de sous-ensembles. Ces derniers sont agglomérés les uns aux autres grâce à des prototypes définis par l'évaluateur lui-même. Elle permet ainsi de révéler des dimensions perceptuelles qui ne sont ni approximées (à l'échelle du panel), ni conditionnées par des prototypes prédéfinis. Dans ce chapitre, cette procédure de recueil des données de Tri Séquentiel Agglomératif est détaillée. Une première étude ayant vocation à tester les performances du Tri Séquentiel Agglomératif est également présentée.

CE CHAPITRE EST INSPIRÉ DE L'ARTICLE :

Brard, M., & Lê, S. The Sequential Agglomerative Sorting Task, a new sensory methodology to characterize large sets of products. *Submitted*.

Sommaire

1	Recueil des données	33
1.1	Réalisation de la tâche de caractérisation perceptuelle par l'évaluateur : description du protocole de Tri Séquentiel Agglomératif	34
1.1.1	Première phase : évaluation des premiers stimuli au travers d'une tâche de tri libre verbalisée	35
1.1.2	Deuxième phase : évaluation des stimuli restants au travers de tâches de tri orientées verbalisées	36
1.2	Organisation de l'ensemble des stimuli par l'expérimentateur	37
1.2.1	Choix du nombre de stimuli par sous-ensemble évalué lors du Tri Séquentiel Agglomératif	38
1.2.2	Choix des stimuli composant chaque sous-ensemble évalué lors du Tri Séquentiel Agglomératif	40
2	Format des données	41
3	Validation du protocole de recueil des données	42
3.1	Mise en place des deux méthodes de caractérisation concurrentes	43
3.1.1	Produits	43
3.1.2	Évaluateurs	43
3.1.3	Déroulement des tâches de caractérisation	43
3.1.3.1	Tri Séquentiel Agglomératif	43
3.1.3.2	Tri libre verbalisé	44
3.2	Comparaison des performances des deux méthodes de caractérisation concurrentes	45
3.2.1	Mesures permettant d'établir la qualité de la caractérisation des produits fournie	46
3.2.1.1	Mesure de la capacité à fournir une caractérisation discriminée	46
3.2.1.2	Mesure de la capacité à fournir une caractérisation consensuelle	47
3.2.1.3	Mesure de la capacité à fournir une caractérisation détaillée	48
3.2.1.4	Mesure de la capacité à fournir une caractérisation répétable	48
3.2.2	Mesures permettant d'établir la faisabilité de réalisation par les évaluateurs	50
3.2.2.1	Mesure de la durée de la tâche de caractérisation	50
3.2.2.2	Mesure de la fatigue cognitive entraînée par la tâche de caractérisation	50
3.2.2.3	Mesure de l'appréciation de la tâche de caractérisation par les consommateurs	54
4	Conclusion	54

Dans un premier temps, nous présentons en détails la procédure de recueil des données de Tri Séquentiel Agglomératif (TSA). Pour cela, nous décrivons la tâche qu'il est demandé à chaque évaluateur de réaliser en vue d'obtenir une évaluation perceptuelle de l'ensemble de stimuli. La particularité du protocole repose sur une combinaison judicieuse des deux solutions existantes pour la gestion de grands ensembles de stimuli dans un contexte de tri libre (c.-à-d. l'utilisation de plans de présentation incomplet et l'utilisation de prototypes). Le TSA nécessite que l'expérimentateur organise en amont l'ensemble de stimuli. Nous verrons que cette organisation est une étape cruciale. Les stimuli doivent en effet être présentés aux évaluateurs de manière contrôlée. Dans ce cadre, nous proposons une solution de plan de présentation des stimuli. Celui-ci repose sur un plan d'expérience. À travers les propriétés statistiques intéressantes qu'il présente, il permet de limiter certains effets indésirables. Nous prenons le temps de détailler ces effets et pourquoi ils doivent être évités.

Dans un second temps, nous introduisons rapidement le format des données de TSA. Nous verrons que ces dernières s'organisent comme une suite de partitions sur les stimuli, chaque partition correspondant aux évaluations fournies par un évaluateur à une étape du TSA.

Dans un troisième et dernier temps, nous présentons les résultats d'un premier terrain dédié à la validation du protocole de recueil des données de TSA. Celui-ci repose sur la comparaison des performances du TSA avec ceux d'une approche concurrente plus classique : le tri libre verbalisé. Les deux méthodes de caractérisation ont été utilisées sur un même ensemble de 40 parfums de gels douche. Cet espace produit présentait la particularité d'être, *a priori*, facile à évaluer. L'objectif était de s'assurer des bonnes performances du TSA avant de poursuivre le processus de validation, et ce en démontrant qu'il présentait des performances au moins aussi bonnes que celles de la méthode de référence dans une situation *a priori* facile. Nous verrons que les résultats de ce premier terrain de validation sont concluants.

1 Recueil des données

L'utilisation de plans de présentation incomplets et l'utilisation de prototypes sont les deux principales techniques permettant de gérer des grands ensembles de stimuli dans un contexte de tri libre. Chacune présente à la fois des avantages et des inconvénients. Nous proposons de les combiner harmonieusement pour développer un protocole de recueil de données adapté à la caractérisation perceptuelle de grands ensembles de stimuli.

En utilisant un plan de présentation incomplet, chaque évaluateur n'évalue qu'un sous-ensemble des stimuli. Ceci réduit les problèmes de fatigue et de saturation de mémoire et permet de limiter la difficulté de la tâche de caractérisation. En

revanche, parce que chaque évaluateur ne caractérise pas l'intégralité des stimuli, il ne peut fournir que des dimensions perceptuelles approximées. Pour dépasser cette limite, nous proposons de présenter l'intégralité des stimuli sous la forme de sous-ensembles successifs. Cette approche permet de conserver les avantages de l'utilisation d'un plan de présentation incomplet tout en supprimant ses inconvénients. Elle permet en effet de collecter une caractérisation perceptuelle de tous les stimuli de la part de chaque évaluateur tout en réduisant la difficulté de la tâche.

Se pose alors la question de savoir comment les stimuli des différents sous-ensembles peuvent être comparés les uns aux autres par l'évaluateur. Pour répondre à cette interrogation, nous proposons de nous inspirer de l'utilisation de prototypes prédéfinis. Notre approche consiste à réaliser une tâche de tri libre verbalisée sur un premier sous-ensemble de stimuli. Chaque groupe de stimuli est ainsi synthétisé à travers un ensemble de caractéristiques perceptuelles qui constitue par la suite un prototype. Les stimuli des sous-ensembles restants sont agglomérés à ces prototypes. Cette approche permet de tirer partie des avantages de l'utilisation de prototypes. La difficulté de la tâche est effectivement limitée puisque les comparaisons sont effectuées entre stimuli et prototypes. Le nombre de prototypes étant plus faible que le nombre total de stimuli, le nombre de comparaisons à effectuer est moindre. Cette approche permet cependant de supprimer les inconvénients de l'utilisation de prototypes prédéfinis. En laissant l'évaluateur libre de définir et d'utiliser ses propres prototypes lors de la tâche de caractérisation, cette procédure permet de révéler des dimensions perceptuelles qui ne sont dépendantes d'aucun prototype prédéfini.

En résumé, notre proposition consiste à organiser les stimuli en sous-ensembles. Ceux-ci sont présentés séquentiellement à l'évaluateur et agglomérés les uns aux autres. Ce raisonnement est au cœur du Tri Séquentiel Agglomératif (TSA).

La procédure de recueil des données correspondante est présentée en détails dans cette section. Comme pour toute méthode de caractérisation sensorielle, nous distinguons ici deux volets : un volet expérimentateur et un volet évaluateur. D'un point de vue pédagogique, il nous a paru plus approprié de présenter d'abord la tâche réalisée par l'évaluateur. Ceci permet d'identifier plus facilement les paramètres qui doivent être considérés au préalable par l'expérimentateur pour organiser l'ensemble de stimuli.

1.1 Réalisation de la tâche de caractérisation perceptuelle par l'évaluateur : description du protocole de Tri Séquentiel Agglomératif

Imaginons ici que l'ensemble de stimuli a déjà été divisé par l'expérimentateur. Différents sous-ensembles (ici 3) ont ainsi été constitués (voir Figure 9).

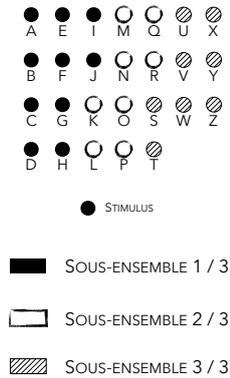


Figure 9: Exemple d'un ensemble de 26 stimuli divisé en trois sous-ensembles. Ces derniers contiennent respectivement 10, 8 et 8 stimuli.

1.1.1 Première phase : évaluation des premiers stimuli au travers d'une tâche de tri libre verbalisée

Pendant la première phase, les stimuli du premier sous-ensemble sont évalués. Cette évaluation repose dans un premier temps sur la réalisation d'une tâche de tri libre. Tous les stimuli du sous-ensemble sont présentés simultanément et les instructions données à l'évaluateur sont les suivantes : 'Veuillez tester au préalable chaque stimulus une fois. Veuillez ensuite constituer des groupes de stimuli homogènes : deux stimuli sont dans un même groupe (resp. dans des groupes différents) s'ils présentent, selon vous, des caractéristiques perceptuelles similaires (resp. différentes). Vous devez réaliser cette tâche selon vos propres critères, et vous n'êtes pas obligé(e) de spécifier ces critères. Il n'y a pas de bonnes ou de mauvaises réponses puisque nous sommes intéressés par votre propre perception. Vous êtes libre de former autant de groupes que vous le souhaitez. Il n'y a aucune contrainte sur le nombre de stimuli qu'un groupe peut contenir'.

Cette tâche de tri libre est suivie par une tâche de verbalisation. Les instructions données à l'évaluateur sont les suivantes : 'Veuillez maintenant décrire les groupes de stimuli que vous avez formés. Pour cela, veuillez caractériser chaque groupe à l'aide d'un ensemble de caractéristiques perceptuelles qui, selon vous, décrivent le groupe de stimuli en question'.

À la fin de cette première phase, les stimuli du premier sous-ensembles sont ainsi agencés en groupes. Chacun de ces groupes de stimuli est synthétisé à travers ses caractéristiques perceptuelles. Nous verrons par la suite comment ces ensembles de caractéristiques constituent des points d'ancrage qui sont utilisés par l'évaluateur pendant le reste du protocole de TSA.

Ces actions sont réalisées par chaque évaluateur du panel. Chacun fournit, *in fine*, une partition labellisée des stimuli du sous-ensemble telle que celle présentée sur la

Figure 10.

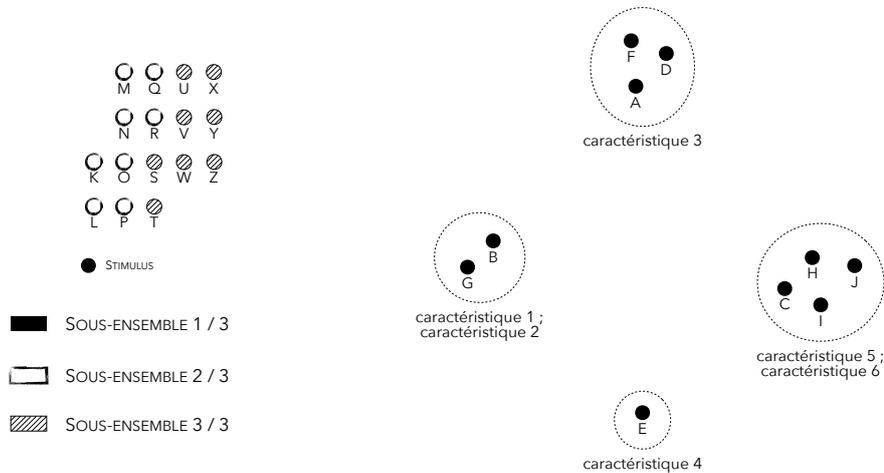


Figure 10: Exemple d'un résultat de la première phase du Tri Séquentiel Agglomératif. Celle-ci est réalisée sur le premier sous-ensemble de stimuli par un évaluateur.

1.1.2 Deuxième phase : évaluation des stimuli restants au travers de tâches de tri orientées verbalisées

Pendant la deuxième phase, les sous-ensembles de stimuli restants sont évalués de manière séquentielle. Les différents stimuli sont alors agglomérés aux groupes de stimuli déjà formés.

Pour un sous-ensemble donné, tous les stimuli qui le constituent sont présentés simultanément. Les instructions données à l'évaluateur sont les suivantes : 'Veillez considérer chaque nouveau stimulus un à un. Comparez-le aux groupes de stimuli que vous avez déjà formés. Cette comparaison doit être réalisée sur la seule base des ensembles de caractéristiques associés aux groupes. Vous ne devez donc pas réévaluer les stimuli des groupes précédents. La comparaison du stimulus avec les groupes vous laisse face à deux types d'actions : (1) si le stimulus présente les mêmes caractéristiques qu'un groupe alors agglomérez-le à ce groupe ; (2) s'il ne correspond à aucun groupe alors créez un nouveau groupe et décrivez-le grâce à un ensemble de caractéristiques'.

Cette tâche s'apparente à une tâche de tri orientée dans laquelle les catégories ont été prédéfinies par l'évaluateur. Chaque catégorie correspond à un ensemble de caractéristiques. Elle joue ici le rôle de point d'ancrage. Cette approche s'inspire de l'une des stratégies utilisées par le cerveau humain lors de l'exécution d'une tâche de catégorisation : la stratégie orientée caractéristiques. Dans ce contexte, chaque catégorie est représentée comme une somme de caractéristiques.

Un stimulus est attribué à une catégorie si elle partage les mêmes caractéristiques que cette catégorie (Homa, 1984). Cette approche est intéressante dans le cadre de la caractérisation d'un grand nombre de stimuli. L'évaluateur effectue en effet ses comparaisons entre un stimulus et un groupe de stimuli, et non pas entre deux stimuli entre eux. Le nombre de groupes étant inférieur au nombre total de stimuli, les comparaisons à effectuer sont moins nombreuses. Cette particularité de la deuxième phase du TSA a pour effet de réduire les problèmes de saturation de mémoire et de fatigue. Elle permet ainsi de réduire la difficulté de la tâche de caractérisation.

Une fois que tous les stimuli du sous-ensemble ont été agglomérés aux groupes déjà formés, l'évaluateur est libre d'en réévaluer certains. Ces éventuelles réévaluations à la marge peuvent avoir plusieurs objectifs : changer un stimulus de groupe, séparer et/ou combiner des groupes, ou modifier un ensemble de caractéristiques. Les terrains menés dans le cadre de cette thèse ont montré que ces réévaluations sont fréquentes. Elles sont primordiales puisqu'elles permettent à l'évaluateur d'affiner son évaluation perceptuelle de l'ensemble de stimuli. La considération d'un trop grand nombre de nouveaux stimuli à chaque étape peut pénaliser ces réévaluations à la marge. C'est pourquoi nous avons fait le choix d'agencer les stimuli sous la forme de sous-ensembles lors de cette phase du protocole. De prime abord, ce choix aurait pourtant pu paraître inutile puisque les stimuli sont évalués l'un après l'autre pendant cette phase.

Ces différentes actions sont réalisées par chaque évaluateur du panel. Elles sont répétées sur tous les sous-ensembles de stimuli restants. Chaque évaluateur considère ainsi chaque stimulus. Il fournit, *in fine*, une partition labellisée de l'intégralité des stimuli telle que celle présentée sur la Figure 11.

Lors des terrains réalisés dans le cadre de cette thèse, les différents sous-ensembles de stimuli étaient tous évalués lors de la même session. Il pourrait pourtant être envisagé de présenter les sous-ensembles lors de sessions différentes. Les points d'ancrage permettent en effet de limiter les problèmes de mémoire : grâce à eux, l'évaluateur peut se rappeler, par exemple, que tel groupe est 'sucré' et que tel autre est 'puissant'. Ils permettent ainsi à l'évaluateur de ne pas réévaluer tous les stimuli d'une session à l'autre. Ceci peut, par exemple, être intéressant pour caractériser un grand nombre de produits saturants (p. ex. des alcools). Des travaux de recherche supplémentaires sont cependant nécessaires pour tester et valider cette approche.

1.2 Organisation de l'ensemble des stimuli par l'expérimentateur

Deux types de paramètres doivent être considérés par l'expérimentateur pour organiser l'ensemble de stimuli : (1) des paramètres relatifs aux caractéristiques 'quantitatives' des sous-ensembles (c.-à-d. la taille des sous-ensembles, celle-ci définissant le nombre total de sous-ensembles), et (2) des paramètres relatifs aux caractéristiques 'qualitatives' des sous-ensembles (c.-à-d. les stimuli composant les sous-ensembles).

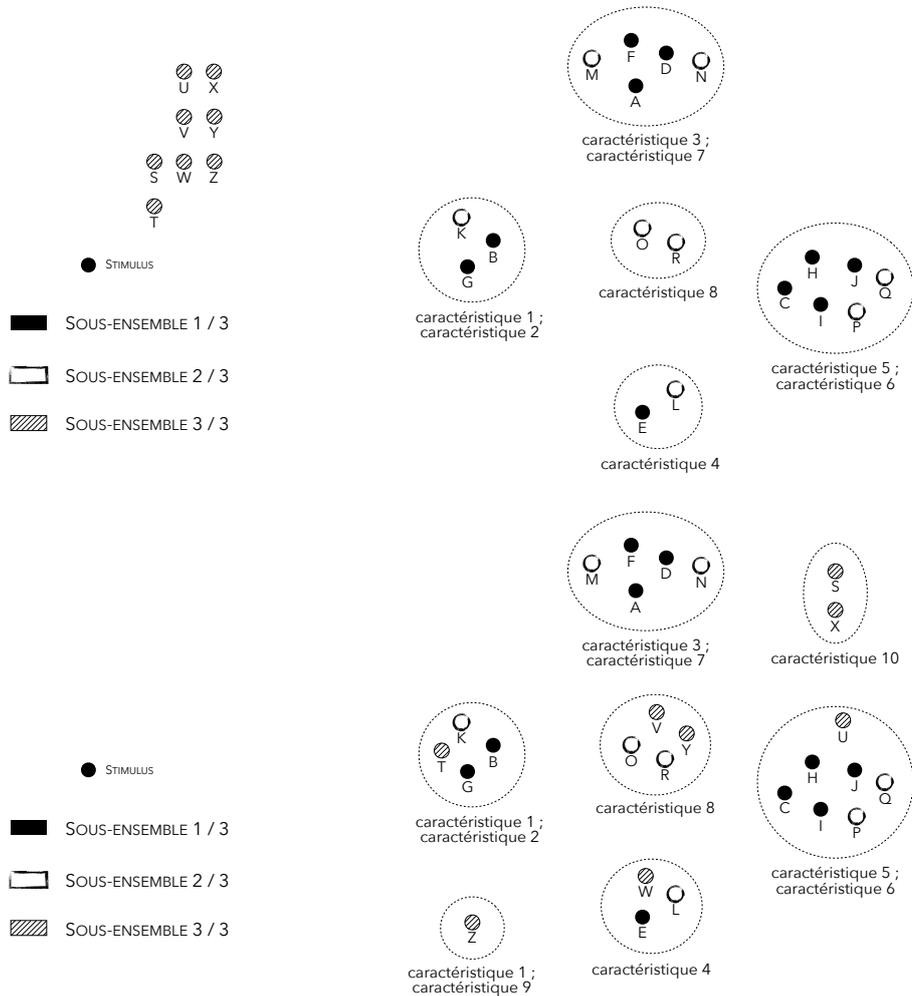


Figure 11: Exemple d'un résultat de la deuxième phase du Tri Séquentiel Agglomératif. Celle-ci est réalisée sur les deuxième et troisième (dernier) sous-ensembles par un évaluateur.

1.2.1 Choix du nombre de stimuli par sous-ensemble évalué lors du Tri Séquentiel Agglomératif

Ces paramètres sont directement liés aux types de tâches cognitives qui sont requises de la part de l'évaluateur lors de la réalisation du TSA.

Dans un premier temps, l'expérimentateur doit choisir le nombre de stimuli constituant le premier sous-ensemble. Celui-ci est évalué lors de la première phase du TSA. Cette phase est importante car c'est celle qui permet de définir les premiers points d'ancrage de l'évaluateur. C'est pourquoi le nombre de stimuli qui constituent le premier sous-ensemble doit être défini avec précautions. Il doit être

suffisamment élevé pour présenter à l'évaluateur une variabilité perceptuelle suffisante permettant de faire émerger des points d'ancrage intéressants. Cependant, il ne doit pas être trop élevé pour ne pas mettre en péril la bonne réalisation de la tâche de caractérisation. Rappelons que la tâche cognitive requise de la part de l'évaluateur lors de la première phase du TSA est une tâche de catégorisation. Pour optimiser la qualité de cette phase, nous proposons donc que le nombre de stimuli dans le premier sous-ensemble corresponde au nombre 'optimal' de stimuli défini dans un contexte de tri libre. Cependant, ce nombre 'optimal' dépend de la nature des stimuli évalués. C'est pourquoi il est impossible de définir de manière absolue et fixe le nombre de stimuli à inclure dans le premier sous-ensemble. Ce nombre doit être choisi par l'expérimentateur en fonction de sa connaissance experte des stimuli. Si l'expérimentateur considère que les stimuli présentent des différences fines alors il est conseillé de choisir un nombre réduit de stimuli. Ces derniers risquent en effet d'être difficiles à évaluer. Au contraire, si les différences entre les stimuli sont plus saillantes, il est possible d'inclure un nombre plus important de stimuli dans le premier sous-ensemble.

Dans un deuxième temps, l'expérimentateur doit choisir le nombre de stimuli constituant les sous-ensembles restants. Ceux-ci sont évalués lors de la deuxième phase du TSA. Rappelons que la tâche cognitive requise ici de la part de l'évaluateur diffère de celle réalisée lors de la première phase. Il s'agit d'une tâche de tri orientée dans laquelle les catégories ont été prédéfinies par l'évaluateur. Relativement à la tâche réalisée lors de la première phase, cette tâche cognitive est moins difficile pour l'évaluateur. Ceci a été confirmé lors d'échanges avec les évaluateurs qui ont participé aux terrains réalisés dans le cadre de cette thèse. Cette réduction de la difficulté est due au fait que les évaluations sont réalisées entre stimuli et groupes de stimuli, et non pas entre stimuli entre eux comme cela est le cas lors de la première phase. Le nombre d'évaluations à réaliser est donc plus faible. Il est alors possible de choisir plus de stimuli dans les sous-ensembles de la deuxième phase que dans le premier sous-ensemble. Cependant, les terrains réalisés dans le cadre de cette thèse nous ont montré qu'il ne faut pas utiliser plus de 20 stimuli dans chacun des sous-ensembles de la deuxième phase. La considération d'un nombre plus important a en effet tendance à limiter les réévaluations à la marge des stimuli. Enfin, nous recommandons d'équilibrer le nombre de stimuli inclus dans les sous-ensembles de la deuxième phase. Cela permet de ne pas perturber l'évaluateur en changeant leur configuration.

L'ensemble de ces recommandations définit nécessairement le nombre de stimuli dans les sous-ensembles. Prenons l'exemple d'une étude dont l'objectif est de caractériser 50 stimuli. Imaginons que l'expérimentateur a décidé d'utiliser 12 stimuli dans le premier sous-ensemble, ce choix ayant été motivé par son expertise des stimuli. Il reste alors 38 stimuli à inclure dans les sous-ensembles de la deuxième phase. Ce nombre est trop élevé pour constituer un seul sous-ensemble. L'expérimentateur est alors amené à constituer deux autres sous-ensembles composés de 19 stimuli cha-

cun. Finalement, le protocole correspondant à cette tâche de TSA impliquera trois sous-ensembles de 12, 19 et 19 stimuli respectivement.

1.2.2 Choix des stimuli composant chaque sous-ensemble évalué lors du Tri Séquentiel Agglomératif

En TSA, l'ordre avec lequel sont présentés les stimuli aux évaluateurs est primordial. Ceci n'est pas une caractéristique propre au TSA puisqu'elle est partagée avec toutes les méthodes de caractérisation sensorielle. Il est effectivement toujours nécessaire qu'un même stimulus soit évalué à des rangs différents par les évaluateurs. Un stimulus toujours évalué en première position fera en effet l'objet de plus de considération de la part de l'évaluateur, ce dernier étant plus attentif et plus concentré au début de la tâche de caractérisation. Inversement, le traitement d'un stimulus toujours évalué en dernière position sera peut-être moins fiable. Ces aspects peuvent avoir des effets sur les résultats fournis par les méthodes de caractérisation. Dans un contexte de tri libre, il a par exemple été prouvé que l'ordre de présentation des stimuli peut impacter la partition fournie par l'évaluateur (Courcoux et al., 2015).

Dans l'objectif de limiter ces aspects indésirables, deux stratégies issues du domaine de la planification expérimentale sont généralement utilisées par les expérimentateurs. Elles permettent de définir des ordres de présentation qui diffèrent d'un évaluateur à l'autre. La première consiste à choisir l'ordre de présentation de manière aléatoire. La seconde repose sur l'utilisation d'un plan d'expérience complet. Contrairement à la précédente, cette stratégie assure à l'expérimentateur certaines propriétés intéressantes. Elle permet par exemple de s'assurer que chaque stimulus est évalué à chaque rang de présentation de manière homogène sur la globalité du panel d'évaluateurs. C'est pourquoi c'est cette seconde stratégie qui est généralement privilégiée par les expérimentateurs.

Dans un contexte de TSA, cette problématique de planification expérimentale est plus délicate à régler. L'utilisation d'un plan d'expérience complet est là encore nécessaire mais non suffisante. Au contraire des autres méthodes de caractérisation sensorielle, les stimuli sont ici agencés en sous-ensembles successifs. La problématique principale n'est donc pas que l'ordre de présentation des stimuli diffère selon l'évaluateur. Il s'agit plus globalement que la composition des sous-ensembles de stimuli diffère selon l'évaluateur. Une fois l'ordre de présentation défini pour chaque évaluateur grâce au plan d'expérience complet, les rangs de présentation sont alors affectés aux différents sous-ensembles de stimuli. Cette affectation est réalisée en fonction de la taille des sous-ensembles déjà définie en amont par l'expérimentateur. Un tel plan de présentation permet de s'assurer de certaines propriétés intéressantes. Chaque stimulus apparaît dans les différents sous-ensembles de manière homogène sur la globalité du panel. Chaque couple de stimuli est également évalué dans les différents sous-ensembles de manière homogène.

Reprenons l'exemple précédent impliquant trois sous-ensembles de 12, 19 et 19 stimuli. Un plan d'expérience complet est défini par l'expérimentateur sur la globalité du panel d'évaluateurs. Chaque évaluateur est ensuite associé à un plan de présentation des stimuli. Pour un évaluateur donné, les stimuli correspondant aux rangs 1-12 dans ce plan sont évalués dans le premier sous-ensemble, les stimuli des rangs 13-31 correspondent au deuxième sous-ensemble, et les stimuli des rangs 32-50 correspondent au troisième sous-ensemble.

2 Format des données

On note S le nombre de stimuli, P le nombre d'évaluateurs, et I le nombre de sous-ensembles. Chaque évaluateur fournit une partition labellisée des stimuli au terme de chaque présentation d'un sous-ensemble de stimuli. Pour un évaluateur donné, les données de TSA sont donc les suivantes : pour le s -ème stimulus présenté dans le i -ème sous-ensemble, l'évaluation par ce p -ème évaluateur est identifiée sous la forme de la valeur notée x_{sip} . Cette dernière correspond à la description (c.-à-d. l'ensemble des caractéristiques) du groupe auquel appartient le stimulus du sous-ensemble pour cet évaluateur (p. ex. 'doux, sucré'). Chaque évaluateur est donc assimilé à une matrice contenant S lignes et I colonnes (voir Figure 12). Dans cette matrice, chaque colonne i correspond à une variable catégorielle à G_{ip} modalités, où G_{ip} représente le nombre de groupes de stimuli formés par l'évaluateur p à la fin de la présentation du i -ème sous-ensemble. Comme nous le verrons dans le Chapitre 2 de cette thèse, la présentation des stimuli sous la forme de sous-ensembles entraîne certaines spécificités au sein des données. Celles-ci ne sont pas abordées dans ce Chapitre 1.

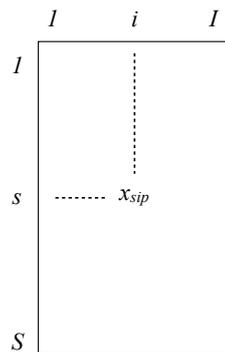


Figure 12: Données recueillies à l'échelle d'un évaluateur dans le cadre de l'implémentation d'une tâche de Tri Séquentiel Agglomératif. La valeur x_{sip} correspond à la description du groupe auquel appartient le stimulus s présenté dans le i -ème sous-ensemble pour cet évaluateur p .

À l'échelle du panel d'évaluateurs, les données de TSA sont mises sous la forme

d'un tableau multiple à S lignes et $P \times I$ colonnes. Ce tableau correspond à la concaténation des P différentes matrices individuelles de dimensions $S \times I$ (voir Figure 13).

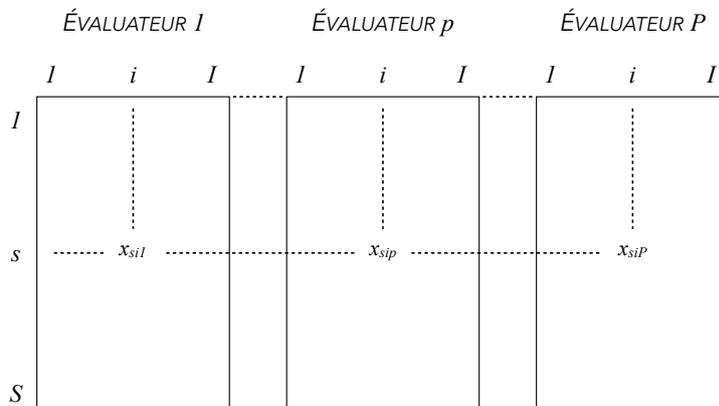


Figure 13: Données recueillies à l'échelle du panel d'évaluateurs dans le cadre de l'implémentation d'une tâche de Tri Séquentiel Agglomératif. La valeur x_{sip} correspond à la description du groupe auquel appartient le stimulus s présenté dans le i -ème sous-ensemble pour cet évaluateur p .

3 Validation du protocole de recueil des données

Le processus de validation consiste à tester le TSA sur plusieurs ensembles de stimuli présentant des niveaux de difficulté de caractérisation croissants. En parallèle, les mêmes ensembles de stimuli sont soumis à la procédure classique : le tri libre - ici verbalisé. La première étape consiste à prouver que les performances du TSA sont - au moins - aussi bonnes que celles de la méthode concurrente dans une situation impliquant un ensemble de stimuli facile à caractériser. Si cet aspect n'est pas prouvé, le processus de validation du TSA s'arrête et la méthode n'est pas éprouvée. Dans le cas contraire, le processus de validation continue. La deuxième étape consiste alors à prouver que les performances du TSA sont bonnes dans une situation impliquant un ensemble de stimuli difficile à caractériser alors que la méthode concurrente, elle, échoue.

Dans cette section, nous présentons les résultats du premier terrain mis en place. Le TSA et le tri libre ont été testés sur un ensemble de 40 parfums de gels douche présentant des différences sensorielles marquées. Celles-ci permettaient de faciliter la réalisation de la tâche de caractérisation. Différents critères de performance ont été calculés pour chaque méthode de caractérisation. Ils ont ensuite été comparés.

3.1 Mise en place des deux méthodes de caractérisation concurrentes

3.1.1 Produits

Lors de cette étude, des parfums de gels douche ont été sélectionnés pour constituer l'espace produit à caractériser. Cette sélection a été mise en place par l'équipe développement de l'entreprise Givaudan. Elle a été réalisée de manière à obtenir des produits présentant des différences sensorielles saillantes. Cela permettait d'obtenir un espace produit 'grossier' facile à caractériser par les évaluateurs. Pour cela, des univers sensoriels hétérogènes et très différents (c.-à-d. sans recouvrement) ont été identifiés. Les produits ont été ensuite choisis au sein de ces univers sensoriels. Au final, 40 parfums de gels douche ont été choisis. Certains - majoritairement des parfums destinés aux hommes - relevaient d'un univers olfactif boisé. D'autres - majoritairement des parfums destinés aux femmes - présentaient des notes olfactives lactées, fruitées, ou relevant d'un univers alimentaire.

Les 40 parfums ont été insérés dans des petites bouteilles blanches en plastique opaque. Une simple pression du doigt sur la bouteille permettait de libérer le parfum. Chaque parfum a ensuite été associé à un code aléatoire à trois chiffres.

3.1.2 Évaluateurs

Cent-vingt consommateurs ont été recrutés pour participer à cette étude. Il s'agissait majoritairement de femmes (55.8 %) et de jeunes personnes (âge moyen = 27.08 ans ; amplitude = 19-62 ans). Ces consommateurs n'ont été soumis à aucun entraînement sensoriel avant de participer à cette expérience.

Une première moitié des consommateurs (c.-à-d. 60) a constitué un premier panel. Ce dernier a été amené à évaluer les parfums selon la procédure du TSA. La deuxième moitié des consommateurs a constitué un deuxième panel qui a été amené à évaluer les parfums selon la procédure du tri libre - ici verbalisé. Les caractéristiques des consommateurs (âge, genre) étaient globalement similaires d'un panel à l'autre.

3.1.3 Déroulement des tâches de caractérisation

3.1.3.1 Tri Séquentiel Agglomératif

L'espace produit était divisé en 3 sous-ensembles. Nous aborderons plus loin dans cette section le nombre de stimuli composant chacun d'entre eux. Un plan d'expérience complet a été utilisé pour définir, pour chaque consommateur, le plan de présentation des parfums.

Lors de la première phase du TSA (évaluation du premier sous-ensemble), les instructions données au consommateur étaient les suivantes : 'Dans un premier

temps, veuillez sentir au moins une fois chaque échantillon. Pour cela, veuillez suivre le plan de présentation défini. Dans un deuxième temps, veuillez constituer des groupes d'échantillons selon leur similarité sensorielle. Pour cela, focalisez-vous uniquement sur les caractéristiques des échantillons et non sur l'appréciation que vous en avez. Vous êtes libre de former les groupes selon n'importe quels critères, et vous n'êtes pas obligé(e) de spécifier ces critères. Vous pouvez former autant de groupes que vous le souhaitez. Vous pouvez mettre autant d'échantillons que vous le désirez dans chaque groupe. Dans un troisième et dernier temps, veuillez associer chaque groupe d'échantillons que vous avez formé à un post-it. Sur chaque post-it, veuillez indiquer une ou plusieurs caractéristique(s) sensorielle(s) qui décrivent, selon vous, le groupe en question.' À la fin de l'évaluation du premier sous-ensemble, le consommateur était amené à indiquer sa partition labellisée des stimuli sur une feuille d'enregistrement.

Lors de la deuxième phase du TSA (évaluation des deuxième et troisième sous-ensembles), les instructions données au consommateur étaient les suivantes : 'Vous n'êtes maintenant plus obligé(e) de sentir tous les échantillons avant de commencer. Veuillez évaluer un à un les nouveaux échantillons. Pour cela, veuillez suivre le plan de présentation défini. À chaque fois que vous sentez un échantillon, veuillez le comparer aux groupes que vous avez déjà formés. Pour réaliser cette comparaison, référez-vous aux post-its associés aux groupes. Vous avez alors deux possibilités :

1. Vous considérez que l'échantillon présente les mêmes caractéristiques qu'un groupe déjà formé. Dans ce cas, ajoutez-le à ce groupe.
2. Vous considérez que l'échantillon ne correspond à aucun des groupes déjà formés. Dans ce cas, créez un nouveau groupe avec cet échantillon et décrivez-le avec une ou plusieurs caractéristique(s) en utilisant un post-it.

Une fois toutes les évaluations réalisées, vous êtes libre d'effectuer des réaffectations à la marge. Vous pouvez changer un échantillon de groupe (échantillons des sous-ensembles précédents inclus), combiner ou diviser des groupes, changer les caractéristiques sensorielles inscrites sur les post-its, etc.' À la fin de l'évaluation de chaque sous-ensemble, le consommateur était amené à indiquer sa partition labellisée des stimuli sur une feuille d'enregistrement.

3.1.3.2 Tri libre verbalisé

Un plan d'expérience complet a été utilisé pour définir, pour chaque consommateur, le plan de présentation des parfums.

Les instructions données au consommateur étaient les suivantes : 'Dans un premier temps, veuillez sentir au moins une fois chaque échantillon. Pour cela, veuillez suivre le plan de présentation défini. Dans un deuxième temps, veuillez constituer des groupes d'échantillons selon leur similarité sensorielle. Pour cela, focalisez-vous uniquement sur les caractéristiques des échantillons et non sur l'appréciation que

vous en avez. Vous êtes libre de former les groupes selon n'importe quels critères, et vous n'êtes pas obligé(e) de spécifier ces critères. Vous pouvez former autant de groupes que vous le souhaitez. Vous pouvez mettre autant d'échantillons que vous le désirez dans chaque groupe. Dans un troisième et dernier temps, veuillez associer chaque groupe d'échantillons que vous avez formé à un post-it. Sur chaque post-it, veuillez indiquer une ou plusieurs caractéristiques sensorielles qui décrivent, selon vous, le groupe en question.' À la fin de cette évaluation, le consommateur était amené à indiquer sa partition labellisée des stimuli sur une feuille d'enregistrement.

3.2 Comparaison des performances des deux méthodes de caractérisation concurrentes

Une méthode de caractérisation est jugée performante dès lors qu'elle satisfait tout d'abord aux exigences suivantes : une bonne capacité à fournir une évaluation discriminée, consensuelle, détaillée et répétable des produits. Ces caractéristiques mesurent la qualité de la caractérisation des produits fournie par la méthode. Elles sont généralement recherchées pour toute méthode de caractérisation sensorielle et ne sont pas spécifiques à la caractérisation d'un nombre important de produits. Ce contexte entraîne pourtant certaines spécificités. Il est par exemple nécessaire de s'assurer que la méthode de caractérisation n'est pas trop longue et fatigante pour l'évaluateur. Aux mesures 'classiques' s'ajoutent alors des mesures permettant d'évaluer la faisabilité de réalisation par les évaluateurs.

Nous détaillons ici chacune des mesures de performance établies lors de ce terrain. Celles-ci ont été recueillies à différents moments de la réalisation de la tâche par le consommateur (voir la Figure 14 présentant un schéma récapitulatif du protocole mis en place dans le cadre de cette étude). L'analyse statistique de ces mesures et les résultats associés sont également présentés. Toutes les analyses ont été réalisées avec le logiciel R. La fonction `fast` du package `SensoMineR` à été utilisée pour réaliser certaines analyses.

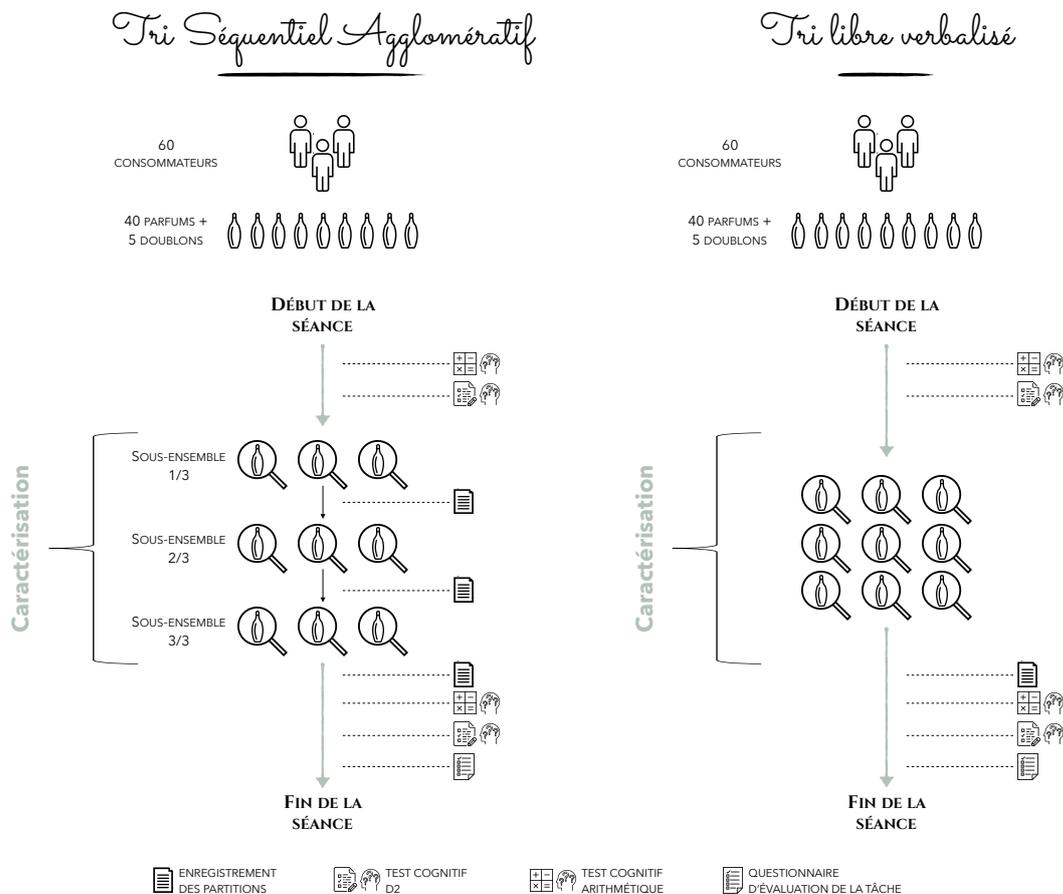


Figure 14: Protocole utilisé pour le terrain de validation du Tri Séquentiel Agglomératif.

3.2.1 Mesures permettant d'établir la qualité de la caractérisation des produits fournie

3.2.1.1 Mesure de la capacité à fournir une caractérisation discriminée

En ce qui concerne les données de TSA, seul le tableau de dimension 40×60 composé des partitions labellisées finales fournies par les consommateurs a été considéré. Cela permettait de comparer ces données avec celles de tri libre verbalisé. Les deux tableaux de données ont été soumis à une procédure d'Analyse des Correspondances Multiples (ACM) couplée à une procédure de rééchantillonnage. Cette approche permet d'obtenir une représentation des parfums agrémentée d'ellipses de confiance. Elle est décrite plus en détails dans le Chapitre 2 de cette thèse. Développée par Cadoret, Lê, and Pagès (2009), elle permet d'évaluer la capacité des consommateurs à discriminer les parfums : les ellipses de confiance obtenues sont d'autant plus petites que la discrimination des parfums est importante. Une méthode de caractérisation permettant d'obtenir de petites ellipses de confiance est donc performante pour obtenir une caractérisation discriminée de l'ensemble de produits.

Les représentations des parfums et des ellipses de confiance sont présentées en Figure 15. Dans les deux cas, les ellipses de confiance associées aux parfums sont petites. Ceci indique une discrimination importante des parfums par les consommateurs. La méthode du tri libre et la méthode du TSA ne présentent pas de différence concernant cette mesure de performance.

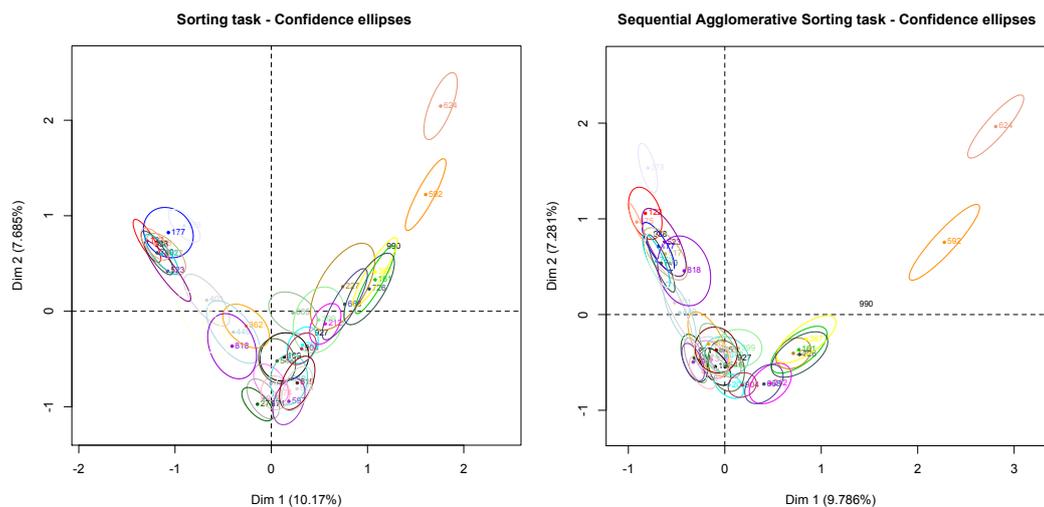


Figure 15: Représentations issues de l'Analyse des Correspondances Multiples des parfums et des ellipses de confiance correspondantes. Représentations obtenues à partir des données de tri libre (à gauche) et des données de Tri Séquentiel Agglomératif (à droite).

Plus globalement, les deux représentations des parfums sont similaires ($RV = 0.95$). Les deux méthodes fournissent donc des caractérisations perceptuelles très proches.

3.2.1.2 Mesure de la capacité à fournir une caractérisation consensuelle

La première valeur propre de chaque ACM a été utilisée comme critère reflétant le degré d'accord observé entre les consommateurs. Une première valeur propre de 1 indique en effet que chaque variable qualitative (c.-à-d. chaque partition sur les parfums fournie par un consommateur) est très liée à la première dimension de l'ACM. Cette première valeur propre reflète alors le niveau de concordance entre les partitions sur les parfums fournies par les consommateurs. Une forte concordance indique un accord important (c.-à-d. un consensus) au sein du panel.

Les premières valeurs propres des ACM sont présentées dans la Table 3. Celles-ci sont proches pour les deux méthodes. La méthode du tri libre et la méthode du TSA mènent à des caractérisations des produits autant consensuelles l'une que l'autre.

Table 3: Premières valeurs propres de l'Analyse des Correspondances Multiples.

Tri libre	Tri Séquentiel Agglomératif
0.65	0.61

3.2.1.3 Mesure de la capacité à fournir une caractérisation détaillée

Le niveau de détails de la caractérisation des produits peut être analysé au travers des caractéristiques sensorielles utilisées par les consommateurs pour décrire leurs groupes de produits. La richesse du vocabulaire utilisé a été évaluée au travers de deux indices numériques : (1) le nombre total de caractéristiques sensorielles différentes citées, et (2) le nombre de caractéristiques sensorielles différentes citées au moins 5 fois. Les 20 caractéristiques les plus utilisées ont également été identifiées pour enrichir la comparaison du vocabulaire utilisé dans le cadre des deux méthodes de caractérisation.

Les nombres totaux de caractéristiques sensorielles utilisées lors des deux méthodes sont présentés dans la Table 4. Les mesures sont proches pour les deux méthodes. La méthode du tri libre et la méthode du TSA permettent d'obtenir un vocabulaire d'une richesse similaire de la part des consommateurs.

Table 4: Richesse du vocabulaire sensoriel utilisé pour décrire les produits.

	Tri libre	Tri Séquentiel Agglomératif
Nombre total de caractéristiques citées	350	351
Nombre de caractéristiques citées plus de 5 fois	54	51

Les caractéristiques sensorielles les plus utilisées lors de la réalisation des deux méthodes sont présentées en Figure 16. Celles-ci sont similaires d'une méthode à l'autre. En particulier, les caractéristiques 'sucré', 'fruité' et 'doux' sont les plus citées dans les deux cas.

3.2.1.4 Mesure de la capacité à fournir une caractérisation répétable

Quarante paires de parfums ont été créées. Chacune était constituée du parfum de référence et de son doublon. Lors de la mise en place de la tâche de caractérisation, chaque consommateur a en réalité évalué 45 parfums (et non 40) : les 40 parfums de référence et 5 doublons. Dans le cadre du TSA, les trois sous-ensembles étaient respectivement composés de 12, 16 et 17 parfums. Les doublons évalués n'étaient pas les mêmes d'un consommateur à l'autre. Ils ont été choisis selon un plan d'expérience optimal.

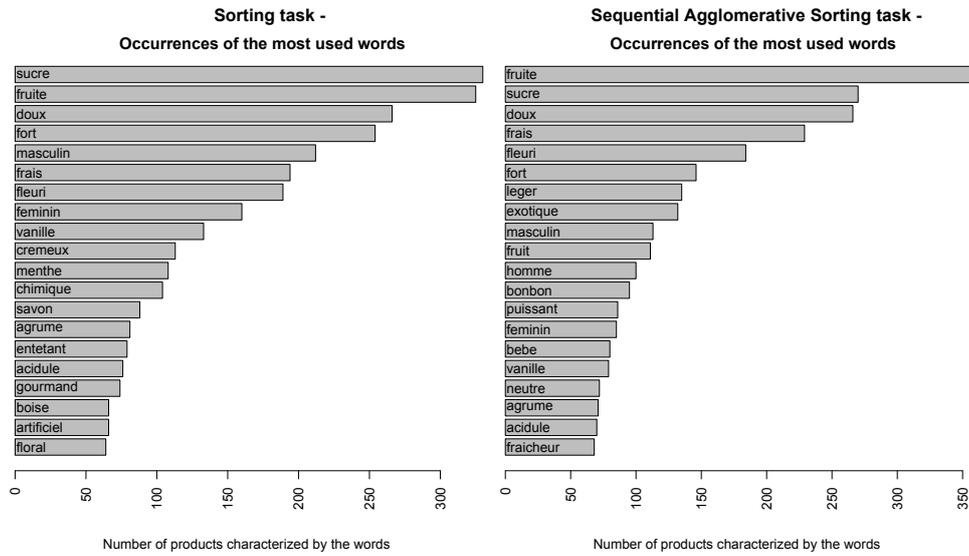


Figure 16: Caractéristiques sensorielles les plus citées. Caractéristiques correspondant aux données de tri libre (à gauche) et aux données de Tri Séquentiel Agglomératif (à droite).

Pour chaque consommateur, le nombre de paires ‘parfum de référence - doublon’ correctement reconstituées a été calculé. Une paire était jugée comme reconstituée si les deux échantillons étaient groupés ensemble par le consommateur. Un nombre de paires reconstituées élevé indique une bonne répétabilité de la caractérisation sensorielle établie. Une analyse de la variance a ensuite été réalisée pour tester l’effet de la méthode de caractérisation sur le nombre de paires reconstituées.

Les nombres moyens de paires de parfums correctement identifiées sont présentés dans la Table 5. Ceux-ci avoisinent une valeur de 3/5 et sont proches pour les deux méthodes. Malgré une légère supériorité du TSA, le test global de l’analyse de la variance n’a pas révélé d’effet significatif de la méthode sur cette valeur (p -value = 0.7045). La méthode du tri libre et la méthode du TSA mènent à des niveaux de répétabilité similaires.

Table 5: Nombre moyen de paires de parfums correctement reconstituées.

Tri libre	Tri Séquentiel Agglomératif
2.88/5	2.97/5

3.2.2 Mesures permettant d'établir la faisabilité de réalisation par les évaluateurs

3.2.2.1 Mesure de la durée de la tâche de caractérisation

La durée totale de réalisation de la tâche a été enregistrée (en minutes) pour chaque consommateur. Elle correspondait au temps écoulé entre le moment où le consommateur avait fini de lire les instructions et le moment où il avait fini de remplir sa dernière feuille d'enregistrement. Lors de l'implémentation du tri libre, le consommateur était amené à remplir une seule feuille d'enregistrement (à la fin de l'expérience) tandis que lors de l'implémentation du TSA, le consommateur était amené à en remplir trois (une après le premier sous-ensemble, une après le deuxième sous-ensemble, et une à la fin de l'expérience). Pour prendre en compte ce déséquilibre, 10 minutes ont été retirées du temps enregistré pour le TSA. Une analyse de la variance a été réalisée pour tester l'effet de la méthode de caractérisation sur la durée totale de l'expérience.

Les durées moyennes de réalisation des deux tâches de caractérisation sont présentées dans la Table 5. Celles-ci varient entre 31 minutes pour le TSA et 37 minutes pour le tri libre. Le test global de l'analyse de la variance a mis en évidence un effet significatif de la méthode sur cette valeur (p -value = 0.0038). La méthode du TSA est associée à une caractérisation sensorielle plus rapide qu'avec la méthode de référence du tri libre.

Table 6: Durée moyenne de la réalisation de la tâche de caractérisation.

Tri libre	Tri Séquentiel Agglomératif
36.85 minutes	31.10 minutes

3.2.2.2 Mesure de la fatigue cognitive entraînée par la tâche de caractérisation

Résultats au test cognitif d2

Le test d2 créé par Brickenkamp and Zillmer (1998) est un test cognitif chronométré qui permet de mesurer l'attention et la performance de concentration d'un individu. Il repose sur une série de lettres 'd' et 'p'. Ces dernières sont associées à quatre tirets, disposés individuellement ou par paires au-dessus ou au-dessous de la lettre. Elles sont agencées dans le test sous la forme de 14 rangées composées chacune de 47 lettres (voir Figure 17). Chaque consommateur était amené à balayer chaque ligne pour identifier et rayer chaque 'd' associé à deux tirets (voir Figure 18). Cette tâche a été réalisée dans un temps imparti de 20 secondes par rangée (un signal sonore retentissait toutes les 20 secondes pour informer le consommateur qu'il devait passer à la rangée suivante). Une fois le test réalisé par le consommateur, l'expérimentateur calculait différents paramètres (voir Table 7).

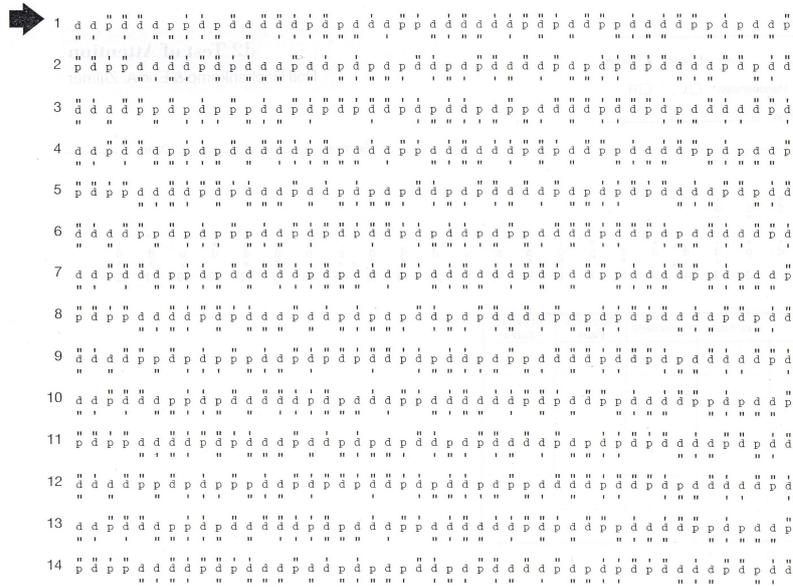


Figure 17: Exemple d'une feuille du test cognitif d2.

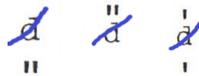


Figure 18: Lettres devant être rayées dans le cadre du test cognitif d2.

Table 7: Paramètres calculés pour chaque test d2.

Paramètre	Signification	Calcul
<i>TN</i> (Total Number)	Nombre de lettres rayées	
<i>OE</i> (Omission Errors)	Nombre de lettres qui n'ont pas été rayées à tort	
<i>CE</i> (Confusion Errors)	Nombre de lettres qui ont été rayées à tort	
<i>E</i> (Errors)	Nombre d'erreurs	$OE + CE$
<i>CR</i> (Correct Responses)	Nombre de réponses correctes	$TN - E$
<i>CP</i> (Concentration Performance)	Performance de concentration	$CR - CE$

Lors de cette étude, chaque consommateur a passé le test d2 une fois avant la tâche de caractérisation. Il l'a de nouveau passé après la tâche. Ce type de passation permet de mesurer le delta des différents critères de performance grâce à un calcul 'mesure après - mesure avant'. Ces derniers permettent de mesurer l'impact de la tâche de caractérisation sur les performances cognitives des consommateurs. L'obtention de deltas positifs est attendue. Ceci est dû au fait que le consommateur est habitué au test lors de la deuxième passation alors que ce n'est pas le cas lors de la première. La valeur du critère lors de la passation après est donc plus importante que lors de la passation avant. La valeur des deltas est d'autant plus

faible que la tâche de caractérisation a été fatigante pour le consommateur. Pour chaque consommateur, les deltas des deux critères principaux du test d2 CR (nombre de réponses correctes) et CP (performance de concentration) ont ainsi été calculés. Deux analyses de la variance ont été réalisées pour tester l'effet de la méthode de caractérisation sur ces valeurs.

Les deltas sont présentés dans la Table 8. Les valeurs sont comparables d'une méthode à l'autre. Le test global de l'analyse de la variance n'a pas mis en évidence d'effet significatif de la méthode sur la valeur de ces critères de fatigue cognitive (p -value = 0.8549 pour le critère CR ; p -value = 0.5826 pour le critère CP).

Table 8: Deltas moyens associés aux deux principaux critères du test d2. Un faible delta témoigne d'une importante fatigue cognitive entraînée par la tâche.

	Tri libre	Tri Séquentiel Agglomératif
CR	82.52	81.16
CP	86.30	82.62

Résultats à un test cognitif arithmétique

Lors de cette étude, un test arithmétique a également été utilisé pour mesurer la fatigue cognitive des consommateurs. Ce test était composé de 200 opérations mathématiques. Ces dernières étaient soit des additions soit des soustractions de deux chiffres. Cette tâche devait être réalisée dans un temps imparti de 1 minute et 30 secondes. Une fois le test réalisé par le consommateur, l'expérimentateur calculait différents paramètres (voir Table 9).

Table 9: Paramètres calculés pour chaque test arithmétique.

Paramètre	Signification	Calcul
TN (Total Number)	Nombre d'opérations effectuées	
FO (False Operations)	Nombre d'opérations fausses	
CO (Correct Operations)	Nombre d'opérations correctes	$TN - FO$
$CO\%$ (Correct Operations Rate)	Taux d'opérations correctes	$(TN - FO)/TN$

Chaque consommateur a passé ce test avant et après réalisation de la tâche de caractérisation. Comme pour le test d2, un delta a été calculé entre les deux passations pour les deux critères principaux CO (nombre d'opérations correctes) et $CO\%$ (taux d'opérations correctes). Comme pour le test d2, la valeur de ces deltas est d'autant plus faible que la tâche de caractérisation a été fatigante pour le consommateur. Deux analyses de la variance ont été réalisées pour tester l'effet de la méthode de caractérisation sur ces valeurs.

Les deltas associés aux critères du test arithmétique sont présentés dans la Table 10. Les valeurs sont globalement meilleures pour la méthode du TSA. Cependant, le test global de l'analyse de la variance n'a pas révélé d'effet significatif de la méthode sur la valeur de ces critères de fatigue cognitive (p -value = 0.09798 pour le critère CO ; p -value = 0.0710 pour le critère $CO\%$).

Table 10: Deltas moyens associés aux deux principaux critères du test arithmétique. Un faible delta témoigne d'une importante fatigue cognitive entraînée par la tâche.

	Tri libre	Tri Séquentiel Agglomératif
CO	-0.233	3.283
$CO\%$	-0.008	0.004

Évaluation de la difficulté par l'évaluateur

À la fin de la séance, chaque consommateur était amené à évaluer la difficulté de la tâche. Pour cela, il devait indiquer s'il avait trouvé cette tâche *pas du tout difficile*, *pas difficile*, *peu difficile*, *difficile* ou *très difficile*. Un test du χ^2 a été réalisé. Il a permis de tester l'indépendance entre la méthode de caractérisation utilisée et l'évaluation par les consommateurs de la difficulté de la tâche.

Les évaluations par les consommateurs sont présentées dans la Table 11. Les deux méthodes ont été globalement perçues comme difficiles à réaliser et le test du χ^2 n'a pas révélé d'effet significatif de la méthode (p -value = 0.8138).

Table 11: Table de contingence présentant les évaluations de la difficulté de la tâche de caractérisation en fonction de la méthode utilisée.

	Tri libre	Tri Séquentiel Agglomératif
Pas difficile du tout	0	1
Pas difficile	7	6
Un peu difficile	18	21
Difficile	32	30
Très difficile	3	2

Évaluation de la fatigue par l'évaluateur

À la fin de la séance, chaque consommateur était également amené à évaluer la fatigue ressentie lors de la tâche. Pour cela, il devait indiquer s'il avait trouvé cette tâche *pas du tout fatigante*, *pas fatigante*, *peu fatigante*, *fatigante* ou *très fatigante*. Un test du χ^2 a été réalisé pour tester l'indépendance entre la méthode de caractérisation utilisée et l'évaluation par les consommateurs de la fatigue entraînée par la tâche.

Les évaluations par les consommateurs sont présentées dans la Table 12. Les deux méthodes ont été globalement perçues comme fatigantes et le test du χ^2 n'a pas permis de mettre en évidence un effet significatif de la méthode (p -value = 0.6476).

Table 12: Table de contingence présentant les évaluations de la fatigue de la tâche de caractérisation en fonction de la méthode utilisée.

	Tri libre	Tri Séquentiel Agglomératif
Pas fatigant du tout	3	4
Pas fatigant	9	5
Un peu fatigant	24	27
Fatigant	24	23
Très fatigant	0	1

3.2.2.3 Mesure de l'appréciation de la tâche de caractérisation par les consommateurs

À la fin de la séance, chaque consommateur était enfin amené à évaluer son appréciation de la tâche. Pour cela, il devait indiquer s'il *n'avait pas du tout aimé*, *n'avait pas aimé*, *avait peu aimé*, *avait aimé* ou *avait beaucoup aimé* la tâche. Un test du χ^2 a là encore été réalisé pour tester l'indépendance entre la méthode de caractérisation utilisée et l'évaluation de l'appréciation de la tâche par les consommateurs.

Les évaluations par les consommateurs sont présentées dans la Table 13. Les deux méthodes ont été globalement appréciées et le test du χ^2 n'a pas permis de mettre en évidence un effet significatif de la méthode sur l'appréciation de la tâche (p -value = 0.3346).

Table 13: Table de contingence présentant les évaluations de la fatigue de la tâche de caractérisation en fonction de la méthode utilisée.

	Tri libre	Tri Séquentiel Agglomératif
Très apprécié	10	6
Apprécié	39	43
Un peu apprécié	11	9
Pas apprécié	0	2

4 Conclusion

L'étude de la représentation perceptuelle d'un stimulus est l'objectif le plus courant des analystes sensoriels. Celle-ci peut être mise en place à travers de nombreuses méthodes de caractérisation. Certaines de ces méthodes sont spécifiquement adaptées à une utilisation par des évaluateurs non-experts. En permettant d'analyser

la perception des consommateurs (évaluateurs non-experts), elles présentent un intérêt majeur dans un contexte de R&D. Parmi ces approches, la méthode du tri libre est probablement l'une des plus utilisées en analyse sensorielle. Sa facilité d'utilisation peut expliquer cette popularité. Elle repose effectivement sur une approche utilisée naturellement par le cerveau humain pour aborder un ensemble de stimuli : la catégorisation. En outre, elle permet de recueillir la perception brute des consommateurs. Elle est en effet basée sur une évaluation holistique des stimuli (c.-à-d. une évaluation globale et non une évaluation spécifique par attribut sensoriel). Elle permet ainsi de faire émerger des leviers opérationnels directement actionnables dans un processus de développement de produit.

Malgré l'efficacité et l'intérêt de la méthode du tri libre, son application à des grands ensembles de stimuli est jusqu'alors restée une problématique en suspens. Certaines réflexions scientifiques ont pourtant été entreprises pour déterminer de quelle manière il est possible de considérer des grands ensembles de stimuli dans ce contexte. L'utilisation de plans de présentation des stimuli incomplets ou le recours à des prototypes prédéfinis sont les deux principales solutions développées dans ce cadre. Nous avons cependant vu que ces approches ne sont pas optimales.

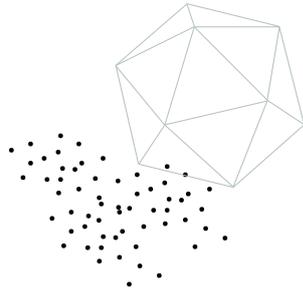
Dans ce chapitre, nous avons proposé une procédure de recueil de données de caractérisation qui peut être une solution à ce problème. Celle-ci repose sur une division des stimuli en plusieurs sous-ensembles de taille restreinte. Ces derniers sont présentés de manière séquentielle à l'évaluateur et agglomérés les uns aux autres. L'évaluation du premier sous-ensemble de stimuli est réalisée sous la forme d'une tâche de tri libre verbalisée classique. En cela, le TSA peut être vu comme une variante du tri libre verbalisé. L'évaluation des sous-ensembles suivants repose, elle, sur une approche utilisée naturellement par le cerveau humain dans un contexte de catégorisation. Tout en préservant les caractéristiques appréciées du tri libre verbalisé, cette procédure permet de gérer un grand nombre de stimuli. Le raisonnement derrière cette procédure est en effet que l'évaluateur travaille au niveau du groupe et non pas au niveau du stimulus. Ceci présente l'avantage de faciliter la réalisation de la tâche de caractérisation en réduisant les problèmes de saturation et de fatigue cognitive. Le nombre de groupes est effectivement beaucoup plus petit que le nombre de stimuli. Nous avons vu que cette approche permet de dépasser les limites des techniques classiquement utilisées pour gérer de grands ensembles de stimuli dans un contexte de tri libre. Elle repose en effet sur une évaluation de l'ensemble de stimuli et sur l'utilisation de points d'ancrage définis par l'évaluateur lui-même lors de la tâche.

Nous avons ensuite présenté brièvement le format des données de caractérisation recueillies. Celui-ci sera étudié de manière plus approfondie dans le Chapitre 2 de cette thèse.

Enfin, nous avons présenté les principaux résultats d'un premier terrain réalisé dans

le cadre du processus de validation du TSA. Les résultats ont révélé que les caractérisations sensorielles obtenues grâce au TSA sont comparables à celles obtenues grâce à la méthode du tri libre. Certaines mesures de performance semblent cependant meilleures pour le TSA que pour le tri libre. La tâche est en effet plus rapide à réaliser par le consommateur. La répétabilité des évaluations semble également meilleure. En outre, le TSA est associé à de meilleurs résultats au test arithmétique. Ces derniers semblent souligner l'apparition d'une fatigue cognitive moindre. Ces résultats ne mettent pourtant en évidence que des tendances puisque l'impact de la méthode de caractérisation sur ces mesures n'est pas significatif. Ce terrain a démontré que, dans un cas d'étude 'facile', le TSA fait - au moins - aussi bien que la méthode classique du tri libre verbalisé. Ces premiers résultats sont encourageants car ils permettent de poursuivre le processus de validation. Dans ce cadre, un deuxième terrain doit être mis en place. Il aura pour vocation à tester si, dans un cas d'étude 'difficile', le TSA fait mieux que la méthode concurrente. Un tel terrain sera réalisé sur un espace produit composé de stimuli aux différences fines. Ces dernières permettront de rendre la réalisation de la tâche de caractérisation difficile.

Ces différentes propositions méthodologiques ont donné naissance à une nouvelle méthode d'évaluation sensorielle : le 'Tri Séquentiel Agglomératif' (TSA). Nous pensons que cette approche est en mesure de répondre aux attentes actuelles du domaine de l'analyse sensorielle. Nous espérons qu'elle deviendra un outil utile pour les analystes sensoriels qui souhaitent étudier la représentation perceptuelle d'un grand ensemble de stimuli.



Chapitre 2

ANALYSE STATISTIQUE DES DONNÉES DE TRI SÉQUENTIEL AGGLOMÉRATIF

RÉSUMÉ : En Tri Séquentiel Agglomératif, les résultats permettant d’analyser la structure de variabilité perceptuelle observée au sein de l’ensemble de stimuli sont primordiaux. Au-delà de ces premiers résultats, ceux permettant de comprendre la façon avec laquelle les évaluateurs ont perçu l’ensemble de stimuli sont également importants. Dans ce chapitre, une procédure d’analyse statistique des données orientée stimuli et orientée évaluateurs permettant d’obtenir de tels résultats est proposée. Une étude du format des données de Tri Séquentiel Agglomératif permet de souligner leur originalité : elles présentent les caractéristiques usuelles des données de tri libre verbalisé ainsi que des caractéristiques spécifiques. Concernant l’approche orientée stimuli, un point actualisé des techniques dédiées à l’analyse des données de tri libre verbalisé est proposé. Une combinaison judicieuse de certaines de ces techniques fournit une compréhension harmonieuse des caractéristiques usuelles des données. Elle consiste à combiner des approches factorielles dédiées à l’analyse de données catégorielles et des approches dédiées à l’analyse de données textuelles. La question de la prise en compte des spécificités des données de Tri Séquentiel Agglomératif est ensuite abordée. Basée sur des approches factorielles dédiées à l’analyse de données de comptage, la démarche proposée permet d’étudier la cinétique de caractérisation perceptuelle des stimuli. Concernant l’approche orientée évaluateurs, la procédure proposée repose sur la définition d’un indice de dissimilarité classiquement utilisé dans un contexte de tri libre pour quantifier la dissemblance entre deux partitions. La procédure statistique proposée est motivée par son application sur des données réelles de caractérisation de produits industriels.

CE CHAPITRE EST INSPIRÉ DE L’ARTICLE :

Brard, M., & Lê, S. The Sequential Agglomerative Sorting Task, a new sensory methodology to characterize large sets of products. *Submitted*.

Sommaire

1	Format des données	60
2	Analyse statistique des données orientée stimuli	63
2.1	Analyse des caractéristiques partagées avec les données de tri libre verbalisé : prise en compte des partitions finales sur les stimuli	63
2.1.1	Première description statistique	63
2.1.2	Définition de la distance inter-stimuli et représentation euclidienne associée	64
2.1.3	Représentation euclidienne des caractéristiques perceptuelles et mise en relation avec la représentation euclidienne des stimuli	66
2.2	Analyse des caractéristiques spécifiques aux données de Tri Séquentiel Agglomératif : prise en compte des partitions intermédiaires sur les stimuli	71
2.2.1	Première description statistique	71
2.2.2	Extraction d'informations supplémentaires sur les stimuli	72
2.2.2.1	Étude de la variabilité perceptuelle à chaque étape du Tri Séquentiel Agglomératif	72
2.2.2.2	Évolution de la variabilité perceptuelle lors de la réalisation du Tri Séquentiel Agglomératif	78
3	Analyse statistique des données orientée évaluateurs	80
3.1	Définition de la distance inter-évaluateurs	80
3.2	Création du dendrogramme et stratégie de coupure	80
3.3	Interprétation de la partition d'évaluateurs	81
3.3.1	Interprétation des clusters grâce aux covariables sur les évaluateurs	82
3.3.1.1	Covariable continue sur les évaluateurs	82
3.3.1.2	Covariable catégorielle sur les évaluateurs	83
3.3.2	Interprétation des clusters grâce à leur caractérisation des stimuli	84
4	Illustration sur des données réelles de caractérisation de produits industriels : données 'parfums de gels douche'	85
4.1	Données	86
4.2	Résultats de l'analyse statistique orientée parfums	86
4.3	Résultats de l'analyse statistique orientée consommateurs	89
5	Conclusion	91

Dans un premier temps, nous rappelons le format des données que la méthodologie du Tri Séquentiel Agglomératif (TSA) fournit. Contrairement à la présentation des données réalisée dans le Chapitre 1, nous prenons cette fois-ci le temps de détailler leurs principales caractéristiques. Comme nous le verrons, les données de TSA présentent un format particulier et inédit en analyse sensorielle. Celles-ci sont en effet composées d'une succession chronologique de deux types de partitions labellisées sur les stimuli. Les premières sont les partitions fournies par les évaluateurs à la fin de la tâche de caractérisation. Elles sont dénommées dans ce chapitre 'partitions finales' et correspondent à des données usuelles de tri libre verbalisé. Les secondes sont les partitions fournies par les évaluateurs au cours de la tâche de caractérisation. Elles sont dénommées dans ce chapitre 'partitions intermédiaires' et contiennent des données manquantes dans la mesure où certains stimuli n'ont pas été évalués par certains évaluateurs lors de certaines étapes. C'est sur ces partitions intermédiaires que repose l'originalité des données de TSA.

Dans un second temps, nous abordons la manière avec laquelle il est possible d'analyser les données de TSA selon une approche orientée stimuli. Nous nous focalisons tout d'abord sur l'ensemble des partitions finales. Ces données pouvant être assimilées à des données de tri libre verbalisé, nous proposons une synthèse des approches statistiques utilisées dans ce cadre. Ces différentes approches sont issues de domaines de recherche variés et ont été développées de manière déconnectée. Dans ce chapitre, nous proposons de les combiner judicieusement afin d'obtenir une compréhension harmonieuse des caractéristiques usuelles des données. Nous verrons comment la procédure proposée permet d'obtenir une représentation euclidienne des stimuli enrichie de la projection des caractéristiques perceptuelles utilisées par les évaluateurs lors du TSA. Cette représentation est obtenue grâce à une approche factorielle dédiée à l'analyse de données catégorielles. Nous verrons comment la combinaison de cette dernière avec des techniques dédiées au traitement de données textuelles permet de pallier à certains problèmes de visualisation. Nous présentons ensuite les résultats de cette procédure lorsqu'elle est appliquée aux données de TSA. Nous verrons qu'ils permettent de comprendre la variabilité perceptuelle observée au sein de l'ensemble de stimuli à la fin de la tâche de caractérisation. Cependant, ils ne sont basés que sur les seules partitions finales et occultent donc totalement la spécificité des données. C'est pourquoi nous abordons également dans ce chapitre la question de savoir dans quelle mesure les partitions intermédiaires peuvent également être utilisées et valorisées. Dans ce cadre, nous proposons une démarche basée sur une technique dédiée à l'analyse de données de comptage. Elle est appliquée à chaque ensemble de partitions intermédiaires (chaque ensemble correspond à une étape du TSA). Ces derniers étant agencés de manière chronologique, il est possible d'étudier l'évolution dans le temps des résultats. Nous verrons comment cette étude permet de mettre en évidence la spécificité perceptuelle de certains stimuli.

Dans un troisième temps, nous abordons la manière avec laquelle il est possible d’analyser les données de TSA selon une approche orientée évaluateurs. Du fait de la diversité probable de perception, les partitions sur les stimuli fournies par les évaluateurs peuvent être sujettes à de la variabilité. Cette dernière est l’expression d’un désaccord entre les évaluateurs. L’objectif est alors d’analyser la manière avec laquelle ces désaccords sont structurés. Dans ce cadre, nous nous limitons à la seule prise en compte des partitions finales fournies par les évaluateurs, l’intégration des partitions intermédiaires n’ayant pas été réalisée à ce jour. La procédure proposée repose sur la définition d’une distance inter-évaluateurs basée sur un indice de dissimilarité entre partitions. Celle-ci reflète de manière spécifique un désaccord de perception entre les évaluateurs. Elle est utilisée comme base à la construction d’un dendrogramme. Une fois ce dendrogramme coupé grâce à un indice de coupure spécifique, il fournit une partition des évaluateurs selon leur perception de l’ensemble de stimuli. L’interprétation des clusters (c.-à-d. groupes homogènes) d’évaluateurs identifiés est facilitée grâce à l’intégration de covariables véhiculant des informations sur les évaluateurs. Elle est également facilitée grâce à la définition d’une partition sur les stimuli dite ‘consensus’ au sein de chaque cluster d’évaluateurs. Il est ainsi possible de déterminer sur quelles spécificités de perception des stimuli reposent ces différents clusters d’évaluateurs.

Dans un quatrième et dernier temps, nous présentons l’application des deux procédures statistiques proposées à des données réelles de TSA. Ce sont des données de caractérisation sensorielle obtenues sur un ensemble de 40 parfums de gels douche, la tâche de caractérisation ayant été réalisée par 60 consommateurs.

1 Format des données

On note S le nombre de stimuli, P le nombre d’évaluateurs, et I le nombre de sous-ensembles. Chaque évaluateur fournit une partition labellisée des stimuli au terme de chaque présentation d’un sous-ensemble de stimuli. Pour un évaluateur donné, les données de TSA sont donc les suivantes : pour le s -ème stimulus présenté dans le i -ème sous-ensemble, l’évaluation par ce p -ème évaluateur est identifiée sous la forme de la valeur notée x_{sip} . Cette dernière correspond à la description (c.-à-d. l’ensemble des caractéristiques) du groupe auquel appartient le stimulus du sous-ensemble pour cet évaluateur (p. ex. ‘doux, sucré’). Chaque évaluateur est donc assimilé à une matrice contenant S lignes et I colonnes (voir Figure 19).

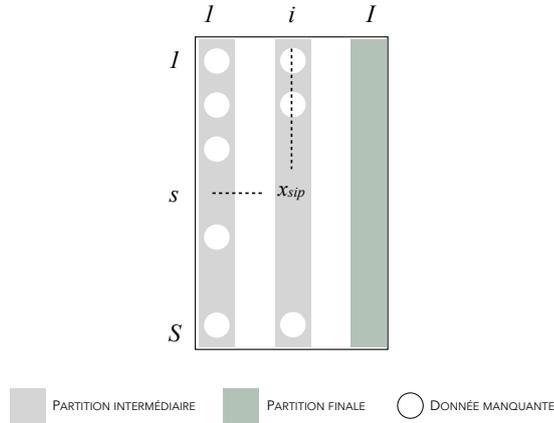


Figure 19: Données recueillies à l'échelle d'un évaluateur dans le cadre de l'implémentation d'une tâche de Tri Séquentiel Agglomératif. La valeur x_{sip} correspond à la description du groupe auquel appartient le stimulus s présenté dans le i -ème sous-ensemble pour cet évaluateur p .

Dans cette matrice, chaque colonne i correspond à une variable catégorielle à G_{ip} modalités, où G_{ip} représente le nombre de groupes de stimuli formés par l'évaluateur p à la fin de la présentation du i -ème sous-ensemble. Les stimuli étant présentés à l'évaluateur sous la forme de sous-ensembles successifs, tous les stimuli ne sont pas évalués à chaque étape. Ces absences d'évaluations sont illustrées par le biais de données manquantes au sein de la matrice $S \times I$ de l'évaluateur.

Les $I - 1$ premières colonnes correspondent aux partitions intermédiaires fournies par l'évaluateur. Elles définissent l'originalité et la spécificité des données de TSA. La I -ème colonne correspond, elle, à la partition finale sur les stimuli. Cette dernière partition est la seule qui ne contient pas de données manquantes. Elle correspond à des données usuelles de tri libre verbalisé.

À l'échelle du panel d'évaluateurs, les données de TSA sont mises sous la forme d'un tableau multiple à S lignes et $P \times I$ colonnes. Ce tableau correspond à la concaténation des P différentes matrices individuelles de dimensions $S \times I$ (voir Figure 20).

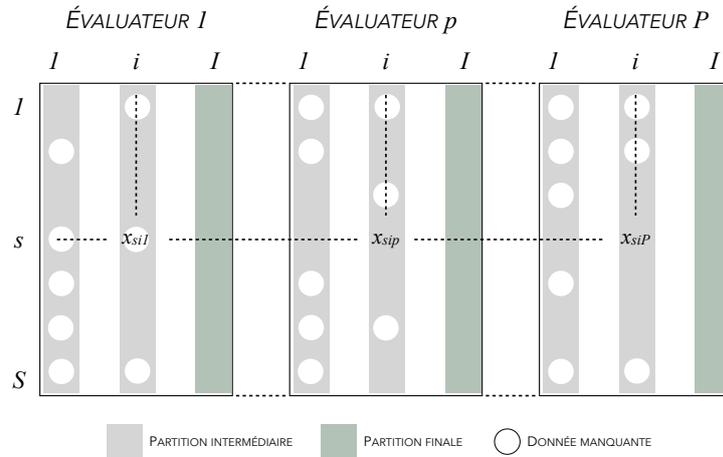


Figure 20: Données recueillies à l'échelle du panel d'évaluateurs dans le cadre de l'implémentation d'une tâche de Tri Séquentiel Agglomératif. La valeur x_{sip} correspond à la description du groupe auquel appartient le stimulus s présenté dans le i -ème sous-ensemble pour cet évaluateur p . Si le stimulus s n'était pas dans le i -ème sous-ensemble de cet évaluateur p , x_{sip} est une donnée manquante.

Un jeu de données issu d'une expérience fictive pédagogique est utilisé pour illustrer l'implémentation des différentes approches présentées dans ce chapitre. Il est présenté dans la Figure 21.

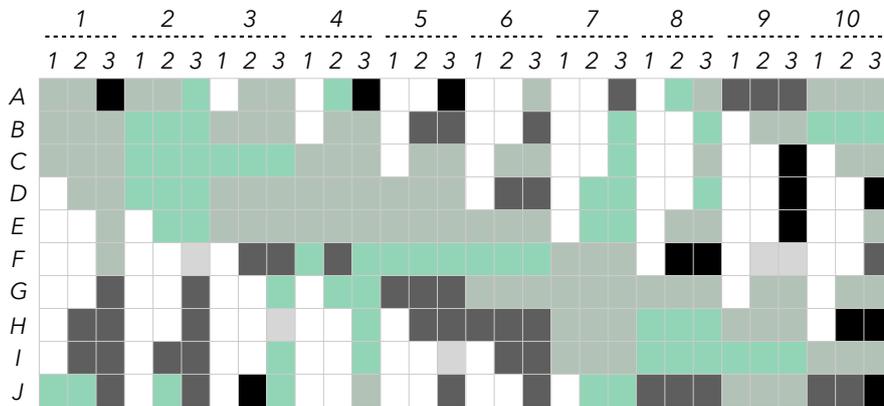


Figure 21: Expérience fictive pédagogique : Données de Tri Séquentiel Agglomératif issues de la caractérisation de 10 stimuli par 10 évaluateurs lors de 3 étapes successives. Une case blanche correspond à la présence d'une donnée manquante. Pour chaque évaluateur, les différents groupes de stimuli formés sont indiqués par un code couleur. Chaque case contient l'ensemble de caractéristiques utilisé par l'évaluateur pour décrire le groupe de stimuli (labellisation non indiquée ici).

2 Analyse statistique des données orientée stimuli

2.1 Analyse des caractéristiques partagées avec les données de tri libre verbalisé : prise en compte des partitions finales sur les stimuli

Dans un premier temps, la variabilité perceptuelle existant au sein de l'ensemble de stimuli est déduite des données recueillies à la fin de la tâche de TSA. C'est pourquoi nous nous intéressons ici à la matrice contenant les partitions finales sur les stimuli fournies par les évaluateurs. Celle-ci contient en lignes les stimuli et en colonnes les évaluateurs. À l'intersection d'une ligne et d'une colonne se trouve l'ensemble de caractéristiques utilisé par l'évaluateur à la fin du TSA pour décrire le groupe auquel appartient le stimulus (voir Figure 22). Ces données sont assimilables à des données issues d'une tâche de tri libre verbalisé.

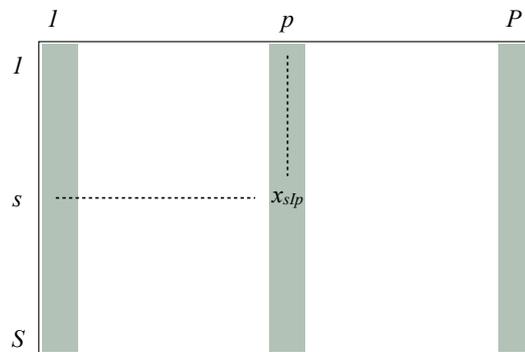


Figure 22: Matrice contenant les partitions finales sur les stimuli fournies par les évaluateurs. La valeur x_{sIp} correspond à la description du groupe auquel appartient le stimulus s pour l'évaluateur p à la fin de la tâche de Tri Séquentiel Agglomératif (c.-à-d. lors de la présentation du l -ème sous ensemble).

2.1.1 Première description statistique

Les partitions finales sont dans un premier temps synthétisées au travers de diagrammes en bâtons. Ceux-ci permettent de mieux comprendre le comportement des évaluateurs lors de la dernière étape du TSA. Ils représentent le nombre de groupes de stimuli constitués par évaluateur ainsi que le nombre de stimuli au sein d'un groupe. Ils indiquent si les évaluateurs ont formé beaucoup de groupes de stimuli à la fin du TSA, et si ces groupes sont composés de beaucoup de stimuli ou non. Ces diagrammes permettent par exemple d'identifier si certains stimuli ont été isolés dans des groupes. Cet aspect révèle la spécificité de ces stimuli.

Un résumé des associations entre stimuli est fourni par une matrice de cooccurrences. Cette matrice de dimension $S \times S$ comptabilise, pour chaque coupe de stimuli (s, s') , le nombre d'évaluateurs ayant placé les stimuli s et s' dans un même

groupe. Elle permet d'identifier les associations les plus fréquentes entre stimuli.

En outre, les partitions finales présentent la particularité d'être labellisées. Il est donc également possible d'établir un bilan sur les caractéristiques perceptuelles utilisées par les évaluateurs pour décrire leurs groupes de stimuli. Un bilan de l'utilisation de ces caractéristiques est réalisé au travers de deux diagrammes en bâtons. Le premier correspond au nombre de caractéristiques perceptuelles utilisées pour décrire un groupe. Le second représente le nombre de stimuli décrits par les quinze caractéristiques perceptuelles majoritairement utilisées à l'échelle du panel. Ces diagrammes permettent d'identifier les caractéristiques perceptuelles majoritairement présentes au sein de l'ensemble de stimuli.

Ces différentes représentations ne sont pas illustrées ici mais seront présentées lors de l'analyse de jeux de données réels dans la section 4 de ce chapitre ainsi que dans le Chapitre 3.

2.1.2 Définition de la distance inter-stimuli et représentation euclidienne associée

Dans un contexte de tri libre verbalisé, l'expérimentateur peut utiliser deux familles de techniques statistiques pour comprendre comment ont été perçus les stimuli par les évaluateurs. Les premières relèvent des techniques factorielles et les secondes des techniques de classification (Faye, Courcoux, Qannari, & Giboreau, 2011). Les techniques factorielles fournissent une représentation euclidienne des stimuli, et permettent d'identifier les distances entre les stimuli et de comprendre les caractéristiques qui les différencient. Dans ce cadre, la procédure statistique usuelle consiste à agréger les données recueillies dans un tableau de cooccurrences qui rassemble le nombre d'associations pour chaque couple de stimuli. Il est utilisé pour établir une matrice de dissimilarités. Celle-ci est ensuite analysée par MultiDimensional Scaling (MDS), une méthode dédiée à l'analyse de matrices de dissimilarités (Kruskal, 1964) : Chollet et al. (2014); Courcoux et al. (2015); Faye et al. (2004); Lawless (1989). Dans ce type de procédure, les données sont prises en compte dans leur ensemble (c.-à-d. sur la globalité du panel) et non du point de vue de chaque évaluateur. Il est ainsi impossible d'étudier la diversité des évaluations que ces derniers ont fournies.

Plusieurs approches permettent de contourner cette limite. C'est par exemple le cas de DISTATIS (Abdi, Valentin, Chollet, & Chr ea, 2007). Cette g n ralisation du MDS est d di e   l'analyse des matrices individuelles de dissimilarit s sur les stimuli fournies par les  valuateurs. Elle permet d'obtenir une repr sentation 'compromis' des stimuli ainsi qu'une repr sentation des  valuateurs. Une solution alternative consiste   consid rer chaque  valuateur comme une variable cat gorielle (d finie sur les stimuli) dont chaque modalit  correspond   un groupe de stimuli. Ceci sugg re d'analyser ces donn es ayant en lignes les stimuli et en colonnes les  valuateurs par Analyse des Correspondances Multiples (ACM), une m thode

dédiée à l'analyse de matrices de données catégorielles (Greenacre, 1984; Lebart, 1975) : Abdi and Valentin (2007); Cadoret et al. (2009); Van der Kloot and Van Herck (1991); Takane (1980). Comme nous le voyons dans la section suivante, l'ACM présente l'avantage de fournir d'autres types de représentations euclidiennes. Ces dernières permettent de mieux comprendre les distances inter-stimuli établies. C'est pour cette raison que nous faisons le choix de nous concentrer sur cette approche pour analyser les partitions finales fournies par les évaluateurs dans un contexte de TSA.

Dans le cadre de l'ACM, les données sont prises en compte à travers un tableau dit *disjonctif complet*. Dans celui-ci, chaque évaluateur est assimilé à autant de vecteurs binaires qu'il a créé de groupes de stimuli. Un stimulus prend la valeur de 1 pour le vecteur correspondant au groupe auquel il appartient, et la valeur de 0 pour les autres (voir Figure 23).

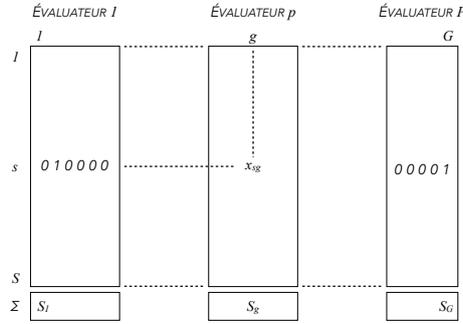


Figure 23: Tableau disjonctif complet dans le cas de données de Tri Séquentiel Agglomératif (partitions finales) : x_{sg} vaut 1 si le s -ème stimulus appartient au groupe g et 0 sinon. Le 1-er évaluateur a créé 6 groupes de stimuli et le P -ème évaluateur a créé 5 groupes de stimuli à la fin de la tâche.

Les stimuli évoluent alors dans un espace à G dimensions, où G représente le nombre de groupes de stimuli formés à la fin du TSA sur la globalité du panel d'évaluateurs (c.-à-d. la somme des nombres de groupes formés par les évaluateurs). Dans cet espace, la distance entre deux stimuli s et s' est définie comme suit :

$$d_{ACM}^2(s, s') = \frac{S}{P} \sum_{g=1}^G \frac{S}{S_g} (x_{sg} - x_{s'g})^2 \quad (2.1)$$

où S représente le nombre total de stimuli évalués pendant la tâche du TSA ; P représente le nombre total d'évaluateurs ayant participé à l'expérience ; G représente le nombre de groupes de stimuli formés à la fin du TSA sur la globalité du panel d'évaluateurs.

Cette distance inter-stimuli de l'ACM est d'autant plus faible que les stimuli s et

s' ont été placés un grand nombre de fois dans un même groupe à la fin de la tâche sur la globalité du panel. L'appartenance à un même groupe à l'issue de la tâche de caractérisation témoigne en effet d'une forte similitude des deux stimuli.

La Figure 24 présente la représentation euclidienne des stimuli fournie par l'ACM dans le cadre du jeu de données pédagogique. La représentation oppose principalement le stimulus F aux autres stimuli (voir la première dimension). Son isolement dans un groupe à la fin du TSA par la majorité des évaluateurs explique cette opposition. Sept évaluateurs sur dix ont en effet placé ce stimulus seul dans un groupe. Cet aspect témoigne d'une dissimilarité entre le stimulus F et le reste de l'ensemble de stimuli. Les fortes proximités entre d'autres stimuli s'expliquent par le fait qu'ils ont été placés dans un même groupe par un grand nombre d'évaluateurs. C'est notamment le cas des stimuli C et E, placés ensemble par 9 évaluateurs sur 10.

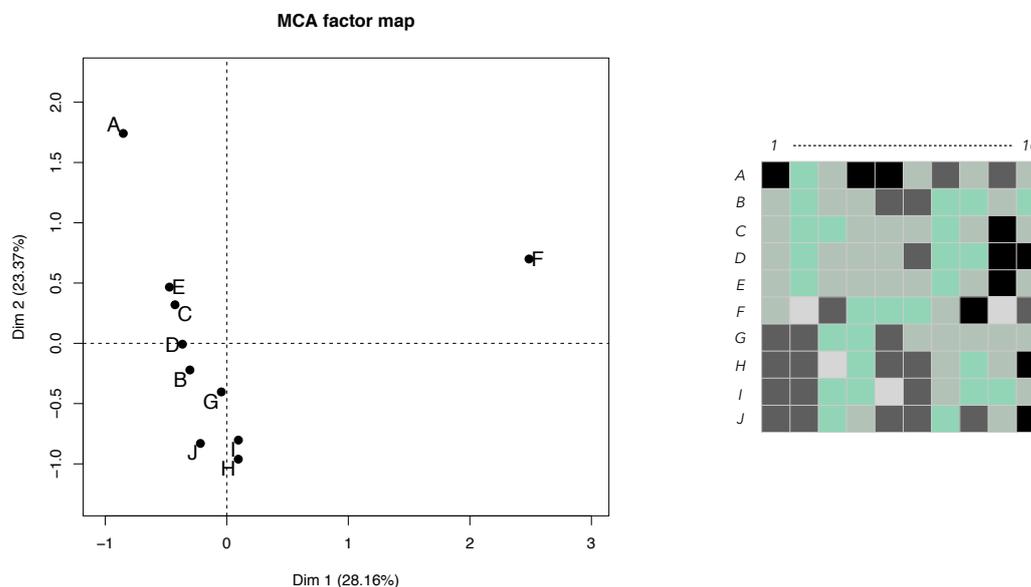


Figure 24: Expérience fictive pédagogique : Représentation des stimuli fournie par l'Analyse des Correspondances Multiples appliquée sur les partitions finales.

2.1.3 Représentation euclidienne des caractéristiques perceptuelles et mise en relation avec la représentation euclidienne des stimuli

Au-delà de la représentation des stimuli, l'ACM fournit également une représentation euclidienne des modalités des variables catégorielles. Dans un contexte de TSA, chaque modalité correspond à un ensemble de caractéristiques utilisé par un évaluateur (p. ex. 'doux, sucré'). Chaque évaluateur est ainsi associé à autant de points qu'il a créé de groupes à la fin du TSA. Sur cette représentation, deux ensembles de caractéristiques sont d'autant plus proches (resp. distants) qu'ils ont

été fréquemment (resp. rarement) utilisés pour décrire les mêmes stimuli.

Cette représentation présente l'avantage d'être liée à la représentation des stimuli par des propriétés barycentriques. Il est ainsi possible de superposer ces deux types de représentations. Sur la représentation superposée, un stimulus est au barycentre des ensembles de caractéristiques qui lui ont été associés. De la même façon, un ensemble de caractéristiques est au barycentre des stimuli qu'il caractérise. Cette représentation permet ainsi de mieux comprendre les distances inter-stimuli établies par l'ACM. Pour souligner son intérêt, la Figure 25 présente la représentation superposée obtenue à partir du jeu de données pédagogique. Grâce à cette représentation, il est possible de déterminer que le stimulus F a été décrit comme écœurant, boisé, avec des notes de fruits rouges. La présence de ces caractéristiques dans ce stimulus peut expliquer son caractère particulier au sein de l'ensemble de stimuli.

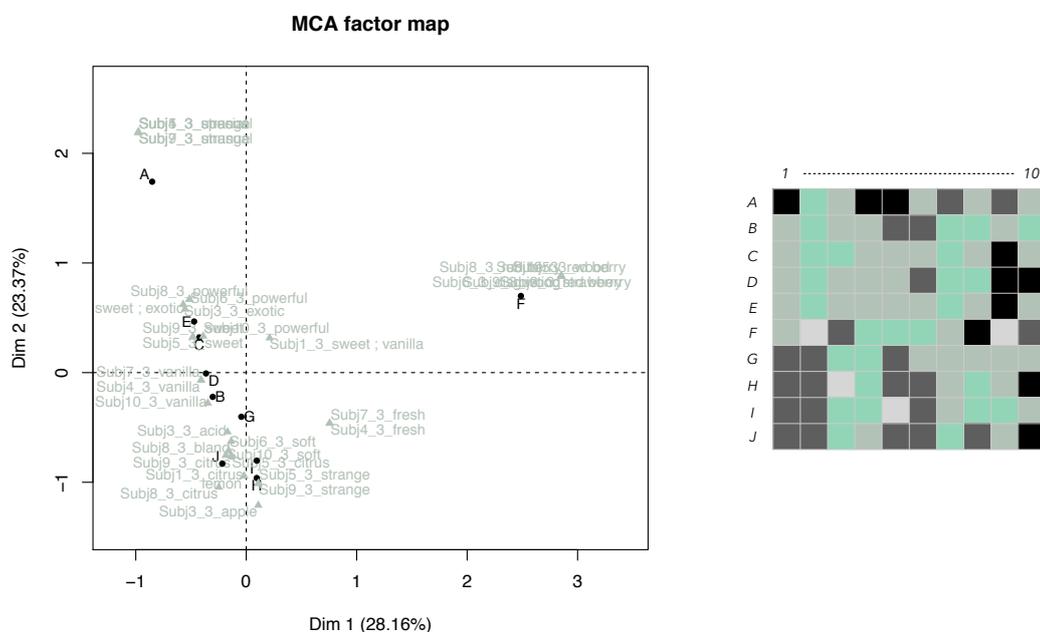


Figure 25: Expérience fictive pédagogique : Représentation des stimuli et des ensembles de caractéristiques fournie par l'Analyse des Correspondances Multiples appliquée sur les partitions finales.

Malgré l'intérêt de la représentation euclidienne des modalités, son interprétation peut s'avérer délicate. Chaque point est en effet associé à un évaluateur donné. Il est ainsi possible d'obtenir deux points pour un même ensemble de caractéristiques. Sur la Figure 25, c'est par exemple le cas de l'ensemble 'frais' qui est représenté par un point correspondant à l'évaluateur 4 et par un point correspondant à l'évaluateur 7. Cet ensemble de caractéristiques a effectivement été utilisé par ces deux évaluateurs pour décrire l'un de leurs groupes à la fin du TSA. En

outre, chaque point correspond à un ensemble des caractéristiques et non à des caractéristiques seules. Sur la Figure 25, la caractéristique ‘vanille’ apparaît parfois seule mais également associée à la caractéristique ‘doux’. C’est le cas pour le point associé à l’évaluateur 1 : ‘doux ; vanille’.

Dans le but de surmonter cette limite, une solution consiste à placer une caractéristique seule au barycentre des ensembles de caractéristiques qui la contiennent. Les coordonnées d’une caractéristique sont ainsi obtenues grâce à la formule suivante :

$$\text{coord}B_d(c) = \frac{1}{N} \sum_{n=1}^N \text{coord}_d(c_n) \quad (2.2)$$

où $\text{coord}B_d(c)$ représente la coordonnée de la caractéristique c sur la d -ème dimension de l’ACM ; N est le nombre total de fois où la caractéristique c a été utilisée pour décrire un groupe de stimuli à l’échelle du panel ; et $\text{coord}_d(c_n)$ représente la coordonnée du n -ème ensemble de caractéristiques qui contient la caractéristique c sur la d -ème dimension de l’ACM.

La Figure 26 expose la représentation résultant de cette approche. Comparativement à la représentation de base présentée en Figure 25, elle présente une meilleure lisibilité. L’interprétation est donc facilitée.

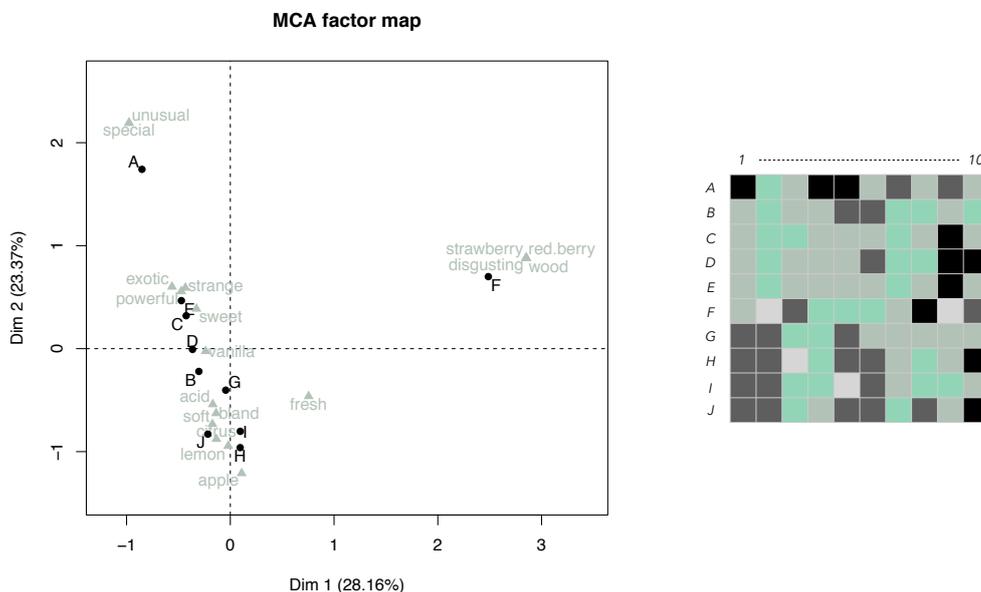


Figure 26: Expérience fictive pédagogique : Représentation des stimuli et des caractéristiques fournie par l’Analyse des Correspondances Multiples appliquée sur les partitions finales.

Cependant, cette approche n'a de sens que si l'on considère qu'une caractéristique perceptuelle a la même signification pour tous les évaluateurs qui l'ont utilisée. Cette hypothèse est pourtant discutable (Cadoret, 2010). Pour dépasser cette limite, une procédure complémentaire consiste à n'afficher sur la représentation euclidienne que les caractéristiques dites 'consensuelles'. Ces caractéristiques représentent la même réalité pour les évaluateurs au sens où elles sont associées aux mêmes stimuli par la majorité d'entre eux. Cette stratégie a d'abord été développée dans le cadre de l'analyse statistique de commentaires libres (Kostov, Bécue-Bertaut, & Husson, 2014). Elle a par la suite été étendue à l'analyse des descriptions fournies lors d'une tâche de tri libre verbalisé.

Elle consiste à calculer le degré de consensus d'utilisation de chaque caractéristique. La variabilité des coordonnées des ensembles de caractéristiques qui la contiennent est un indicateur de ce degré de consensus. Cette variabilité est effectivement d'autant plus faible que la caractéristique a été utilisée pour décrire les mêmes stimuli. La mesure du degré de consensus d'utilisation d'une caractéristique est alors réalisée au travers de l'inertie intra qui lui est associée :

$$WI(c) = \sum_{d=1}^D \frac{1}{N_c} \sum_{n=1}^{N_c} (\text{coord}_d(c_n) - \text{coord}B_d(c))^2 \quad (2.3)$$

où $WI(c)$ représente l'inertie intra de la caractéristique c ; N_c est le nombre total de fois où la caractéristique c a été utilisée pour décrire un groupe de stimuli à l'échelle du panel (c.-à-d. le nombre d'ensemble de caractéristiques qui contiennent c) ; $\text{coord}_d(c_n)$ représente la coordonnée du n -ème ensemble de caractéristiques qui contient c sur la d -ème dimension de l'ACM ; et $\text{coord}B_d(c)$ est la coordonnée du barycentre de la caractéristique c sur la d -ème dimension de l'ACM.

Un test statistique est ensuite implémenté. Il permet de déterminer si la caractéristique c n'est pas consensuelle (hypothèse nulle) ou consensuelle (hypothèse alternative). Rappelons que l'inertie intra de la caractéristique c (c.-à-d. $WI(c)$) est une mesure naturelle de son degré de consensus d'utilisation. Celle-ci est donc utilisée comme statistique de test. Sa distribution sous l'hypothèse nulle est approchée par une méthode de Monte-Carlo. Cette procédure consiste à choisir aléatoirement un échantillon de N_c ensembles de caractéristiques parmi l'intégralité des ensembles de caractéristiques utilisés à l'échelle du panel. Cette opération est répétée un grand nombre de fois. La probabilité critique associée au test est définie comme la proportion d'échantillons aléatoires ayant une inertie intra inférieure ou égale à l'inertie intra de la caractéristique c . Lorsque l'hypothèse nulle est rejetée, la caractéristique c est considérée comme consensuelle. Elle est alors affichée sur la représentation euclidienne.

La Table 14 présente les résultats de ces tests statistiques dans le cadre du jeu

de données pédagogique. La Figure 27 présente la représentation euclidienne correspondante.

Table 14: Expérience fictive pédagogique : Résultats de la procédure de rééchantillonnage permettant l'identification des caractéristiques consensuelles. Seules les caractéristiques utilisées au moins deux fois à l'échelle du panel sont considérées ici.

	Inertie intra $WI(c)$	p -value
sweet	0.378	0.004
citrus	0.616	0.004
powerful	0.146	0.010
red.berry	0.000	0.010
vanilla	1.002	0.032
unusual	0.000	0.052
wood	0.000	0.056
exotic	0.090	0.070
fresh	0.000	0.076
soft	0.296	0.086
strange	6.485	0.936

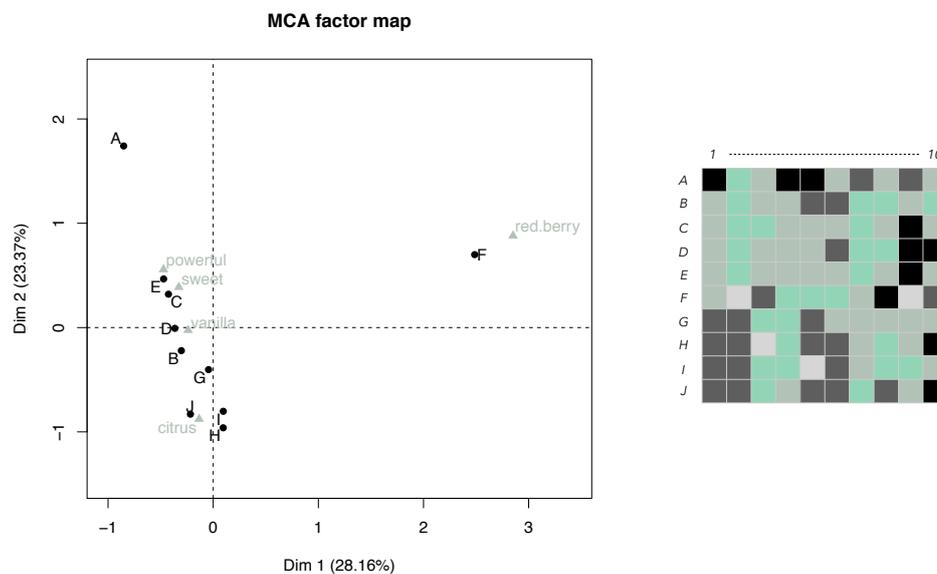


Figure 27: Expérience fictive pédagogique : Représentation des stimuli et des caractéristiques consensuelles fournie par l'Analyse des Correspondances Multiples appliquée sur les partitions finales. Le seuil de significativité associé à la procédure de rééchantillonnage est fixé à 5%.

Dans cette section, nous avons proposé d'utiliser les procédures statistiques dédiées au traitement des données de tri libre verbalisé pour analyser les données de TSA.

Celles-ci ne sont pourtant pas optimales. Étant basées uniquement sur les partitions finales fournies par les évaluateurs, elles occultent la spécificité des données de TSA.

2.2 Analyse des caractéristiques spécifiques aux données de Tri Séquentiel Agglomératif : prise en compte des partitions intermédiaires sur les stimuli

Nous nous attachons ici à proposer une procédure permettant de prendre en compte les partitions intermédiaires fournies par les évaluateurs. Chaque étape du TSA est analysée indépendamment. Pour chaque étape, la matrice contenant en lignes les stimuli et en colonnes les évaluateurs est considérée. À l'intersection d'une ligne et d'une colonne se trouve l'ensemble de caractéristiques utilisé par l'évaluateur à la fin de cette étape pour décrire le groupe auquel appartient le stimulus (voir Figure 28).

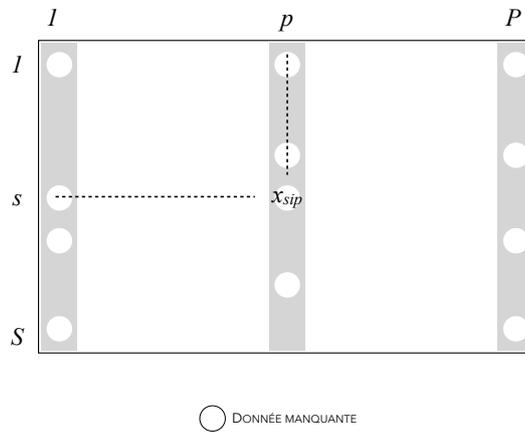


Figure 28: Matrice contenant les partitions intermédiaires sur les stimuli fournies par les évaluateurs à la fin d'une étape du Tri Séquentiel Agglomératif. La valeur x_{sip} correspond à la description du groupe auquel appartient le stimulus s pour l'évaluateur p à la fin de l'étape. Si le stimulus s n'était pas dans le i -ème sous-ensemble de cet évaluateur p , x_{sip} est une donnée manquante.

2.2.1 Première description statistique

Tout comme les partitions finales, les partitions intermédiaires peuvent être synthétisées au travers de diagrammes en bâtons. Pour chaque étape du TSA, des diagrammes sont créés pour représenter le nombre de groupes de stimuli constitués par évaluateur ainsi que le nombre de stimuli au sein d'un groupe. Les diagrammes des différentes étapes peuvent ensuite être mis en relation. Cette analyse permet de mieux comprendre le comportement des évaluateurs lors de l'intégralité de la tâche de TSA. Il est par exemple possible de déterminer si l'une des étapes du TSA a mené

à une augmentation (ou à une diminution) du nombre de groupes de stimuli formés.

En outre, tout comme les partitions finales, les partitions intermédiaires présentent la particularité d'être labellisées. Il est donc également possible d'établir un bilan sur les caractéristiques perceptuelles utilisées par les évaluateurs pour décrire leurs groupes de stimuli lors des différentes étapes du TSA. Un bilan de l'utilisation de ces caractéristiques est réalisé au travers de deux diagrammes en bâtons. Le premier correspond au nombre de caractéristiques perceptuelles utilisées pour décrire un groupe. Le second représente le nombre de stimuli décrits par les quinze caractéristiques perceptuelles majoritairement utilisées à l'échelle du panel. La mise en relation des diagrammes des différentes étapes permet d'identifier si la présentation d'un sous-ensemble de stimuli a entraîné une différence de verbalisation des groupes.

Ces différentes représentations ne sont pas illustrées ici mais seront présentées lors de l'analyse de jeux de données réels dans la section 4 de ce chapitre ainsi que dans le Chapitre 3.

2.2.2 Extraction d'informations supplémentaires sur les stimuli

Le TSA implique plusieurs étapes liées de manière chronologique. Il est alors possible d'analyser la cinétique associée à la caractérisation perceptuelle des stimuli. Nous proposons ici une procédure permettant d'aborder cette problématique. Dans un premier temps, la variabilité perceptuelle des stimuli à la fin de chaque étape du TSA est étudiée. Cette approche permet de comprendre les similitudes et dissemblances perçues entre les stimuli à chaque étape du TSA. Il est ainsi possible de déterminer si chaque stimulus a plutôt été perçu comme un stimulus 'moyen' ou comme un stimulus 'singulier'. Cette spécificité de chaque stimulus est synthétisée au travers d'un indicateur numérique. Dans un second temps, l'évolution au cours du temps de cet indicateur est analysée. Cette approche permet d'étudier la cinétique associée à la caractérisation perceptuelle des stimuli. Il est par exemple possible de déterminer si un stimulus jugé singulier lors des premières étapes a été jugé moyen à la fin du TSA. Comme nous le voyons, ces informations permettent de mieux comprendre la représentation perceptuelle des stimuli issue des partitions finales seules. Elles permettent de nuancer certaines conclusions tirées de cette représentation.

2.2.2.1 Étude de la variabilité perceptuelle à chaque étape du Tri Séquentiel Agglomératif

Représentation euclidienne des stimuli à chaque étape du Tri Séquentiel Agglomératif

Les évaluations perceptuelles des stimuli à la fin de chaque étape du TSA sont synthétisées dans une matrice de la forme de celle présentée en Figure 28. Ces données présentent la particularité de contenir des données manquantes. Chaque

stimulus n'est en effet évalué que par un sous-ensemble d'évaluateurs à chaque étape.

Précédemment, nous avons vu que l'ACM peut être utilisée pour étudier la variabilité perceptuelle des stimuli à la fin de la dernière étape du TSA. Cependant, la présence de données manquantes dans les matrices correspondant aux étapes intermédiaires limite son utilisation. L'ACM prend en effet en compte les données manquantes en créant une modalité supplémentaire dans les variables catégorielles qui en contiennent. Dans notre cas précis, l'ACM crée un groupe de stimuli supplémentaire pour chaque évaluateur. Ce groupe est constitué des stimuli qui ne lui ont pas été présentés. Il est raisonnable de supposer que ceci peut avoir pour effet de modifier les distances entre les stimuli évalués.

Pour dépasser cette limite, nous proposons d'agréger les données dans un tableau de cooccurrences qui rassemble le nombre d'associations pour chaque couple de stimuli. Les données considérées sont donc des données de comptage qui contiennent, à la fois en lignes et en colonnes, les stimuli (voir Figure 29).

	I	s''	S	Σ
I				S^+_I
s		$x^+_{ss''}$		S^+_s
S				S^+_S
Σ	S^+_I	$S^+_{s''}$	S^+_S	

Figure 29: Matrice de cooccurrences obtenue à la fin d'une étape du Tri Séquentiel Agglomératif. La valeur $x^+_{ss''}$ correspond au nombre d'évaluateurs ayant placé les stimuli s et s'' dans un même groupe à la fin de cette étape. La valeur S^+_s est une valeur marginale. Elle est égale au nombre d'associations impliquant le s -ème stimulus sur la globalité du panel additionné au nombre d'évaluateurs ayant évalué ce stimulus lors de cette étape (valeur sur la première diagonale de la matrice).

L'obtention d'un tel tableau de cooccurrences est courant dans un contexte de tri libre. Il est usuellement transformé en une matrice de dissimilarité puis analysé par MultiDimensional Scaling (MDS). Une procédure alternative consiste à utiliser une approche dédiée à l'analyse de données de comptage : l'Analyse Factorielle des Correspondances (AFC) (Benzécri, 1992). Celle-ci est appliquée directement sur le tableau de cooccurrences. Son application à des données similaires a été récemment proposée par Cariou and Qannari (2018). Tout comme l'ACM, l'AFC est basée sur une décomposition en valeurs singulières. C'est pourquoi nous faisons ici le choix de nous concentrer sur cette approche.

En AFC, les stimuli évoluent dans un espace à S dimensions, où S représente le nombre de stimuli évalués. Dans cet espace, la distance entre deux stimuli s et s' est définie comme suit :

$$d_{AFC}^2(s, s') = \frac{S}{P} \sum_{s''=1}^S \frac{1}{S_{s''}^+} (x_{ss''}^+ - x_{s's''}^+)^2 \quad (2.4)$$

où S représente le nombre total de stimuli évalués ; P représente le nombre total d'évaluateurs ayant participé à l'expérience ; $S_{s''}^+$ représente la valeur marginale associée au s'' -ème stimulus ; et $x_{ss''}^+$ représente le nombre de fois où le s -ème stimulus a été associé avec le s'' -ème stimulus sur la globalité du panel.

Cette distance inter-stimuli de l'AFC est d'autant plus faible que les valeurs $x_{ss''}^+$ et $x_{s's''}^+$ sont similaires. En d'autres termes, elle est d'autant plus faible que les deux stimuli présentent des profils semblables. En outre, le tableau de cooccurrences présente ici la particularité d'être symétrique selon la première diagonale. De plus, les lignes et les colonnes du tableau de cooccurrences font référence aux mêmes entités : les stimuli. Par conséquent, les représentations euclidiennes des lignes et des colonnes sont identiques. Sur la représentation euclidienne correspondante, un stimulus est attiré par les stimuli avec lesquels il a été associé dans un même groupe un grand nombre de fois.

La Figure 30 présente la représentation euclidienne des stimuli correspondant à la deuxième étape du TSA dans le cadre de l'expérience fictive pédagogique.

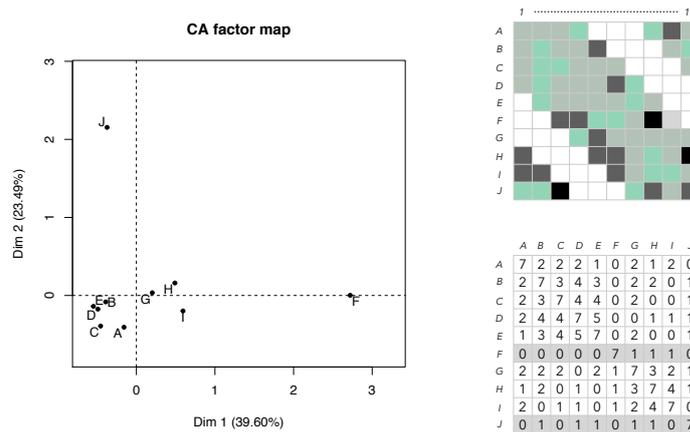


Figure 30: Expérience fictive pédagogique : Représentation des stimuli obtenue à partir des partitions intermédiaires correspondant à l'étape 2. La particularité des stimuli F et J est mise en évidence dans le tableau de cooccurrences (à droite).

L'utilisation d'un plan de présentation des stimuli tel que celui présenté dans le Chapitre 1 assure que chaque stimulus est évalué lors de chaque étape au niveau du panel d'évaluateurs. De la même façon, il assure que les couples de stimuli sont évalués de manière homogène au sein de chaque sous-ensemble de stimuli. Cet équilibre peut pourtant s'avérer imparfait : certains couples de stimuli peuvent être évalués moins souvent que d'autres lors de certaines étapes du TSA. Il est possible de s'affranchir de l'impact de ce déséquilibre lors de la création de la représentation euclidienne des stimuli. Une solution à cette limite consiste en effet à appliquer l'AFC sur le tableau de cooccurrences dont chaque terme a été pondéré par le nombre de présentations effectuées.

La Figure 31 expose les cooccurrences calculées à la fin de la deuxième étape du TSA dans le cadre de l'expérience fictive pédagogique. Elle présente également le nombre de fois où chaque couple de stimuli a été présenté lors de cette étape à l'échelle du panel. Il apparaît par exemple que le couple F-G été évalué 6 fois alors que les couples F-H et F-I ont été évalués respectivement 5 et 4 fois. Chacun de ces trois couples a par ailleurs été associé dans un même groupe une fois sur la globalité du panel. La ressemblance du couple F-G apparaît pourtant plus faible que celle des deux autres couples. La proportion d'évaluateurs les ayant jugés similaires est effectivement plus faible. Le tableau de cooccurrences pondéré permet de prendre en compte cet aspect (voir Figure 32).

	A	B	C	D	E	F	G	H	I	J
A	7	2	2	2	1	0	2	1	2	0
B	2	7	3	4	3	0	2	2	0	1
C	2	3	7	4	4	0	2	0	1	0
D	2	4	4	7	5	0	0	1	1	1
E	1	3	4	5	7	0	2	0	0	1
F	0	0	0	0	0	7	1	1	1	0
G	2	2	2	0	2	1	7	3	2	1
H	1	2	0	1	0	1	3	7	4	1
I	2	0	1	1	0	1	2	4	7	0
J	0	1	0	1	1	0	1	1	0	7

	A	B	C	D	E	F	G	H	I	J
A	7	6	5	4	4	4	4	4	5	6
B	6	7	6	5	4	4	4	4	4	5
C	5	6	7	6	5	4	4	4	4	4
D	4	5	6	7	6	5	4	4	4	4
E	4	4	5	6	7	6	5	4	4	4
F	4	4	4	5	6	7	6	5	4	4
G	4	4	4	4	5	6	7	6	5	4
H	4	4	4	4	4	5	6	7	6	5
I	5	4	4	4	4	4	5	6	7	6
J	6	5	4	4	4	4	4	5	6	7

Figure 31: Expérience fictive pédagogique : Tableau de cooccurrences à la fin de l'étape 2 (à gauche) et tableau du nombre de présentations réalisées lors de l'étape 2 (à droite).

	A	B	C	D	E	F	G	H	I	J
A	1	0.33	0.40	0.50	0.25	0	0.50	0.25	0.40	0
B	0.33	1	0.50	0.80	0.75	0	0.50	0.50	0	0.20
C	0.40	0.50	1	0.67	0.80	0	0.50	0	0.25	0
D	0.50	0.80	0.67	1	0.83	0	0	0.25	0.25	0.25
E	0.25	0.75	0.80	0.83	1	0	0.40	0	0	0.25
F	0	0	0	0	0	1	0.17	0.20	0.25	0
G	0.50	0.50	0.50	0	0.40	0.17	1	0.50	0.40	0.25
H	0.25	0.50	0	0.25	0	0.20	0.50	1	0.67	0.20
I	0.40	0	0.25	0.25	0	0.25	0.40	0.67	1	0
J	0	0.20	0	0.25	0.25	0	0.25	0.20	0	1

Figure 32: Expérience fictive pédagogique : Tableau de cooccurrences pondéré à la fin de l'étape 2.

La Figure 33 présente la représentation euclidienne des stimuli fournie par l'AFC appliquée sur le tableau de cooccurrences pondéré. Le déséquilibre du plan de présentation étant léger, celle-ci est très proche de la représentation euclidienne obtenue directement à partir du tableau de cooccurrences présentée en Figure 30.

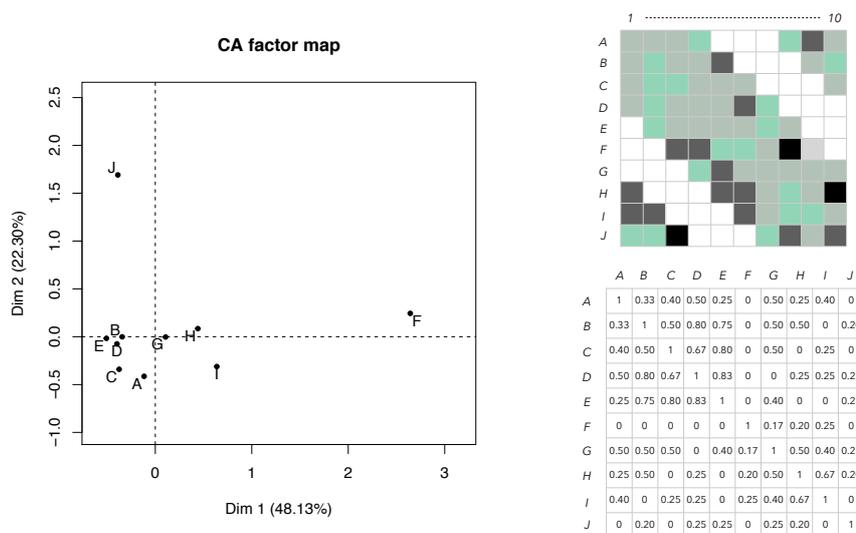


Figure 33: Expérience fictive pédagogique : Représentation des stimuli obtenue à partir des partitions intermédiaires correspondant à l'étape 2 (tableau de cooccurrences pondéré).

Mise en évidence de la singularité des stimuli à chaque étape du Tri Séquentiel Agglomératif

La représentation euclidienne obtenue pour chaque étape du TSA permet de révéler la particularité de certains stimuli. Dans le cadre de l'étape 2 de l'expérience fictive

pédagogique, c'est par exemple le cas des stimuli F et J (voir Figure 33). Ceux-ci s'écartent effectivement nettement du centre de gravité du nuage des stimuli. À cette étape du TSA, ces deux stimuli peuvent alors être définis comme des stimuli 'singuliers' alors que les autres peuvent être considérés comme des stimuli 'moyens'.

La contribution d'un stimulus à l'inertie des axes de la représentation factorielle est un indicateur de sa singularité. Un stimulus contribue en effet d'autant plus à la construction d'un axe factoriel que sa projection sur cet axe est éloignée du centre de gravité du nuage des stimuli. Pour un axe factoriel donné, les stimuli présentant une forte (resp. une faible) contribution sont des stimuli singuliers (resp. moyens). Pour quantifier la singularité perceptuelle d'un stimulus lors d'une étape du TSA, nous proposons d'utiliser l'indicateur numérique suivant :

$$sing(s) = \sum_{d=1}^2 contrib_d(s) = \sum_{d=1}^2 \frac{coord_d(s)^2 \times \frac{n_s^+}{n}}{\iota_d} \quad (2.5)$$

où $contrib_d(s)$ représente la contribution du stimulus s à la construction de la d -ème dimension de l'AFC. $coord_d(s)$ correspond à la coordonnée du stimulus s sur d -ème dimension de l'AFC, n_s^+ représente la valeur de la marge line associée au stimulus s dans le tableau de contingence, n représente la somme des valeurs du tableau de contingence, et ι_d correspond à l'inertie associée à la d -ème dimension de l'AFC. Seules les deux premières dimensions sont considérées ici, puisqu'il s'agit de celles qui reflètent les principaux axes de variabilité perceptuelle de l'ensemble de stimuli. Il est également important de noter que la définition de la contribution est différente dans le cas de l'ACM.

La distribution des valeurs $sing(s)$ est ensuite utilisée pour déterminer quels stimuli sont singuliers et lesquels sont moyens pour cette étape du TSA. Les stimuli dont l'indicateur $sing(s)$ dépasse le quantile d'ordre 85 % de cette distribution sont définis comme des *outliers*. Leur caractère singulier est jugé extrême. Ils sont alors considérés comme des stimuli singuliers. Les stimuli restants sont, eux, considérés comme moyens.

La Table 15 présente les résultats de cette approche dans le cadre de l'expérience fictive pédagogique.

Table 15: Expérience fictive pédagogique : Particularité des stimuli lors des différentes étapes du Tri Séquentiel Agglomératif.

	Étape 1		Étape 2		Étape 3	
	<i>sing(s)</i>	type	<i>sing(s)</i>	type	<i>sing(s)</i>	type
A	2.66	moyen	8.36	moyen	57.58	singulier
B	13.07	moyen	3.24	moyen	1.98	moyen
C	9.63	moyen	9.65	moyen	4.00	moyen
D	14.03	moyen	4.63	moyen	1.75	moyen
E	9.06	moyen	6.63	moyen	6.37	moyen
F	22.83	singulier	69.09	singulier	89.07	singulier
G	8.37	moyen	0.29	moyen	2.60	moyen
H	17.63	moyen	4.50	moyen	14.76	moyen
I	21.58	moyen	11.90	moyen	10.34	moyen
J	81.14	singulier	81.71	singulier	11.55	moyen

2.2.2.2 Évolution de la variabilité perceptuelle lors de la réalisation du Tri Séquentiel Agglomératif

La particularité de chaque stimulus peut évoluer ou non au cours de la réalisation du TSA. C'est l'étude de cette cinétique que nous proposons d'aborder ici. Quatre cas de figure peuvent effectivement apparaître. Ceux-ci sont exposés dans la Table 16. Leur interprétation est également détaillée.

Table 16: Évolution de la particularité des stimuli lors du Tri Séquentiel Agglomératif.

Type d'évolution	Interprétation
Évolution Moyen (lors de toutes les étapes intermédiaires) → Moyen (à l'étape finale)	Stimulus dont le caractère moyen est absolu : il ne dépend pas des autres entités composant l'ensemble de stimuli
Évolution Singulier (lors d'au moins une étape intermédiaire) → Moyen (à l'étape finale)	Stimulus dont le caractère moyen est relatif : il dépend des autres entités composant l'ensemble de stimuli
Évolution Moyen (lors de toutes les étapes intermédiaires) → Singulier (à l'étape finale)	Stimulus dont le caractère singulier est relatif : il dépend des autres entités composant l'ensemble de stimuli
Évolution Singulier (lors d'au moins une étape intermédiaire) → Singulier (à l'étape finale)	Stimulus dont le caractère singulier est absolu : il ne dépend pas des autres entités composant l'ensemble de stimuli

La Figure 34 présente l'évolution de la particularité des stimuli A, B, F et J dans le cadre de l'expérience fictive pédagogique. L'analyse des partitions finales avait mis en évidence le fait que A et F étaient très particuliers au sein de l'ensemble de stimuli (voir Figure 24). L'analyse de l'évolution de leur particularité au cours de la tâche de caractérisation tempère pourtant cette conclusion. Elle indique en effet que ces deux stimuli ne sont pas aussi particuliers l'un que l'autre. Alors que le stimulus F présente une singularité absolue, le stimulus A présente une singularité relative. Dans un contexte de développement de produit, il arrive que l'expérimentateur soit

amené à sélectionner le produit le plus particulier d'une gamme. Dans une telle situation, l'expérimentateur devra alors choisir le produit F. D'autre part, l'analyse des partitions finales avait fait ressortir les stimuli B et J comme des stimuli moyens (voir Figure 24). Là encore, l'analyse de l'évolution de leur particularité permet de nuancer ces résultats. Alors que le caractère moyen du stimulus B est absolu, celui du stimulus J est relatif. Ce dernier a en effet été perçu comme un stimulus singulier lors des premières étapes de caractérisation. Si l'objectif est ici de sélectionner le produit le plus représentatif (c.-à-d. le plus moyen) d'une gamme, il faudra sélectionner le produit B.

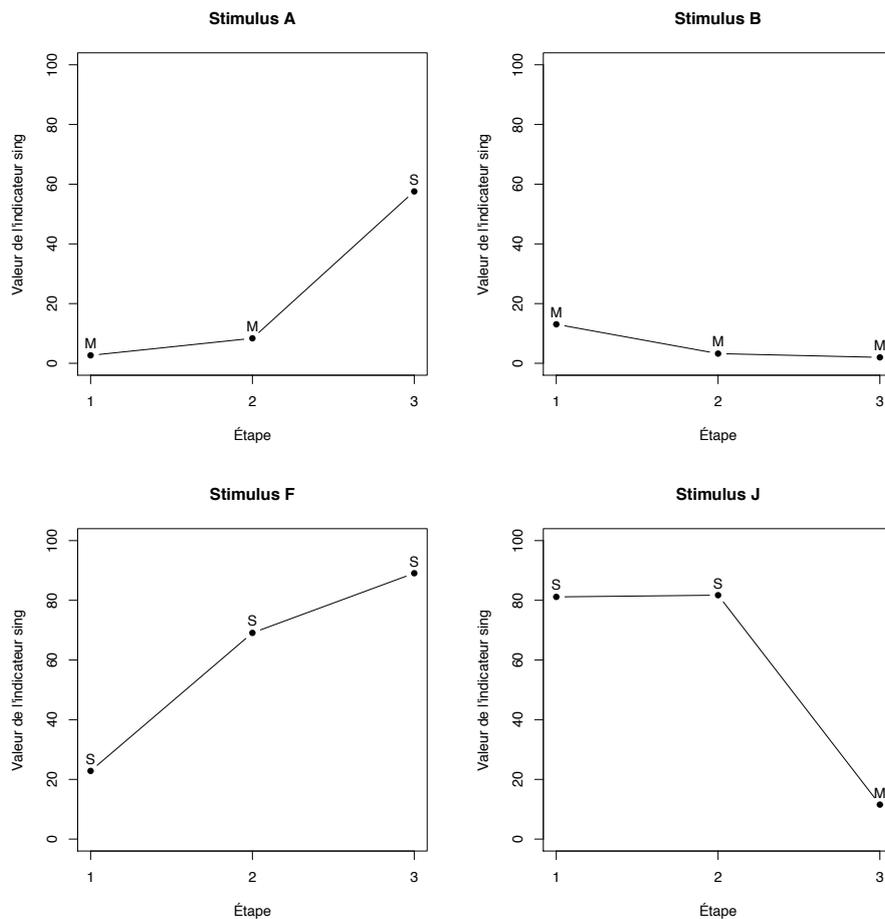


Figure 34: Expérience fictive pédagogique : Évolution de la caractérisation perceptuelle des stimuli A, B, F et J au cours de la tâche. 'M' indique que le stimulus a été perçu comme moyen à cette étape. 'S' indique que le stimulus a été perçu comme singulier à cette étape.

3 Analyse statistique des données orientée évaluateurs

L'objectif est ici de mettre en évidence les différences de perception des stimuli observées au sein du panel d'évaluateurs. Dans cette section, nous proposons une procédure de classification non supervisée pour aborder cette problématique. Nous verrons que l'algorithme de classification hiérarchique est appliqué sur une matrice de dissimilarité qui révèle les désaccords de perception entre évaluateurs. L'indice de dissimilarité inter-évaluateurs le plus naturel dans ce cadre repose sur une mesure de concordance entre les partitions finales sur les stimuli fournies par les évaluateurs. Cette procédure permet d'obtenir un dendrogramme qui représente la structure de désaccord observée au sein du panel d'évaluateurs. Une fois coupé, il fournit une partition des évaluateurs. Celle-ci est composée de clusters (c.-à-d. groupes homogènes) d'évaluateurs. Ces derniers sont assimilables à des types de perception de l'ensemble de stimuli. Nous verrons comment l'interprétation de ces clusters peut être enrichie. Dans ce cadre, des covariables véhiculant des informations sur les évaluateurs sont utilisées. Une partition 'consensus' des stimuli est également définie pour chaque cluster d'évaluateurs. Elle permet de mieux comprendre le type de perception des stimuli véhiculée par le cluster d'évaluateurs.

3.1 Définition de la distance inter-évaluateurs

Nous nous intéressons ici uniquement aux configurations des stimuli finales fournies par les évaluateurs. La matrice considérée contient donc en lignes les stimuli et en colonnes les évaluateurs. Dans ce cadre, chaque évaluateur est assimilable à une partition sur les stimuli. Une possibilité pour étudier l'accord inter-évaluateurs est alors d'étudier la concordance entre les deux partitions qu'ils ont fournies. Cette mesure de concordance peut être réalisée par l'intermédiaire de l'indice de Rand ajusté, noté *ARI* (Hubert & Arabie, 1985). Cet indice est une version corrigée de l'indice de Rand *RI* développé par Rand (1971). Ce dernier est permis d'établir la similarité entre deux partitions d'un même ensemble d'objets. La correction réalisée permet de rendre l'indice moins sensible au nombre de clusters et à la taille des clusters dans les deux partitions comparées (Courcoux, Faye, & Qannari, 2014).

La procédure usuelle consiste à créer une matrice de dissimilarité entre évaluateurs en considérant l'indice $1 - ARI$ comme indice de dissimilarité (Faye et al., 2004). Cette matrice correspond à la matrice de désaccord entre évaluateurs.

3.2 Création du dendrogramme et stratégie de coupure

Les algorithmes de classification ascendante hiérarchique sont privilégiés ci-après pour extraire la structure de désaccord de la matrice de désaccord. Ceux-ci fournissent un dendrogramme représentant une séquence de partitions imbriquées des évaluateurs, partant d'une structure de désaccord avec un cluster d'évaluateurs jusqu'à la situation extrême d'un désaccord complet avec autant de clusters que d'évaluateurs. Parmi les algorithmes de classification ascendante hiérarchique,

l'algorithme basé sur le critère d'agglomération de Ward (Ward, 1963) est largement utilisé. Ce critère d'agglomération permet de produire un dendrogramme dans lequel les partitions présentent une dispersion intra-cluster minimale (Murtagh & Legendre, 2014). Dans la suite, l'algorithme de classification basé sur ce critère d'agglomération est utilisé pour déduire la structure de désaccord au sein du panel d'évaluateurs. Pour un niveau de coupure donné, le dendrogramme obtenu fournit une segmentation du panel en clusters homogènes d'évaluateurs. Deux évaluateurs sont dans le même cluster s'ils ont perçu l'ensemble de stimuli de la même façon.

L'indice de McClain et Rao (McClain & Rao, 1975) est utilisé pour déterminer le niveau de coupure optimal du dendrogramme. Il est calculé pour chaque partition d'évaluateurs obtenue à chaque niveau de coupure. Il est défini grâce à la formule suivante :

$$mc = \frac{\overline{d_{intra}}/N_{d_{intra}}}{\overline{d_{inter}}/N_{d_{inter}}} \quad (2.6)$$

où $\overline{d_{intra}}$ (resp. $\overline{d_{inter}}$) représente la moyenne des distances intra-clusters (resp. des distances inter-clusters) au sein de la partition ; et $N_{d_{intra}}$ (resp. $N_{d_{inter}}$) représente le nombre de distances intra-clusters (resp. de distances inter-clusters). Lorsqu'un cluster contient un unique évaluateur, d_{intra} est fixée à 0 et n'est pas comptabilisée dans $N_{d_{intra}}$.

La valeur minimale de l'indice est utilisée pour indiquer le nombre optimal de clusters.

3.3 Interprétation de la partition d'évaluateurs

Par la suite, nous supposons que le panel $\mathcal{C} = [1; P]$ d'évaluateurs, où $P \geq 2$ est le nombre d'évaluateurs, a été partitionné en K clusters $\mathcal{C}_k, k=1, \dots, K$ avec $1 \leq K \leq P$, $\bigcup_k \mathcal{C}_k = \mathcal{C}$ et, pour tout $k \neq k'$, $\mathcal{C}_k \cap \mathcal{C}_{k'} = \emptyset$. Le nombre d'évaluateurs dans le cluster \mathcal{C}_k est dorénavant noté P_k , avec $P_k \geq 2$. Les P_k partitions finales sur les stimuli fournies par les évaluateurs du cluster \mathcal{C}_k sont notées U_1, U_2, \dots, U_{P_k} .

Plusieurs informations sont facilement disponibles sur les clusters d'évaluateurs issus de la procédure de classification non supervisée. Chaque cluster est décrit par la proportion d'évaluateurs du panel qu'il contient. Il peut également être décrit par son paragon. Ce dernier correspond à l'évaluateur le plus proche du centre du cluster et peut être considéré comme l'évaluateur le plus représentatif.

Dans certaines situations, des covariables supplémentaires dont on connaît les valeurs sur la population de P évaluateurs sont disponibles. Elles sont récoltées par l'expérimentateur en parallèle de l'implémentation du TSA. Ces covariables représentent généralement des caractéristiques signalétiques des évaluateurs (p. ex.

genre, âge, profession). Elles peuvent également correspondre à des caractéristiques d'habitude d'usage de la gamme de stimuli évalués lors de l'expérience. L'un des objectifs de l'expérimentateur est alors de déterminer si un cluster d'évaluateurs peut être relié à une typologie de personnes. Nous illustrons ici la flexibilité de la procédure de classification non supervisée en présentant comment les clusters d'évaluateurs peuvent être interprétés grâce à ces covariables. Nous démontrons aussi comment il est possible d'établir, pour chaque cluster, une partition sur les stimuli dite 'consensus'. Celle-ci permet d'identifier les spécificités de perception des stimuli au sein du cluster.

3.3.1 Interprétation des clusters grâce aux covariables sur les évaluateurs

La mise en relation entre les clusters et les covariables sur les évaluateurs est basée sur les données présentées en Figure 35. Elle consiste à comparer, pour chaque covariable, deux types de distributions : sa distribution au sein de chaque cluster d'évaluateurs d'une part, et sa distribution au sein du panel d'autre part. Cette comparaison est réalisée au moyen de tests statistiques. Ceux-ci dépendent de la nature de la covariable sur les évaluateurs (covariable continue ou covariable catégorielle).

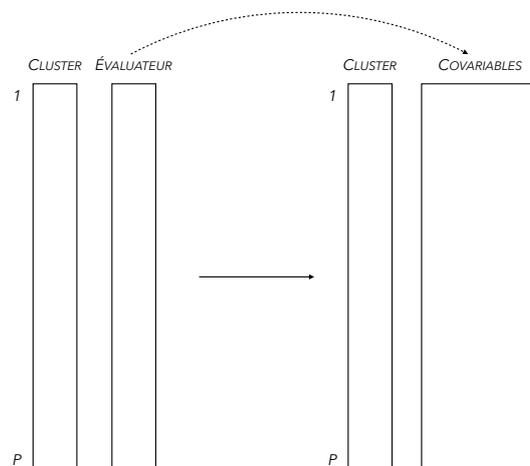


Figure 35: Jeu de données utilisé pour la description des clusters par des covariable sur les évaluateurs. Il présente les clusters des évaluateurs et les valeurs prises par les covariables disponibles sur les évaluateurs.

3.3.1.1 Covariable continue sur les évaluateurs

Comme proposé par Husson, Lê, and Pagès (2011), l'effet global de la variable catégorielle *cluster* sur la covariable continue d'intérêt est étudié au travers du test global d'un modèle d'analyse de variance à un facteur. Si l'effet global de

la variable *cluster* sur la covariable continue est significatif, cela signifie que les valeurs moyennes de la covariable continue diffèrent d'un cluster d'évaluateurs à l'autre.

Dans l'objectif d'analyser la nature de ces différences, une *v*-test est calculée comme suit pour chaque cluster \mathcal{C}_k :

$$v\text{-test} = \frac{\bar{x}_k - \bar{x}}{\sqrt{\frac{s^2}{P_k} \left(\frac{P - P_k}{P - 1} \right)}}$$

où \bar{x}_k représente la moyenne des valeurs prises par la covariable continue X dans le cluster \mathcal{C}_k , \bar{x} représente la moyenne des valeurs prises par la covariable à l'échelle de la globalité du panel, P_k représente le nombre d'évaluateurs dans le cluster \mathcal{C}_k , et P représente le nombre total d'évaluateurs au sein du panel.

Cette valeur *v*-test est utilisée pour tester l'hypothèse nulle suivante : *les valeurs de la covariable continue prises par les évaluateurs du cluster \mathcal{C}_k sont sélectionnées de manière aléatoire parmi toutes les valeurs possibles de la covariable continue*. Entre autres choses, une *p*-value peut être associée à ce test statistique. Si l'hypothèse nulle est rejetée, cela signifie que la covariable continue caractérise le cluster \mathcal{C}_k . Il est ensuite possible de déterminer si les valeurs moyennes de la covariable au sein du cluster \mathcal{C}_k sont élevées ou faibles, et ce relativement aux observations faites sur la globalité du panel d'évaluateurs.

3.3.1.2 Covariable catégorielle sur les évaluateurs

Comme proposé par Husson et al. (2011), l'effet global de la variable catégorielle *cluster* sur la covariable catégorielle d'intérêt est étudié au moyen d'un test d'indépendance du χ^2 . Si la liaison entre la variable *cluster* et la covariable catégorielle est significative, cela signifie que les valeurs prises par la covariable catégorielle diffèrent d'un cluster d'évaluateurs à l'autre.

Pour chaque cluster \mathcal{C}_k et pour chaque modalité de la variable catégorielle, l'objectif est alors de déterminer si le cluster \mathcal{C}_k est caractérisé par cette modalité donnée de la covariable catégorielle. Cet objectif peut être appréhendé au travers de la comparaison des deux proportions suivantes : P_{kq}/P_k , la proportion d'évaluateurs du cluster \mathcal{C}_k prenant la modalité q pour la covariable catégorielle ; et P_q/P , la proportion d'évaluateurs du panel prenant la modalité q pour la covariable catégorielle.

Sous l'hypothèse nulle de l'indépendance, ces deux proportions sont égales en moyenne. P_k évaluateurs étant choisis de manière aléatoire parmi les P évaluateurs du panel, la variable aléatoire P_{kq} suit, sous l'hypothèse nulle, la distribution hypergéométrique $\mathcal{H}(P, P_q, P_k)$. La probabilité d'avoir une valeur plus extrême que la valeur observée peut donc être calculée. Si l'hypothèse nulle est rejetée, cela signifie que la modalité q de la covariable catégorielle caractérise le cluster \mathcal{C}_k . Il

est ensuite possible de déterminer si la modalité q de la covariable catégorielle est sur-représentée au sein du cluster \mathcal{C}_k ou sous-représentée, et ce relativement aux observations faites sur la globalité du panel d'évaluateurs.

3.3.2 Interprétation des clusters grâce à leur caractérisation des stimuli

Pour chaque cluster \mathcal{C}_k , une partition consensus (aussi 'partition moyenne' ou 'partition centrale') sur les stimuli est définie. Celle-ci, notée par la suite U_k , résume l'ensemble des partitions individuelles $[U_1, U_2, \dots, U_{P_k}]$ du cluster \mathcal{C}_k . Les P_k partitions finales sur les stimuli fournies par les évaluateurs du cluster \mathcal{C}_k sont notées U_1, U_2, \dots, U_{P_k} .

Elle est définie comme la partition qui minimise la distance avec les P_k partitions individuelles. En d'autres termes, elle correspond à la partition qui maximise l'accord avec les P_k partitions individuelles. Comme proposé par Courcoux et al. (2014), la partition consensus du cluster \mathcal{C}_k est alors la partition U_k qui maximise le critère suivant :

$$C(U_k) = \sum_{p=1}^{P_k} ARI(U_k, U_p)$$

Ces auteurs ont proposé un algorithme permettant de trouver une partition des stimuli consensus avec un nombre fixé de classes L . Il consiste à partir d'une partition initiale des stimuli en L classes. À chaque étape de l'algorithme, l'impact du réassignement d'un stimulus à une classe de la partition U_k autre que celle à laquelle il appartient est évalué. Le stimulus est alors réassigné à la classe pour laquelle une amélioration de la valeur du critère $C(U_k)$ est observée. Cette opération est répétée pour chaque stimulus jusqu'à convergence de l'algorithme.

Le déroulement de l'algorithme et les résultats qu'il fournit dépendent de la partition initiale des stimuli. Pour déterminer cette partition initiale, Courcoux et al. (2014) proposent d'implémenter une procédure ascendante hiérarchique particulière. Celle-ci part d'une partition avec autant de classes que de stimuli (c.-à-d. S) jusqu'à une partition à L classes. À chaque étape de la procédure, l'impact de la fusion de chaque couple de classes est évalué. Plus précisément, la partition avec les deux classes fusionnées est provisoirement considérée comme une partition consensus et le critère $C(U_k)$ est calculé. Finalement, c'est la fusion qui mène à la meilleure amélioration de la valeur du critère $C(U_k)$ qui est retenue. Cette opération est répétée jusqu'à ce qu'une partition des stimuli en L classes soit obtenue. Celle-ci est utilisée comme partition initiale de l'algorithme décrit précédemment.

Le nombre de classes de stimuli dans la partition consensus U_k étant inconnu, l'algorithme est implémenté pour différentes valeurs de L . La partition consensus U_k choisie est celle qui présente le meilleur critère $C(U_k)$.

Par la suite, nous supposons que l'ensemble de stimuli, où $S \geq 2$ est le nombre de stimuli, a été partitionné en L classes dans la partition consensus U_k du cluster d'évaluateurs \mathcal{C}_k . \mathcal{C}_l $l = 1, \dots, L$ avec $1 \leq L \leq S$, $\bigcup_l \mathcal{C}_l = \mathcal{C}$ et, pour tout $l \neq l'$, $\mathcal{C}_l \cap \mathcal{C}_{l'} = \emptyset$. Le nombre de stimuli dans la classe \mathcal{C}_l est dorénavant noté S_l , avec $S_l \geq 2$.

Chaque classe de stimuli \mathcal{C}_l de la partition consensus U_k est caractérisée par les caractéristiques perceptuelles qui lui ont été associées par les évaluateurs du cluster \mathcal{C}_k . Pour chaque classe de stimuli \mathcal{C}_l et pour chaque caractéristique c , l'objectif est de déterminer si la classe de stimuli est caractérisée par cette caractéristique. Comme proposé par S. Lê and Worch (2015) dans le cadre de la caractérisation de produits par des données textuelles, cet objectif peut être appréhendé au travers de la comparaison des deux proportions suivantes :

$$\frac{O_{lc}}{O_l} \text{ et } \frac{O_{.c}}{O}$$

où O , $O_{.c}$, O_l et O_{lc} sont spécifiques au cluster d'évaluateurs \mathcal{C}_k . O représente le nombre d'occurrences des caractéristiques utilisées pour décrire tous les stimuli ; $O_{.c}$ représente le nombre d'occurrences de la caractéristique c pour décrire tous les stimuli ; O_l représente le nombre d'occurrences de caractéristiques utilisées pour décrire les S_l stimuli de la classe \mathcal{C}_l ; et O_{lc} représente le nombre d'occurrences de la caractéristique c utilisée pour décrire les S_l stimuli de la classe \mathcal{C}_l .

Sous l'hypothèse nulle de l'indépendance, ces deux proportions sont égales en moyenne. O_l caractéristiques étant choisies de manière aléatoire parmi l'intégralité des caractéristiques O , la variable aléatoire O_{lc} suit, sous l'hypothèse nulle, la distribution hypergéométrique $\mathcal{H}(O, O_{.c}, O_l)$. La probabilité d'avoir une valeur plus extrême que la valeur observée peut donc être calculée. Si l'hypothèse nulle est rejetée, cela signifie que la caractéristique c caractérise la classe de stimuli \mathcal{C}_l . Il est ensuite possible de déterminer si la caractéristique c est sur-utilisée pour décrire la classe \mathcal{C}_l ($O_{lc}/O_l > O_{.c}/O$) ou sous-utilisée ($O_{lc}/O_l < O_{.c}/O$), et ce relativement aux observations faites sur la globalité des stimuli.

4 Illustration sur des données réelles de caractérisation de produits industriels : données 'parfums de gels douche'

Dans l'objectif d'illustrer l'intérêt de la procédure d'analyse statistique introduite précédemment, nous présentons ici les résultats de son application sur des données réelles de TSA. Ces données sont issues de la caractérisation sensorielle d'un grand ensemble de 40 parfums de gels douche.

Lors de cette étude, 40 parfums de gels douche ont été sélectionnés pour constituer l'espace produit à caractériser. Ils correspondent aux parfums présentés dans le Chapitre 1 de cette thèse. Chacun des parfums a été labellisé grâce à un code aléatoire à trois chiffres. Un panel de 60 consommateurs a été recruté pour participer à la tâche de caractérisation sensorielle de ces produits. Les consommateurs n'ont été soumis à aucun entraînement sensoriel avant de participer à cette expérience. Le déroulement du TSA impliquait 3 sous-ensembles de parfums. Ceux-ci contenaient respectivement 12, 14 et 14 parfums. Les instructions données aux consommateurs étaient identiques à celles décrites dans le Chapitre 1.

4.1 Données

Les données recueillies ont été mises en forme dans un jeu de données de dimensions 40×180 présentant en lignes les parfums et en colonnes des blocs de 3 colonnes. Chacun de ces blocs correspondait aux 3 partitions fournies par un consommateur au cours de la tâche. L'intersection de la ligne s et de la colonne i du bloc p contenait l'ensemble de caractéristiques utilisé par le consommateur p pour décrire le groupe associé au parfum s présenté à l'étape i . À ces données de TSA s'ajoutaient 2 covariables véhiculant des informations sur les consommateurs. Il s'agissait de leur genre et de leur tranche d'âge.

4.2 Résultats de l'analyse statistique orientée parfums

La Figure 36 représente le nombre de groupes de parfums formés par les consommateurs lors des différentes étapes du TSA. Il apparaît que la répartition du nombre de groupes de parfums évolue entre la première et la deuxième étape. Ceci indique que les consommateurs ont créé de nouveaux groupes de parfums après la présentation du deuxième sous-ensemble. Au contraire, la répartition du nombre de groupes évolue peu entre la deuxième et la dernière étape. Ceci indique que les consommateurs ont créé peu de nouveaux groupes. Ils se sont globalement contentés de combiner les parfums du troisième sous-ensemble aux groupes formés précédemment. Cette observation peut être le témoin d'une saturation de la variabilité sensorielle de l'espace produit après la deuxième étape. En d'autres termes, 26 parfums (sur les 40 au total) étaient suffisants pour aborder les différentes dimensions de variabilité de cet espace produit.

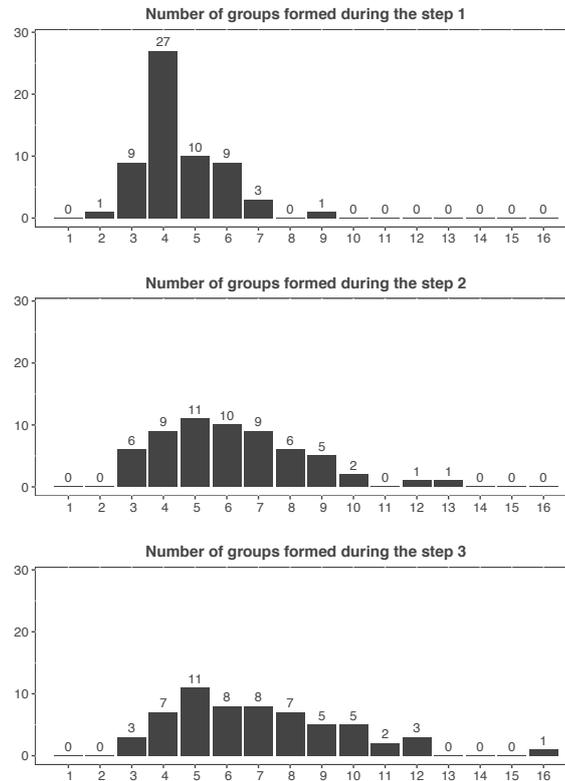


Figure 36: Étude ‘parfums de gels douche’ : Représentation du nombre de groupes de stimuli formés à chaque étape du Tri Séquentiel Agglomératif.

Les deux premières dimensions de l’ACM expriment 17.34% de l’inertie totale des données. Ce pourcentage est élevé compte-tenu du nombre important de lignes et de colonnes dans le jeu de données analysé. Par la suite, les interprétations de la représentation multidimensionnelle des parfums se limiteront à ces deux premières dimensions. Le plan factoriel correspondant est présenté en Figure 37. Une Classification Ascendante Hiérarchique des parfums a également été réalisée. Celle-ci a été appliquée sur les coordonnées des parfums sur les 5 premières dimensions de l’ACM. Un code couleur est utilisé sur la Figure 37 pour mettre en évidence les 3 clusters (c.-à-d. groupes homogènes) de parfums identifiés grâce à cette procédure. La première dimension de l’ACM oppose le cluster 1 au cluster 3. La deuxième dimension oppose, elle, les clusters 1 et 3 au cluster 2. La projection des caractéristiques sensorielles consensuelles sur le plan factoriel permet de comprendre la spécificité de ces clusters. Le cluster 1 est composé de parfums masculins, avec des notes musquées et marines. Le cluster 2 est composé de parfums légers, aux notes florales et fruitées. Le cluster 3, lui, contient des parfums aux notes vanillées, de bonbons et de nourriture. Le principal axe de variabilité sensorielle de cet espace produit réside principalement sur une opposition entre les parfums masculins (cluster 1) et les parfums gourmands (cluster 3).

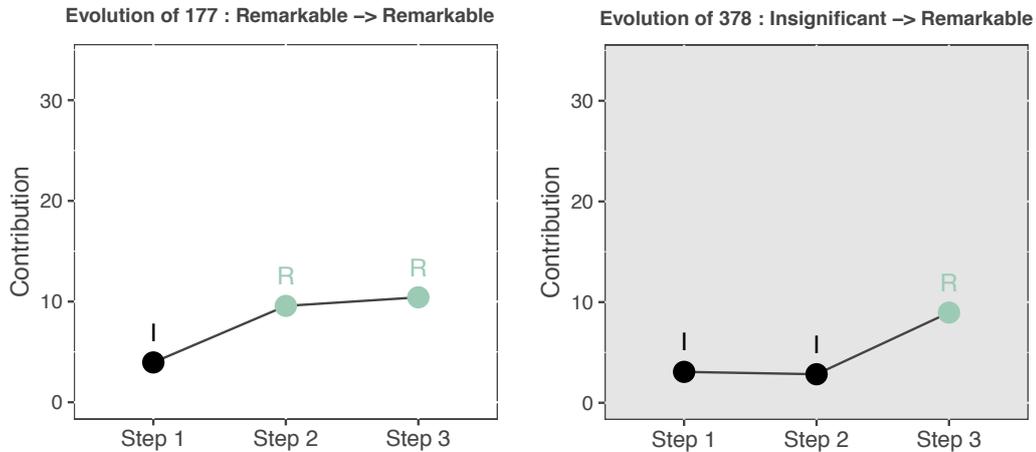


Figure 38: Étude ‘parfums de gels douche’ : Évolution de la caractérisation sensorielle des parfums 177 et 378 au cours de la tâche. ‘I’ (‘insignifiant’ en anglais) indique que le parfum a été perçu comme moyen à cette étape. ‘R’ (‘remarkable’ en anglais) indique que le parfum a été perçu comme singulier à cette étape.

4.3 Résultats de l’analyse statistique orientée consommateurs

La Figure 39 présente les résultats de la classification non supervisée des consommateurs. Ils mettent en évidence l’existence de deux clusters de consommateurs au sein du panel, le niveau du coupure optimal du dendrogramme correspondant à celui fournissant deux clusters. En d’autres termes, il est possible de conclure qu’il existe deux types de perception sensorielle des parfums. Chacun de ces types peut être assimilé à un cluster de consommateurs du panel.

Les Figures 40 et 41 présentent les partitions consensuelles sur les parfums associées aux deux clusters d’évaluateurs. La comparaison de celles-ci permet d’identifier sur quelles différences de perception des parfums reposent les désaccords observés entre les deux clusters d’évaluateurs. Tout d’abord, certaines similitudes existent entre les deux partitions consensuelles. Les deux clusters d’évaluateurs ont identifié un groupe de parfums perçus comme fruités avec des notes d’agrumes ; un groupe de parfums perçus comme sucrés, gourmands et vanillés ; un groupe de parfums perçus comme masculins et puissants ; et un groupe de parfums perçus comme doux et neutres. Ces différents groupes sont globalement composés des mêmes parfums dans les deux partitions consensuelles. Au-delà de ces similitudes, certaines différences peuvent pourtant être notées. Le cluster d’évaluateurs 1 a créé un groupe de parfums persistants, avec une odeur associée au bébé et au savon de Marseille. Le cluster 2 n’a pas formé un groupe de stimuli similaire. Il a, lui, formé un groupe de parfums agressifs et acides ; un groupe de parfums fleuris et poivrés ; et un groupe de parfums frais et vivifiants, avec des notes de menthe.

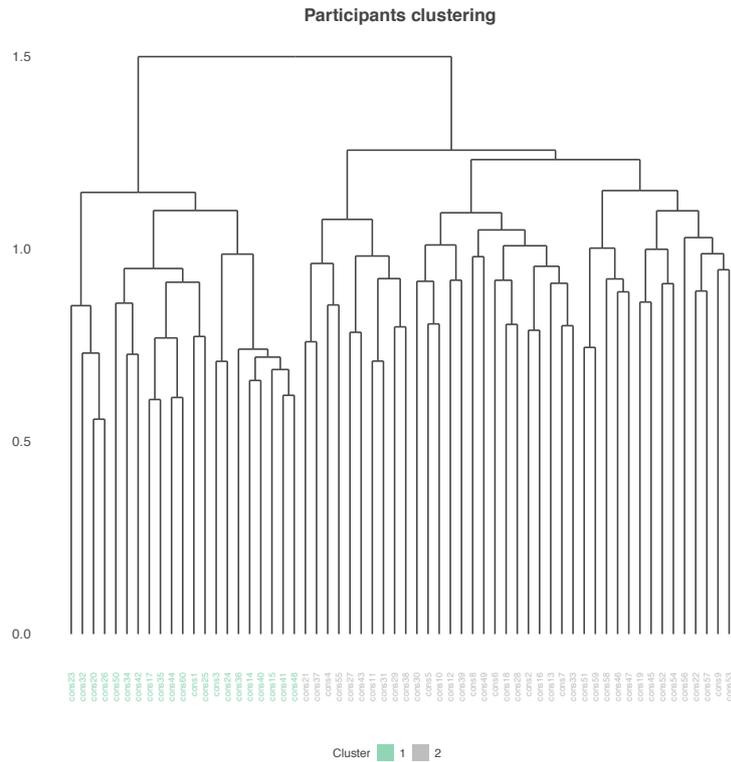


Figure 39: Étude ‘parfums de gels douche’ : Résultats de la classification non supervisée du panel de consommateurs.

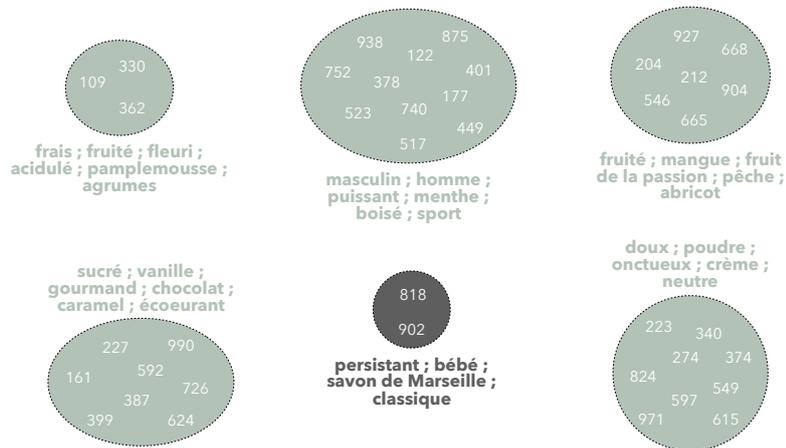


Figure 40: Étude ‘parfums de gels douche’ : Partition des parfums consensuelle associée au cluster 1 de consommateurs. Chaque groupe est décrit par les caractéristiques sensorielles sur-utilisées pour le décrire (relativement à l’ensemble des parfums). Les principales différences avec la partition consensuelle du cluster 2 sont indiquées en gris.

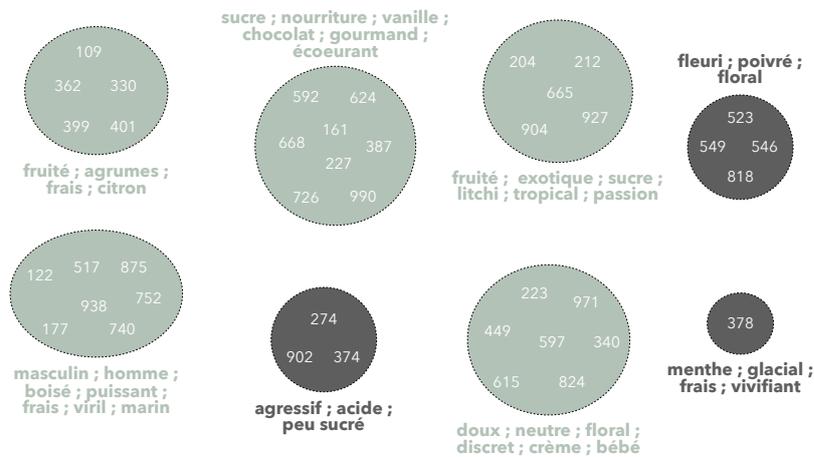


Figure 41: Étude 'parfums de gels douche' : Partition des parfums consensuelle associée au cluster 2 de consommateurs. Chaque groupe est décrit par les caractéristiques sensorielles sur-utilisées pour le décrire (relativement à l'ensemble des parfums). Les principales différences avec la partition consensuelle du cluster 1 sont indiquées en gris.

La caractérisation automatique des clusters de consommateurs à travers leur genre ou leur tranche d'âge n'a pas permis de mettre en évidence un lien significatif. Dans ce cas précis, il est impossible de relier les types de perception sensorielle des parfums à une typologie de personnes en particulier.

5 Conclusion

L'étude de la représentation perceptuelle d'un stimulus est l'objectif le plus courant des analystes sensoriels. Celle-ci peut être mise en place à travers de nombreuses méthodes de caractérisation. Certaines de ces méthodes sont spécifiquement adaptées à une utilisation par des évaluateurs non-experts. En permettant d'analyser la perception des consommateurs (évaluateurs non-experts), elles présentent un intérêt majeur dans un contexte de R&D. Parmi ces approches, la méthode du tri libre est probablement l'une des plus utilisées en analyse sensorielle. Sa facilité d'utilisation peut expliquer cette popularité. Elle repose effectivement sur une approche utilisée naturellement par le cerveau humain pour aborder un ensemble de stimuli : la catégorisation. En outre, elle permet de recueillir la perception brute des consommateurs. Elle est en effet basée sur une évaluation holistique des stimuli (c.-à-d. une évaluation globale et non une évaluation spécifique par attribut sensoriel). Elle permet ainsi de faire émerger des leviers opérationnels directement actionnables dans un processus de développement de produit.

Malgré l'efficacité et l'intérêt de la méthode du tri libre, son application à des grands ensembles de stimuli est jusqu'alors restée une problématique en suspens.

Dans le cadre de cette thèse, une nouvelle méthodologie de recueil de données a été développée pour répondre à cette interrogation : le ‘Tri Séquentiel Agglomératif’ (TSA). Cette approche repose sur une division des stimuli en plusieurs sous-ensembles de taille restreinte. Ces derniers sont présentés de manière séquentielle à l’évaluateur et agglomérés les uns aux autres.

Nous avons vu que cette procédure permet de recueillir des données présentant un format particulier. Elles sont en effet constituées d’une succession de partitions labellisées sur les stimuli, chacune d’entre elles correspondant à la partition sur les stimuli fournie par un évaluateur à une étape du TSA. Nous avons vu que cet ensemble de partitions présentent deux particularités majeures. La première repose sur le fait que certaines partitions contiennent des données manquantes, les stimuli n’ayant pas été évalués par tout le panel dans tous les sous-ensembles. Les partitions contenant des données manquantes constituent les ‘partitions intermédiaires’ fournies par les évaluateurs. Celles n’en contenant pas constituent les ‘partitions finales’. La deuxième particularité des données repose sur le fait que les partitions d’un évaluateur sont liées de manière chronologique. Ces deux particularités font des données de TSA des données inédites en analyse sensorielle. Par conséquent, une procédure d’analyse statistique *ad hoc* devait être développée pour les appréhender. C’est sur cette problématique que nous nous sommes penchés dans le cadre de ce chapitre.

Dans un premier temps, nous avons abordé la manière avec laquelle il est possible d’analyser les données de TSA selon une approche orientée stimuli. Nous avons vu que, sur la seule base des partitions finales, il est possible d’utiliser une procédure statistique fournissant une représentation superposée des stimuli et des caractéristiques perceptuelles consensuelles. Nous avons ensuite proposé une procédure statistique complémentaire permettant d’analyser la cinétique de caractérisation perceptuelle. Elle est déduite de l’analyse des partitions intermédiaires fournies par les évaluateurs. Nous avons démontré que cette démarche complémentaire permet de nuancer certaines conclusions tirées de la représentation euclidienne des stimuli fournie par l’analyse des partitions finales. Cette démonstration a été appuyée par l’utilisation d’un jeu de données fictif pédagogique.

Dans un deuxième temps, nous avons abordé la manière avec laquelle il est possible d’analyser les données de TSA selon une approche orientée évaluateurs. Dans ce cadre, nous avons proposé d’analyser seulement les partitions finales des évaluateurs. Pour cela, nous avons analysé les partitions finales grâce à une démarche statistique classiquement utilisée pour analyser les données de tri libre. Cette procédure permet d’obtenir une classification non supervisée des évaluateurs selon la structure de désaccord de perception observée au sein du panel. Les bases du désaccord mis en évidence sont ensuite illustrées au travers de la détermination d’une partition consensuelles sur les stimuli par cluster d’évaluateurs. Nous avons également vu comment les clusters d’évaluateurs pouvaient être automatiquement

interprétés grâce à l'intégration de covariables. Malgré l'intérêt de la procédure de classification non supervisée présentée ici, il est important de préciser qu'elle pourrait être améliorée. Certains auteurs ont ainsi proposé des solutions pour améliorer la segmentation du panel obtenue. Par exemple, Courcoux et al. (2014) ont proposé de consolider la segmentation grâce à un processus complémentaire. Celui-ci consiste à utiliser la partition sur les évaluateurs issues de l'approche hiérarchique comme partition initiale d'un algorithme de partitionnement. À chaque étape de cet algorithme, l'allocation de chaque évaluateur aux différents clusters est considérée en calculant l'accord entre sa propre partition des stimuli et la partition consensus associée à chaque cluster. En outre, la procédure présentée ici est uniquement basée sur une démarche de comparaison de partitions. Via cette approche, deux évaluateurs appartiennent au même cluster s'ils ont globalement formé les mêmes groupes de stimuli. Il est pourtant possible qu'ils aient créé ces groupes pour des raisons perceptuelles différentes. Une façon de prendre en compte cette information serait de considérer également la labellisation des partitions dans la procédure de classification. Enfin, la procédure orientée évaluateurs présentée dans ce chapitre est uniquement basée sur les partitions finales qu'ils ont fournies. Des travaux complémentaires doivent être entrepris dans l'objectif de déterminer comment les partitions intermédiaires peuvent être analysées pour enrichir cette démarche.

L'intérêt de cette procédure d'analyse statistique des données de TSA a par ailleurs été illustré à travers son application à des données réelles.



Chapitre 3

APPLICATION DU TRI SÉQUENTIEL AGGLOMÉRATIF À LA CARACTÉRISATION DE GESTES CULINAIRES

RÉSUMÉ : Les bonnes performances du Tri Séquentiel Agglomératif ont été démontrées lors d'une étude portant sur la caractérisation sensorielle de 40 parfums de gels douche. La question posée ici est celle de savoir si cette méthodologie peut également être utilisée pour caractériser des gestes culinaires. Ces derniers sont en effet traditionnellement appréhendés de manière qualitative. Leur appréhension de manière quantitative nécessite donc une réflexion méthodologique spécifique. Plus précisément, il s'agit notamment ici de déterminer comment il est possible de recueillir les gestes culinaires, ainsi que d'identifier de quelle manière il est possible de les intégrer à une procédure de Tri Séquentiel Agglomératif. La piste du digital apparaît comme une solution privilégiée à ces différentes interrogations. En outre, se pose la question de savoir si des évaluateurs non-experts sont en mesure de fournir une évaluation sur ces stimuli inhabituels. Dans ce cadre, une étude réalisée au Centre Culinaire Contemporain est présentée. Elle visait à étudier les principaux axes de variabilité perçus par des évaluateurs au sein d'un ensemble de 27 recettes de tarte au citron meringuée. Cette étude démontre l'adaptabilité du Tri Séquentiel Agglomératif pour la caractérisation perceptuelle de grands ensembles de gestes culinaires.

Sommaire

1	Données ‘recettes de tarte au citron meringuée’	97
1.1	Procédure de recueil des données	97
1.1.1	Enregistrement vidéo des recettes de tarte au citron meringuée	97
1.1.1.1	Définition des paramètres à faire varier au sein des recettes	97
1.1.1.2	Captation des vidéos de recettes	98
1.1.1.3	Montage des vidéos de recettes	99
1.1.2	Réalisation de la tâche de caractérisation perceptuelle par le panel d’évaluateurs	99
1.2	Format des données	101
2	Résultats de l’analyse statistique	102
2.1	Résultats orientés recettes	102
2.2	Résultats orientés évaluateurs	107
3	Conclusion	110

L’utilisation du Tri Séquentiel Agglomératif (TSA) n’a de sens que si l’ensemble de stimuli étudié est adapté : il doit être source de variabilité de perception. Ceci se traduit en pratique par (1) la nécessité d’avoir des stimuli présentant des caractéristiques variées et (2) la nécessité de recruter des évaluateurs percevant les stimuli de manière différente. Dans le cadre de l’étude présentée ici, nous avons pris le soin de vérifier ces différents aspects. Nous nous sommes en effet focalisés sur des gestes culinaires dont la réalisation technique est variable d’une personne à l’autre : la création d’une tarte au citron meringuée. Cette variabilité technique permet de s’assurer d’une potentielle diversité de la perception des stimuli.

Dans un premier temps, nous présentons les données ‘recettes de tarte au citron meringuée’. Nous détaillons la manière avec laquelle les gestes culinaires ont été enregistrés sous format vidéo puis mis en forme. Nous voyons en particulier que les différents gestes ont été créés grâce à un plan d’expérience permettant de faire varier différents aspects de la recette. Ils ont ensuite été réalisés par une même personne. Nous présentons ensuite comment s’est déroulée la tâche de caractérisation perceptuelle au sein de cet ensemble de gestes. Pour finir, nous présentons succinctement le format des données recueillies.

Dans un second temps, nous présentons les résultats de la procédure statistique *ad hoc* proposée. Nous voyons comment l’approche orientée stimuli permet d’étudier la variabilité perceptuelle de l’ensemble des recettes. Dans ce cadre, nous voyons que la représentation perceptuelle de la recette de tarte au citron meringuée dépend surtout de la manière avec laquelle le produit fini est dressé. Nous voyons ensuite

comment l'approche orientée évaluateurs permet d'étudier la variabilité de perception des recettes au sein de la population. Nous voyons qu'il existe plusieurs types de perception de ce concept. Nous voyons que ceux-ci ne peuvent pas être reliés de manière significative à des caractéristiques des évaluateurs.

1 Données 'recettes de tarte au citron meringuée'

1.1 Procédure de recueil des données

1.1.1 Enregistrement vidéo des recettes de tarte au citron meringuée

1.1.1.1 Définition des paramètres à faire varier au sein des recettes

Les vidéos de recettes devaient être les plus variables possible. L'objectif était en effet d'obtenir un ensemble de recettes le plus représentatif possible de la diversité existante au sein de la population. Pour obtenir un tel ensemble de recettes, deux options ont été envisagées. La première était basée sur une approche dite 'écologique'. Elle consistait à recruter un large panel d'utilisateurs présentant des niveaux de familiarité avec la recette différents. Ces utilisateurs étaient ensuite filmés pendant leur réalisation habituelle d'une tarte au citron meringuée. Cette option garantissait d'obtenir un ensemble de recettes variable et représentatif de la diversité existante. La deuxième solution reposait, elle, sur une approche 'laboratoire'. Elle consistait à identifier les principaux paramètres de variabilité de réalisation d'une recette de tarte au citron meringuée. Ces différents paramètres étaient ensuite utilisés pour établir plusieurs recettes différentes. Les recettes ainsi créées étaient ensuite réalisées par une même personne. Cette deuxième option était plus risquée, car sa réussite reposait sur la bonne définition des paramètres de variabilité des recettes. Ceux-ci devaient en effet mener à la création de recettes variables mais également représentatives de la diversité existante au sein de la population. Cette option était cependant plus facile et plus rapide à mettre en place que la première. C'est donc celle que nous avons privilégiée dans le cadre de cette étude.

Les différents facteurs de variabilité d'une recette de tarte au citron meringuée ont été identifiés par un chef. Ce dernier avait réalisé plusieurs projets sur cette recette et il était donc informé des différentes possibilités de réalisation d'une tarte au citron meringuée. Les facteurs de variabilité de la recette établis correspondaient aux trois éléments culinaires impliqués : l'ingrédient utilisé dans la crème au citron, le type de pâte utilisé et le type de dressage associé à la meringue. Pour chacun de ces trois facteurs, trois modalités ont été définies. L'ingrédient utilisé dans la crème au citron pouvait ainsi être du citron jaune, du citron combava (c.-à-d. un citron vert exotique) ou du citron confit. La pâte pouvait être une pâte brisée, une pâte constituée de biscuits émiettés ou une pâte à texture modifiée composée de texturants. Enfin, le dressage associé à la meringue était un dressage type 'tartelette', un dressage type 'dôme' ou un dressage type 'macaron'. Ces différents paramètres sont répertoriés sur la Figure 42.

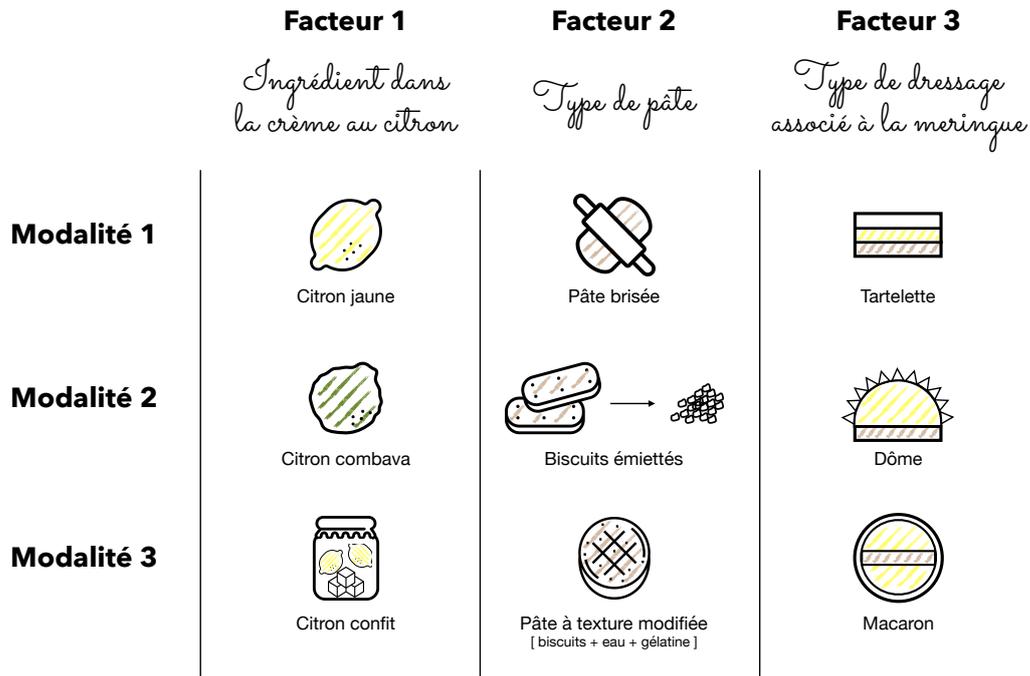


Figure 42: Étude ‘recettes de tarte au citron meringuée’ : Paramètres de variabilité des recettes.

1.1.1.2 Captation des vidéos de recettes

Les différents paramètres de variabilité ont été utilisés pour définir un plan d’expérience complet 3^3 . Vingt-sept recettes de tarte au citron meringuée ont ainsi été créées. Elles ont ensuite été réalisées par le chef et enregistrées sous format vidéo.

Cet enregistrement des vidéos a été réalisé de manière standardisée. Cela permettait de s’assurer que la perception des recettes ne soit pas impactée par le contexte de tournage. Pour cela, tous les accessoires présents sur la vidéo étaient harmonisés. En guise d’exemple, la Figure 43 présente des extraits des deux séquences correspondant à la réalisation de la crème au citron jaune et de la crème au citron combava. Enfin, le visage du chef n’était jamais visible.



Figure 43: Étude ‘recettes de tarte au citron meringuée’ : Tournage d’une séquence correspondant à la réalisation de la crème au citron jaune (à gauche) et de la crème au citron combava (à droite).

Une fois le chef installé, il lui était demandé de réaliser la recette qui lui était indiquée. Le lieu de tournage était équipé d’une caméra. Pour chaque séquence du tournage, elle était placée de manière à capter l’ensemble des gestes réalisés par le chef.

1.1.1.3 Montage des vidéos de recettes

Afin d’éviter que la tâche future de TSA ne dure trop longtemps, chacune des 27 vidéos obtenues a été accélérée. Au final, chaque vidéo durait environ une minute (durée moyenne = 59.4 secondes). Cette durée était homogène d’une vidéo à l’autre (amplitude = 56-61 secondes ; écart-type = 1 seconde). Cette homogénéité permettait de ne pas influencer les évaluateurs pour la tâche future de caractérisation. En effet, de grosses différences de durée auraient pu être perçues comme des axes de variabilité des recettes.

Enfin, le son a été retiré de chacune des vidéos. Cela permettait de s’assurer que la perception des recettes ne soit pas impactée par le discours du chef. Chacune des 27 vidéos obtenues a été associée à un code aléatoire à trois chiffres.

1.1.2 Réalisation de la tâche de caractérisation perceptuelle par le panel d’évaluateurs

Soixante-six évaluateurs ont été recrutés pour caractériser les recettes de tarte au citron meringuée. Ce panel d’évaluateurs était majoritairement composé de femmes (68.2 %), de jeunes personnes (78.8 % de personnes âgées de 18 à 25 ans), et de personnes avec un niveau auto-évalué en cuisine intermédiaire (13.6 % avec un niveau extrêmement faible, 15.1 % avec un niveau faible, 51.1 % avec un niveau intermédiaire, et 19.6% avec un bon niveau). Les évaluateurs n’ont été soumis à aucun entraînement particulier avant de participer à cette étude.

Le TSA impliquait trois sous-ensembles contenant chacun 9 recettes. Un plan aléatoire a été utilisé pour déterminer la composition de ces sous-ensembles pour

chaque évaluateur. La session de caractérisation perceptuelle s’est déroulée comme suit. Chaque évaluateur était placé dans un box individuel qui contenait une tablette numérique et un écran d’ordinateur. La tablette numérique présentait l’espace dédié à la réalisation de la tâche de caractérisation. Ce dernier contenait une zone dédiée à l’affichage des codes des vidéos pour chaque sous-ensemble de recettes. Lors de la présentation de chaque sous-ensemble, seuls les codes correspondant aux recettes de ce sous-ensemble étaient affichés. Cet espace de caractérisation contenait également des zones de texte permettant de renseigner les caractéristiques perceptuelles de chaque groupe de recettes. Il était enfin constitué d’un espace dédié à la constitution des groupes. L’écran d’ordinateur était, lui, connecté à la plateforme Youtube et présentait les différentes vidéos de recettes. La Figure 44 présente une visualisation schématique d’un box d’un évaluateur.

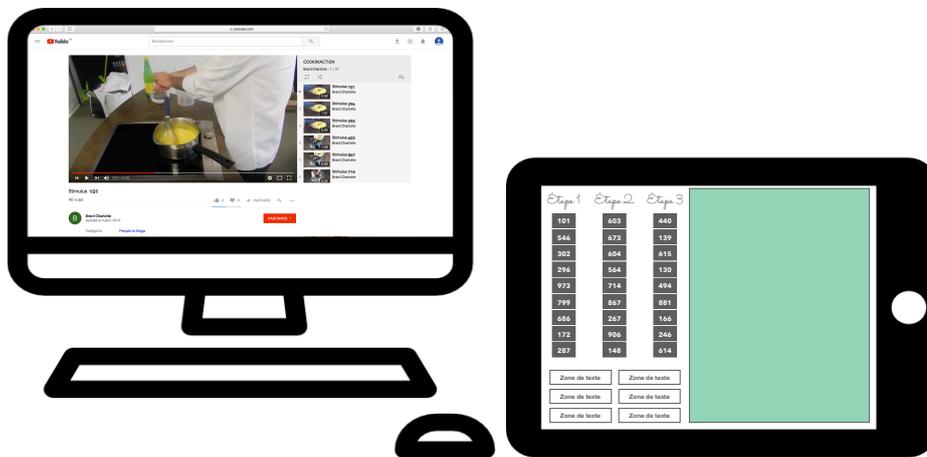


Figure 44: Étude ‘recettes de tarte au citron meringuée’ : Ordinateur et tablette tactile mis à la disposition de chaque évaluateur pour réaliser la tâche de Tri Séquentiel Agglomératif.

Lors de la première phase du TSA (c.-à.d. l’évaluation du premier sous-ensemble de recettes), les instructions suivantes étaient fournies à l’évaluateur : ‘Dans un premier temps, veuillez regarder au moins une fois chaque vidéo. Dans un deuxième temps, veuillez constituer des groupes de recettes selon leur similarité. Vous êtes libre de former les groupes selon n’importe quels critères, et vous n’êtes pas obligé(e) de spécifier ces critères. Vous pouvez former autant de groupes que vous le souhaitez. Vous pouvez mettre autant de recettes que vous le désirez dans chaque groupe. Dans un troisième et dernier temps, veuillez associer chaque groupe de recettes que vous avez formé à une zone de texte. Sur chaque zone de texte, veuillez indiquer une ou plusieurs caractéristique(s) qui décrivent, selon vous, le groupe en question. Essayez au maximum d’utiliser des caractéristiques relatives au caractère innovant des recettes.’ Le fait de demander à l’évaluateur d’utiliser majoritairement ce type de caractéristiques permettait

d'éviter que la labellisation de la partition des recettes soit directement liée au plan d'expérience utilisé pour établir ces recettes. À la fin de l'évaluation du premier sous-ensemble, l'expérimentateur effectuait une capture de l'écran de la tablette sous format image. Cela permettait d'enregistrer la partition labellisée des recettes.

Lors de la deuxième phase du TSA (c.-à.d. l'évaluation des deuxième et troisième sous-ensembles), les instructions données à l'évaluateur étaient les suivantes : 'Vous n'êtes maintenant plus obligé(e) de visualiser toutes les recettes avant de commencer. Veuillez évaluer une à une les nouvelles recettes. À chaque fois que vous visualisez une recette, veuillez la comparer aux groupes que vous avez déjà formés. Pour réaliser cette comparaison, référez-vous aux zones de texte associés aux groupes. Vous avez alors deux possibilités :

1. Vous considérez que la recette présente les mêmes caractéristiques qu'un groupe déjà formé. Dans ce cas, ajoutez-la à ce groupe.
2. Vous considérez que la recette ne correspond à aucun des groupes déjà formés. Dans ce cas, créez un nouveau groupe avec cette recette et décrivez-la avec une ou plusieurs caractéristique(s) en utilisant une zone de texte.

Une fois toutes les évaluations réalisées, vous êtes libre d'effectuer des réaffectations à la marge. Vous pouvez changer une recette de groupe (recettes des sous-ensembles précédents incluses), combiner ou diviser des groupes, changer les caractéristiques inscrites sur les zones de texte, etc.' À la fin de l'évaluation de chaque sous-ensemble, l'expérimentateur effectuait une capture de l'écran de la tablette sous format image.

À la suite de cette tâche, chaque évaluateur était amené à remplir un questionnaire pour indiquer des informations le concernant (genre, âge, niveau d'expertise culinaire auto-évalué, etc.).

1.2 Format des données

Les données recueillies ont été mises en forme dans un jeu de données de dimensions 27×198 présentant en lignes les recettes de tarte au citron meringuée et en colonnes des blocs de 3 colonnes. Chacun de ces blocs correspondait aux 3 partitions fournies par un évaluateur au cours de la tâche. L'intersection de la ligne s et de la colonne i du bloc p contenait l'ensemble de caractéristiques utilisé par l'évaluateur p pour décrire le groupe associé à la recette s présentée à l'étape i . À ces données de TSA s'ajoutaient 3 covariables véhiculant des informations sur les évaluateurs. Il s'agissait de leur genre, de leur tranche d'âge, et de leur expertise culinaire auto-évaluée.

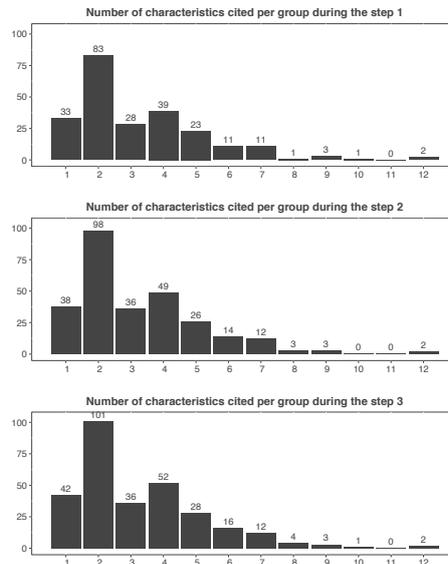


Figure 46: Étude ‘recettes de tarte au citron meringuée’ : Représentation du nombre de caractéristiques perceptuelles par groupe lors de chaque étape du Tri Séquentiel Agglomératif.

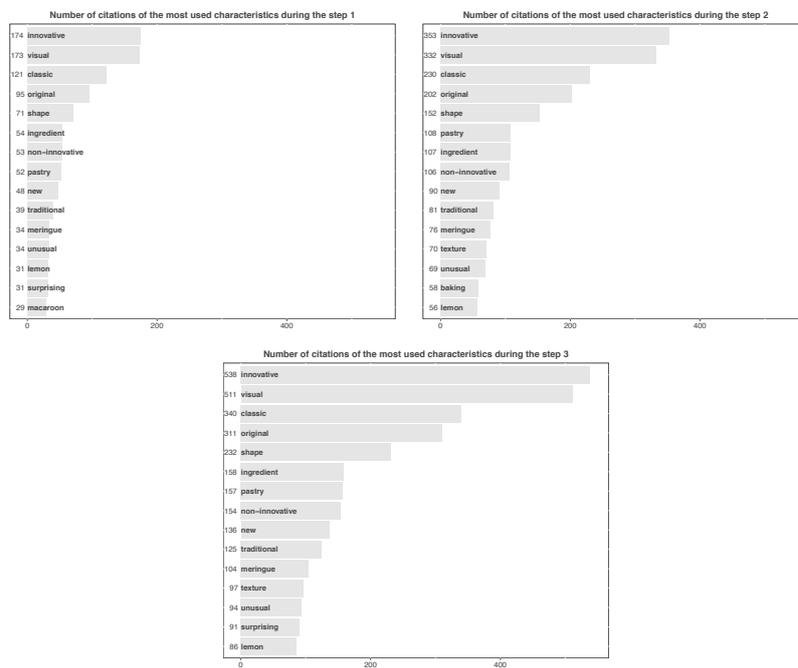


Figure 47: Étude ‘recettes de tarte au citron meringuée’ : Caractéristiques perceptuelles les plus utilisées lors de chaque étape du Tri Séquentiel Agglomératif.

La décroissance des valeurs propres de l'Analyse des Correspondances Multiples (ACM) utilisée dans le cadre de la procédure statistique orientée stimuli est présentée en Figure 48. Elle indique une forte diminution des valeurs propres après la troisième dimension de l'ACM. Seuls les deux premiers plans factoriels, composés des dimensions 1-2 et 1-3, seront donc interprétés par la suite pour comprendre la variabilité perceptuelle des recettes.

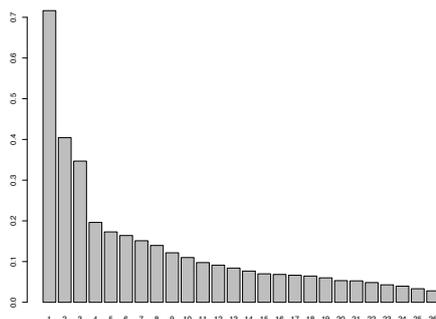


Figure 48: Étude ‘recettes de tarte au citron meringuée’ : Décroissance des valeurs propres de l'Analyse des Correspondances Multiples.

Les deux plans factoriels correspondants sont présentés en Figure 49. Une Classification Ascendante Hiérarchique des recettes a également été réalisée. Celle-ci a été appliquée sur les coordonnées des recettes sur les 5 premières dimensions de l'ACM. Un code couleur est utilisé sur la Figure 49 pour mettre en évidence les 3 clusters (c.-à-d. groupes homogènes) de recettes identifiés grâce à cette procédure.

La projection des caractéristiques perceptuelles consensuelles sur le plan factoriel permet de comprendre la spécificité de ces clusters. Cette interprétation est également facilitée par les résultats des tests statistiques visant à identifier les caractéristiques perceptuelles caractéristiques de chaque cluster de recettes et par les résultats d'une description automatique des dimensions de l'ACM (résultats non présentés ici).

La première dimension de l'ACM oppose principalement le cluster 1 aux clusters 2 et 3. La deuxième dimension oppose, elle, les clusters 1 et 3 au cluster 2. Le cluster 1 est composé de recettes impliquant un dressage de la meringue de type macaron. Ces recettes ont été jugées innovantes, déstructurées, créatives et originales. Le cluster 2 est composé de recettes impliquant un dressage de la meringue de type dôme, celles-ci étant jugées hétérogènes, festives, mignonnes, contemporaines et enfantines. Le cluster 3 est, lui, composé de recettes impliquant un dressage de la meringue de type tartelette. Ces recettes ont été jugées classiques, non-innovantes, homogènes et traditionnelles. Le principal axe de variabilité perceptuelle des recettes est donc basé sur la perception du type de dressage de la meringue (voir

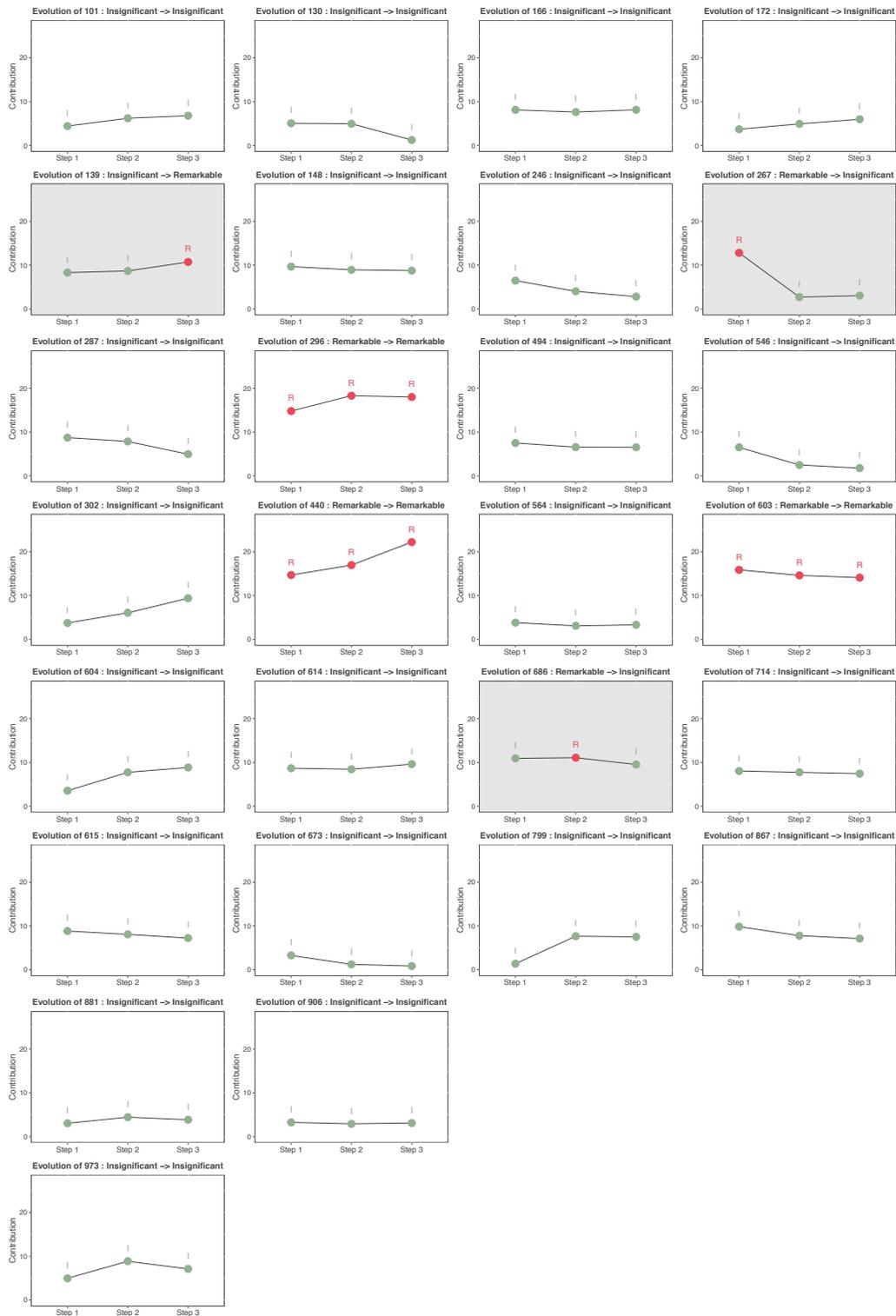


Figure 50: Étude ‘recettes de tarte au citron meringuée’ : Évolution de la caractérisation perceptuelle des recettes au cours de la tâche de Tri Séquentiel Agglomératif.

2.2 Résultats orientés évaluateurs

La Figure 51 présente les résultats de la classification non supervisée des évaluateurs. Ils mettent en évidence l'existence de deux clusters d'évaluateurs au sein du panel, le niveau du coupure optimal du dendrogramme correspondant à celui fournissant deux clusters. En d'autres termes, il est possible de conclure qu'il existe deux types de perception des recettes de tarte au citron meringuée. Chacun de ces types peut être assimilé à un cluster d'évaluateurs du panel.

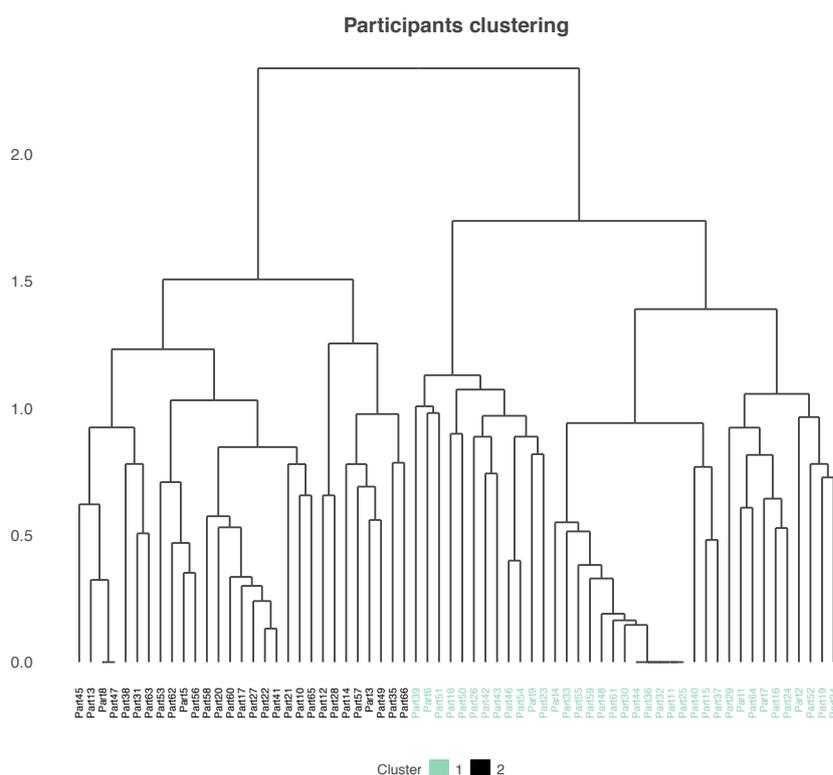


Figure 51: Étude 'recettes de tarte au citron meringuée' : Résultats de la classification non supervisée du panel d'évaluateurs.

Les Figures 52 et 53 présentent les partitions consensuelles sur les recettes associées aux deux clusters d'évaluateurs. La comparaison de celles-ci permet d'identifier sur quelles différences de perception des recettes reposent les désaccords observés entre les deux clusters d'évaluateurs. Tout d'abord, certaines similitudes existent entre les deux partitions consensuelles. Les deux clusters d'évaluateurs ont identifié un groupe de recettes perçues comme créatives, innovantes, design et élaborées. Ce groupe est composé des mêmes recettes dans les deux partitions consensuelles. Les recettes en question sont celles impliquant un dressage de la meringue de type macaron. Au-delà de cette similitude, certaines différences peuvent pourtant être notées. Le cluster d'évaluateurs 1 a également créé un groupe composé des

recettes impliquant un dressage de type dôme et un groupe de recettes impliquant un dressage de type tartelette. Ce cluster d'évaluateurs a donc perçu uniquement le facteur de variabilité 'type du dressage de la meringue' au sein de l'ensemble des recettes. Le cluster 2 n'a pas formé de tels groupes de recettes. Il a, lui, formé un groupe composé des recettes impliquant une pâte brisée. Ces recettes ont été décrites comme traditionnelles et basiques. Il a ensuite créé un groupe composé des recettes impliquant les autres types de pâtes (c.-à-d. la pâte composée d'émiettés de biscuits et la pâte à texture modifiée). Il a jugé ces recettes nouvelles et rapides à mettre en place. Contrairement au cluster 1, le cluster 2 a donc également perçu le facteur de variabilité 'type de pâte' au sein de l'ensemble des recettes.

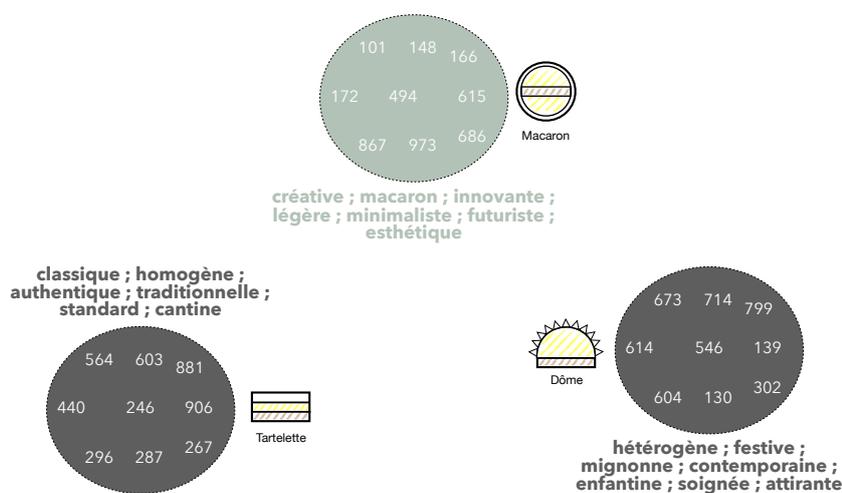


Figure 52: Étude 'recettes de tarte au citron meringuée' : Partition des recettes consensuelle associée au cluster 1 d'évaluateurs. Chaque groupe est décrit par les caractéristiques perceptuelles sur-utilisées pour le décrire (relativement à l'ensemble des recettes). Les principales différences avec la partition consensuelle du cluster 2 sont indiquées en gris.

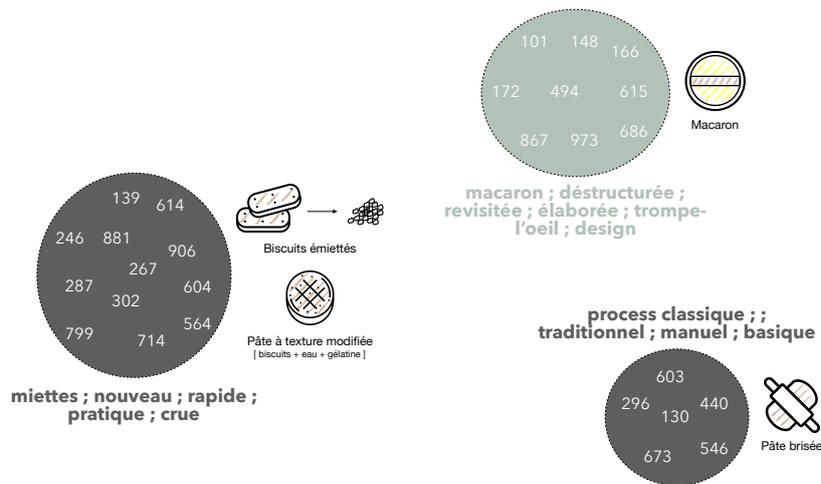


Figure 53: Étude 'recettes de tarte au citron meringuée' : Partition des recettes consensuelle associée au cluster 2 d'évaluateurs. Chaque groupe est décrit par les caractéristiques perceptuelles sur-utilisées pour le décrire (relativement à l'ensemble des recettes). Les principales différences avec la partition consensuelle du cluster 1 sont indiquées en gris.

Les caractéristiques des évaluateurs au sein des deux clusters sont présentées dans les Tables 17 et 18. Le cluster 1 présente une forte proportion de personnes âgées de 36 à 55 ans. Il présente également une forte proportion de personnes ayant un niveau d'expertise culinaire faible ou intermédiaire. Il apparaît donc que les personnes ayant perçu uniquement le facteur de variabilité 'type de dressage' sont majoritairement des personnes ayant plus de 35 ans et des personnes avec un faible niveau d'expertise culinaire. Le cluster 2, lui, présente une forte proportion de personnes âgées de 18 à 35 ans. Plus important, il présente une forte proportion de personnes ayant un bon niveau d'expertise culinaire. Ces personnes sont celles qui ont perçu, en plus du facteur 'type de dressage', le facteur 'type de pâte'. Cependant, ces observations ne restent que des tendances. La caractérisation automatique des clusters d'évaluateurs n'a en effet pas mis en évidence de lien significatif entre le cluster et les caractéristiques individuelles des évaluateurs. Dans ce cas précis, il est impossible de relier de manière significative les types de perception des recettes à une typologie de personnes en particulier.

Table 17: Étude ‘recettes de tarte au citron meringuée’ : Caractéristiques individuelles des évaluateurs du cluster 1.

Cluster 1		
<i>Number of raters</i>	37 (56% of the panel)	
<i>Covariates about the raters</i>	<i>Proportion in the cluster (%)</i>	<i>Proportion in the panel (%)</i>
Age=46-55	10.81	6.06
Age=36-45	10.81	6.06
Cooking.level=medium	54.05	51.51
Cooking.level=weak	16.27	15.15
Gender=M	32.43	31.82
Cooking.level=extremely.weak	13.51	13.64
Gender=F	67.57	68.18
Age=18-25	75.68	78.79
Cooking.level=good	16.22	19.70
Age=26-35	2.70	6.06
Age=-18	0.00	3.03

Table 18: Étude ‘recettes de tarte au citron meringuée’ : Caractéristiques individuelles des évaluateurs du cluster 2.

Cluster 2		
<i>Number of raters</i>	29 (44% of the panel)	
<i>Covariates about the raters</i>	<i>Proportion in the cluster (%)</i>	<i>Proportion in the panel (%)</i>
Age=-18	6.90	3.03
Age=26-35	10.34	6.06
Cooking.level=good	24.14	19.70
Age=18-25	82.76	78.79
Gender=F	68.97	68.18
Cooking.level=extremely.weak	13.78	13.64
Gender=M	31.03	31.82
Cooking.level=weak	13.79	15.15
Cooking.level=medium	48.28	51.51
Age=46-55	0.00	6.06
Age=36-45	0.00	6.06

3 Conclusion

Dans ce chapitre, nous avons proposé une application du Tri Séquentiel Agglomératif (TSA) pour la caractérisation de 27 recettes de tarte au citron meringuée. Cette application montre tout l’intérêt du TSA pour la caractérisation perceptuelle d’un grand nombre de gestes culinaires.

D’une part, la procédure de recueil des données a été globalement appréciée par les évaluateurs. Bien que la tâche ait été perçue comme difficile et fatigante, 80 % des évaluateurs ont dit avoir été très intéressés par l’expérience. Plus important encore, ce terrain a permis de démontrer que des évaluateurs non-experts sont capables de caractériser des gestes culinaires.

D'autre part, la procédure d'analyse statistique des données développée dans le cadre de cette thèse a fourni des résultats intéressants. En ce qui concerne la représentation perceptuelle des recettes évaluées, les résultats ont montré que celle-ci dépend principalement de l'aspect visuel du produit créé grâce à la recette. En ce qui concerne la perception des recettes par les évaluateurs, les résultats ont montré que celle-ci peut prendre différentes formes. Deux clusters d'évaluateurs ont en effet été mis en évidence par la procédure de segmentation du panel. Ces clusters sont assimilables à des types de perception des recettes de tarte au citron meringuée. Les résultats ont notamment montré que les personnes présentant un faible niveau d'expertise culinaire (cluster 1) se focalisent sur le type de dressage utilisé pour présenter le produit fini. Les personnes qui présentent un bon niveau d'expertise culinaire se focalisent, elles, également sur des caractéristiques plus techniques de la recette telles que le type de pâte utilisé. L'utilisation de certains types de pâte représente pour elles une certaine praticité et une certaine rapidité de mise en œuvre. Contrairement au cluster 1, ce cluster 2 prend également en compte les propriétés d'usage de la recette. Cependant, ces observations sont à nuancer. Aucun lien significatif entre le cluster d'appartenance et les caractéristiques individuelles des personnes n'a en effet été établi.

Enfin, les résultats ont montré qu'une saturation de la variabilité perceptuelle des recettes apparaissait rapidement au cours du TSA. Le fait que les recettes soient issues d'un plan d'expérience peut expliquer cet aspect. Il était effectivement probable que les évaluateurs aient accès à la majorité des 9 paramètres de variation dès la première phase du TSA. À l'avenir, il sera nécessaire de tester le TSA sur des gestes 'spontanés' recueillis dans des conditions écologiques. Un tel terrain mis en place dans un autre contexte est présenté dans le Chapitre 6 de cette thèse.



Annexe

IMPLÉMENTATION DANS R DE L'ANALYSE STATISTIQUE DES DONNÉES DE TRI SÉQUENTIEL AGGLOMÉRATIF

RÉSUMÉ : Une procédure *ad hoc* d'analyse statistique des données de Tri Séquentiel Agglomératif a été introduite dans cette thèse. Dans cette annexe, son implémentation dans le logiciel libre R est présentée. Celle-ci a été réalisée à travers le développement du package **SeqAgSorting**. Ce package repose sur une fonction principale qui fournit à l'utilisateur tous les résultats numériques et graphiques utiles à la compréhension des données de Tri Séquentiel Agglomératif. Des fonctions complémentaires permettent à l'utilisateur de personnaliser les sorties graphiques, voire de les rendre interactives pour faciliter leur interprétation et leur diffusion.

Sommaire

1	Fonction AnalyseSAS	115
1.1	Arguments en entrée	115
1.2	Sorties	118
2	Fonction plot.AnalyseSAS	120
2.1	Arguments en entrée	120
2.2	Sorties	121

Le choix du logiciel R pour implémenter la démarche statistique est motivé par plusieurs aspects. Tout d’abord, il s’agit d’un logiciel libre et gratuit. Aussi, R tend à devenir le logiciel de statistique de référence. L’étendue de ses fonctionnalités et son essor permanent peuvent expliquer cet engouement. Il présente également l’avantage d’être utilisé à la fois dans le monde académique et dans le monde de l’entreprise. Enfin, il propose de plus en plus de solutions pour faciliter la communication des résultats statistiques. Parmi celles-ci, nous pouvons citer la création de graphiques esthétiques et dynamiques grâce à des packages dédiés comme `ggplot2` (Wickham, 2016) et `plotly` (Sievert et al., 2017), ou la création d’interfaces web grâce au package `RShiny`. Le logiciel R est ainsi accessible au plus grand nombre : des universitaires, des membres d’industries, des statisticiens, des ingénieurs R&D, des débutants ou des utilisateurs confirmés et expérimentés.

Le package `SeqAgSorting`, développé dans le cadre de ce travail de recherche, est dédié à l’analyse statistique des données de Tri Séquentiel Agglomératif (TSA). Il est disponible gratuitement sur la plateforme GitHub : <https://github.com/MargotBr/SeqAgSorting>. À partir de cette plateforme, l’utilisateur peut télécharger le package.

À travers les fonctions R qu’il contient, le package implémente les procédures statistiques orientée stimuli et orientée évaluateurs (voir Chapitre 2). L’implémentation de l’approche orientée stimuli a été construite autour des fonctions `MCA` du package `FactoMineR` (S. Lê, Josse, & Husson, 2008), et des fonctions `fast` et `ConsensualWords` du package `SensoMineR` (Husson, Lê, & Cadoret, 2017). L’implémentation de l’approche orientée évaluateurs est, elle, implémentée en utilisant les fonctions de base de R pour la mise en oeuvre de la procédure de classification non supervisée. En outre, elle fait appel à la fonction `NbClust` du package `NbClust` pour définir le niveau de coupure optimal du dendrogramme (Charrad, Ghazzali, Boiteau, & Niknafs, 2014). Elle utilise également la fonction `RandIndex` du package `FreeSortR` pour la construction de la matrice de désaccord entre évaluateurs (Courcoux, 2014). La fonction `ConsensusPartition` du même package est également utilisée pour établir la partition sur les stimuli consensus de chaque cluster d’évaluateurs.

Le package `SeqAgSorting` contient des données fictives pédagogiques. Celles-ci permettent à l'utilisateur de comprendre rapidement le fonctionnement des différentes fonctions du package. Des données réelles sont également disponibles. Elles correspondent aux données 'parfums de gel douche' introduites dans le Chapitre 2. L'utilisateur peut également utiliser les données 'recettes de tarte au citron meringuée' présentées dans le Chapitre 3.

Dans cette annexe, nous détaillons succinctement les fonctions de ce package R.

1 Fonction AnalyseSAS

La fonction `AnalyseSAS` est la fonction principale du package. C'est elle qui permet d'implémenter la procédure statistique des données. Son appel dans R est réalisé grâce aux lignes de commande suivantes :

```
AnalyseSAS(dta, sast.parameters, sep.part.step = "_", sep.character =
" ; ", nbtimes.consens.character = 5, proba.consens.character = 0.05,
id.info.stim = NULL, type.info.stim = NULL, id.info.part = NULL,
type.info.part = NULL, graph = TRUE, ext.dev.Rstudio = FALSE)
```

1.1 Arguments en entrée

La fonction `AnalyseSAS` admet douze arguments majeurs en entrée :

- `dta` : Jeu de données de TSA de dimensions $S \times (I \times P)$. S définit le nombre de stimuli, I le nombre d'étapes (c.-à-d. de sous-ensembles de stimuli) du TSA, et P le nombre d'évaluateurs. Il peut être supplémenté par des covariables sur les stimuli (en colonnes) ou sur les évaluateurs (en lignes). Un exemple schématique est présenté sur la Figure 54.
- `sast.parameters` : Nombre de stimuli par sous-ensembles.
- `sep.part.step` : Caractère utilisé pour séparer l'identifiant des évaluateurs (= participants) et l'identifiant des étapes dans le nom des colonnes de `dta`. Par défaut, cet argument est fixé à "_" et correspond à une situation similaire à celle présentée en Figure 54.
- `sep.character` : Caractère utilisé pour séparer les caractéristiques perceptuelles dans les cellules de `dta`. Par défaut, cet argument est fixé à " ; " et correspond à une situation similaire à celle présentée en Figure 54.
- `nbtimes.consens.character` : Nombre minimum de fois où une caractéristique perceptuelle doit être citée à l'échelle du panel pour être considérée dans les tests statistiques dédiés à la détermination des caractéristiques consensuelles. Par défaut, cet argument est fixé à 5.

- `proba.consens.character` : Seuil utilisé pour accepter ou rejeter l'hypothèse nulle des tests statistiques dédiés à la détermination des caractéristiques consensuelles. Par défaut, cet argument est fixé à 0.05.
- `id.info.stim` : Identifiants des colonnes de `dta` correspondant aux covariables sur les stimuli. Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les stimuli.
- `type.info.stim` : Nature des covariables sur les stimuli. 'cont' est utilisé pour indiquer une covariable continue et 'cat' est utilisé pour indiquer une covariable catégorielle. Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les stimuli.
- `id.info.part` : Identifiants des lignes de `dta` correspondant aux covariables sur les évaluateurs (= participants). Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les évaluateurs.
- `type.info.part` : Nature des covariables sur les évaluateurs. 'cont' est utilisé pour indiquer une covariable continue et 'cat' est utilisé pour indiquer une covariable catégorielle. Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les évaluateurs.
- `graph` : Choix d'afficher ou non les sorties graphiques produites par la fonction (booléen). Si cette valeur est égale à `TRUE` alors les graphiques sont affichés. Par défaut, cet argument est fixé à `TRUE`.
- `ext.dev.Rstudio` : Choix d'afficher ou non les sorties graphiques à l'extérieur de la fenêtre de RStudio (booléen). Si cette valeur est égale à `TRUE` alors les graphiques sont affichés à l'extérieur. Par défaut, cet argument est fixé à `FALSE`.

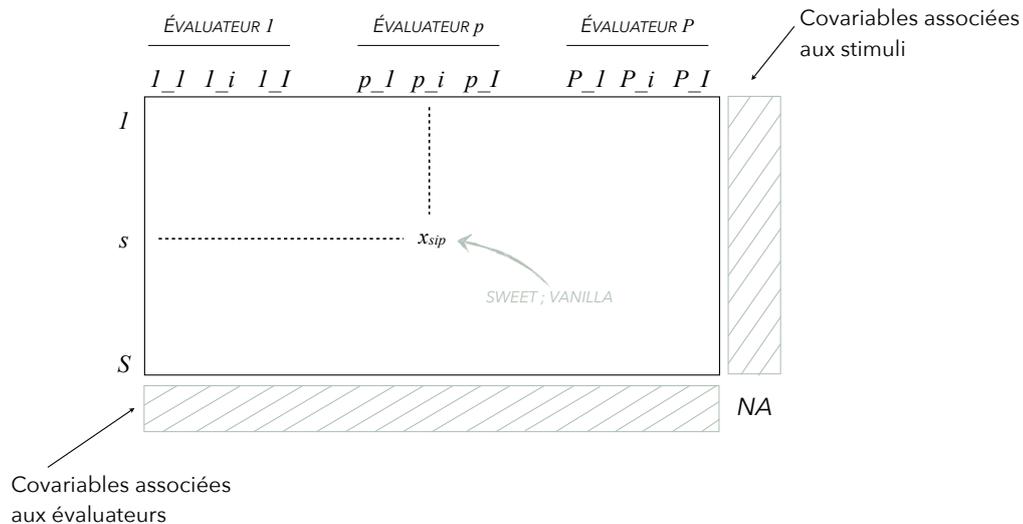


Figure 54: Format du jeu de données `dta` utilisé dans la fonction `AnalyseSAS`.

Les dix derniers arguments de la fonction `AnalyseSAS` présentent des valeurs par défaut. Si celles-ci conviennent à l'utilisateur, il suffit alors d'entrer uniquement le nom du jeu de données - sans covariables sur les stimuli ou les évaluateurs - et le nombre de stimuli par sous-ensemble dans la fonction pour lancer l'analyse statistique.

Exemple

Le package `SeqAgSorting` contient un petit exemple fictif pédagogique : `pedagdata`. Ce jeu de données correspond à des données de TSA relatives à un ensemble de 10 stimuli caractérisés par un panel de 10 évaluateurs. La tâche de TSA impliquait trois sous-ensembles contenant respectivement 4, 3 et 3 stimuli. Ces données de TSA sont supplémentées par des covariables sur les évaluateurs et sur les stimuli. Deux covariables sont disponibles sur les évaluateurs : genre et tranche d'âge. Elles sont toutes les deux catégorielles. Une covariable est disponible sur les stimuli : type de packaging ('packaging actuel' ou 'packaging prototype'). Elle est également catégorielle.

L'utilisation de ce jeu de données peut être réalisée grâce aux commandes suivantes :

```
> data(pedagdata)
> res.pedag <- AnalyseSAS(pedagdata,
+   sast.parameters = c(4, 3, 3),
+   nbtimes.consens.character = 2,
+   id.info.stim = 31,
+   type.info.stim = "cat",
+   id.info.part = 11 : nrow(pedagdata),
```

```
+ type.info.part = rep("cat", 2))
> res.pedag
```

1.2 Sorties

La fonction `AnalyseSAS` retourne à l'utilisateur une liste contenant les éléments suivants :

- `res.mca` : Résultats de l'Analyse des Correspondances Multiples (ACM) réalisée dans le cadre de la procédure statistiques orientée stimuli. Grâce à cet élément, il est possible d'accéder aux valeurs propres de l'ACM, aux coordonnées des stimuli, etc.
- `coocc.stim` : Matrice des cooccurrences entre stimuli. Pour chaque couple de stimuli, le nombre d'évaluateurs les ayant placés dans un même groupe à la fin du TSA est comptabilisé.
- `consensual.charact.stim` : Résultats des tests statistiques dédiés à la détermination des caractéristiques perceptuelles consensuelles.
- `partition.stim` : Partition des stimuli issue d'une Classification Ascendante Hiérarchique réalisée sur les coordonnées des stimuli sur les 5 premiers axes de l'ACM.
- `charact.clust.stim` : Caractérisation des clusters de stimuli de la partition précédente grâce aux caractéristiques perceptuelles.
- `remark.stim` : Évolution de la caractérisation perceptuelle des stimuli au cours du TSA.
- `ARI.part` : Matrice de similarité entre évaluateurs.
- `partition.part` : Partition des évaluateurs issue de la procédure statistiques orientée évaluateurs.
- `cons.partition.stim.clust` : Partition sur les stimuli consensus associée à chaque cluster d'évaluateurs. Chaque cluster de stimuli au sein de la partition est ensuite caractérisé grâce aux caractéristiques perceptuelles.
- `charact.clust.part` : Description automatique des clusters d'évaluateurs, notamment grâce aux covariables disponibles.

En parallèle de cette liste, la fonction `AnalyseSAS` retourne à l'utilisateur plusieurs sorties graphiques. Celles-ci sont réalisées grâce au package `ggplot2`. Elles sont présentées en Figure 55.

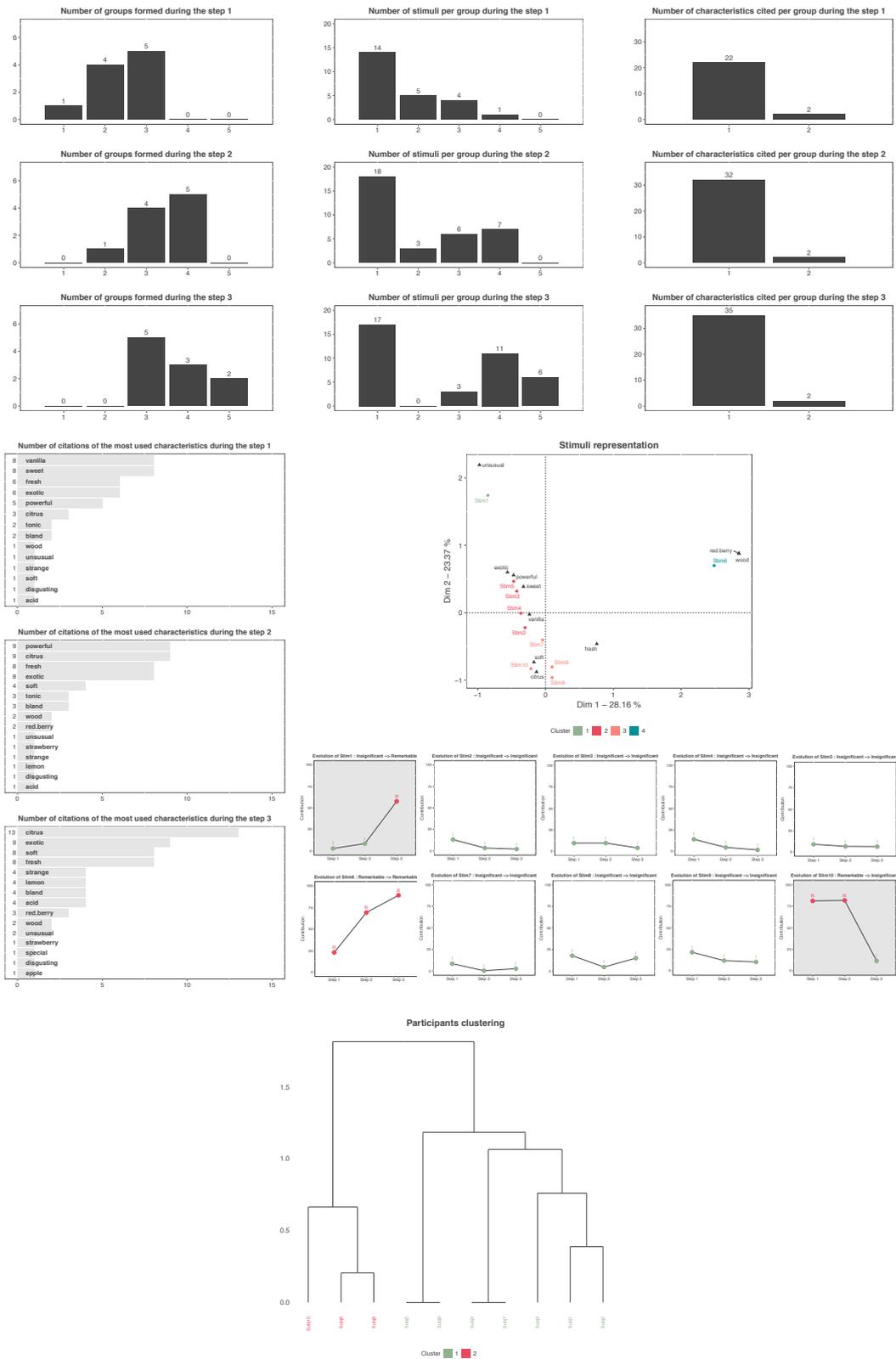


Figure 55: Jeu de données pédagogique du package SeqAgSorting : Sorties graphiques de la fonction AnalyseSAS. Les résultats du haut sont ceux de la procédure statistique orientée stimuli et les résultats du bas sont ceux de la procédure statistique orientée évaluateurs.

2 Fonction `plot.AnalyseSAS`

La fonction `plot.AnalyseSAS` est une fonction secondaire du package. Elle permet d'afficher les représentations graphiques issues de la fonction `AnalyseSAS`. Elle prend donc en entrée l'objet résultant de cette fonction. Elle permet d'afficher certains graphiques en les modifiant si besoin. Il est ainsi possible de modifier les couleurs ou de changer les axes des analyses factorielles affichés. Enfin, si l'utilisateur le souhaite, cette fonction peut afficher les graphiques sous un format interactif grâce au package `plotly`. L'appel de cette fonction se fait grâce aux lignes de commande suivantes :

```
plot.AnalyseSAS(res, choice = "all", interact = FALSE,  
col.clust.stim = NULL, col.clust.part = NULL, axis = c(1, 2),  
ext.dev.Rstudio = FALSE)
```

2.1 Arguments en entrée

La fonction `plot.AnalyseSAS` admet six arguments majeurs en entrée :

- `res` : Objet retourné par la fonction `AnalyseSAS`.
- `choice` : Type(s) de graphique(s) à créer. `'stim'` indique le souhait d'afficher la représentation multidimensionnelle des stimuli (procédure statistique orientée stimuli). `'part'` indique le souhait d'afficher le dendrogramme et la partition représentant la structure de désaccord au sein du panel (procédure statistique orientée évaluateurs). `'all'` indique que l'intégralité de ces graphiques doit être affichée. Par défaut, cet argument est fixé à `'all'`.
- `interact` : Choix de créer ou non les graphiques sous un format interactif (booléen). Si cette valeur est égale à `TRUE` alors les graphiques créés sont interactifs. Par défaut, cet argument est fixé à `FALSE`.
- `col.clust.stim` : Couleurs associées aux clusters de stimuli sur l'ACM de la procédure statistique orientée stimuli. Les éléments de cet argument peuvent être spécifiés à travers une chaîne de caractères correspondant à chaque couleur (ex. `'pink'`) ou grâce au code hexadécimal associé à chacune (ex. `'#B52E57'`).
- `col.clust.part` : Couleurs associées aux clusters d'évaluateurs (= de participants) issus de la procédure statistique orientée évaluateurs. Les éléments de cet argument peuvent être spécifiés à travers une chaîne de caractères correspondant à chaque couleur (ex. `'pink'`) ou grâce au code hexadécimal associé à chacune (ex. `'#B52E57'`).
- `axis` : Axes factoriels à afficher sur la représentation multidimensionnelle des stimuli. Par défaut, cet argument est fixé à `axis = c(1, 2)` et les deux premiers axes de l'ACM sont affichés.

- `ext.dev.Rstudio` : Choix d'afficher ou non les sorties graphiques à l'extérieur de la fenêtre de RStudio (booléen). Si cette valeur est égale à `TRUE` alors les graphiques sont affichés à l'extérieur. Par défaut, cet argument est fixé à `FALSE`.

2.2 Sorties

Les lignes de commande suivantes permettent de changer les couleurs des graphiques (un exemple de sortie est présenté sur la Figure 56) :

```
> plot.AnalyseSAS(res.pedag,
+ choice = "stim",
+ col.clust.stim = c("#B3C2B7", "#92D5B6", "#737575", "#A2A590"))
```

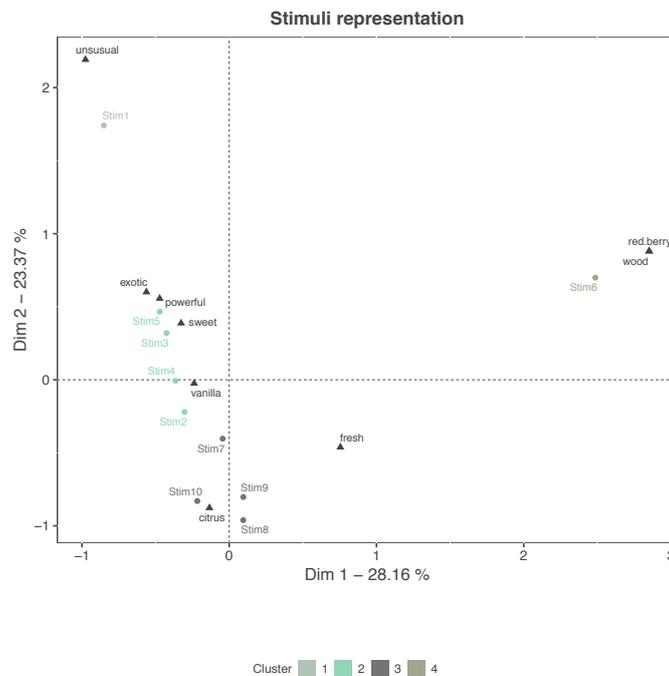


Figure 56: Jeu de données pédagogique du package `SeqAgSorting` : Modification des couleurs de la sortie graphique de la procédure statistique orientée stimuli grâce à la fonction `plot.AnalyseSAS`.

En plus de changer les couleurs des graphiques, les lignes de commande suivantes permettent de les afficher sous format interactif. Sur la représentation multidimensionnelles des stimuli, un simple déplacement du curseur de la souris de la part de l'utilisateur permet d'afficher des informations complémentaires sur les stimuli

ou sur les caractéristiques perceptuelles (un exemple de sortie est présenté sur la Figure 57) :

```
> plot.AnalyseSAS(res.pedag,
+ interact = TRUE,
+ choice = "stim",
+ col.clust.part = c("#B3C2B7", "#92D5B6", "#737575", "#A2A590"))
```

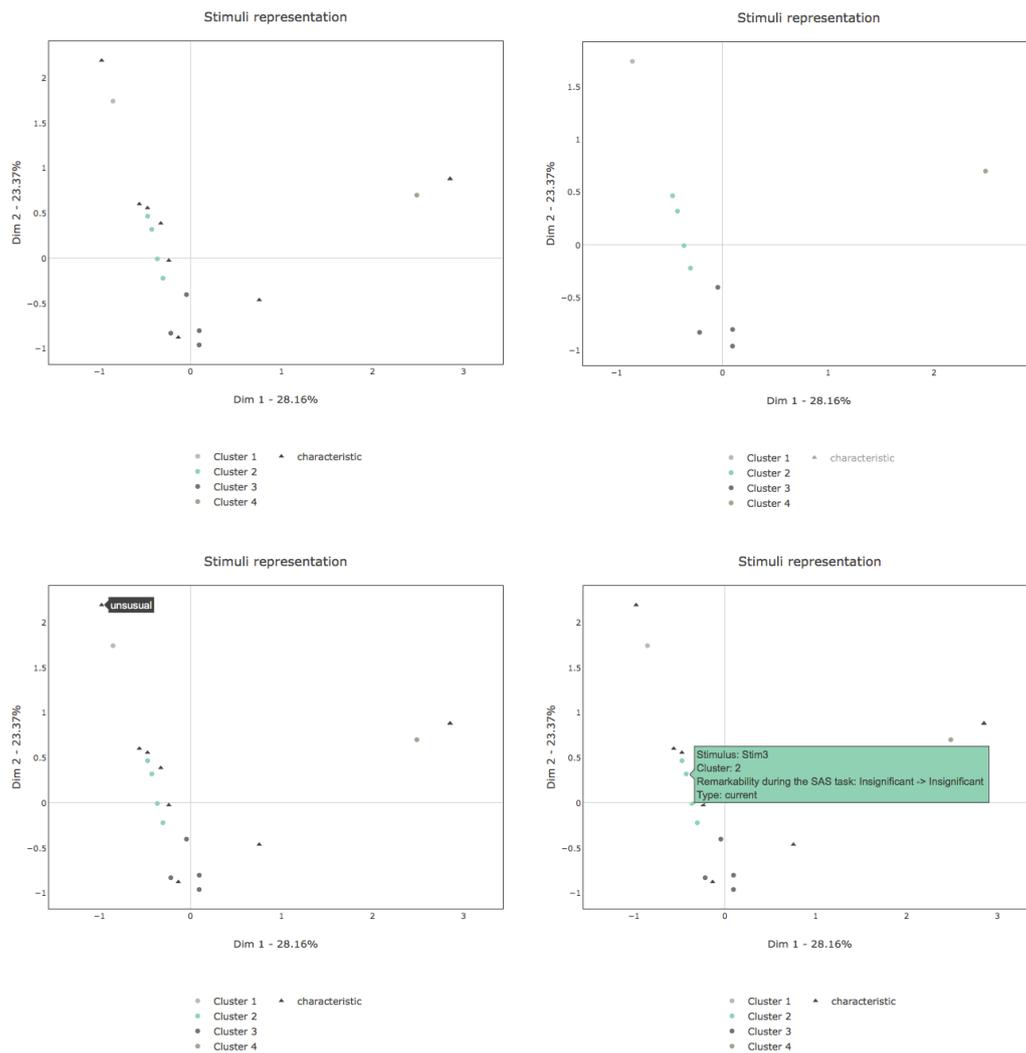


Figure 57: Jeu de données pédagogique du package `SeqAgSorting` : Modification des couleurs des sorties graphiques de la procédure statistique orientée stimuli et affichage sous format interactif grâce à la fonction `plot.AnalyseSAS`.

Le même type de sorties interactives peut également être obtenue pour la procédure orientée évaluateurs (voir Figure 58).

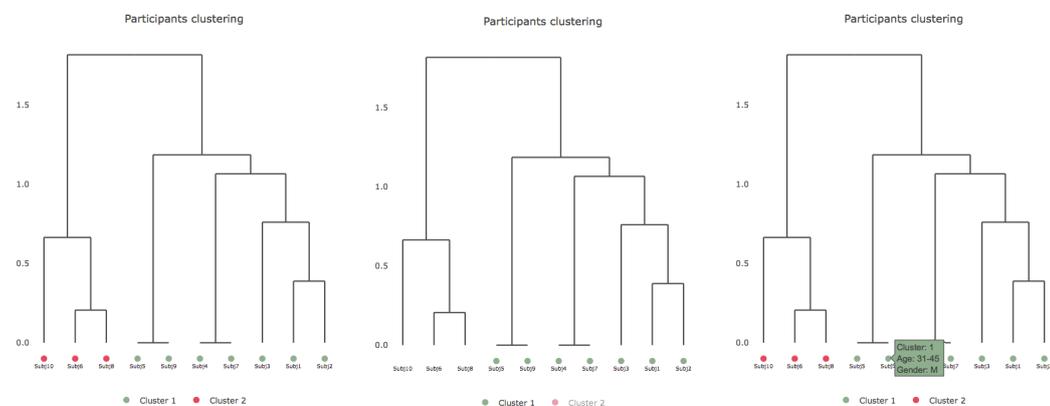


Figure 58: Jeu de données pédagogique du package `SeqAgSorting` : Affichage de la sortie graphique de la procédure statistique orientée évaluateurs sous format interactif grâce à la fonction `plot.AnalyseSAS`.

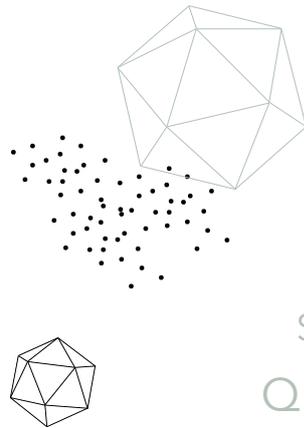
Dans l'objectif de rendre les fonctionnalités de ce package accessibles au plus grand nombre d'utilisateurs - y compris des utilisateurs non familiers avec l'environnement de R - une application Shiny doit également être développée. Cette application web permettra à l'utilisateur, grâce à une interface ergonomique, d'accéder à toutes les sorties du package R de manière dynamique et interactive.

Partie II :

Caractérisation conceptuelle

La seconde partie de cette thèse présente notre proposition méthodologique pour l'analyse statistique des données de Q-méthodologie binaire. Elle présente également son application à des données de caractérisation d'un grand nombre de gestes culinaires.

Dans un premier temps, le protocole de recueil des données de Q-méthodologie binaire est présenté. Une étude approfondie du format des données permet de révéler leurs spécificités. Comme nous le verrons, celles-ci sont au nombre de deux. Une procédure d'analyse statistique permettant d'identifier les différences et similitudes conceptuelles qui existent au sein de l'ensemble de stimuli est proposée. Elle permet de dépasser les limites des approches utilisées dans la littérature, notamment dans l'article originel de Jaffré et al. (2009). Elle prend en effet en compte les deux spécificités des données de Q-méthodologie binaire, et non uniquement une comme cela est habituellement le cas. Ces différentes notions sont abordées dans le Chapitre 4. Dans le Chapitre 5, une procédure d'analyse statistique permettant de comprendre comment les évaluateurs ont perçu les stimuli et, éventuellement, d'identifier des clusters d'évaluateurs aux perceptions différentes est introduite. Cette problématique n'a jamais été abordée dans un contexte de Q-méthodologie binaire. Dans le Chapitre 6, l'étude de l'applicabilité de la Q-méthodologie binaire à des vidéos de gestes culinaires est abordée. Pour cela, les résultats d'un terrain réalisé au Centre Culinaire Contemporain sont exposés. Enfin, cette partie se conclut par une Annexe dans laquelle est présentée l'implémentation dans R de la procédure d'analyse statistique des données de Q-méthodologie binaire.



Chapitre 4

RECUEIL & ANALYSE STATISTIQUE ORIENTÉE STIMULI DES DONNÉES DE Q-MÉTHODOLOGIE BINAIRE

RÉSUMÉ : Dans un contexte de Q-méthodologie binaire, les résultats permettant d'analyser la structure de variabilité conceptuelle observée au sein de l'ensemble de stimuli sont primordiaux. Des solutions ont été mises en place pour obtenir de tels résultats. Celles-ci sont dédiées à l'analyse des données de tri libre. Elles permettent d'obtenir une distance euclidienne inter-stimuli qui reflète de manière spécifique une différence de représentation du concept. Des jeux de données fictifs pédagogiques sont utilisés pour démontrer que ces techniques classiques ne sont pourtant pas optimales. Elles occultent en effet une partie de l'information véhiculée par les données. Une présentation du protocole de recueil des données de Q-méthodologie binaire ainsi qu'une étude approfondie de leur format permet d'identifier quelles sont ces informations. Les données présentent deux spécificités majeures. L'une est commune avec les données obtenues dans le cadre d'une tâche de tri libre ; l'autre est commune avec les données obtenues dans le cadre d'une tâche d'association. L'échec des techniques classiques est dû au fait qu'elles ne prennent en compte que la première lors de la définition de la distance inter-stimuli. Dans ce chapitre, une procédure d'analyse statistique permettant de prendre en compte ces deux spécificités est proposée. Celle-ci est implémentée sous la forme de la combinaison d'une approche d'analyse factorielle de tableaux multiples et d'une approche originale de visualisation des résultats fournis par la précédente. La procédure statistique proposée est motivée par son application sur des données réelles de caractérisation de produits industriels.

CE CHAPITRE EST INSPIRÉ DE L'ARTICLE :

Brard, M., & Lê, S. (2018). Adaptation of the Q-methodology for the characterization of a complex concept through a set of products: from the collection of the data to their analysis. *Food Quality and Preference*, 67, 77-86.

Sommaire

1	Recueil des données	129
1.1	Réalisation de la tâche de caractérisation conceptuelle par l'évaluateur	129
1.2	Format des données	130
2	Analyse statistique des données orientée stimuli	131
2.1	Définition de la distance inter-stimuli et représentation euclidienne associée	131
2.1.1	Distance prenant en compte la première spécificité des données : focus sur l'Analyse des Correspondances Multiples	131
2.1.2	Distance prenant en compte la seconde spécificité des données : focus sur l'Analyse Factorielle des Correspondances	137
2.1.3	Proposition d'une distance globale <i>ad hoc</i> : focus sur l'Analyse Factorielle Multiple	140
2.2	Amélioration de la visualisation de la représentation euclidienne des stimuli	148
3	Illustration sur des données réelles de caractérisation de produits industriels : données 'concept d'innovation et parfums de shampoings'	151
3.1	Données	152
3.2	Résultats de l'analyse statistique orientée parfums	152
4	Conclusion	154

Dans un premier temps, nous présentons brièvement la procédure de recueil des données mise en place lors de l'implémentation de la Q-méthodologie binaire. Nous décrivons la tâche qu'il est demandé à chaque évaluateur de réaliser en vue d'obtenir une évaluation de la représentation du concept au sein de l'ensemble de stimuli. Le rappel de cette procédure nous amène par la suite à introduire le format des données recueillies.

Dans un second temps, nous abordons la manière avec laquelle il est possible d'analyser ces données selon une approche orientée stimuli. Dans le cadre de cette méthodologie, comme pour la plupart des procédures sensorielles, les résultats d'une telle approche sont généralement présentés sous la forme d'une représentation euclidienne des stimuli. Sur ce type de représentation, la distance euclidienne entre deux stimuli est d'autant plus importante que ces stimuli représentent le concept étudié d'une façon différente. C'est sur cette problématique de l'obtention d'une telle distance inter-stimuli que nous proposons de nous pencher. Les techniques statistiques classiquement utilisées dans ce cadre ne sont pas optimales. C'est le cas de celle proposée par Jaffré et al. (2009) qui utilise une distance qui ne restitue

pas l'intégralité des informations sur les stimuli véhiculées par les données. Des jeux de données fictifs pédagogiques sont utilisés pour identifier ces informations. Ceci nous permet de mettre en évidence le fait que les données de Q-méthodologie binaire présentent deux spécificités. La première est commune avec les données obtenues dans le cadre d'une tâche de tri libre ; la seconde est commune avec les données obtenues dans le cadre d'une tâche d'association. Les techniques statistiques classiques ne prennent en compte que la première spécificité. Elles occultent totalement la seconde. Nous présentons ensuite une approche statistique permettant de prendre en compte cette seconde spécificité des données. Ceci nous amène, par la suite, à proposer une distance inter-stimuli *ad hoc*. Celle-ci est basée sur une combinaison des deux distances obtenues grâce aux approches précédentes, chacune d'entre elles prenant en compte une spécificité des données. L'approche proposée repose sur une technique d'analyse factorielle de tableaux multiples. L'adéquation de la distance obtenue est démontrée sur les jeux de données fictifs pédagogiques. Cependant, nous verrons que l'interprétation de la représentation euclidienne des stimuli obtenue peut s'avérer délicate. Pour dépasser cette limite, nous achevons ce chapitre sur la proposition d'une approche de visualisation originale. Celle-ci permet d'interpréter efficacement la structure de variabilité conceptuelle existant au sein de l'ensemble de stimuli.

Dans un troisième et dernier temps, nous présentons l'application de cette procédure statistique à des données réelles de Q-méthodologie binaire. Celles-ci sont des données d'évaluation de la représentation du concept d'innovation au sein d'un ensemble de 39 parfums de shampoings.

1 Recueil des données

1.1 Réalisation de la tâche de caractérisation conceptuelle par l'évaluateur

Chaque stimulus est affecté à l'une des deux catégories définies en amont par l'expérimentateur : soit il est évalué comme 'non-représentatif du concept', soit il est au contraire évalué comme 'représentatif du concept'. Tous les stimuli sont présentés simultanément. Les instructions données à l'évaluateur sont les suivantes : 'Pour le concept dont il est question lors de cette étude, veuillez trier les stimuli dans l'une de ces deux catégories : les stimuli que vous jugez être non-représentatifs du concept et les stimuli que vous jugez être représentatifs du concept. Vous pouvez vous appuyer sur une comparaison des stimuli pour réaliser cette tâche, et cette comparaison doit être réalisée selon vos propres critères. Il n'y a pas de bonnes ou de mauvaises réponses puisque nous sommes intéressés par votre propre perception. Il n'y a aucune contrainte sur le nombre de stimuli qu'une catégorie peut contenir à la fin de la tâche'. Cette tâche est réalisée par chaque évaluateur du panel. Chacun fournit, *in fine*, une partition labellisée des stimuli telle que celle présentée sur la Figure 59.

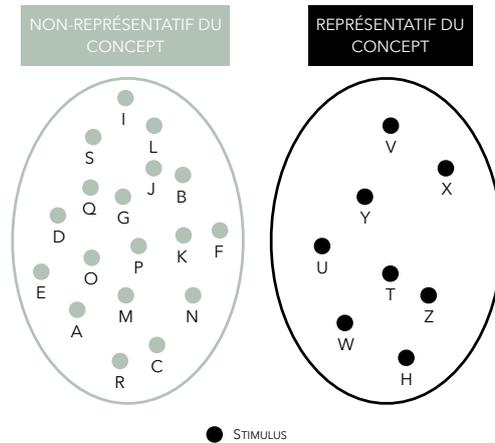


Figure 59: Exemple d'un tri basé sur deux catégories prédéfinies (l'une représentant le concept et l'autre représentant sa négation) réalisée par un évaluateur sur 26 stimuli.

1.2 Format des données

Les données recueillies dans le cadre de l'implémentation d'une Q-méthodologie binaire sont les suivantes : pour le s -ème stimulus, l'évaluation par le p -ème évaluateur est une variable binaire Y_{sp} qui est égale à 1 si le stimulus a été jugé comme représentatif du concept par l'évaluateur (évaluation positive), et à 0 dans le cas contraire (évaluation négative). Ces données peuvent être combinées dans une matrice Y à S lignes et P colonnes ; chaque ligne s correspondant à un stimulus, chaque colonne p correspondant à un évaluateur, et à l'intersection de la ligne s et de la colonne p la valeur y_{sp} (voir Figure 60).

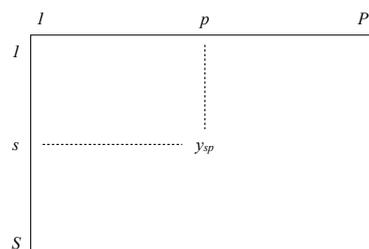


Figure 60: Données recueillies dans le cadre de l'implémentation de la Q-méthodologie binaire.

Dans cette matrice, le s -ème stimulus est défini par le profil d'évaluations binaires (Y_{s1}, \dots, Y_{sP}) , profil correspondant aux différentes évaluations que les P évaluateurs lui ont associées. C'est de l'analyse statistique de ces profils associés aux stimuli dont il est question dans la suite de ce chapitre. De la même façon, le p -ème évaluateur est défini par le profil d'évaluations binaires (Y_{1p}, \dots, Y_{Sp}) , profil

correspondant aux différentes évaluations qu'il a associées aux S stimuli. L'étude de ces profils associés aux évaluateurs fait l'objet d'une réflexion méthodologique spécifique présentée dans le Chapitre 5.

2 Analyse statistique des données orientée stimuli

2.1 Définition de la distance inter-stimuli et représentation euclidienne associée

2.1.1 Distance prenant en compte la première spécificité des données : focus sur l'Analyse des Correspondances Multiples

Classiquement, les données de Q-méthodologie binaire sont analysées comme des données de tri libre. C'est le cas dans l'article originel de Jaffré et al. (2009). Pour rappel, la tâche de tri libre consiste à demander à l'évaluateur de former des groupes de stimuli en fonction de leur ressemblance (voir Figure 61). L'évaluateur est libre de former autant de groupes qu'il le souhaite. Il doit utiliser ses propres critères d'évaluation pour réaliser la tâche.

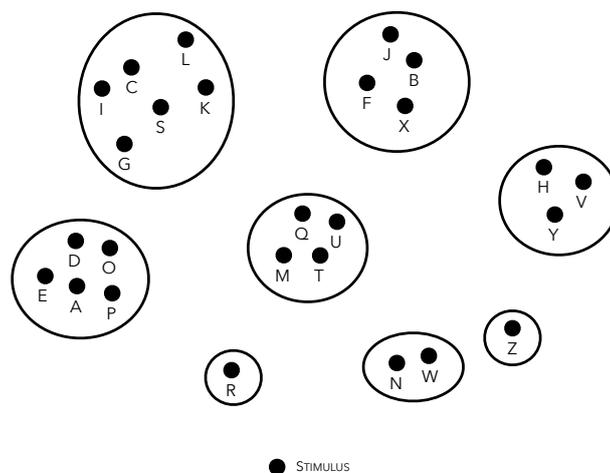


Figure 61: Exemple d'une tâche de tri libre réalisée par un évaluateur sur 26 stimuli. Cet évaluateur a créé 8 groupes de stimuli.

Comme décrit dans la Partie 1 de cette thèse (voir Section 2.1.2 du Chapitre 2), l'expérimentateur peut utiliser plusieurs techniques statistiques pour comprendre comment ont été perçus les stimuli par les évaluateurs. L'une d'elle a fait l'objet d'une thèse réalisée à AGROCAMPUS OUEST : Cadoret (2010). Elle consiste à analyser les données ayant en lignes les stimuli et en colonnes les évaluateurs par Analyse des Correspondances Multiples (ACM), une méthode dédiée à l'analyse

de matrices de données catégorielles (Greenacre, 1984; Lebart, 1975) : Abdi and Valentin (2007); Cadoret et al. (2009); Van der Kloot and Van Herck (1991); Takane (1980). Tout comme les autres procédures présentées dans ce chapitre, cette approche est basée sur une décomposition en valeurs singulières. C'est pourquoi nous faisons ici le choix de nous concentrer sur celle-ci.

Dans le cadre de l'ACM, les données sont prises en compte à travers un tableau dit *disjonctif complet*. Dans celui-ci, chaque évaluateur est représenté par autant de vecteurs binaires qu'il a créé de groupes de stimuli. Un stimulus prend la valeur de 1 pour le vecteur correspondant au groupe auquel il appartient, et la valeur de 0 pour les autres (voir Figure 62).

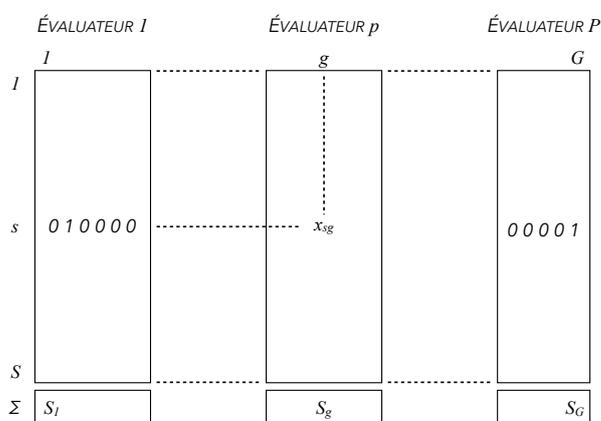


Figure 62: Tableau disjonctif complet dans le cas de données de tri libre : x_{sg} vaut 1 si le s -ème stimulus appartient au groupe g et 0 sinon. Le 1-er évaluateur a créé 6 groupes de stimuli et le P -ème évaluateur a créé 5 groupes de stimuli.

Les stimuli évoluent alors dans un espace à G dimensions, où G représente le nombre de groupes de stimuli formés sur la globalité du panel d'évaluateurs (c.-à-d. la somme des nombres de groupes formés par les évaluateurs). Dans cet espace, la distance entre deux stimuli s et s' est définie comme suit :

$$d_{ACM}^2(s, s') = \frac{S}{P} \sum_{g=1}^G \frac{S}{S_g} (x_{sg} - x_{s'g})^2 \quad (4.1)$$

où S représente le nombre total de stimuli évalués pendant la tâche de tri libre ; P représente le nombre total d'évaluateurs ayant participé à l'expérience ; G représente le nombre de groupes de stimuli formés sur la globalité du panel d'évaluateurs.

Cependant, par rapport aux données de tri libre, les données de Q-méthodologie binaire présentent une spécificité puisque les groupes de stimuli sont définis *a priori* par l'expérimentateur. Leur nombre est effectivement fixé (c.-à-d. 2). De la même façon, leur label sont fixés (c.-à-d. 'représentatif du concept' et 'non-représentatif du concept'). Le tableau disjonctif complet est alors présenté sous une forme légèrement différente de celle présentée précédemment (voir Figure 63).

	ÉVALUATEUR 1		ÉVALUATEUR p		ÉVALUATEUR P	
	1	0	1	0	1	0
1						
s	1	0	x_{sp}^1	x_{sp}^0	0	1
S						
Σ	S^1	S^0	S^1	S^0	S^1	S^0

Figure 63: Tableau disjonctif complet dans le cas de données de Q-méthodologie binaire : x_{sp}^1 vaut 1 et x_{sp}^0 vaut 0 (resp. x_{sp}^1 vaut 0 et x_{sp}^0 vaut 1) si le s -ème stimulus a été trié dans la catégorie 'représentatif du concept' (resp. dans la catégorie 'non-représentatif du concept') par le p -ème évaluateur.

Dans le contexte particulier de la Q-méthodologie binaire, la distance de l'ACM entre deux stimuli s et s' peut alors être redéfinie comme suit :

$$d_{ACM}^2(s, s') = \frac{S}{P} \sum_{p=1}^P \left[\frac{S}{S_p^1} (x_{sp}^1 - x_{s'p}^1)^2 + \frac{S}{S_p^0} (x_{sp}^0 - x_{s'p}^0)^2 \right] \quad (4.2)$$

où S représente le nombre total de stimuli évalués pendant la tâche de Q-méthodologie binaire ; P représente le nombre total d'évaluateurs ayant participé à l'expérience ; S_p^1 (resp. S_p^0) représente le nombre de stimuli triés dans la catégorie 'représentatif du concept' (resp. 'non-représentatif du concept') par le p -ème évaluateur ; $x_{sp}^1 = 1$ et $x_{sp}^0 = 0$ (resp. $x_{sp}^1 = 0$ et $x_{sp}^0 = 1$) si le s -ème stimulus a été trié dans la catégorie 'représentatif du concept' (resp. 'non-représentatif du concept') par le p -ème évaluateur.

Cette distance inter-stimuli de l'ACM est d'autant plus faible que les situations dans lesquelles ($x_{sp}^1 = 1$ et $x_{s'p}^1 = 1$) ou ($x_{sp}^0 = 1$ et $x_{s'p}^0 = 1$) sont observées un grand nombre de fois. En d'autres termes, elle est d'autant plus faible que

les deux stimuli ont été placés un grand nombre de fois dans un même groupe sur la globalité du panel. Ce comportement est en concordance avec l'esprit du tri libre. Dans ce contexte, de telles situations témoignent effectivement d'une forte similitude des deux stimuli. Cependant, ce comportement n'est pas optimal dans un contexte de Q-méthodologie binaire. En effet, cette distance traite de manière totalement confondue et indifférenciée deux types de situations : celles dans lesquelles les deux stimuli ont été regroupés dans la catégorie 'représentatif du concept' ($x_{sp}^1 = 1$ et $x_{s'p}^1 = 1$), et celles dans lesquelles ils ont été regroupés dans la catégorie 'non-représentatif du concept' ($x_{sp}^0 = 1$ et $x_{s'p}^0 = 1$). Cet aspect s'avère particulièrement limitant. Rappelons en effet qu'il s'agit ici d'obtenir une distance reflétant les différences entre les stimuli, mais ceci relativement au concept étudié.

Pour illustrer cette limite des techniques statistiques classiques, trois expériences fictives pédagogiques sont définies. Celles-ci représentent des ensembles de configurations de 8 stimuli fournies par 20 évaluateurs dans un contexte de Q-méthodologie binaire (voir Figure 64). Dans l'expérience 1, le panel sépare clairement les stimuli A, B, C et D d'une part ; et les stimuli E, F, G et H, d'autre part. Cependant, la première moitié des évaluateurs pense que A, B, C et D ne sont pas représentatifs du concept et que E, F, G et H sont représentatifs du concept ; tandis que le reste des évaluateurs pense le contraire. En ce sens, la perception globale des stimuli en termes de représentation du concept pourrait être qualifiée d'ambiguë : elle est négative la première moitié du temps et positive l'autre moitié du temps. Dans l'expérience 2, le panel sépare là aussi clairement A, B, C et D d'une part ; et E, F, G et H, d'autre part. Cependant, trois quarts des évaluateurs pensent que A, B, C et D sont représentatifs du concept et que E, F, G et H sont non-représentatifs du concept ; tandis que le reste des évaluateurs pense le contraire. En ce sens, la perception globale des stimuli A, B, C et D en termes de représentation du concept pourrait être qualifiée comme plutôt positive ; alors que celle des stimuli E, F, G et H pourrait être qualifiée comme plutôt négative. Enfin, dans l'expérience 3, le panel ne sépare pas clairement A, B, C et D de E, F, G, et H. Néanmoins, il semble que la perception globale des stimuli A, B, C et D en termes de représentation du concept soit plutôt positive ; alors que celle des stimuli E, F, G et H soit plutôt négative.

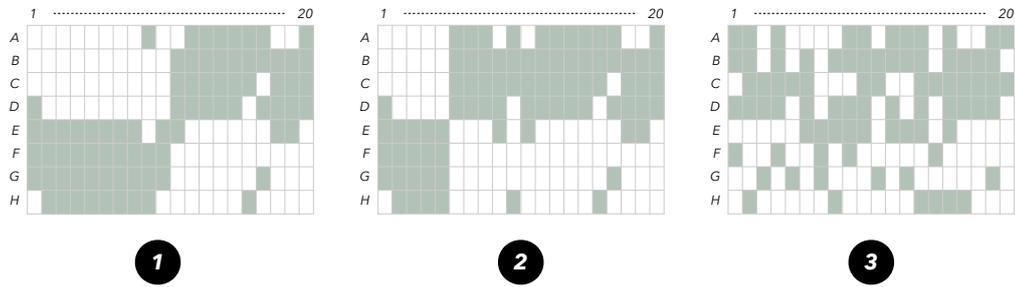


Figure 64: Expériences fictives pédagogiques. Une case colorée correspond à une évaluation binaire positive du stimulus par l'évaluateur (c.-à-d. 'représentatif du concept') et une case blanche correspond à une évaluation binaire négative du stimulus par l'évaluateur (c.-à-d. 'non-représentatif du concept').

La Figure 65 présente les représentations euclidiennes correspondantes. Elles sont basées sur les distances inter-stimuli fournies par les trois ACM appliquées aux données brutes. Il apparaît que ces représentations sont identiques pour l'expérience 1 et pour l'expérience 2. Ceci appuie bien le fait que la distance inter-stimuli de l'ACM fournit un point de vue incomplet des données. En effet, bien que ces expériences reflètent des réalités différentes, cette dernière fournit des représentations euclidiennes similaires d'une expérience à l'autre.

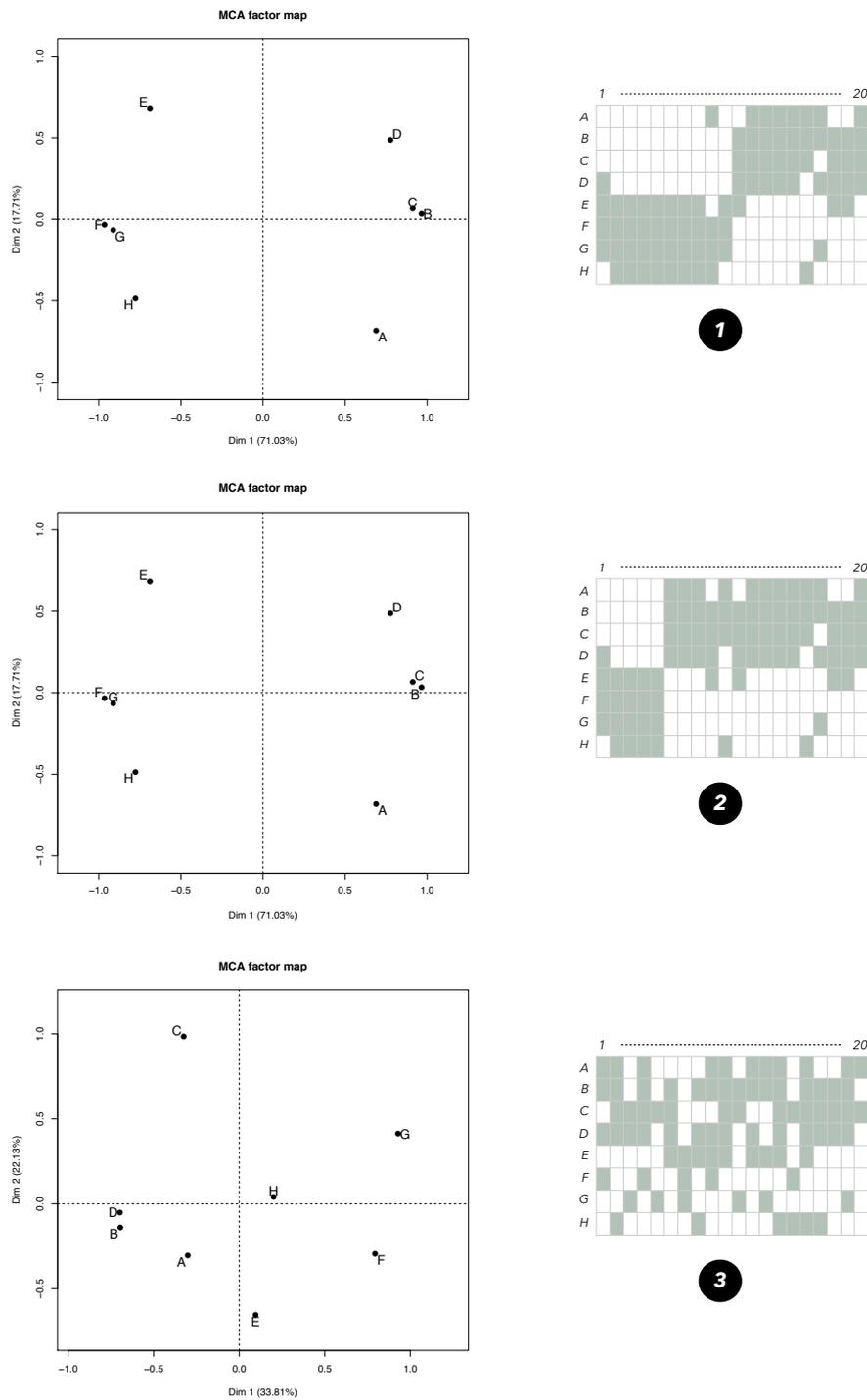


Figure 65: Expériences fictives pédagogiques : Représentations des stimuli fournies par l'Analyse des Correspondances Multiples appliquée sur les données brutes.

Les procédures statistiques dédiées au traitement des données de tri libre ne sont donc pas adaptées pour analyser les données de Q-méthodologie binaire. Elles fournissent en effet une distance inter-stimuli qui ne prend en compte qu'une partie de l'information véhiculée par les données. Plus précisément, elles prennent uniquement en compte la manière dont ont été partitionnés les stimuli, et occultent totalement la manière avec laquelle ces stimuli ont été associés aux deux labels prédéfinies représentant le concept et sa négation. Cette information est pourtant primordiale dans un contexte de Q-méthodologie binaire.

2.1.2 Distance prenant en compte la seconde spécificité des données : focus sur l'Analyse Factorielle des Correspondances

Nous nous attachons ici à proposer une procédure qui fournit une distance inter-stimuli reflétant le degré avec lequel les stimuli représentent le concept évalué. Nous verrons plus tard comment il est possible de combiner cette distance avec celle de l'ACM, dans l'objectif d'obtenir une distance globale qui restitue les deux informations sur les stimuli véhiculées par les données.

Nous nous concentrons ici uniquement sur la manière avec laquelle un évaluateur a associé les stimuli aux deux labels (c.-à-d. catégories) prédéfinis lors d'une tâche de Q-méthodologie binaire. Cela revient à assimiler la configuration sur les stimuli qu'il a fournie à une configuration issue d'une tâche d'association. De manière générale, une telle tâche d'association consiste à demander à l'évaluateur d'associer chaque stimulus aux labels qui lui correspondent le plus (voir Figure 66).

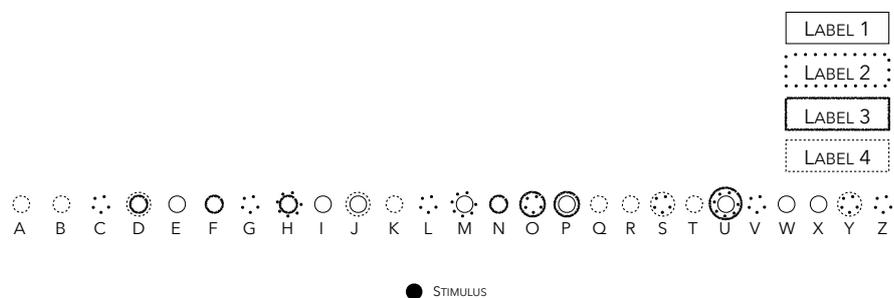


Figure 66: Exemple d'une tâche d'association à quatre labels réalisée par un évaluateur sur 26 stimuli. L'évaluateur a associé certains stimuli à plus de un label, comme cela est autorisé par la tâche.

La procédure statistique usuelle permettant de comprendre comment ont été perçus les stimuli par les évaluateurs consiste à agréger les données dans un tableau de comptage. Celui-ci rassemble le nombre d'associations observées sur la globalité du panel pour chaque couple stimulus-label (voir Figure 67).

	l	l	L
l			
s	4 x_{sl}^+	30
s			
Σ	$S_{\cdot l}$	$S_{\cdot l}$	$S_{\cdot L}$

Figure 67: Tableau de comptage dans le cas de données d'association : x_{sl}^+ est égal au nombre de fois où le s -ème stimulus a été associé avec le l -ème label sur la globalité du panel. Ici, le stimulus s a été associé au 1-er label par 4 évaluateurs.

Ceci suggère d'analyser ces données ayant en lignes les stimuli et en colonnes les différents labels par Analyse Factorielle des Correspondances (AFC), une méthode dédiée à l'analyse de matrices de données de comptage (Benzécri, 1992). En AFC, les stimuli évoluent dans un espace à L dimensions, où L représente le nombre de labels prédéfinis utilisés lors de la tâche. Dans cet espace, la distance entre deux stimuli s et s' est définie comme suit :

$$d_{AFC}^2(s, s') = \frac{S}{P} \sum_{l=1}^L \frac{1}{S_l^+} (x_{sl}^+ - x_{s'l}^+)^2 \quad (4.3)$$

où S représente le nombre total de stimuli évalués pendant la tâche d'association ; P représente le nombre total d'évaluateurs ayant participé à l'expérience ; L représente le nombre de labels prédéfinis par l'expérimentateur pour la tâche d'association ; S_l^+ représente le nombre de fois où le l -ème label a été associé avec un stimulus sur la globalité du panel ; x_{sl}^+ représente le nombre de fois où le s -ème stimulus a été associé avec le l -ème label sur la globalité du panel.

Cependant, par rapport aux données d'association, les données de Q-méthodologie binaire présentent une spécificité puisque les différents labels prédéfinis sont au nombre de deux. Le tableau de comptage ne contient donc que deux colonnes. Deuxièmement, chaque stimulus ne peut être associé qu'à un seul label : soit le label positif (c.-à-d. 'représentatif du concept'), soit le label négatif (c.-à-d. 'non-représentatif du concept'). Ainsi, les marges lignes du tableau de comptage sont toutes égales au nombre d'évaluateurs ayant réalisé la tâche. Le tableau de comptage est alors présenté sous une forme légèrement différente de celle présentée précédemment (voir Figure 68).

	1	0	Σ
I	⋮	⋮	
s	x_s^{1+}	x_s^{0+}	P
S			
Σ	S^{1+}	S^{0+}	

Figure 68: Tableau de comptage dans le cas de données de Q-méthodologie binaire : x_s^{1+} (resp. x_s^{0+}) est égal au nombre de fois où le s -ème stimulus a été associé à la catégorie ‘représentatif du concept’ (resp. ‘non-représentatif du concept’) sur la globalité du panel.

Dans le contexte particulier de la Q-méthodologie binaire, la distance de l’AFC entre deux stimuli s et s' peut par conséquent être redéfinie comme suit :

$$d_{AFC}^2(s, s') = \frac{S}{P} \left[\frac{1}{S^{1+}} (x_s^{1+} - x_{s'}^{1+})^2 + \frac{1}{S^{0+}} (x_s^{0+} - x_{s'}^{0+})^2 \right] \quad (4.4)$$

où S représente le nombre total de stimuli évalués pendant la tâche de Q-méthodologie binaire ; P représente le nombre total d’évaluateurs ayant participé à l’expérience ; S^{1+} (resp. S^{0+}) représente le nombre de stimuli triés dans la catégorie ‘représentatif du concept’ (resp. ‘non-représentatif du concept’) sur la globalité du panel ; x_s^{1+} (resp. x_s^{0+}) est égal au nombre de fois où le s -ème stimulus a été associé à la catégorie ‘représentatif du concept’ (resp. ‘non-représentatif du concept’) sur la globalité du panel.

Cette distance inter-stimuli de l’AFC est d’autant plus faible que les valeurs x_s^{1+} et $x_{s'}^{1+}$ sont similaires. En d’autres termes, elle est d’autant plus faible que les deux stimuli ont été associés un même nombre de fois à la catégorie ‘représentatif du concept’ sur la globalité du panel. Cette situation témoigne en effet d’une certaine similitude entre des deux stimuli concernant le degré avec lequel ils représentent le concept évalué. Cette distance fournie par l’AFC apparaît alors comme une solution pour prendre en compte la spécificité des données occultée par les techniques classiques.

Pour confirmer cet aspect, une AFC est réalisée sur les tableaux de comptage construits à partir des trois jeux de données fictifs pédagogiques. Rappelons que les expériences 2 et 3 présentent des similitudes quant à la manière avec laquelle les stimuli ont été associés aux deux labels prédéfinies représentant le concept et sa négation. Les représentations des stimuli correspondantes sont présentées sur la Figure

69. Il apparaît qu'elles sont identiques pour l'expérience 2 et pour l'expérience 3. Ceci appuie bien le fait que la distance inter-stimuli de l'AFC permet de prendre en compte cette seconde spécificité des données.

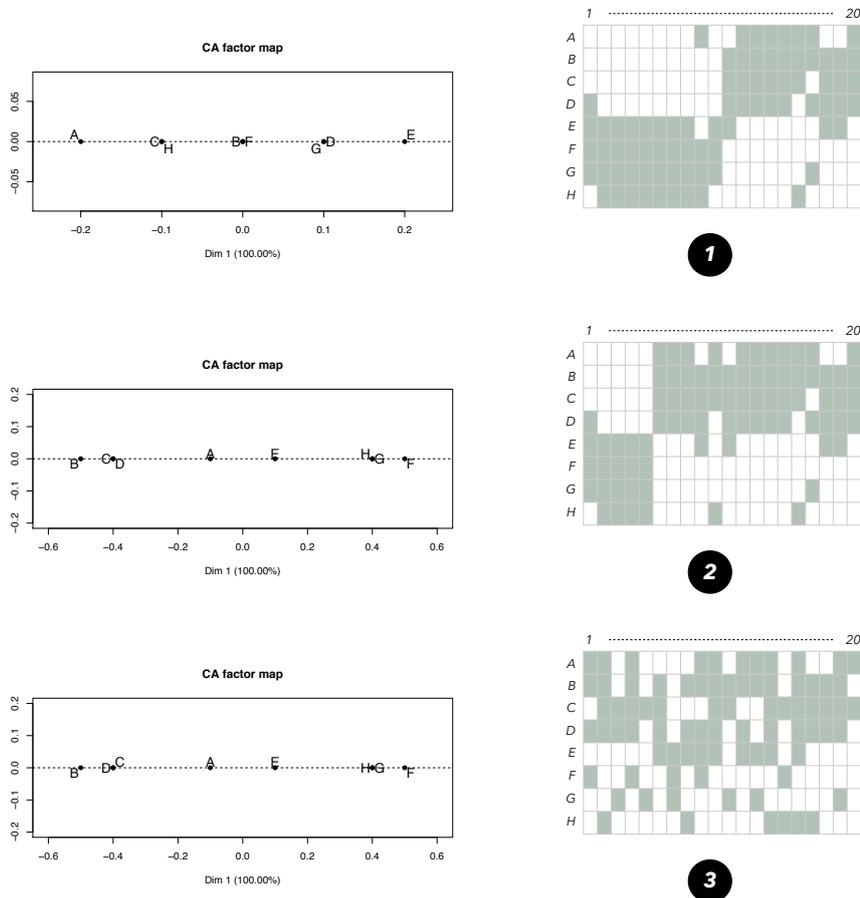


Figure 69: Expériences fictives pédagogiques : Représentations des stimuli fournies par l'Analyse Factorielle des Correspondances réalisée sur le tableau de comptage dérivé des données brutes.

2.1.3 Proposition d'une distance globale *ad hoc* : focus sur l'Analyse Factorielle Multiple

Dans les sections précédentes, nous avons introduit deux distances inter-stimuli. Dans un premier temps, nous avons présenté celle de l'ACM lorsque cette procédure factorielle est appliquée sur les données brutes de Q-méthodologie binaire. Nous avons vu qu'elle permet de prendre en compte la première spécificité des données (c.-à-d. la manière dont ont été partitionnés les stimuli par les évaluateurs). Dans un second temps, nous avons présenté celle de l'AFC lorsque cette procédure factorielle est réalisée sur un tableau de comptage particulier déduit des données.

Nous avons vu qu'elle permet de prendre en compte la seconde spécificité des données (c.-à-d. la manière avec laquelle les stimuli ont été associés par les évaluateurs aux deux catégories prédéfinies représentant le concept et sa négation). Nous proposons maintenant de combiner ces deux distances à travers une distance globale. La vocation d'une telle distance est de prendre en compte l'intégralité des informations véhiculées sur les stimuli par les données.

Les deux tableaux de données abordés ici peuvent être concaténés sous la forme d'un tableau multiple (voir Figure 70).

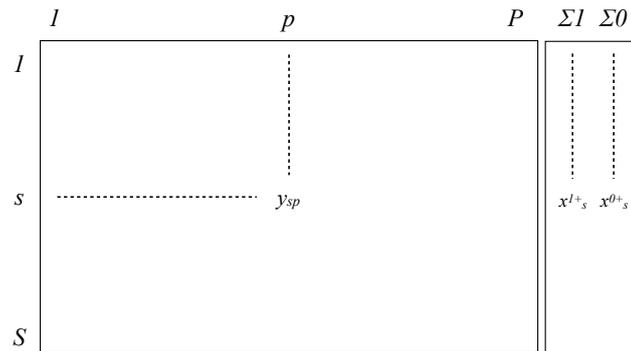


Figure 70: Tableau multiple prenant en compte les deux spécificités des données de Q-méthodologie binaire.

Certaines techniques statistiques sont spécifiquement dédiées à l'analyse de tels tableaux. En pratique, elles permettent de prendre en compte des ensembles de données de natures différentes à travers le calcul d'une distance entre deux individus statistiques globale. L'Analyse Procustéenne Généralisée (Gower, 1975), le modèle Inscal (Carroll & Chang, 1970) et l'Analyse Factorielle Multiple (AFM ; (Pagès, 2014)) font partie de ces approches.

Le tableau multiple obtenu ici présente la particularité de contenir des données mixtes. Il contient des données catégorielles d'une part, et des données de comptage d'autre part. À notre connaissance, l'AFM est la seule approche d'analyse de tableaux multiples qui permet de traiter ce type de données. Une extension de l'AFM a en effet été développée dans un contexte de traitement de données d'enquête pour aborder cette problématique. Les données d'enquête sont effectivement souvent composées de données catégorielles correspondant à des réponses à des questions fermées, et de données de comptage correspondant à des réponses à des questions ouvertes (Bécue-Bertaut & Pagès, 2005, 2007).

La distance inter-stimuli de l'AFM repose sur une combinaison linéaire des distances obtenues séparément à partir des différents groupes de variables. Dans le cas des données de Q-méthodologie binaire, ces dernières correspondent à la distance de l'ACM d'une part (notée $d_{ACM}^2(s, s')$) et à la distance de l'AFC d'autre

part ($d_{AFC}^2(s, s')$). La distance inter-stimuli de l'AFM est alors définie de la manière suivante :

$$d_{AFM}^2(s, s') = a.d_{ACM}^2(s, s') + b.d_{AFC}^2(s, s') \quad (4.5)$$

Les coefficients a et b peuvent être aussi bien vus comme des poids de variables (identiques au sein d'un même groupe de variables) que comme des coefficients d'homothétie des nuages de stimuli. Ils sont choisis de façon à harmoniser les contributions des deux distances inter-stimuli (celle de l'ACM et celle de l'AFC). En AFM, ces coefficients a et b sont tels que l'inertie axiale maximum de chaque nuage des stimuli est égale à 1. Pour obtenir cette propriété, en notant λ_1^{ACM} (resp. λ_1^{AFC}) la première valeur propre de l'ACM (resp. de l'AFC), on choisit les coefficients qui conduisent à la distance inter-stimuli globale suivante :

$$d_{AFM}^2(s, s') = \frac{1}{\lambda_1^{ACM}} \cdot d_{ACM}^2(s, s') + \frac{1}{\lambda_1^{AFC}} \cdot d_{AFC}^2(s, s') \quad (4.6)$$

$$d_{AFM}^2(s, s') = \frac{1}{\lambda_1^{ACM}} \cdot \frac{S}{P} \sum_{p=1}^P \left[\frac{S}{S_p^1} (x_{sp}^1 - x_{s'p}^1)^2 + \frac{S}{S_p^0} (x_{sp}^0 - x_{s'p}^0)^2 \right] + \quad (4.7)$$

$$\frac{1}{\lambda_1^{AFC}} \cdot \frac{S}{P} \left[\frac{1}{S^{1+}} (x_s^{1+} - x_{s'}^{1+})^2 + \frac{1}{S^{0+}} (x_s^{0+} - x_{s'}^{0+})^2 \right] \quad (4.8)$$

Cette distance inter-stimuli globale est d'autant plus faible que les deux stimuli s et s' ont été (1) placés dans le même groupe par la majorité des évaluateurs et (2) associés à la catégorie 'représentatif du concept' un même nombre de fois.

Comme précisé précédemment, le choix des coefficients $a = \frac{1}{\lambda_1^{ACM}}$ et $b = \frac{1}{\lambda_1^{AFC}}$ permet de fournir un premier axe de l'AFM qui s'appuie harmonieusement sur les deux types de variables. Ceci est en concordance avec le point de vue selon lequel les deux spécificités des données de Q-méthodologie binaire fournissent des informations de même importance.

En revanche, il est généralement impossible en AFM de prévoir l'impact que les deux types de variables auront sur la construction des axes suivants. C'est le groupe présentant la plus faible décroissance des valeurs propres qui aura une influence prépondérante dès le deuxième axe. Dans un contexte de Q-méthodologie binaire, on est pourtant en mesure d'identifier quel est le groupe de variables qui va être prépondérant dès le deuxième axe. En effet, l'AFC présente une très forte décroissance des valeurs propres. Ceci est dû au fait qu'elle ne fournit en réalité qu'un seul 'vrai' axe (voir Figure 69). Le groupe de variables catégorielles présente ainsi nécessairement une faible décroissance des valeurs propres relativement à ce groupe de variables de comptage. C'est donc lui qui est prépondérant dès le

deuxième axe de l'AFM. Ce deuxième axe a ainsi tendance à éloigner des stimuli qui n'ont pas été mis dans le même groupe par la majorité des évaluateurs du panel. Prenons l'exemple de deux stimuli qui sont proches sur le premier axe de l'AFM mais éloignés sur le deuxième. Leur proximité sur le premier axe témoigne d'une similarité de représentation conceptuelle. Leur éloignement sur le deuxième axe témoigne, lui, du fait qu'ils n'ont pas été mis dans le même groupe par la majorité des évaluateurs du panel. Une telle situation peut par exemple apparaître si une partie des évaluateurs a mis l'un des stimuli dans la catégorie 'représentatif du concept' et l'autre dans la catégorie 'non-représentatif du concept', et que l'autre partie a fourni une configuration opposée. L'interprétation du deuxième axe de l'AFM permet ainsi de fournir une première vision du désaccord qu'il existe éventuellement au sein du panel.

Pour s'assurer que cette distance inter-stimuli est adaptée à la nature des données de Q-méthodologie binaire, une AFM est réalisée sur les trois tableaux multiples obtenus à partir des expériences fictives pédagogiques (voir Figure 71). Les représentations des stimuli correspondantes sont présentées sur la Figure 72.

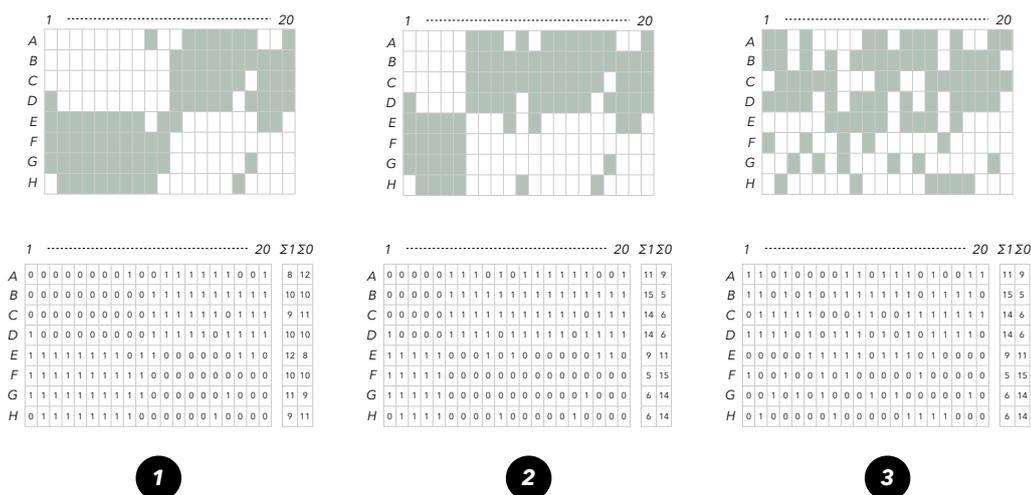


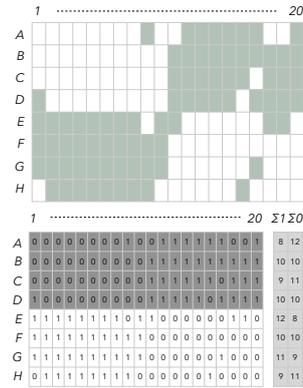
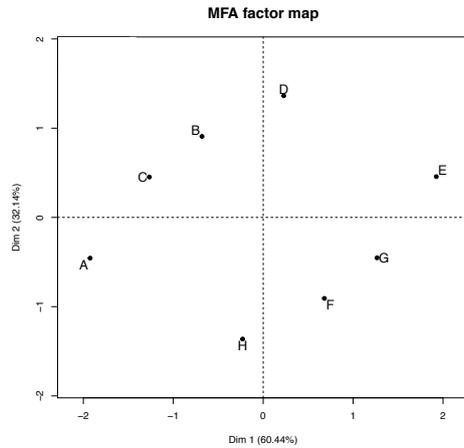
Figure 71: Expériences fictives pédagogiques : Tableaux multiples.

Concernant l'expérience 1, l'ACM appliquée aux données catégorielles fournissait une structure claire des stimuli en deux groupes distincts : les stimuli A, B, C, D d'une part ; et les stimuli E, F, G, H d'autre part (voir Figure 65). Ce point de vue apparaissait pourtant comme étant optimiste. Il ne prenait pas en compte le fait que les stimuli présentaient des degrés de représentation du concept similaires. Tous les stimuli représentaient en effet le concept de manière neutre (c.-à-d. ni vraiment représentatifs, ni vraiment non-représentatifs). L'AFM semble prendre en compte cette deuxième information. Elle fournit une représentation sur laquelle l'opposition entre les deux groupes de stimuli semble atténuée, et

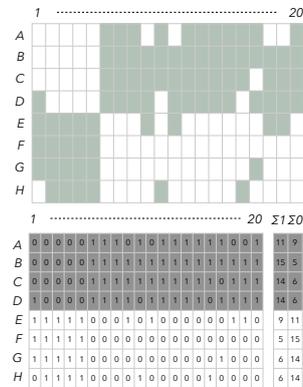
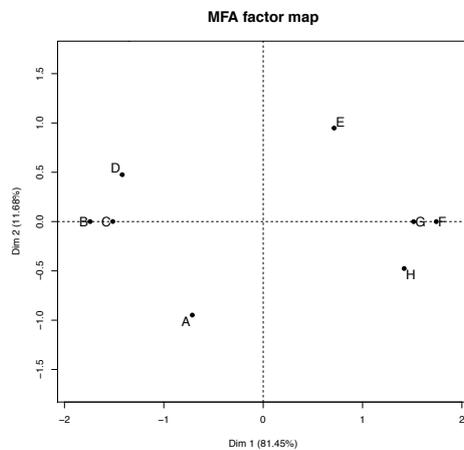
sur laquelle l'inertie portée par le premier axe est moins importante (voir Figure 72).

Concernant l'expérience 3, l'ACM fournissait une structure confuse des stimuli (voir Figure 65). Ceci témoignait d'une certaine absence de structure au sein de l'ensemble de stimuli. Ce point de vue apparaissait pourtant comme étant pessimiste. Il ne prenait pas en compte le fait que les stimuli présentaient une certaine structure de variabilité en termes de degré de représentation du concept. Les stimuli A, B, C, D étaient en effet plutôt représentatifs du concepts ; et les stimuli E, F, G, H étaient plutôt non-représentatifs. L'AFM semble prendre en compte cette deuxième information. Elle fournit une représentation sur laquelle une opposition entre les deux groupes de stimuli semble apparaître, et sur laquelle l'inertie portée par le premier axe est plus importante (voir Figure 72).

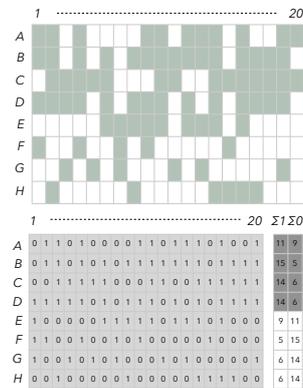
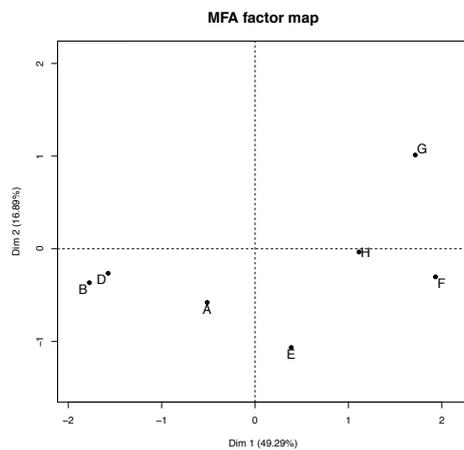
Enfin, concernant l'expérience 2, l'ACM fournissait une structure claire des stimuli en deux groupes distincts : les stimuli A, B, C, D d'une part ; et les stimuli E, F, G, H d'autre part (voir Figure 65). Ce point de vue apparaissait comme étant proche de la réalité. En effet, cette structure en deux groupes de stimuli existait aussi en termes de degré de représentation du concept : les stimuli A, B, C, D étaient plutôt représentatifs du concepts ; et les stimuli E, F, G, H étaient plutôt non-représentatifs. L'AFM semble prendre en compte cette deuxième information. Elle fournit une représentation sur laquelle l'opposition entre les deux groupes de stimuli semble accentuée, et sur laquelle l'inertie portée par le premier axe est plus importante (voir Figure 72).



1



2



3

Figure 72: Expériences fictives pédagogiques : Représentations des stimuli fournies par l'Analyse Factorielle Multiple appliquée sur les tableaux de données multiples. Des codes couleurs sont utilisés pour mettre en évidence la variabilité des stimuli véhiculée par le tableau de données catégorielles d'une part et le tableau de données de comptage d'autre part. Le code blanc/gris foncé indique la présence de variabilité alors que le code gris clair indique l'absence de variabilité.

Dans cette section, l'AFM a été introduite comme une procédure adaptée à l'analyse orientée stimuli des données de Q-méthodologie binaire. Elle fournit une distance inter-stimuli basée sur la combinaison de deux distances prenant chacune en compte une spécificité des données. L'AFM permet ainsi d'analyser la structure de variabilité conceptuelle observée au sein de l'ensemble de stimuli tout en prenant en compte la nature des données collectées. Le caractère opérationnel de l'AFM a été démontré sur des jeux de données fictifs pédagogiques.

En outre, l'analyse des données de Q-méthodologie binaire par AFM présente d'autres avantages. Cette procédure fournit en effet des représentations euclidiennes complémentaires. Celles-ci facilitent la compréhension des distances inter-stimuli observées. Plus précisément, l'AFM fournit une représentation des variables du tableau de comptage. Il s'agit ici d'une représentation des deux catégories : 'représentatif du concept' et 'non-représentatif du concept'. En AFM, cette représentation des deux catégories prédéfinies est liée à celle des stimuli par des propriétés barycentriques. Pour cette raison, il est possible de les superposer. Sur la représentation superposée obtenue, une catégorie est attirée (resp. repoussée) par les stimuli avec lesquels elle a été la plus (resp. la moins) associée. Celle-ci permet alors de distinguer aisément les stimuli représentatifs du concept des autres.

Ces représentations superposées obtenues à partir des trois expériences fictives pédagogiques sont présentées sur la Figure 73. Concernant l'expérience 1, la représentation permet de conclure qu'aucune tendance sur les stimuli ne se dessine concernant leur représentation du concept. Malgré cela, il est possible de voir que deux groupes de stimuli (A, B, C, D et E, F, G, H) ont majoritairement été séparés par les évaluateurs. Concernant l'expérience 2, il est possible de déterminer qu'un premier groupe de stimuli (A, B, C, D) a été considéré comme plutôt représentatif du concept alors qu'un deuxième groupe (E, F, G, H) a plutôt été considéré comme non-représentatif. Enfin, concernant l'expérience 3, la représentation superposée indique qu'un premier groupe de stimuli (A, B, C, D) a été considéré comme plutôt représentatif du concept alors qu'un deuxième groupe (E, F, G, H) a plutôt été considéré comme non-représentatif. Il est par ailleurs possible de voir qu'il n'existe pas de partition nette des stimuli en deux groupes distincts.

Il apparaît que cette représentation superposée des stimuli et des catégories prédéfinies permet de comprendre avec plus de finesse les distances inter-stimuli obtenues. Malgré cela, les possibilités d'interprétation qui lui sont associées peuvent s'avérer limitées. Dans le cas de l'expérience 3 par exemple, la comparaison des stimuli A et C peut être délicate. Il est difficile de déterminer lequel des deux est le plus représentatif du concept.

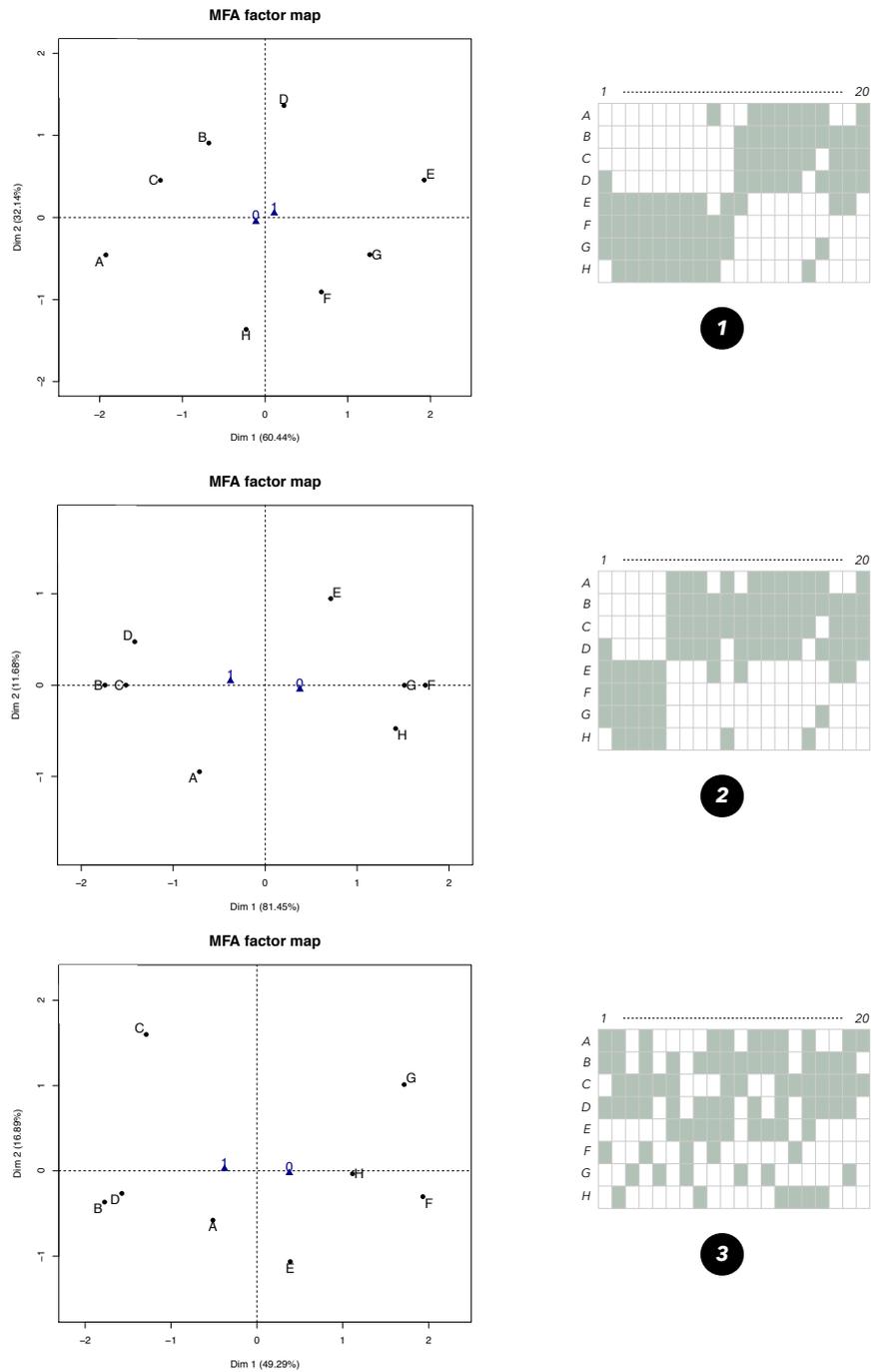


Figure 73: Expériences fictives pédagogiques : Représentations des stimuli et des deux catégories prédéfinies fournies par l'Analyse Factorielle Multiple appliquée sur les tableaux de données multiples. Le label 1 correspond à la catégorie 'représentatif du concept' et le label 0 correspond à la catégorie 'non-représentatif du concept'.

2.2 Amélioration de la visualisation de la représentation euclidienne des stimuli

Nous introduisons ici une procédure originale de visualisation associée à la représentation euclidienne des stimuli. Elle permet de faciliter la compréhension des distances inter-stimuli observées. Contrairement à la représentation superposée des stimuli et des catégories présentée dans la section précédente, elle est plus élégante et plus facile à interpréter. Basée sur la projection de zones colorées, elle permet de détecter rapidement les zones du plan factoriel de l'AFM plutôt associées à une représentation du concept et celles plutôt associées à une non-représentation du concept.

Cette procédure de visualisation de données repose sur les grandes étapes suivantes (voir Figure 74 pour une version schématique) :

1. Le plan factoriel de l'AFM est discrétisé. Ceci permet d'obtenir les coordonnées de centaines de points. Ces derniers sont assimilés par la suite à des stimuli fictifs.
2. Pour chaque évaluateur du panel, un modèle logistique est construit. Il permet de modéliser la probabilité $\mathbb{P}(Y_{sp} = 1 | X = (F_1(s), F_2(s)))$ qu'un stimulus réel s reçoive une évaluation positive (c.-à-d. 'représentatif du concept') de la part de l'évaluateur p , et ce conditionnellement aux coordonnées de ce stimulus.

$$\text{logit} [\mathbb{P}(Y_{sp} = 1 | X = (F_1(s), F_2(s)))] = \beta_0 + \beta_1 F_1(s) + \beta_2 F_2(s) + \beta_3 F_1(s) F_2(s) \quad (4.9)$$

où $F_1(s)$ (resp. $F_2(s)$) représente la coordonnée du s -ème stimulus sur le premier axe de l'AFM d'intérêt (resp. sur le deuxième axe de l'AFM d'intérêt) ;

3. Pour chaque évaluateur du panel, l'évaluation de chaque stimulus fictif est prédite grâce au modèle logistique individuel (4.9).
4. Au niveau du panel, le pourcentage de fois où chaque stimulus fictif a reçu une évaluation prédite positive est calculé. Une surface de réponse correspondant à cette valeur en fonction des coordonnées du stimulus fictif est créée. Cette surface de réponse est ensuite projetée sur le plan factoriel de l'AFM et des zones de niveaux sont définies.

Les représentations des stimuli obtenues grâce à cette procédure de visualisation dans le cadre des trois expériences fictives pédagogiques sont présentées en Figure 75. Celles-ci sont plus faciles à interpréter que les représentations classiques présentées en Figure 73. Elles permettent en effet de relever certaines subtilités sur la manière dont été évalués les stimuli en termes de représentation de concept. Si nous reprenons la représentation correspondant à l'expérience 3, il est maintenant possible de comparer les stimuli A et C en termes de représentation du concept.

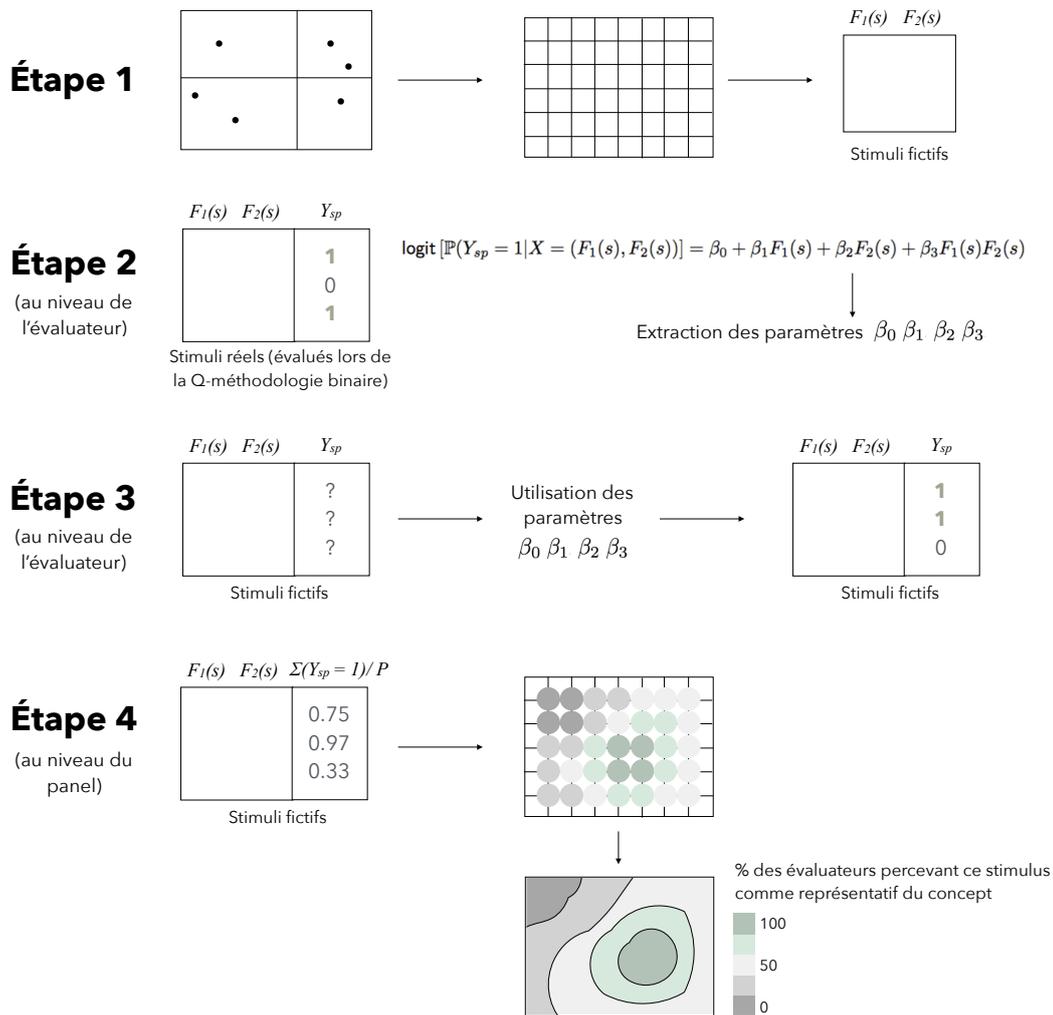


Figure 74: Procédure de visualisation facilitant l'interprétation de la représentation euclidienne des stimuli.

L'approche présentée ici consiste à modéliser comment chaque évaluateur associe les stimuli au concept en fonction des dimensions de variabilité fournies par l'AFM. Puisque l'AFM prend déjà en compte cette information dans le calcul de la distance inter-stimuli, cette approche peut paraître discutable. Cela ne pose pourtant pas de problème particulier dès lors qu'il s'agit ici de proposer un outil de visualisation. Celui-ci ne doit être utilisé que pour faciliter l'interprétation et la communication des résultats.

Cette représentation est similaire à certaines approches développées en sensométrie. C'est le cas des représentations de type cartographie des préférences (*preference mapping*) telles que proposées par Danzart, Sieffermann, and Delarue (2004).

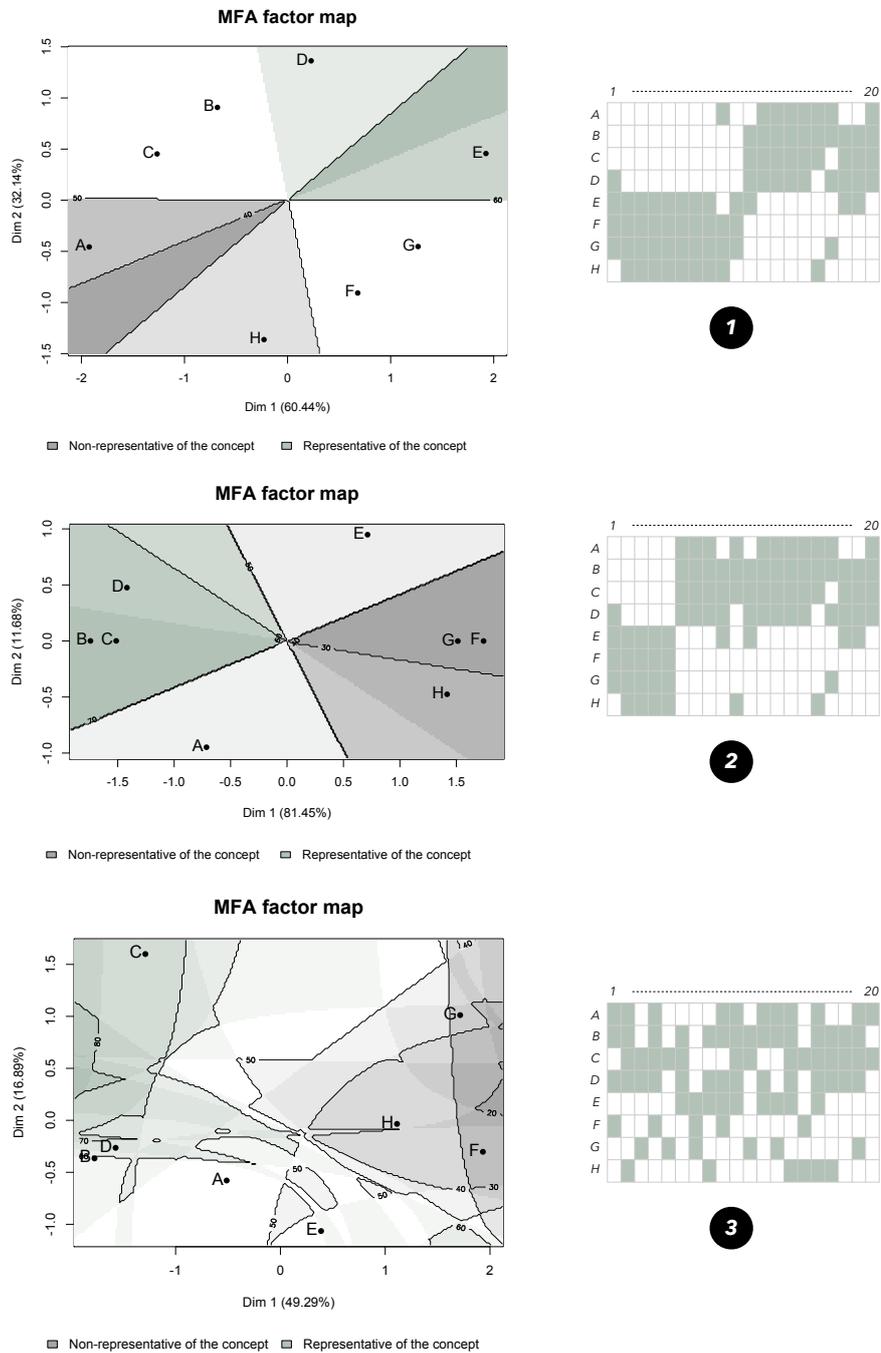


Figure 75: Expériences fictives pédagogiques : Représentations des stimuli obtenues avec la procédure de visualisation de données.

Celles-ci sont basées sur une mise en relation de la représentation des stimuli avec des données continues externes (c.-à-d. données non prises en compte lors du calcul

de la distance inter-stimuli). Ces dernières correspondent à des données hédoniques (c.-à-d. d'appréciation) obtenues grâce à une échelle graduée. Ces représentations sont utilisées par les analystes sensoriels pour extraire de nouvelles connaissances sur l'espace produit. Elles permettent notamment d'identifier les caractéristiques sensorielles qui entraînent une appréciation élevée des stimuli. L'approche que nous proposons ici apporte un regard nouveau sur ce type de représentations graphiques. Nous démontrons par exemple qu'elles pourraient être utilisées dans un autre cadre. En liant la représentation des stimuli à des données internes (c.-à-d. données déjà prises en compte lors du calcul de la distance inter-stimuli), elles pourraient être utilisées comme des outils de visualisation. En outre, nous démontrons qu'il est possible de remplacer les données hédoniques externes par des données binaires externes. En analyse sensorielle et sciences du consommateur, l'obtention de telles données binaires est fréquente. Elles peuvent par exemple correspondre à des réponses de type oui/non à une question d'intention d'achat. Grâce à l'approche que nous proposons, il est désormais possible d'identifier les caractéristiques sensorielles qui entraînent une importante intention d'achat.

3 Illustration sur des données réelles de caractérisation de produits industriels : données 'concept d'innovation et parfums de shampoings'

Dans l'objectif d'illustrer l'intérêt de la procédure d'analyse statistique introduite précédemment, nous présentons ici les résultats de son application sur des données réelles de Q-méthodologie binaire. Ces données sont issues de la caractérisation du concept 'd'innovation' au sein d'un grand ensemble de parfums de shampoings.

Lors de cette étude, 39 parfums de shampoings ont été sélectionnés pour constituer l'espace produit à caractériser. Cette sélection a été mise en place par l'équipe développement de l'entreprise Givaudan. Ils ont été choisis de manière à ce qu'ils soient aussi variables que possible. Les 39 parfums ont été insérés dans des petites bouteilles blanches en plastique opaque. Une simple pression du doigt sur la bouteille permettait de libérer le parfum. Chacune des bouteilles ainsi obtenues a été labellisée grâce à un code aléatoire à trois chiffres. Un panel de 120 consommateurs a été recruté pour participer à la tâche de caractérisation conceptuelle de ces produits. Ce panel de consommateurs était majoritairement composé de femmes (72.5 %) et de jeunes personnes (âge moyen = 21.24 ans ; amplitude = 17-26 ans). Les consommateurs n'ont été soumis à aucun entraînement sensoriel avant de participer à cette expérience. La session de caractérisation conceptuelle consistait à demander aux consommateurs d'évaluer le concept d'innovation au sein de l'ensemble de stimuli.

3.1 Données

Les données recueillies ont été mises en forme dans un jeu de données binaires de dimensions 39×120 présentant en lignes les parfums et en colonnes les consommateurs. L'intersection de la ligne s et de la colonne p contenait une valeur binaire y_{sp} égale à 1 si le parfum avait été jugé innovant par le consommateur (évaluation positive) et à 0 sinon (évaluation négative). À ces données de Q-méthodologie binaire s'ajoutaient 27 covariables véhiculant des informations sur les parfums de shampoings. Il s'agissait des 27 caractéristiques sensorielles les plus utilisées par les consommateurs pour décrire les parfums de shampoings de manière spontanée. Chaque valeur de l'une de ces covariables correspondait au nombre de fois où la caractéristique avait été utilisée pour décrire le parfum sur la globalité du panel. Enfin, les données étaient complétées de 2 covariables véhiculant des informations sur les consommateurs. Il s'agissait de leur genre et de leur tranche d'âge. Ces covariables sur les stimuli et les évaluateurs ne seront pas utilisées ici mais seront valorisées dans les chapitres suivants de cette Partie 2.

3.2 Résultats de l'analyse statistique orientée parfums

Les deux premières dimensions de l'AFM expriment 41.5% de l'inertie totale des données (voir Table 19). Exprimant à elle seule 34.6% de l'inertie totale, la première dimension est nettement prédominante. Par la suite, les interprétations de la représentation multidimensionnelle des parfums se limiteront à ces deux premières dimensions.

Table 19: Étude 'concept d'innovation et parfums de shampoings' :
Extrait des valeurs propres de l'Analyse Factorielle Multiple dans le cadre de la procédure statistique orientée stimuli.

	Valeurs propres	Pourcentage de variance cumulé (%)
Dim 1	1.98	34.6
Dim 2	0.40	41.5
Dim 3	0.24	45.7
Dim 4	0.22	49.5
Dim 5	0.19	52.7
Dim 6	0.17	55.7
Dim 7	0.16	58.6
Dim 8	0.16	61.3
Dim 9	0.14	63.9
Dim 10	0.14	66.3
...
Dim 38	0.02	100.0

Le plan factoriel correspondant est présenté en Figure 76. La première dimension oppose les parfums 942, 639, 407, 379 et 688 (à droite) aux parfums 963, 666, 935,

713 et 766 (à gauche). La deuxième dimension oppose les parfums 407 et 996 (en haut) au parfum 436 (en bas). Les zones colorées obtenues grâce à la procédure de visualisation permettent de déterminer aisément que la partie droite du plan factoriel est une zone d'innovation alors que la partie gauche est une zone de non-innovation.

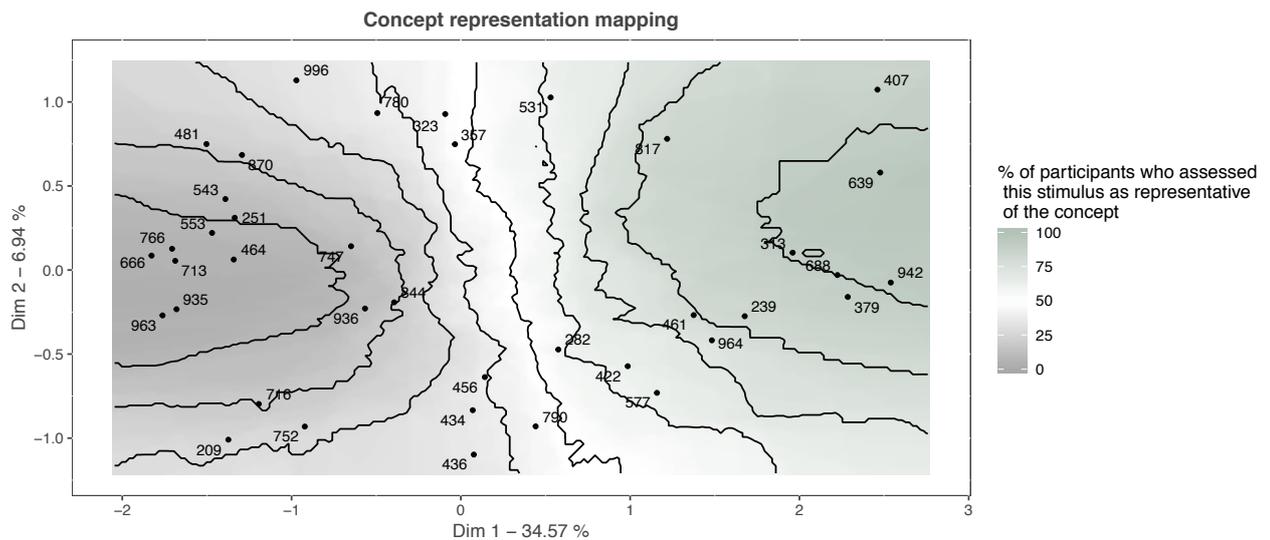


Figure 76: Étude ‘concept d’innovation et parfums de shampoings’ : Représentation des parfums de shampoings issue de l’Analyse Factorielle Multiple dans le cadre de la procédure statistique orientée stimuli.

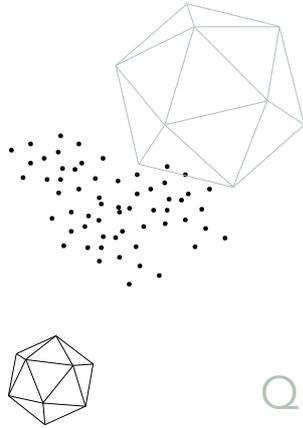
La proximité entre deux parfums est à mettre en relation avec (1) le nombre de fois où ils ont été mis dans un même groupe par les consommateurs et (2) le nombre de fois où ils ont été perçus comme innovants par les consommateurs. Par exemple, les parfums 963 et 935 (tous deux à gauche) sont proches sur le plan factoriel parce qu’ils ont été perçus comme similaires : ils ont été rassemblés dans un même groupe par de nombreux consommateurs (85 sur 120) et ils ont été considérés comme innovants un nombre similaire de fois (23 fois pour le parfum 963 et 24 fois pour le parfum 935). Prenons maintenant l’exemple des parfums 323 (en haut) et 436 (en bas). Ces deux parfums sont éloignés l’un de l’autre sur le plan factoriel. Bien qu’ils aient été considérés comme innovants un même nombre de fois (53 fois pour le parfum 323 et 52 fois pour le parfum 436), ils n’ont été rassemblés dans un même groupe que par un faible nombre de consommateurs (39 sur 120). Soulignons ici que le seul calcul d’un score d’innovation par parfum n’aurait pas pu mettre en évidence cette caractéristique. Celui-ci montre pourtant une certaine différence entre ces deux parfums. Cet aspect souligne la présence potentielle de clusters de consommateurs en désaccord au sein du panel. Nous abordons cette problématique dans le Chapitre 5.

4 Conclusion

L'étude de la représentation conceptuelle d'un stimulus est de plus en plus courante en analyse sensorielle. Celle-ci génère depuis quelques années de nouvelles questions de recherche concernant les méthodologies de recueil et d'analyse statistique de données à mettre en place pour aborder cette problématique. Dans ce cadre, Jaffré et al. (2009) ont introduit une nouvelle méthodologie dédiée à la caractérisation conceptuelle de grands ensembles de stimuli. Elle peut être vue comme une version binaire de la Q-méthodologie utilisée en psychologie pour étudier la subjectivité humaine. Elle consiste à recueillir, de la part de chaque évaluateur, un profil d'évaluations binaires. Chaque valeur correspond à l'évaluation donnée à un stimulus : représentatif du concept (évaluation positive) ou non-représentatif du concept (évaluation négative). Le protocole permettant de recueillir ces données a été détaillé dans ce chapitre.

Malgré l'efficacité du protocole de recueil de données, l'analyse statistique des profils d'évaluations binaires associés aux stimuli qu'il permet de recueillir est restée inachevée. Les réflexions menées jusqu'alors reposent sur l'utilisation de techniques statistiques dédiées à l'analyse des données de tri libre. Parmi celles-ci, nous avons cité l'Analyse des Correspondances Multiples. Grâce à des jeux de données fictifs pédagogiques, nous avons pourtant démontré dans ce chapitre que ces procédures ne sont pas adaptées à la nature des données collectées. Elles ne permettent pas de prendre en compte le degré avec lequel les stimuli représentent le concept évalué lors de l'étude. Cette conclusion nous a amenés à proposer l'Analyse Factorielle des Correspondances comme une solution pour prendre en compte cette information occultée par les techniques classiques. Nous avons ensuite proposé l'Analyse Factorielle Multiple comme méthode permettant de prendre en compte l'intégralité des informations sur les stimuli véhiculées par les données. Toujours grâce à l'appui des jeux de données fictifs pédagogiques, nous avons prouvé que celle-ci permettait de prendre en compte la nature des données collectées. En ce sens, elle permet de comprendre de manière optimale la structure de variabilité conceptuelle observée au sein de l'ensemble de stimuli. Pour finir, nous avons vu comment cette approche pouvait être associée à une procédure de visualisation. Cette combinaison originale fournit une représentation euclidienne des stimuli esthétique et facilement interprétable. Ces différentes propositions ont donné naissance à une nouvelle procédure statistique *ad hoc* pour l'analyse orientée stimuli des données de Q-méthodologie binaire.

L'intérêt de cette procédure a par ailleurs été illustré à travers son application à des données réelles.



Chapitre 5

ANALYSE STATISTIQUE ORIENTÉE ÉVALUATEURS DES DONNÉES DE Q-MÉTHODOLOGIE BINAIRE

RÉSUMÉ : Dans un contexte de Q-méthodologie binaire, il peut exister une diversité de perception du concept au sein de l'ensemble de stimuli. En conséquence, les profils d'évaluations binaires fournis par les évaluateurs peuvent être sujets à de la variabilité. Cette variabilité est l'expression d'un désaccord de perception entre les évaluateurs. L'analyse de la structure de ces désaccords peut permettre de comprendre la diversité de perception du concept. Cette problématique n'a jamais été abordée, bien que certaines techniques statistiques utilisées en psychologie permettent d'analyser des données binaires similaires selon une approche orientée évaluateurs. C'est le cas des techniques dédiées à l'analyse de l'accord inter-évaluateurs (*inter-rater agreement* en anglais). En occultant une partie de l'information véhiculée par les données, elles ne permettent cependant pas d'analyser la structure de désaccord au sein du panel de manière optimale. Dans ce chapitre, une procédure d'analyse statistique permettant de dépasser cette limite est proposée. Celle-ci est basée sur une approche modèle. En plus de fournir une analyse de la structure de désaccord adaptée à la nature des données, le cadre modèle présente des caractéristiques intéressantes supplémentaires. Il peut être utilisé pour dériver une distance inter-évaluateurs et, de manière plus globale, une procédure de classification non supervisée des évaluateurs. Une telle approche permet d'identifier la potentielle existence de clusters d'évaluateurs au sein du panel. La procédure statistique proposée est motivée par son application sur des données réelles de caractérisation de produits industriels.

CE CHAPITRE EST INSPIRÉ DE L'ARTICLE :

Brard, M., Lê, S., & Causeur, D. A latent class regression framework for modeling agreement: clustering of multivariate profiles of binary ratings based on a new distance and a model selection strategy. *Submitted*.

Sommaire

1	Format des données	157
2	Analyse statistique des données orientée évaluateurs	158
2.1	Analyse de la structure de désaccord au sein d'un ensemble de profils d'évaluations binaires	159
2.1.1	Techniques statistiques classiques et limites associées : focus sur les indices de désaccord inter-évaluateurs	159
2.1.2	Proposition d'une procédure permettant de dépasser les limites des techniques statistiques classiques : modèle de régression logistique à classes latentes	162
2.2	Classification non supervisée de profils d'évaluations binaires	167
2.2.1	Création du dendrogramme représentant la structure de désaccord	168
2.2.1.1	Distance inter-évaluateurs	168
2.2.1.2	Critère d'agglomération	171
2.2.1.3	Stratégie de coupure	172
2.2.1.4	Évaluation des performances globales	175
2.2.2	Consolidation de la partition d'évaluateurs	180
2.2.3	Interprétation de la partition d'évaluateurs : prise en compte de covariables	181
2.2.3.1	Interprétation des clusters grâce aux covariables sur les évaluateurs	181
2.2.3.2	Interprétation des clusters grâce aux covariables sur les stimuli	182
3	Illustration sur des données réelles de caractérisation de produits industriels : données 'concept d'innovation et parfums de shampoings'	184
3.1	Données	185
3.2	Résultats de l'analyse statistique orientée consommateurs	185
4	Conclusion	189
5	Annexe	191

Dans un premier temps, nous rappelons succinctement le format des données que la Q-méthodologie binaire permet de recueillir.

Dans un second temps, nous abordons la manière avec laquelle il est possible d'analyser ces données selon une approche orientée évaluateurs. Dans ce cadre, nous proposons un point actualisé des approches statistiques utilisées en psychologie pour étudier l'accord inter-évaluateurs. Nous verrons que celles-ci ne permettent pas d'analyser de manière optimale la structure de désaccord observée au sein du panel. S'appuyant sur des modèles trop simplistes d'accord, elles peuvent en

effet mener à de mauvaises interprétations de la structure de désaccord. Cette présentation des limites de ces techniques classiques nous amène par la suite à proposer une solution alternative. Celle-ci est fondée sur un modèle de régression logistique à classes latentes ajusté sur l'intégralité des profils d'évaluations binaires. Après avoir détaillé ses caractéristiques et principaux avantages, nous voyons comment cette approche modèle peut être plus globalement utilisée pour obtenir une classification non supervisée des profils d'évaluations binaires. Nous voyons en effet qu'elle permet d'obtenir deux aspects principaux d'une telle approche de classification. Elle permet tout d'abord de définir une distance inter-évaluateurs qui reflète un désaccord de perception. Cette distance peut être utilisée comme base à la construction d'un dendrogramme. Cette approche modèle permet également de définir une stratégie de sélection de modèle. Celle-ci peut être utilisée pour déterminer le niveau de coupure optimal du dendrogramme, définissant ainsi le nombre de clusters d'évaluateurs au sein du panel. Cette approche de classification permet de déterminer si les évaluateurs ont perçu le concept au sein de l'ensemble de stimuli de manière consensuelle. Si ce n'est pas le cas, elle permet de mettre en évidence des clusters d'évaluateurs pouvant être assimilés à des types de perception du concept. Grâce à une étude comparative basée sur des simulations, nous montrons que notre procédure de classification présente de meilleures performances que des approches concurrentes plus classiques. Nous démontrons également la flexibilité de cette procédure grâce à l'ajout de processus complémentaires. Dans ce cadre, un algorithme de partitionnement permet de consolider la partition d'évaluateurs obtenue. L'intégration de covariables sur les évaluateurs ou les stimuli permet, elle, de faciliter l'interprétation des clusters d'évaluateurs identifiés. Afin de servir la compréhension de ces différentes approches, des jeux de données pédagogiques sont utilisés.

Dans un troisième et dernier temps, nous présentons l'application de cette procédure statistique *ad hoc* à des données réelles de Q-méthodologie binaire. Celles-ci correspondent aux données utilisées précédemment dans le Chapitre 4. Ce sont des données d'évaluation de la représentation du concept d'innovation au sein d'un ensemble de 39 parfums de shampoings, cette tâche de caractérisation conceptuelle ayant été réalisée par 120 consommateurs.

1 Format des données

Pour rappel, les données recueillies dans le cadre de l'implémentation d'une Q-méthodologie binaire sont les suivantes : pour le s -ème stimulus, l'évaluation par le p -ème évaluateur est une variable binaire Y_{sp} qui est égale à 1 si le stimulus a été jugé comme représentatif du concept par l'évaluateur (évaluation positive), et à 0 dans le cas contraire (évaluation négative). Ces données peuvent être combinées dans une matrice Y à S lignes et P colonnes ; chaque ligne s correspondant à un stimulus, chaque colonne p correspondant à un évaluateur, et à l'intersection de la ligne s et de la colonne p la valeur y_{sp} (voir Figure 77).

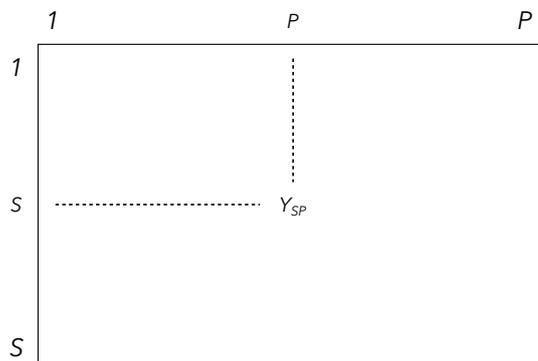


Figure 77: Données recueillies dans le cadre de l'implémentation de la Q-méthodologie binaire.

Dans cette matrice, le s -ème stimulus est défini par le profil d'évaluations binaires (Y_{s1}, \dots, Y_{sP}) , profil correspondant aux différentes évaluations que les P évaluateurs lui ont associées. L'analyse statistique de ces profils associés aux stimuli a fait l'objet d'une réflexion méthodologique spécifique dans le Chapitre 4. De la même façon, le p -ème évaluateur est défini par le profil d'évaluations binaires (Y_{1p}, \dots, Y_{Sp}) , profil correspondant aux différentes évaluations qu'il a associées aux S stimuli. C'est de l'analyse statistique de ces profils associés aux évaluateurs dont il est question ici.

2 Analyse statistique des données orientée évaluateurs

Nous proposons dans un premier temps de nous pencher sur la problématique de l'analyse de la structure de désaccord observée au sein d'un ensemble de profils d'évaluations binaires. Nous voyons comment des techniques statistiques classiques peuvent être utilisées pour répondre à cette problématique. Cependant, nous démontrons qu'elles échouent à analyser de manière optimale la façon avec laquelle les désaccords entre évaluateurs sont structurés. Ceci nous amène par la suite à proposer une approche basée sur un modèle de régression logistique à classes latentes. Dans un second temps, nous voyons comment ce cadre modèle proposé peut être utilisé de manière plus globale dans un contexte de classification non supervisée des profils d'évaluations binaires. Grâce à des simulations, nous démontrons que cette procédure de classification permet de dépasser les limites des approches plus classiques. L'ensemble de cette procédure *ad hoc* permet de fournir une segmentation des évaluateurs dans un contexte de Q-méthodologie binaire.

Des jeux de données issus d'expériences fictives pédagogiques sont utilisés pour illustrer l'implémentation des différentes approches présentées ici. Ils sont présentés sur la Figure 78.

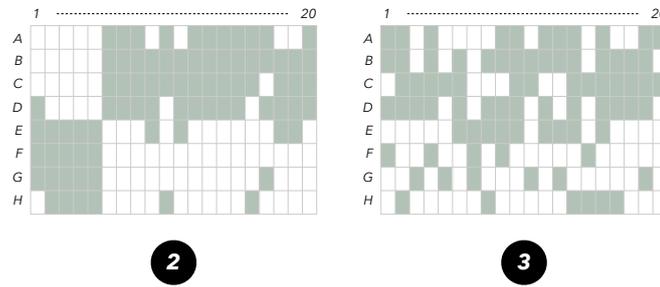


Figure 78: Expériences fictives pédagogiques. Une case colorée correspond à une évaluation binaire positive du stimulus par l'évaluateur (c.-à-d. 'représentatif du concept') et une case blanche correspond à une évaluation binaire négative du stimulus par l'évaluateur (c.-à-d. 'non-représentatif du concept').

2.1 Analyse de la structure de désaccord au sein d'un ensemble de profils d'évaluations binaires

2.1.1 Techniques statistiques classiques et limites associées : focus sur les indices de désaccord inter-évaluateurs

L'obtention de données binaires similaires à celles abordées dans le cadre de cette thèse est courante en psychologie. C'est par exemple le cas lorsqu'un certain nombre de psychologues ou de psychiatres interrogent un même groupe de patients en utilisant une échelle dichotomique pour déterminer la présence ou l'absence d'un symptôme. Dans un tel cadre, l'observation d'une certaine variabilité entre les profils d'évaluations binaires fournis par les évaluateurs est attendue. Un problème d'intérêt constant en psychologie est d'analyser cette variabilité et d'exprimer les résultats de manière concise. Cette problématique est généralement abordée à travers l'analyse de l'accord observé entre les évaluateurs. Par définition, l'accord inter-évaluateurs (*inter-rater agreement* en anglais) représente le degré d'accord observé entre deux évaluateurs. Celui-ci est généralement mesuré par le biais d'un score numérique quantifiant le consensus entre les deux profils d'évaluations binaires. Des mesures d'accord élevées indiquent un consensus entre les deux évaluateurs dans le diagnostic.

Différents indices d'accord inter-évaluateurs dédiés aux évaluations binaires ont été développés. De nombreux états de l'art sur le sujet sont disponibles dans la littérature : Adejumo, Heumann, and Toutenburg (2004); Banerjee, Capozzoli, McSweeney, and Sinha (1999); Fleiss (1975); Fleiss, Levin, and Paik (2003); Gisev, Bell, and Chen (2013); Goodman and Kruskal (1954); Warrens (2008). Tous ces indices d'accord inter-évaluateurs présentent la particularité d'être déduits d'une table de contingence 2×2 appelée *table d'accord* (*pairwise agreement table* en anglais). Cette table de contingence présente les profils d'évaluations binaires croisés fournis par les

deux évaluateurs correspondants (Warrens, 2015). Un exemple est présenté dans la Table 20.

Table 20: Table d'accord pour profils d'évaluations binaires : '1' représente une évaluation positive et '0' une évaluation négative.

		Évaluateur 2		
		1	0	Σ
Évaluateur 1	1	a	b	p_1
	0	c	d	q_1
		Σ	p_2	q_2
				1

Les premières approches développées se sont concentrées sur la proportion observée d'évaluations en accord (Goodman & Kruskal, 1954). L'indice d'accord inter-évaluateurs le plus simple et le plus intuitif est appelé dans la littérature la *proportion d'accord global* ou *simple matching coefficient* SM, avec $SM = a + d$. Dans le cas de l'expérience fictive pédagogique 2, la Figure 79 montre le résultat d'une procédure de MultiDimensional Scaling (MDS, voir Kruskal (1964)), l'une des approches statistiques les plus utilisées pour représenter de manière multidimensionnelle une matrice de dissimilarité. Cette procédure est ici appliquée à la matrice de désaccord considérant comme terme générique la proportion observée d'évaluations en désaccord $d^{SM} = 1 - SM$. Cette représentation multidimensionnelle met clairement en évidence une structure modulaire de désaccord. Celle-ci repose sur des sous-ensembles d'évaluateurs. La perception des stimuli est homogène au sein d'un sous-ensemble, et hétérogène d'un sous-ensemble à l'autre.

Cependant, l'approche précédente ne tient pas compte des différences potentiellement importantes parmi les propensions des deux évaluateurs à fournir une évaluation positive. Dans le contexte de l'analyse de la structure de désaccord, cette dernière question est parfois exprimée comme le fait qu'une certaine dose d'accord peut être attendue sous le seul effet de la chance. Elle peut se produire même s'il n'y a pas de tendance systématique à évaluer de la même manière les stimuli (Banerjee et al., 1999). Dans le but de résoudre ce problème, Cohen (1960) a proposé l'un des premiers indices d'accord inter-évaluateurs permettant de s'affranchir de cette dose d'accord attendue sous le seul effet de la chance : le *kappa de Cohen*. À l'heure actuelle, cet indice est certainement l'un des *gold standards* pour évaluer l'accord inter-évaluateurs observé entre deux profils d'évaluations binaires. Il est défini comme la proportion d'accord observée affranchie de la proportion d'accord attendue sous le seul effet de la chance. Il suppose que la proportion d'accord attendue sous le seul effet de la chance est directement liée à la propension des évaluations à fournir une évaluation positive $p_e = p_1 p_2 + q_1 q_2$:

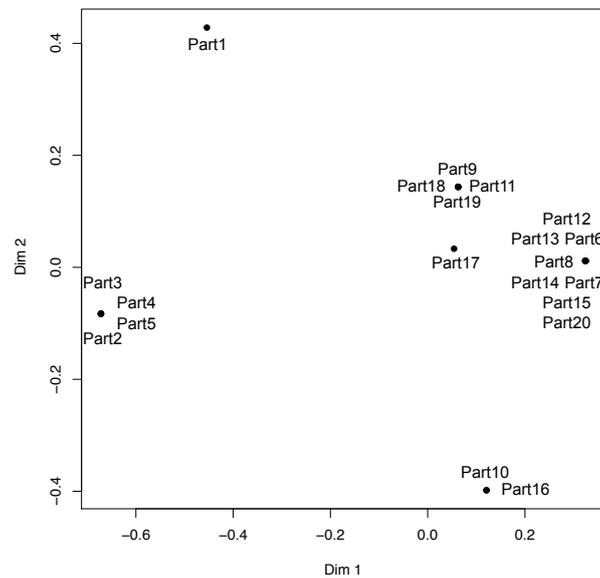


Figure 79: Expérience fictive pédagogique 2 : Représentation multidimensionnelle de la structure de désaccord obtenue par le biais d’une procédure de MultiDimensional Scaling appliquée à la matrice de dissimilarité entre évaluateurs en considérant la proportion d’évaluations en désaccord comme indice de dissimilarité.

$$\kappa = \frac{(a + d) - p_e}{1 - p_e}$$

Certains auteurs ont démontré que l’indice κ pouvait donner des résultats contre-intuitifs dans des situations particulières (voir Cicchetti and Feinstein (1990), Di Eugenio and Glass (2004) et Gwet (2008)). La première de ces situations est appelée le *paradoxe de prévalence*. Elle se produit si les deux distributions marginales dans la table d’accord sont symétriquement déséquilibrées (c.-à-d. si une notation, soit 0 soit 1, est largement sur-utilisée par rapport à l’autre). Cette situation tend à diminuer la valeur de κ . La deuxième de ces situations est appelée le *paradoxe de biais*. Elle se produit si les deux distributions marginales dans la table d’accord sont asymétriquement déséquilibrées (c.-à-d. si les deux évaluateurs présentent des propensions vraiment différentes à fournir une évaluation positive). Cette situation tend à augmenter la valeur de κ . En conséquence, Gwet (2014) a proposé une alternative résistante aux paradoxes de l’indice κ . De même que précédemment, l’indice AC1 proposé par Gwet (2014) est corrigé pour l’accord attendu sous le seul effet de la chance. La différence entre celui-ci et l’indice κ repose sur la manière avec laquelle la proportion d’accord attendue sous l’effet de la chance est calculée :

$$AC1 = \frac{[a + d] - [2q(1 - q)]}{1 - [2q(1 - q)]}$$

où $q = (2a + b + c)/2$.

Ces trois indices d'accord (à savoir SM, κ et AC1) constituent raisonnablement une vue d'ensemble des indices d'accord inter-évaluateurs généralement utilisés dans la littérature. Ils sont par la suite utilisés comme bases pour calculer les trois indices de désaccord correspondants : $d^{SM} = 1 - SM$, $d^{\kappa} = 1 - \kappa$, et $d^{AC1} = 1 - AC1$.

Cependant, étudier la structure de désaccord en utilisant des indices de désaccord obtenus par le biais de tables de contingence 2×2 peut être inadéquat. Ce type d'approches occulte en effet une grande partie de l'information véhiculée par les profils d'évaluations binaires. Tout d'abord, les informations sur les stimuli sont totalement occultées par la construction de la table de contingence. Il est alors impossible de définir les principaux axes de désaccord observés à l'échelle du panel d'évaluateurs. Il est également impossible de prendre en compte les propensions des stimuli à obtenir une évaluation positive. Certains stimuli sont en effet plus difficiles à évaluer positivement que d'autres. La prise en compte de cette information pour analyser la structure de désaccord est importante. Ensuite, même s'il est vrai que les indices de désaccord d^{κ} et d^{AC1} prennent en compte les différences de propension des évaluateurs à fournir une évaluation positive, celles-ci ne sont calculées que sur la seule base des profils des deux évaluateurs et non sur l'intégralité des profils du panel. Pour finir, il est impossible avec ce type d'approches de formuler des hypothèses quant à la possible structure complexe de désaccord. Une telle complexité renvoie à des situations dans lesquelles des sous-ensembles d'évaluateurs ne sont en désaccord que sur des sous-ensembles de stimuli seulement.

2.1.2 Proposition d'une procédure permettant de dépasser les limites des techniques statistiques classiques : modèle de régression logistique à classes latentes

Nous proposons ici un cadre général de modélisation de l'ensemble des données pour analyser la structure de désaccord observée au sein d'un ensemble de profils d'évaluations binaires. Contrairement aux techniques classiques, cette approche modèle permet de ne pas occulter les informations sur les stimuli. Elle permet également de considérer à la fois des scores de propension des stimuli et des évaluateurs. Ce procédé étend le principe de la correction du désaccord attendu sous le seul effet de la chance introduit précédemment pour les indices de désaccord d^{κ} et d^{AC1} . Cette fois-ci, les distributions marginales des évaluations sont évaluées sur l'intégralité des évaluateurs et des stimuli. Le cadre général de modélisation proposé stipule également l'existence d'une structure modulaire de désaccord latente introduite dans le modèle comme une interaction *stimulus* \times *évaluateur* particulière.

Paramètres de régression

Par la suite, nous supposons que le panel $\mathcal{C} = [1; P]$ d'évaluateurs, où $P \geq 2$ est le nombre d'évaluateurs, est partitionné en K clusters $\mathcal{C}_k, \overline{=} 1, \dots, K$ avec $1 \leq K \leq P$, $\bigcup_k \mathcal{C}_k = \mathcal{C}$ et, pour tout $k \neq k'$, $\mathcal{C}_k \cap \mathcal{C}_{k'} = \emptyset$. Le nombre d'évaluateurs dans le cluster \mathcal{C}_k est dorénavant noté P_k , avec $P_k \geq 2$. Soit $Y_{spk} = 0$ ou 1 l'évaluation binaire du s -ème stimulus, avec $s = 1, \dots, S$, par le p -ème évaluateur appartenant au cluster \mathcal{C}_k . Le modèle de régression logistique à classes latentes suivant est supposé pour la probabilité $\pi_{spk} = \mathbb{P}(Y_{spk} = 1)$ d'observer une évaluation positive :

$$\text{logit}\pi_{spk} = \mu + \alpha_s + \beta_{p|k} + \delta_k + (\alpha\delta)_{sk}, \quad (5.1)$$

où α_s et $\beta_{p|k}$ sont les paramètres de régression correspondant aux effets principaux des stimuli et des évaluateurs respectivement, la notation $p|k$ étant ici utilisée pour indiquer que l'effet *évaluateur* est emboîté dans l'effet *cluster* (c.-à-d. dans l'effet *classe latente*). Les paramètres $(\alpha_s)_{s=1, \dots, S}$ peuvent être interprétés comme des scores de propension des stimuli à obtenir une évaluation positive. De la même manière, pour tout k , les paramètres $(\beta_{p|k})_{p=1, \dots, P_k}$ peuvent être interprétés comme des scores de propension des évaluateurs spécifiques au cluster à fournir une évaluation positive. Les paramètres δ_k peuvent, eux aussi, être interprétés comme des scores de propension des clusters à fournir une évaluation positive. Discuter la pertinence de ces différents scores de propension dans le modèle fait l'objet d'une discussion spécifique dans la suite de ce chapitre. Enfin, les paramètres $(\alpha\delta)_{sk}$ représentent l'effet de l'interaction *stimulus* \times *cluster* et permettent de mesurer de quelle manière l'évaluation d'un stimulus varie d'un cluster à l'autre. Ces derniers paramètres de régression $(\alpha\delta)_{sk}$ permettent de capturer la possible complexité de la structure de désaccord.

Notons que le modèle additif suivant est obtenu comme un cas particulier du modèle (5.1) lorsqu'aucune structure en classes latentes (c'est-à-dire $K = 1$) n'est supposée :

$$\text{logit}\pi_{sp} = \mu + \alpha_s + \beta_p \quad (5.2)$$

Ce modèle sans classe latente (5.2) assume que l'évaluation d'un stimulus par un évaluateur ne dépend que du score de propension du stimulus et du score de propension de l'évaluateur. Il présente des similitudes avec le modèle de Rasch (Rasch, 1961). Ce dernier est utilisé dans la théorie de la réponse à un item pour décrire comment les réponses aux items d'un test psychologique dépendent à la fois d'un paramètre spécifique à l'item et d'un paramètre spécifique à la personne ayant réalisé le test. Certains auteurs ont d'ailleurs souligné l'adaptabilité d'un tel modèle pour étudier la structure de désaccord entre plusieurs évaluateurs (voir Agresti (1992); Petersen, Larsen, and Kreiner (2010)).

Ce modèle sans classe latente (5.2) stipule l'absence d'un effet d'interaction

stimulus \times *cluster*. Plus globalement, il stipule l'absence d'un effet d'interaction *stimulus* \times *évaluateur*. Ainsi, un parfait ajustement de ce modèle témoigne d'une situation dans laquelle l'évaluation d'un stimulus ne varie pas d'un évaluateur à l'autre. Une telle situation témoigne d'une situation d'accord au sein du panel. Pour cette raison, ce modèle sans classe latente sera par la suite dénoté 'modèle d'accord'. Il est important de préciser que ce modèle permet d'identifier une situation d'accord dans laquelle les évaluateurs auraient tous fourni la même évaluation relative (et non absolue) des stimuli les uns par rapport aux autres.

Le modèle (5.1) peut être vu comme un mélange de 'modèles d'accord' spécifiques aux clusters dans le sens où chaque modèle (5.2) permet de prendre en compte des scores de propension spécifiques au cluster. Par conséquent, le modèle (5.1) s'écarte du 'modèle d'accord' en capturant un éventuel désaccord entre les clusters sur les évaluations fournies au s -ème stimulus par le paramètre d'interaction *stimulus* \times *cluster*.

Analyse de la structure de désaccord

Ces types de modèles supposant que l'évaluation d'un stimulus par un évaluateur ne dépend que d'un effet *stimulus* et d'un effet *évaluateur* sont courants en analyse sensorielle. C'est notamment le cas dans le cadre des études cherchant à comprendre l'appréciation de produits par un panel de participants. En pratique, lors de ces études, chaque participant est amené à fournir une évaluation continue pour chaque produit, cette évaluation continue représentant son degré d'appréciation pour ce dernier (0 : 'je n'aime pas du tout' à 10 : 'j'aime beaucoup'). Pour modéliser les profils d'évaluations continues d'appréciation fournis par les évaluateurs, un modèle d'analyse de la variance à deux facteurs introduisant un effet *produit* et un effet *participant* est généralement utilisé.

Un effet de l'interaction *produit* \times *participant* dans le modèle d'analyse de la variance permettrait d'analyser le désaccord observé au sein des profils d'évaluations continues d'appréciation fournis. Cependant, comme dans un contexte de Q-méthodologie binaire, les données d'appréciation présentent la particularité de ne pas contenir de répliques. Chaque produit est en effet évalué uniquement une fois par chaque participant. En conséquence, la considération d'un effet de l'interaction *produit* \times *participant* dans le modèle est impossible. Cette interaction fait alors partie du terme d'erreur du modèle. C'est pourquoi c'est ce dernier qui est généralement utilisé pour modéliser l'hétérogénéité dans les appréciations des participants (Mammasse & Schlich, 2014). La possible structure d'hétérogénéité au sein des profils d'évaluations est ainsi définie comme la divergence entre les valeurs observées et les valeurs ajustées avec le modèle d'analyse de la variance à deux facteurs.

De la même façon, dans notre contexte binaire, le possible désaccord entre les P évaluateurs du panel peut être vu comme une divergence entre les valeurs observées et

les valeurs ajustées avec le ‘modèle d’accord’ (5.2). Cette divergence est mesurable par les résidus de déviance de ce modèle (5.2) :

$$\hat{\varepsilon}_{sp} = z_{sp} \sqrt{2 \log(1 + \exp(-z_{sp}(\hat{\mu} + \hat{\alpha}_s + \hat{\beta}_p)))} \quad (5.3)$$

où $\hat{\mu}$, $\hat{\alpha}_s$ et $\hat{\beta}_p$ représentent les estimateurs du maximum de vraisemblance des paramètres de régression associés aux effets du modèle (5.2) ; et z_{sp} représente le signe de l’évaluation fournie par le p -ème évaluateur au s -ème stimulus : $z_{sp} = 1$ si cette évaluation est positive et -1 sinon.

Ces résidus de la déviance servent de base pour l’analyse de la structure de désaccord observée au sein du panel d’évaluateurs. Ils peuvent être concaténés dans une matrice de dimensions $P \times S$ (voir Figure 80). Dans cette matrice, chaque ligne p correspond au profil de résidus de la déviance associé au p -ème évaluateur du panel : $\hat{\varepsilon}_p = (\hat{\varepsilon}_{p1}, \dots, \hat{\varepsilon}_{pS})'$.

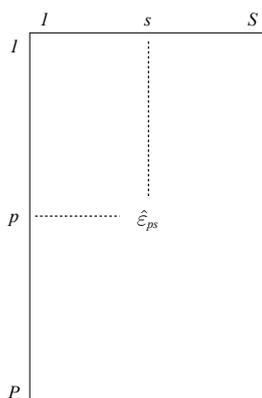


Figure 80: Matrice contenant les profils de résidus de la déviance associés aux évaluateurs : la valeur $\hat{\varepsilon}_{ps}$ correspond au résidu de la déviance associé au p -ème évaluateur et au s -ème stimulus.

Tirant parti de la nature continue et de la normalité multivariée des profils de résidus de la déviance spécifiques aux évaluateurs, des techniques de réduction des données telles que l’Analyse en Composantes Principales (ACP) peuvent être utilisées. Celles-ci permettent de déduire une éventuelle structure modulaire d’interaction révélant des désaccords partiels spécifiques à des sous-ensembles de stimuli. L’ACP présente l’avantage de fournir une représentation multidimensionnelle de la structure de désaccord sur laquelle la distance entre les évaluateurs peut être facilement interprétée à l’aide de la représentation des stimuli.

En effet, l’ACP fournit deux représentations principales : une représentation des individus statistiques (dans notre situation, une représentation des P évaluateurs) et une représentation des variables (dans notre situation, une représentation des

S stimuli). Sur la représentation des évaluateurs, deux évaluateurs sont distants s'ils présentent un fort désaccord. La distance inter-évaluateurs est effectivement d'autant plus grande que les deux profils de résidus de la déviance diffèrent. Ceci indique que les deux évaluateurs contribuent de manière différente à la divergence de la situation 'd'accord', et donc que leur perception du concept au sein de l'ensemble de stimuli n'est pas la même.

La représentation des P évaluateurs présente la particularité d'être liée à la représentation des S stimuli par une formule de transition. La représentation des évaluateurs doit donc être interprétée conjointement à la représentation des stimuli. Ceci peut être exprimé comme suit : un évaluateur est du même côté que les stimuli pour lesquels il a fourni une évaluation positive (c.-à-d. 'représentatif du concept'), et opposé aux stimuli pour lesquels il a fourni une évaluation négative (c.-à-d. 'non-représentatif du concept').

Dans le cas de l'expérience fictive pédagogique 2, la Figure 81 montre le résultat de l'ACP appliquée à la matrice contenant les profils de résidus de la déviance des évaluateurs. Afin de donner plus de poids aux stimuli les plus discriminants dans le calcul de la distance inter-évaluateurs, l'ACP n'est pas standardisée. Comme sur la représentation de la Figure 79 fournie par la procédure de MDS réalisée sur la matrice de désaccord D^{SM} , la représentation des évaluateurs fournie par l'ACP met clairement en évidence une structure modulaire de désaccord. Cependant, la représentation des évaluateurs de l'ACP prend à la fois en compte les différences de propension observées entre les évaluateurs et celles observées entre les stimuli. Enfin, la représentation des évaluateurs de l'ACP permet d'interpréter la structure de désaccord observée entre les évaluateurs en fonction des stimuli évalués. Il est ainsi possible de déterminer que le désaccord observé entre le sous-ensemble d'évaluateurs identifiés de 1 à 5 (à droite du premier axe) et le sous-ensemble d'évaluateurs identifiés de 6 à 20 (à gauche du premier axe) repose sur les axes de désaccord suivants : le sous-ensemble d'évaluateurs identifiés de 1 à 5 a globalement évalué positivement les stimuli E, F, G, H et négativement les stimuli A, B, C, D ; alors que le sous-ensemble d'évaluateurs identifiés de 6 à 20 a globalement fourni une évaluation opposée. Une telle interprétation était impossible en utilisant une procédure de MDS appliquée à la matrice D^{SM} .

L'approche modèle que nous proposons apparaît donc comme une solution permettant d'analyser avec plus de détails la structure de désaccord que les approches usuelles basées sur des indices de désaccord synthétiques. Entre autres, nous avons montré que cette procédure permet de conserver l'information sur les stimuli, de prendre en compte les différences de propension des stimuli et des évaluateurs, et de mettre en évidence des structures complexes de désaccord grâce à une structure d'interaction particulière.

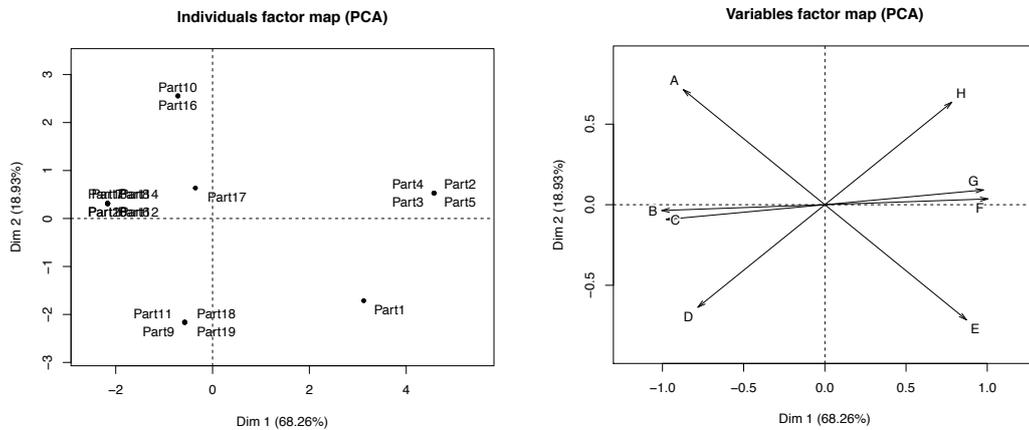


Figure 81: Expérience fictive pédagogique 2 : Représentation multidimensionnelle de la structure de désaccord obtenue par le biais d'une Analyse en Composantes Principales non-normée appliquée à la matrice contenant les profils de résidus de la déviance des évaluateurs. La représentation des individus statistiques correspond à la représentation des évaluateurs ; et la représentation des variables correspond à la représentation des stimuli.

2.2 Classification non supervisée de profils d'évaluations binaires

Dans la section précédente, une approche modèle a été introduite pour analyser la structure de désaccord au sein d'un ensemble de profils d'évaluations binaires. Nous exposons ici comment son utilisation peut être étendue à la problématique plus globale de classification non supervisée des profils d'évaluations binaires. Dans un contexte de Q-méthodologie binaire, une telle procédure de classification est utilisée pour segmenter le panel d'évaluateurs en fonction de leur perception du concept au sein de l'ensemble de stimuli. Si l'utilisation de l'approche modèle dans ce cadre est possible, c'est parce que deux aspects principaux d'une procédure de classification non supervisée peuvent être déduits de celle-ci. D'une part, elle permet de définir une distance inter-évaluateurs qui peut servir de base à la construction d'un dendrogramme. D'autre part, elle permet de définir une stratégie de sélection de modèle qui peut être utilisée pour déterminer le niveau de coupure optimal du dendrogramme obtenu. Nous voyons également comment la distance inter-évaluateurs peut de nouveau être utilisée pour consolider la partition d'évaluateurs obtenue grâce à un algorithme de partitionnement. Nous exposons enfin comment le modèle de régression logistique peut être utilisé pour enrichir l'interprétation des clusters d'évaluateurs identifiés. Dans ce cadre, des covariables sont utilisées. Il s'agit de covariables véhiculant des informations sur les évaluateurs ou sur les stimuli.

2.2.1 Création du dendrogramme représentant la structure de désaccord

2.2.1.1 Distance inter-évaluateurs

Nous avons précédemment démontré que les résidus de la déviance du ‘modèle d’accord’ (5.2) peuvent être utilisés pour étudier la structure de désaccord observé. Deux évaluateurs présentant des profils de résidus de la déviance $\hat{\epsilon}_p = (\hat{\epsilon}_{p1}, \dots, \hat{\epsilon}_{pS})'$ similaires s’écartent du ‘modèle d’accord’ de la même façon, et doivent donc apparaître dans le même cluster de désaccord. Nous proposons alors de considérer la distance euclidienne $d_{pp'}^{\text{res}}$ entre les profils de résidus de la déviance comme un indice de dissimilarité entre les deux évaluateurs p et p' :

$$d_{pp'}^{\text{res}} = \sqrt{\sum_{s=1}^S (\hat{\epsilon}_{sp} - \hat{\epsilon}_{sp'})^2} \quad (5.4)$$

La Table 21 correspond à la matrice de désaccord entre évaluateurs obtenue à partir de l’expérience fictive pédagogique 2. Cette matrice met en évidence deux sous-ensembles d’évaluateurs aux perceptions de l’ensemble de stimuli différentes : le sous-ensemble contenant les évaluateurs identifiés de 1 à 5 d’une part, et le sous-ensemble contenant les évaluateurs identifiés de 6 à 20 d’autre part. Nous retrouvons ici les groupes d’évaluateurs identifiés grâce à la représentation multidimensionnelle de la structure de désaccord présentée en Figure 81.

Table 21: Expérience fictive pédagogique 2 : Matrice de désaccord entre évaluateurs en considérant la distance euclidienne entre deux profils de résidus de la déviance du ‘modèle d’accord’ $d_{pp'}^{\text{res}}$ comme indice de dissimilarité.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
P1	0	3	3	3	3	6	6	6	5	7	5	6	6	6	6	7	5	5	5	6
P2	3	0	0	0	0	7	7	7	6	6	6	7	7	7	7	6	6	6	6	7
P3	3	0	0	0	0	7	7	7	6	6	6	7	7	7	7	6	6	6	6	7
P4	3	0	0	0	0	7	7	7	6	6	6	7	7	7	7	6	6	6	6	7
P5	3	0	0	0	0	7	7	7	6	6	6	7	7	7	7	6	6	6	6	7
P6	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P7	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P8	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P9	5	6	6	6	6	3	3	3	0	5	0	3	3	3	3	5	5	0	0	3
P10	7	6	6	6	6	3	3	3	5	0	5	3	3	3	3	0	5	5	5	3
P11	5	6	6	6	6	3	3	3	0	5	0	3	3	3	3	5	5	0	0	3
P12	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P13	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P14	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P15	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0
P16	7	6	6	6	6	3	3	3	5	0	5	3	3	3	3	0	5	5	5	3
P17	5	6	6	6	6	3	3	3	5	5	5	3	3	3	3	5	0	5	5	3
P18	5	6	6	6	6	3	3	3	0	5	0	3	3	3	3	5	5	0	0	3
P19	5	6	6	6	6	3	3	3	0	5	0	3	3	3	3	5	5	0	0	3
P20	6	7	7	7	7	0	0	0	3	3	3	0	0	0	0	3	3	3	3	0

Le ‘modèle d’accord’, tout comme le modèle de régression logistique à classes latentes, permet de prendre en compte comme paramètres de régression les scores de propension des stimuli et des évaluateurs à être associés à une évaluation positive. Le choix de considérer ou non ces scores de propension dans les modèles de désaccord, ou de façon équivalente le choix de la correction de ces modèles, doit rester à la seule charge de l’expérimentateur. Si ce dernier soupçonne que des différences de scores de propension des évaluateurs reflètent uniquement des différences de comportement lors de la réalisation de la tâche et non des différences de perception des stimuli, alors l’expérimentateur doit considérer ces scores dans le modèle. De cette façon, il s’affranchit de l’impact potentiel de ces différences de propension lors de l’analyse de la structure de désaccord. En annexe de ce chapitre, nous voyons que ces différences de propension peuvent impacter le processus de classification.

Cette problématique est récurrente en analyse sensorielle. Lors des études d’appréciation de produits, certains participants ont tendance à donner systé-

matiquement des notes élevées, et d'autres des notes systématiquement basses. D'après les chercheurs, ces aspects ne soulignent pas nécessairement des différences d'appréciation des produits mais plutôt des différences d'utilisation de l'échelle continue de notation. Pour cette raison, l'expérimentateur considère toujours un effet *participant* dans le modèle d'analyse de la variance. Cette démarche permet de s'affranchir de l'impact potentiel de ces différences d'utilisation de l'échelle. De cette manière, l'expérimentateur ne s'intéresse pas aux évaluations absolues des produits mais plutôt à leur évaluation relative.

Cependant, si l'expérimentateur ne préfère pas s'affranchir de l'impact de ces différences de propension, l'approche modèle offre un contexte flexible. Les modèles de désaccord peuvent en effet être formulés de manière analogue avec $\alpha_s = 0$ pour tout $s = 1, \dots, S$ ou $\beta_p = 0$, pour tous $p = 1, \dots, P$. Dans ce contexte, il existe des similitudes entre les distances inter-évaluateurs dérivées de ces modèles sans propensions et les indices de désaccord inter-évaluateurs communément utilisés dans la littérature. Dans le cas de l'expérience fictive 3, la Figure 82 montre le lien qu'il existe entre notre indice de désaccord d^{res} et les indices plus classiques d^{SM} et d^{κ} . Cette figure montre la correspondance parfaite entre d^{res} et d^{SM} quand aucun score de propension n'est inclus dans le 'modèle d'accord'. Elle montre également la correspondance étroite entre d^{res} et d^{κ} quand les scores de propension des évaluateurs sont inclus dans le 'modèle d'accord'.

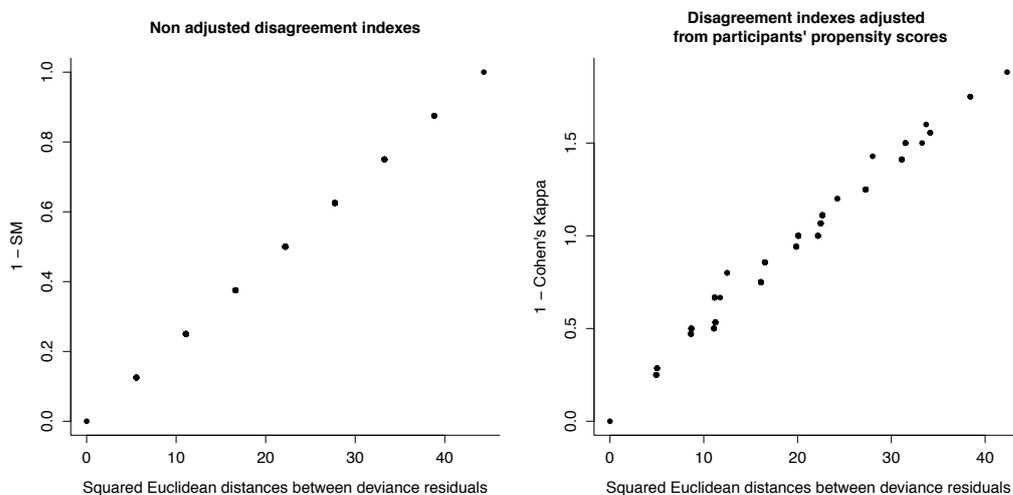


Figure 82: Expérience fictive pédagogique 3 : Correspondance entre d^{res} et d^{SM} quand aucun score de propension n'est inclus dans le modèle (5.2) ; et entre d^{res} et d^{κ} quand les scores de propension des évaluateurs sont inclus dans le modèle.

La relation linéaire parfaite entre les indices de désaccord d^{res} et d^{SM} quand aucun score de propension n'est inclus dans le modèle est généralement vrai, cet aspect étant largement prouvé sur la Figure 82. La légère différence entre l'indice de

désaccord d^{res} et d^{κ} quand les scores de propension des évaluateurs sont inclus dans le modèle est due à la manière avec laquelle l'estimation du facteur de correction de la chance est réalisée. Pour notre indice, les scores de propension des évaluateurs sont calculés en utilisant l'ensemble des données. Pour le kappa de Cohen, il sont évalués à l'aide de comparaisons par paires d'évaluateurs.

2.2.1.2 Critère d'agglomération

Les résidus de la déviance $\hat{\epsilon}_{ps}$ étant approximativement distribués selon une distribution normale standard, tout algorithme de classification appliqué à la matrice de dissimilarité D^{res} avec le terme générique $d_{pp'}^{\text{res}}$ peut être utilisé pour extraire des modules d'évaluateurs partageant le même point de vue sur le concept étudié.

Les algorithmes de classification ascendante hiérarchique sont privilégiés ci-après pour extraire la structure de désaccord en classes latentes de la matrice de dissimilarité D^{res} . Ceux-ci fournissent un dendrogramme représentant une séquence de partitions imbriquées des évaluateurs, partant d'une structure de désaccord sans classe latente (un cluster d'évaluateurs) jusqu'à la situation extrême d'un désaccord complet avec autant de classes latentes que d'évaluateurs (P clusters d'évaluateurs). Parmi les algorithmes de classification ascendante hiérarchique, l'algorithme basé sur le critère d'agglomération de Ward (Ward, 1963) est largement utilisé. Ce critère d'agglomération permet de produire un dendrogramme dans lequel les partitions présentent une dispersion intra-cluster minimale (Murtagh & Legendre, 2014). Dans la suite, l'algorithme de classification basé sur ce critère d'agglomération est utilisé pour déduire la structure de désaccord en classes latentes du panel d'évaluateurs.

Les dendrogrammes représentant les structures de désaccord observées au sein des panels d'évaluateurs dans les expériences fictives pédagogiques 2 et 3 sont présentés en Figure 83. Le dendrogramme correspondant à l'expérience 2 met clairement en évidence une structure de désaccord impliquant deux classes latentes : l'une correspond aux évaluateurs identifiés de 1 à 5, et l'autre correspond aux évaluateurs identifiés de 6 à 20. Le dendrogramme correspondant à l'expérience 3 met, lui, en évidence une structure de désaccord floue sans classes latentes particulières.

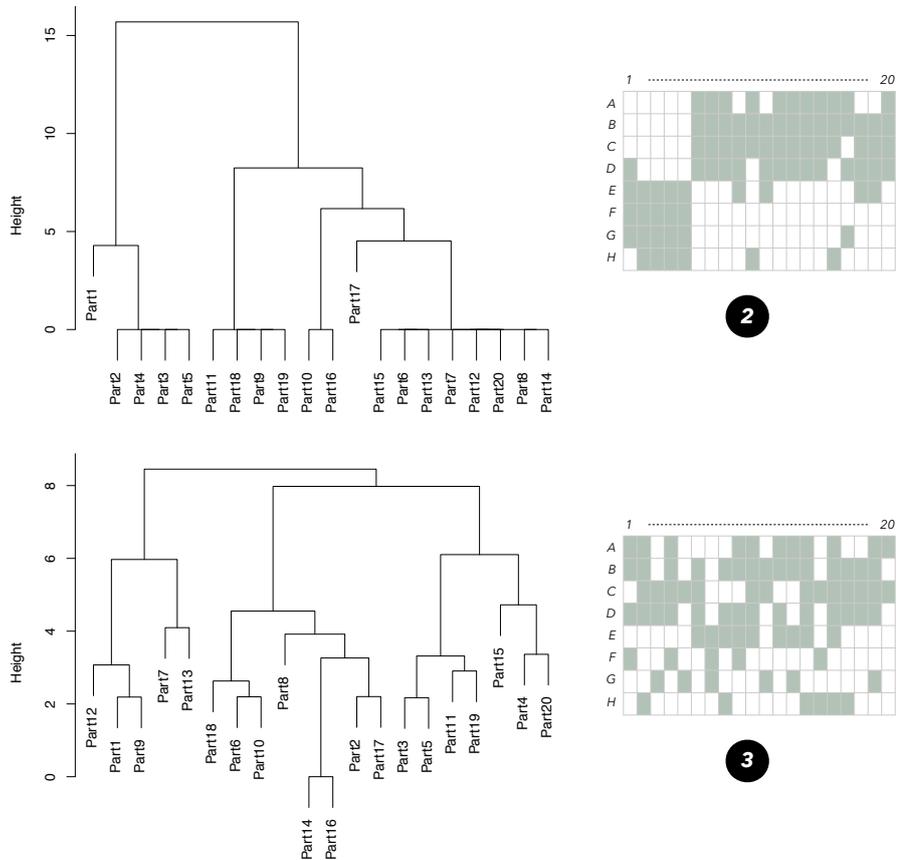


Figure 83: Expériences fictives pédagogiques 2 et 3 : Représentations des structures de désaccord observées au sein des panels d'évaluateurs.

2.2.1.3 Stratégie de coupure

Une fois le dendrogramme construit, l'expérimentateur a la charge d'identifier le niveau où celui-ci doit être coupé. Dans ce cadre, nous proposons ici une approche permettant de détecter le niveau de coupure optimal. Celle-ci est basée sur une procédure d'ajustement des modèles de désaccord (5.1).

Nous proposons de tester de manière séquentielle la significativité d'une structure de désaccord impliquant K clusters d'évaluateurs intégrée dans une super-structure de désaccord impliquant $(K - 1)$ clusters d'évaluateurs. Une telle démarche permet de compléter chaque niveau du dendrogramme par une p -value représentant la significativité de la partition en K clusters (par rapport à la partition en $(K - 1)$ clusters). Elle s'inspire d'approches récentes fournissant des outils statistiques pour estimer le nombre de clusters dans un dendrogramme (voir Bruzzese and Vistocco (2015); Liu, Hayes, Nobel, and Marron (2008); Park, Manjourides, Bonetti, and Pagano (2009); Ryota and Shimodaira (2011)). Ces procédures permettent de

dépasser certaines limites des heuristiques les plus courantes (p. ex. l'observation visuelle du dendrogramme ou le calcul d'indices de coupure à chaque niveau du dendrogramme). Elles protègent notamment l'utilisateur contre la détection erronée d'une structure de désaccord stipulant l'existence de classes latentes dans des situations de désaccord sans classe latente. Elles permettent ainsi de détecter les situations qui ne nécessitent pas de couper le dendrogramme.

Pour un nombre fixé de clusters (disons K), l'ajustement du modèle de désaccord avec K classes latentes (5.1) implique trois étapes :

1. Ajustement par maximum de vraisemblance du 'modèle d'accord' sans classe latente (5.2), et extraction des résidus de la déviance associés.
2. Création du dendrogramme en appliquant l'algorithme de Ward à la matrice de désaccord D^{res} , et extraction de la partition en K clusters d'évaluateurs.
3. Ajustement par maximum de vraisemblance du modèle de désaccord avec K classes latentes (5.1).

Retracer le dendrogramme à partir du nœud supérieur jusqu'aux nœuds inférieurs conduit à une séquence de modèles de désaccord emboîtés avec un nombre croissant de classes latentes. Nous proposons de déterminer le nombre optimal de classes latentes au sein de la structure de désaccord en implémentant de manière séquentielle des tests de comparaison de modèles. Ceux-ci sont basés sur la statistique du rapport de vraisemblance (LRT_K). Ils permettent de déterminer la significativité du modèle de désaccord à K classes latentes par rapport au modèle de désaccord à $(K - 1)$ classes latentes.

Cependant, la distribution de la statistique de test LRT_K sous l'hypothèse nulle d'existence de $(K - 1)$ clusters d'évaluateurs doit être adaptée au fait que la structure d'interaction en classes latentes est déduite des données. Elle doit donc prendre en compte le processus de classification. Pour cette raison, nous proposons d'approcher cette distribution nulle par une méthode de Monte-Carlo consistant à simuler des données sur la base du modèle de désaccord nul à $(K - 1)$ classes latentes (5.1). Lorsque l'hypothèse nulle ne peut plus être rejetée, $(K - 1)$ est utilisé comme le nombre de classes latentes de désaccord au sein du panel. Le dendrogramme est alors coupé au niveau correspondant à la partition en $(K - 1)$ clusters.

Les dendrogrammes représentant les structures de désaccord observées au sein des panels d'évaluateurs dans les expériences fictives pédagogiques 2 et 3 sont présentés en Figure 84. Les p -values correspondant à la significativité de chaque structure de désaccord en K classes latentes (par rapport à une structure de désaccord en $(K - 1)$ classes latentes) sont ajoutées à chaque niveau du dendrogramme. Elles sont indiquées jusqu'à l'apparition d'une p -value supérieure au seuil de 5 %. L'apparition d'une telle p -value témoigne en effet d'une acceptation de

l'hypothèse nulle d'une structure de désaccord en $(K - 1)$ classes latentes et donc d'un arrêt de l'algorithme. Le dendrogramme correspondant à l'expérience 2 confirme l'existence d'une structure de désaccord impliquant deux classes latentes.

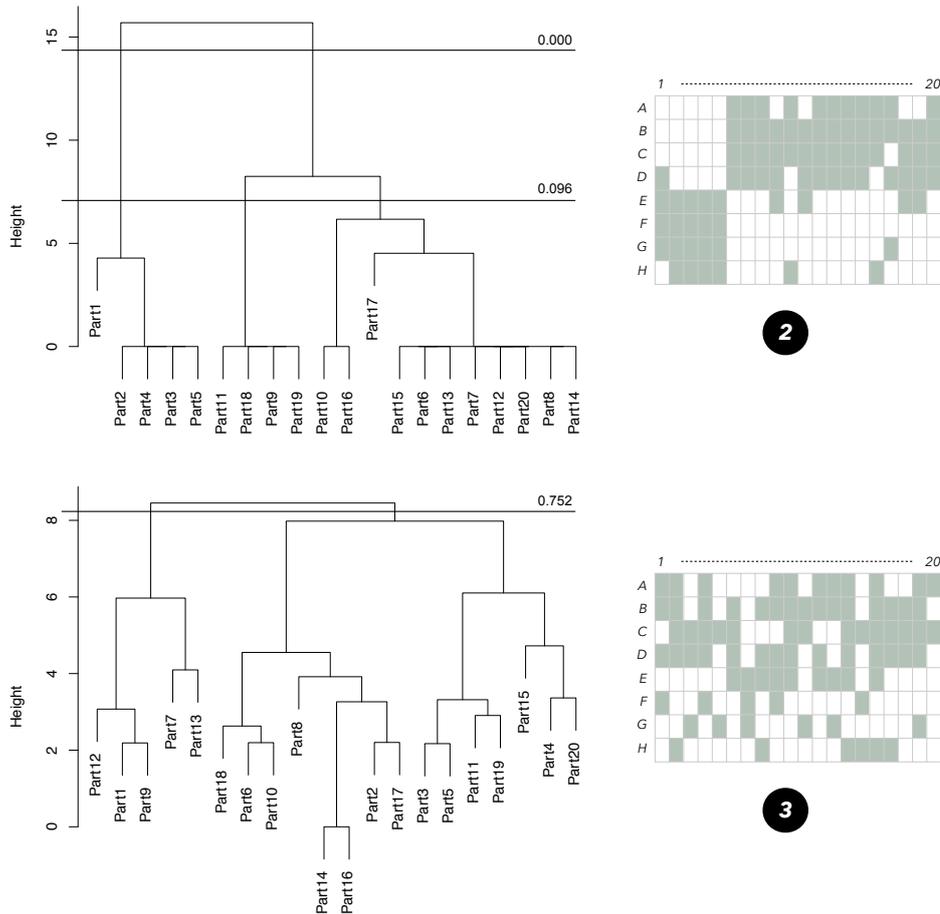


Figure 84: Expériences fictives pédagogiques 2 et 3 : Représentations des structures de désaccord observées au sein des panels d'évaluateurs. Chaque p -value située sur un niveau de coupure horizontal correspond au test de l'existence d'une structure de désaccord en K clusters d'évaluateurs par rapport à une structure de désaccord en $(K - 1)$ clusters d'évaluateurs.

Dans le cas de l'expérience 3, le modèle de désaccord à deux classes latentes n'est pas significativement meilleur que le modèle de désaccord sans classe latente. Ceci indique qu'il n'est pas possible d'identifier de classes latentes de désaccord. Il n'est donc pas nécessaire de couper le dendrogramme. Cependant, cette situation peut être observée dans deux configurations de désaccord différentes : soit (1) les profils d'évaluations binaires sont trop hétérogènes pour mettre en évidence une

structure modulaire de désaccord ; soit (2) les profils d'évaluations binaires sont trop homogènes pour mettre en évidence une structure modulaire de désaccord, signifiant ainsi que le panel d'évaluateurs est en accord total sur leur perception des stimuli.

Cet aspect étant primordial dans l'étude des désaccords au sein d'un panel, nous proposons d'implémenter un test statistique complémentaire pour évaluer la qualité d'ajustement du 'modèle d'accord' (5.2). Ce test, qui n'est réalisé que lorsque la structure de désaccord en deux classes latentes n'est pas significative, repose sur une statistique naturelle de mesure de la qualité d'ajustement de ce modèle : la déviance résiduelle \mathcal{D}_{res} . Sous l'hypothèse nulle d'une structure de désaccord inexistante au sein du panel, la distribution de référence de cette statistique \mathcal{D}_{res} est la distribution du χ^2 à $(S \times P) - (S + P - 1)$ degrés de liberté. Si l'hypothèse nulle de ce test est acceptée, cela signifie que le 'modèle d'accord' (5.2) ajuste bien les données et que, par conséquent, il n'existe pas de désaccord au sein du panel d'évaluateurs.

Le dendrogramme représentant la structure de désaccord observée au sein du panel d'évaluateurs dans l'expérience fictive pédagogique 3 est de nouveau présenté en Figure 85. Contrairement à la représentation précédente, une p -value est indiquée sur le nœud supérieur pour témoigner du bon ajustement ou non du 'modèle d'accord' aux données. La valeur de cette p -value indique ici que le 'modèle d'accord' présente un très mauvais ajustement. Il est ainsi possible de conclure que le panel d'évaluateurs a fourni des profils d'évaluations binaires très hétérogènes.

2.2.1.4 Évaluation des performances globales

Nous comparons ici les performances de la procédure que nous proposons à celles d'approches concurrentes. Nous entendons par 'performance' la capacité à identifier la structure de désaccord signalée au sein d'un ensemble de profils d'évaluations binaires. Dans cette section, les mesures de performances sont réalisées sur des jeux de données binaires créés par simulations dans lesquels la structure de désaccord est connue.

Quatre procédures d'extraction de la structure de désaccord sont testées : notre procédure basée sur l'approche modèle de régression logistique à classes latentes, et trois procédures concurrentes permettant d'extraire des structures de désaccord au sein de jeux de données binaires. Ces procédures concurrentes impliquent le même algorithme de classification ascendante hiérarchique (algorithme de Ward) appliqué à trois matrices de dissimilarité déduites des indices de désaccord $d^{SM} = 1 - SM$, $d^{\kappa} = 1 - \kappa$, et $d^{AC1} = 1 - AC1$. Pour chacune des procédures concurrentes, le nombre optimal de clusters est déterminé en utilisant l'indice de *Ratkowsky-Lance* (voir Ratkowsky and Lance (1978)), indice de coupure particulièrement recommandé dans le cadre de données binaires par Dimitriadou, Dolnicar, and Weingessel

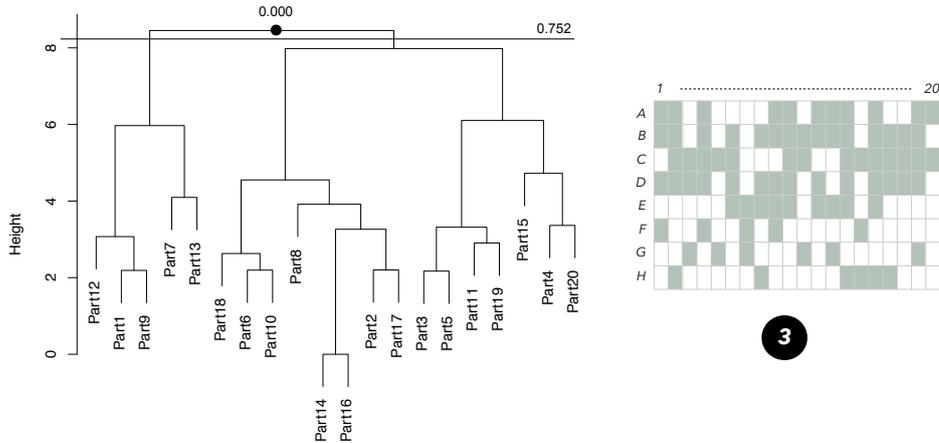


Figure 85: Expérience fictive pédagogique 3 : Représentation de la structures de désaccord observée au sein du panel d'évaluateurs. Chaque p -value située sur un niveau de coupure horizontale correspond au test de l'existence d'une structure de désaccord en K clusters d'évaluateurs par rapport à une structure de désaccord en $(K - 1)$ clusters d'évaluateurs. La p -value située sur le nœud supérieur correspond au test du bon ajustement du 'modèle d'accord' (5.2).

(2002). Les matrices de désaccord sont calculées en utilisant le logiciel R (R Core Team, 2017) ; et plus particulièrement les fonctions `smc` du package `scrim` pour D^{SM} (Schwender, 2015), `cohen` du package `scrim` pour D^{κ} (Schwender, 2015), et le code R fourni par l'auteur et disponible en ligne pour téléchargement gratuit pour D^{AC1} (Gwet, 2017). L'algorithme de Ward est appliqué en utilisant la fonction `hclust` du package `stats`, avec l'argument `method = 'Ward.D2'` (Murtagh & Legendre, 2014). Le niveau de coupure du dendrogramme basé sur l'indice de *Ratkowsky-Lance* est déterminé grâce à la fonction `NbClust` du package `NbClust` (Charrad et al., 2014).

Comme proposé par Hands and Everitt (1987) et utilisé plus récemment par d'autres auteurs tels que Henry, Dymnicki, Mohatt, Allen, and Kelly (2015), un modèle de classes latentes est employé pour générer des jeux de données binaires artificiels représentant différentes structures de désaccord impliquant K clusters. Ces jeux de données sont de dimensions 15×30 (c.-à-d. 15 stimuli évalués par 30 évaluateurs). Le plan de simulation est essentiellement défini par les paramètres suivants :

- K , le nombre de clusters d'évaluateurs.
- $\Theta = \{\Theta_k, k = 1, \dots, K\}$, un vecteur de K proportions d'évaluateurs dans chaque cluster.
- K vecteurs de S probabilités $P_k = \{P_{sk}, s = 1, \dots, S\}$ d'observer une

évaluation positive pour chaque stimulus dans chaque cluster.

Trois valeurs de K sont testées : $K = 1$ avec le vecteur de proportions correspondant $\Theta = 1$; $K = 2$ avec $\Theta = \{0.6, 0.4\}$ (ce qui signifie que 60 % des évaluateurs sont dans le cluster 1 et les 40 % restants dans le cluster 2) ; et $K = 4$ avec $\Theta = \{0.4, 0.3, 0.2, 0.1\}$. Les 30 vecteurs de 15 probabilités d'observer une évaluation positive sont créés selon l'un des quatre schémas suivants :

- Cluster 1 : les neuf premiers stimuli sont positivement évalués avec une probabilité $1 - p$ alors que les six derniers sont positivement évalués avec une probabilité p ; où $p = 1, 0.9, 0.8, 0.7, \text{ ou } 0.6$.
- Cluster 2 : les neuf premiers stimuli sont positivement évalués avec une probabilité p alors que les six derniers sont positivement évalués avec une probabilité $1 - p$.
- Cluster 3 : les stimuli avec indices impairs sont positivement évalués avec probabilité $1 - p$ alors que ceux avec des indices pairs sont positivement évalués avec une probabilité p .
- Cluster 4 : les stimuli avec des indices impairs sont positivement évalués avec une probabilité p alors que ceux avec des indices pairs sont positivement évalués avec une probabilité $1 - p$.

Dans les paramètres de simulation ci-dessus, p peut être considéré comme une mesure de l'évidence de la structure de désaccord. Ses valeurs extrêmes peuvent être interprétées comme suit. Si $p = 1$, la structure de désaccord intra-cluster n'est pas soumise au hasard. Si $p = 0.5$, les modèles de désaccord intra-cluster sont tous uniformes, ce qui signifie que la structure de désaccord en classes latentes elle-même n'existe pas. Notons également que nous avons choisi de simuler des situations dans lesquelles les scores de propension, à la fois pour les stimuli et les évaluateurs, ne sont pas contrastés. En effet, comme mentionné précédemment, les indices de désaccord utilisés présentent des comportements différents face à des différences de propension. S'affranchir de ces comportements différents est primordial dans l'objectif de comparer les performances des procédures.

Pour chaque valeur de K et chaque valeur de p , 1000 jeux de données sont simulés grâce au modèle de classes latentes présenté plus haut. La structure de désaccord de chaque jeu de données est extraite en utilisant les quatre procédures concurrentes. Les performances de chaque procédure sont évaluées sur la base de deux types de critères de performance permettant de mettre l'accent sur différents aspects de la structure de désaccord extraite. Le premier critère de performance est le pourcentage de fois où le nombre de clusters signal est correctement identifié, noté *CorrectK*. Puisque l'indice de *Ratkovsky-Lance* n'est pas capable de trouver moins de deux clusters, notre procédure est également implémentée de manière

restrictive afin de trouver au moins deux clusters. Cependant, pour information, le pourcentage de fois où une structure sans classe latente est identifiée par une implémentation sans restriction de notre procédure, notée *No-K*, est également enregistré. Cependant, ce n'est pas parce que le nombre de clusters signal a été identifié que la structure de désaccord signale a été correctement détectée. Pour cette raison, un deuxième critère de performance est nécessaire. Il s'agit de la mesure de la concordance entre la partition des évaluateurs extraite grâce à la procédure et la partition des évaluateurs signale. Cette mesure de concordance est réalisée par l'intermédiaire de l'indice de Rand ajusté, noté *ARI* (Hubert & Arabie, 1985). Cet indice est une version corrigée de l'indice de Rand *RI* développé par Rand (1971). La correction réalisée permet de rendre l'indice moins sensible au nombre de clusters et à la taille des clusters dans les deux partitions comparées (Courcoux et al., 2014). Ce critère est calculé en utilisant la fonction `RandIndex` du package `FreeSortR` (Courcoux, 2014).

Avec l'augmentation de la mesure de l'évidence de la structure de désaccord p , nous nous attendons à ce que les critères *CorrectK* et *ARI* augmentent, reflétant le fait que la structure de désaccord signale est de plus en plus correctement identifiée. En parallèle, avec la diminution de la mesure de l'évidence de la structure de désaccord p , nous nous attendons à ce que le critère *No-K* augmente, reflétant le fait qu'une structure de désaccord sans classe latente est de plus en plus correctement identifiée. Une procédure de classification présentant ces caractéristiques est considérée comme efficace pour l'extraction d'une structure de désaccord au sein d'un ensemble de profils d'évaluations binaires.

Les résultats de cette étude de comparaison des performances des quatre procédures de classification sont présentés en Figure 86. Cette figure montre les courbes de performances des procédures, c'est-à-dire les valeurs des différents critères de performance *ARI*, *CorrectK*, et *No-K* en fonction de la mesure de l'évidence de la structure de désaccord p .

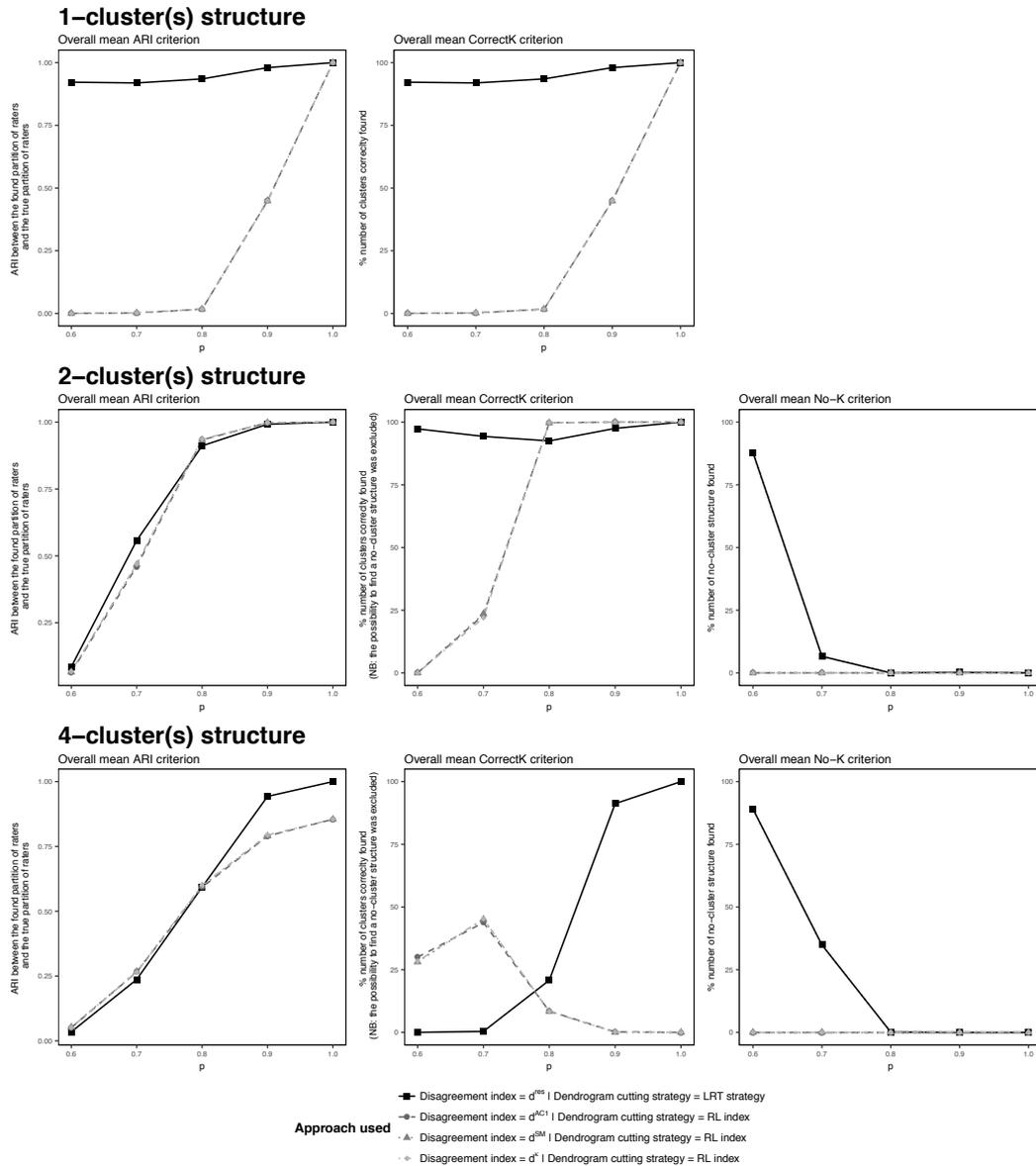


Figure 86: Étude comparative : Précision pour la détection de clusters d'évaluateurs en désaccord (concordance entre la partition d'évaluateurs détectée et la partition signalée *ARI*, proportion de jeux de données simulés pour lesquels le nombre de clusters signal est détecté *CorrectK*, et proportion de jeux de données simulés pour lesquels aucun cluster n'a été détecté *No-K*) en fonction de différentes valeurs de la mesure de l'évidence de la structure de désaccord p et selon différentes structures de désaccord impliquant K clusters d'évaluateurs soumises à quatre procédures d'extraction de clusters : notre procédure basée sur une approche modèle de régression logistique à classes latentes, d^{AC1} associé à l'indice de coupe de Ratkowsky-Lance, d^{SM} associé à l'indice de coupe de Ratkowsky-Lance, et d^k associé à l'indice de coupe de Ratkowsky-Lance.

Notre procédure s'avère être la plus efficace pour identifier le nombre de clusters signal au sein des jeux de données simulés (voir le critère *CorrectK*). Cette observation est particulièrement vraie pour la configuration impliquant quatre clusters d'évaluateurs, configuration qui s'avère la plus délicate à analyser quelle que soit la procédure d'extraction considérée. En effet, dans cette situation, la proportion de jeux de données simulés pour lesquels notre procédure détecte le bon nombre de clusters augmente en parallèle de l'augmentation de p pour atteindre des valeurs élevées pour une mesure de p supérieure à 0.9. Étonnamment, même pour les valeurs élevées de p , les autres procédures concurrentes ne parviennent pas à détecter la présence de quatre clusters d'évaluateurs. Notons au passage que ces trois procédures concurrentes ne différant que selon l'indice de désaccord considéré (c.-à-d. avec ou sans correction) et les paramètres de simulation n'introduisant pas de différences marquées dans les scores de propension, leurs performances sont très similaires. En ce qui concerne la concordance entre les partitions d'évaluateurs signales et les partitions extraites, les résultats de cette étude suggèrent que notre procédure montre une meilleure précision (voir le critère *ARI*). Cet aspect est particulièrement vrai pour les configurations de désaccord sans classe latente et avec quatre classes latentes. Enfin, comme attendu, notre procédure est la seule qui soit capable d'identifier correctement de plus en plus de structures de désaccord sans classe latente lorsque p décroît (voir critère *No-K*). Les résultats de cette étude comparative montrent donc que notre procédure apparaît comme une procédure efficace pour classifier un ensemble de profils d'évaluations binaires en fonction de la structure de désaccord observée.

2.2.2 Consolidation de la partition d'évaluateurs

La procédure de classification non supervisée proposée précédemment étant basée sur un algorithme hiérarchique, elle fournit une suite de partitions emboîtées. Ainsi, le critère d'agglomération n'est pas global mais dépend des clusters déjà obtenus dans les partitions précédentes. Il est donc impossible d'affecter à la marge des évaluateurs à d'autres clusters au cours de l'implémentation de l'algorithme. Dans l'objectif de favoriser ce type de réaffectations individuelles, nous proposons d'implémenter un algorithme de partitionnement à la suite de la procédure décrite jusqu'alors.

La partition des évaluateurs en K clusters obtenue grâce à la procédure de classification hiérarchique est introduite comme partition initiale d'un algorithme des K -means. Sur la base des évaluateurs appartenant à chaque cluster, les K centres des clusters sont calculés à partir de la matrice contenant les profils des résidus de la déviance (voir Figure 80). Chaque cluster peut ainsi être vu comme un profil de résidus de la déviance 'moyen'. À chaque étape de l'algorithme, un évaluateur est affecté au cluster dont le centre est le plus proche (au sens de la distance Euclidienne) et les K nouveaux centres des clusters sont calculés. Cette étape est répétée de manière itérative jusqu'à convergence de l'algorithme.

La partition d'évaluateurs ainsi obtenue est appelée 'partition consolidée'. Elle assure que les clusters sont (un peu) plus cohérents et mieux séparés que ceux obtenus sur la seule base de l'algorithme de classification ascendante hiérarchique (Husson et al., 2011).

2.2.3 Interprétation de la partition d'évaluateurs : prise en compte de covariables

Plusieurs informations sont facilement disponibles sur les clusters d'évaluateurs issus de la procédure de classification non supervisée. Chaque cluster est décrit par la proportion d'évaluateurs du panel qu'il contient. Il peut également être décrit par son paragon. Ce dernier correspond à l'évaluateur le plus proche du centre du cluster et peut être considéré comme l'évaluateur le plus représentatif.

Dans certaines situations, des covariables supplémentaires dont on connaît les valeurs sur la population de P évaluateurs sont disponibles. Elles sont récoltées par l'expérimentateur en parallèle de l'implémentation de la Q-méthodologie binaire. Ces covariables représentent généralement des caractéristiques signalétiques des évaluateurs (p. ex. genre, âge, profession). Elles peuvent également correspondre à des caractéristiques d'habitude d'usage de la gamme de stimuli évalués lors de l'expérience. L'un des objectifs de l'expérimentateur est alors de déterminer si un cluster d'évaluateurs peut être relié à une typologie de personnes.

De la même manière, des covariables supplémentaires dont on connaît les valeurs sur l'échantillon de S stimuli peuvent être disponibles. Dans le cas de stimuli correspondant à des produits industriels, ces covariables peuvent aussi bien représenter des caractéristiques intrinsèques des produits (p. ex. composition, propriétés rhéologiques, caractéristiques sensorielles) que des caractéristiques extrinsèques (p. ex. type de packaging, volumes vendus). Là encore, l'un des objectifs de l'expérimentateur est alors de déterminer si un cluster d'évaluateurs peut être relié à un type de perception des caractéristiques des stimuli.

Nous illustrons ici la flexibilité de notre procédure de classification non supervisée en présentant comment les clusters d'évaluateurs peuvent être interprétés grâce à ces covariables.

2.2.3.1 Interprétation des clusters grâce aux covariables sur les évaluateurs

La mise en relation entre les clusters et les covariables sur les évaluateurs est basée sur les données présentées en Figure 87. Elle consiste à comparer, pour chaque covariable, deux types de distributions : sa distribution au sein de chaque cluster d'évaluateurs d'une part, et sa distribution au sein du panel d'autre part. Cette comparaison est réalisée au moyen de tests statistiques. Ceux-ci dépendent de la nature de la covariable sur les évaluateurs (covariable continue ou covariable

catégorielle). Les procédures étant identiques à celles présentées dans la Partie 1 de cette thèse (voir Section 3.3.1 du Chapitre 2), elles ne seront pas détaillées de nouveau ici.

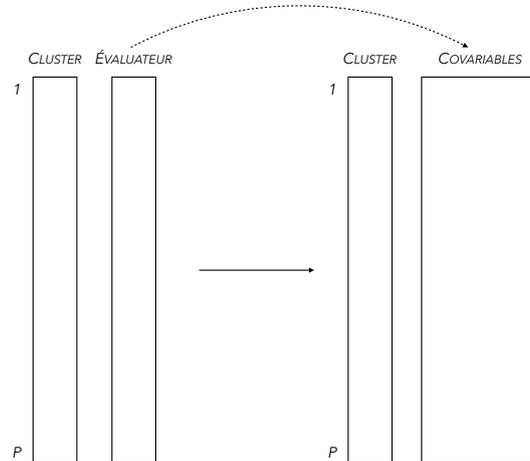


Figure 87: Jeu de données utilisé pour la description des clusters par des covariable sur les évaluateurs. Il présente les clusters des évaluateurs et les valeurs prises par les covariables disponibles sur les évaluateurs.

2.2.3.2 Interprétation des clusters grâce aux covariables sur les stimuli

La mise en relation entre les clusters et les covariables sur les stimuli est basée sur les données présentées en Figure 88. Elle consiste à déterminer, pour chaque covariable, si son effet sur l'évaluation varie d'un cluster d'évaluateurs à l'autre. Cette démarche repose, là encore, sur l'utilisation de tests statistiques. Si l'effet de la covariable sur l'évaluation binaire est significatif, il s'agit alors de définir la nature de cet impact. Plus précisément, l'objectif est de déterminer si, au sein de chaque cluster, la valeur de la covariable tend à favoriser ou à pénaliser la probabilité d'observation d'une évaluation positive. La démarche statistique utilisée dépend de la nature de la covariable sur les stimuli (covariable continue ou covariable catégorielle).

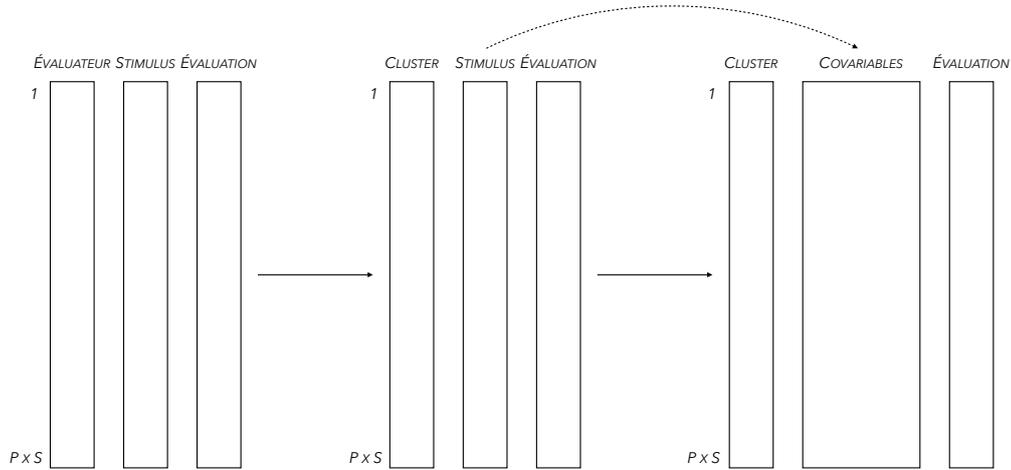


Figure 88: Jeu de données utilisé pour la description des clusters par des covariable sur les stimuli. Il présente les évaluations binaires fournies lors de l'implémentation de la Q-méthodologie binaire en fonction des clusters et des valeurs prises par les covariables disponibles sur les stimuli.

Covariable continue sur les stimuli

L'effet global de l'interaction entre la variable catégorielle *cluster* et la covariable continue sur la probabilité d'observer une évaluation positive est étudié à l'aide du modèle de régression logistique suivant :

$$\text{logit}(\mathbb{P}(Y_{kl} = 1)) = \mu + \delta_k + \beta x_{kl} + \sigma_k x_{kl}$$

$$\text{avec } \sum_{k=1}^K \delta_k = 0 \text{ et } \sum_{k=1}^K \sigma_k = 0$$

où Y_{kl} représente la l -ème répétition de l'évaluation binaire fournie par le cluster \mathcal{C}_k , δ_k représente l'effet spécifique du cluster \mathcal{C}_k , β représente l'effet de la covariable continue X , et σ_k représente l'effet spécifique de la covariable continue pour le cluster \mathcal{C}_k . D'autre part, la relation $\sum_{k=1}^K \delta_k = 0$ indique le fait que des contrastes de somme sur les paramètres de régression sont utilisés dans ce modèle.

Si l'effet global de l'interaction entre la variable *cluster* et la covariable continue est significatif, cela signifie que l'effet de la covariable continue sur l'évaluation binaire diffère d'un cluster d'évaluateurs à l'autre. Dans l'objectif d'analyser la nature de ces différences, les paramètres de régression du modèle σ_k sont ensuite étudiés indépendamment. Si le paramètre σ_k associé au cluster \mathcal{C}_k est significativement non nul, cela signifie que l'effet des valeurs de la covariable continue sur l'évaluation est caractéristique du cluster \mathcal{C}_k . Le signe du paramètre σ_k est ensuite utilisée pour déterminer si, au sein du cluster \mathcal{C}_k , des valeurs élevées de la covariable

continue tendent à favoriser l'observation d'une évaluation positive ($\sigma_k > 0$) ou à pénaliser l'observation d'une évaluation positive ($\sigma_k < 0$), et ce relativement aux observations faites sur la globalité du panel d'évaluateurs.

Covariable catégorielle sur les stimuli

L'effet global de l'interaction entre la variable catégorielle *cluster* et la covariable catégorielle sur la probabilité d'observer une évaluation positive est étudié à l'aide du modèle de régression logistique suivant :

$$\text{logit}(\mathbb{P}(Y_{qkl} = 1)) = \mu + \omega_q + \delta_k + (\omega\delta)_{qk}$$

$$\text{avec } \sum_{q=1}^Q \omega_q = 0, \sum_{k=1}^K \delta_k = 0, \text{ et } \sum_{q=1}^Q \sum_{k=1}^K (\omega\delta)_{qk} = 0$$

où Y_{qkl} représente la l -ème répétition de l'évaluation binaire fournie par le cluster \mathcal{C}_k à un stimulus présentant la modalité q de la covariable catégorielle d'intérêt, ω_q représente l'effet spécifique de la modalité q , δ_k représente l'effet spécifique du cluster \mathcal{C}_k , et $(\omega\delta)_{qk}$ représente l'effet spécifique de la modalité q de la covariable catégorielle pour le cluster \mathcal{C}_k . D'autre part, les relations $\sum_{q=1}^Q \omega_q = 0$, $\sum_{k=1}^K \delta_k = 0$, et

$\sum_{q=1}^Q \sum_{k=1}^K (\omega\delta)_{qk} = 0$ indiquent le fait que des contrastes de somme sur les paramètres de régression sont utilisés dans ce modèle.

Si l'effet global de l'interaction entre la variable *cluster* et la modalité q de la covariable catégorielle est significatif, cela signifie que l'effet de la modalité de la covariable catégorielle sur l'évaluation binaire diffère d'un cluster d'évaluateurs à l'autre. Dans l'objectif d'analyser la nature de ces différences, les paramètres de régression du modèle $(\omega\delta)_{qk}$ sont ensuite étudiés indépendamment. Si le paramètre $(\omega\delta)_{qk}$ associé au cluster \mathcal{C}_k est significativement non nul, cela signifie que l'effet de la caractéristique q sur l'évaluation est caractéristique du cluster \mathcal{C}_k . Le signe du paramètre $(\omega\delta)_{qk}$ est ensuite utilisé pour déterminer si, au sein du cluster \mathcal{C}_k , une présence de la caractéristique q de la covariable catégorielle tend à favoriser l'observation d'une évaluation positive ($(\omega\delta)_{qk} > 0$) ou à pénaliser l'observation d'une évaluation positive ($(\omega\delta)_{qk} < 0$), et ce relativement aux observations faites sur la globalité du panel d'évaluateurs.

3 Illustration sur des données réelles de caractérisation de produits industriels : données 'concept d'innovation et parfums de shampoings'

Dans l'objectif d'illustrer l'intérêt de la procédure d'analyse statistique introduite précédemment, nous présentons ici les résultats de son application sur des données

réelles de Q-méthodologie binaire. Ces données, déjà utilisées dans le Chapitre 4, sont issues de la caractérisation du concept ‘d’innovation’ au sein d’un grand ensemble de parfums de shampoings.

3.1 Données

Pour rappel, les données recueillies ont été mises en forme dans un jeu de données binaires de dimensions 39×120 présentant en lignes les parfums et en colonnes les consommateurs. L’intersection de la ligne s et de la colonne p contenait une valeur binaire y_{sp} égale à 1 si le parfum avait été jugé innovant par le consommateur (évaluation positive) et à 0 sinon (évaluation négative). À ces données de Q-méthodologie binaire s’ajoutaient 27 covariables en colonnes véhiculant des informations sur les parfums de shampoings. Il s’agissait des 27 caractéristiques sensorielles les plus utilisées par les consommateurs pour décrire les parfums de shampoings de manière spontanée. Chaque valeur de l’une de ces covariables correspondait au nombre de fois où la caractéristique avait été utilisée pour décrire le parfum sur la globalité du panel. Enfin, les données étaient complétées de 2 covariables en lignes véhiculant des informations sur les consommateurs. Il s’agissait de leur genre et de leur tranche d’âge.

3.2 Résultats de l’analyse statistique orientée consommateurs

La Figure 89 présente les résultats de la classification non supervisée des consommateurs. Ils mettent en évidence l’existence de deux clusters de consommateurs au sein du panel, la structure de désaccord en trois classes latentes n’étant pas significative par rapport à la structure en deux classes latentes (p -value = 0.12). En d’autres termes, il est possible de conclure qu’il existe deux types de perception du concept d’innovation dans le cadre des parfums pour shampoings, chacun de ces types pouvant être assimilé à un cluster de consommateurs du panel. La procédure de partitionnement réalisée à la suite de la procédure de classification ascendante hiérarchique a permis de réaffecter à la marge une vingtaine de consommateurs dans des clusters différents.

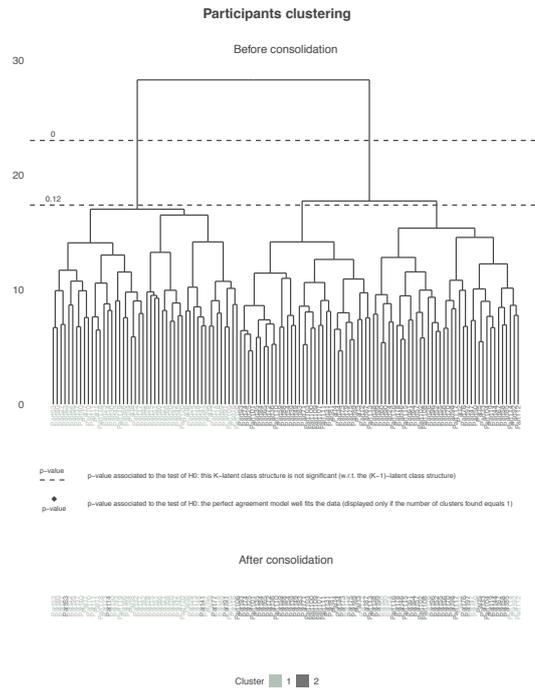


Figure 89: Étude ‘concept d’innovation et parfums de shampoings’ : Résultats de la classification non supervisée du panel de consommateurs. La partition des consommateurs du haut représente le résultat de la procédure de classification ascendante hiérarchique, et la partition des consommateurs du bas représente le résultat de la procédure de partitionnement.

La représentation multidimensionnelle de la structure de désaccord observée entre ces deux clusters de consommateurs est présentée sur la Figure 90. Cette représentation permet d’identifier rapidement les principaux axes de désaccord observés entre les deux clusters. Chaque axe de désaccord est en réalité un parfum pour lequel les clusters présentent un désaccord important de perception. Il est par exemple possible de mettre en évidence le fait que les principaux axes de désaccord observés entre le cluster 1 et le cluster 2 repose sur l’évaluation du parfum 996 (à droite de la première dimension). Plus précisément, le cluster 1 a évalué ce parfum comme étant innovant alors que le cluster 2 a évalué ce parfum comme étant non-innovant.

Multidimensional representation of the structure of disagreement among the panel of participants

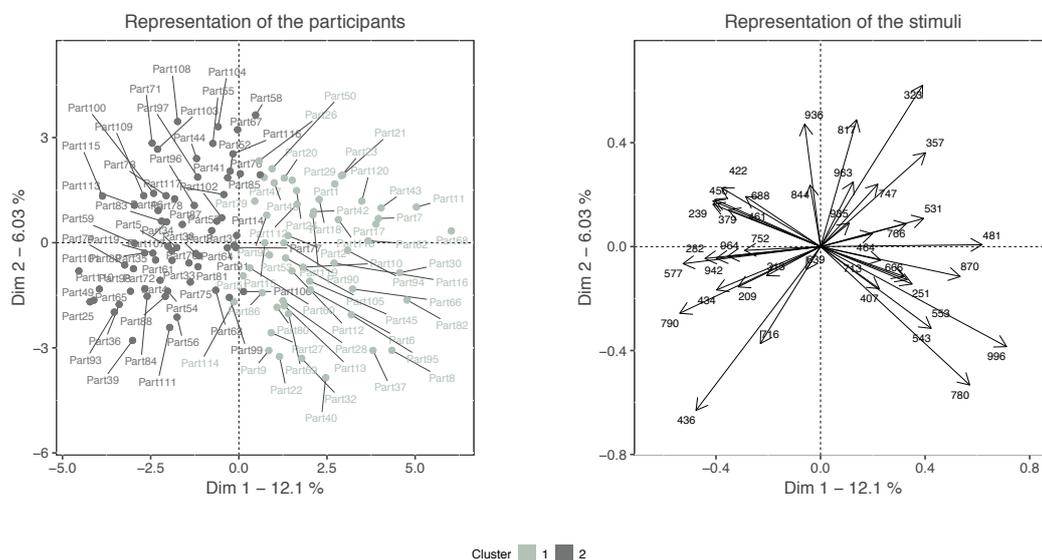


Figure 90: Étude ‘concept d’innovation et parfums de shampoings’ : Représentation multidimensionnelle de la structure de désaccord observée au sein du panel de consommateurs.

Grâce à la caractérisation automatique des clusters de consommateurs, il est possible de quantifier la représentativité de chaque type de perception du concept d’innovation des parfums de shampoings au sein de la population. Plus important encore, et ce grâce aux covariables disponibles sur les parfums, il est possible de comprendre avec plus de finesse sur quels critères des parfums reposent ces différents types de perception du concept d’innovation. De la même manière, grâce aux covariables disponibles sur les consommateurs, il est possible de déterminer si un type de perception du concept d’innovation est majoritairement véhiculé par une typologie de personnes en particulier.

La Table 22 présente la description du cluster 1, qui inclut 45 % des consommateurs du panel. Les deux types de perception du concept d’innovation sont donc globalement autant représentés l’un que l’autre au sein de la population. Grâce aux covariables sur les parfums, il est possible de déterminer que ce type de perception peut être interprété comme suit : un parfum de shampoing innovant est un parfum chimique, qui sent le propre et qui est puissant. Grâce aux covariables sur les consommateurs, il est possible de souligner que ce type de perception est majoritairement véhiculé par des hommes.

Table 22: Étude ‘concept d’innovation et parfums de shampoings’ : Résultats de la caractérisation automatique du cluster 1.

Cluster 1		
<i>Number of consumers</i>	54 (45% of the panel)	
<i>Parangon</i>	Part8	
<i>Covariates about the consumers</i>	<i>Sign of the statistic test</i>	<i>p-value</i>
Gender=H	+	< 0.005
Gender=F	-	< 0.005
<i>Covariates about the perfumes</i>	<i>Sign of the statistic test</i>	<i>p-value</i>
Chemical	+	< 0.005
Clean.scent	+	< 0.005
Strong	+	< 0.005
Honey	+	< 0.005
...
Milk	-	< 0.005
Not strong	-	0.01
Not sweet	-	0.01
Shampoo	-	0.01

Interprétation des statistiques de tests : Pour les covariables sur les évaluateurs : une statistique de test positive (resp. négative) indique que cette caractéristique des évaluateurs est sur-représentée (resp. sous-représentée) dans ce cluster, et ce relativement aux observations faites sur la globalité du panel. Pour les covariables sur les stimuli : une statistique de test positive (resp. négative) indique que la présence de cette caractéristique favorise (resp. pénalise) la probabilité d’observer une évaluation positive, et ce relativement aux observations faites sur la globalité du panel.

La Table 23 présente, elle, la description du cluster 2. Ce dernier inclut 55 % des consommateurs du panel. Grâce aux covariables sur les parfums, il est possible de déterminer que ce type de perception peut être interprété comme suit : un parfum de shampoing innovant est un parfum qui sent vraiment le shampoing, qui n’est pas sucré et qui n’est pas particulièrement puissant. Grâce aux covariables sur les consommateurs, il est possible de souligner que ce type de perception est majoritairement véhiculé par des femmes.

Table 23: Étude ‘concept d’innovation et parfums de shampoings’ : Résultats de la caractérisation automatique du cluster 2.

Cluster 2		
<i>Number of consumers</i>	66 (55% of the panel)	
<i>Parangon</i>	Part72	
<i>Covariates about the consumers</i>	<i>Sign of the statistic test</i>	<i>p-value</i>
Gender=F	+	< 0.005
Gender=H	-	< 0.005
<i>Covariates about the perfumes</i>	<i>Sign of the statistic test</i>	<i>p-value</i>
Shampoo	+	0.01
Not sweet	+	0.01
Not strong	+	0.01
Milk	+	< 0.005
...
Chemical	-	< 0.005
Clean.scent	-	< 0.005
Strong	-	< 0.005
Honey	-	< 0.005

Interprétation des statistiques de tests : Pour les covariables sur les évaluateurs : une statistique de test positive (resp. négative) indique que cette caractéristique des évaluateurs est sur-représentée (resp. sous-représentée) dans ce cluster, et ce relativement aux observations faites sur la globalité du panel. Pour les covariables sur les stimuli : une statistique de test positive (resp. négative) indique que la présence de cette caractéristique favorise (resp. pénalise) la probabilité d’observer une évaluation positive, et ce relativement aux observations faites sur la globalité du panel.

4 Conclusion

L’étude de la représentation conceptuelle d’un stimulus est de plus en plus courante en analyse sensorielle. Celle-ci génère depuis quelques années de nouvelles questions de recherche concernant les méthodologies de recueil et d’analyse statistique de données à mettre en place pour aborder cette problématique. Dans ce cadre, Jaffré et al. (2009) ont introduit une nouvelle méthodologie dédiée à la caractérisation conceptuelle de grands ensembles de stimuli. Elle peut être vue comme une version binaire de la Q-méthodologie utilisée en psychologie pour étudier la subjectivité humaine. Elle consiste à recueillir, de la part de chaque évaluateur, un profil d’évaluations binaires. Chaque valeur correspond à l’évaluation donnée à un stimulus : représentatif du concept (évaluation positive) ou non-représentatif du concept (évaluation négative).

Malgré l’efficacité du protocole de recueil de données, l’analyse statistique des profils d’évaluations binaires associés aux évaluateurs qu’il permet de recueillir est restée inachevée. À l’heure actuelle, aucune réflexion scientifique n’a été entreprise pour déterminer de quelle manière il est possible d’analyser la structure de désac-

cord observée entre les évaluateurs. Dans ce chapitre, nous avons présenté une première solution potentielle à ce problème. Celle-ci est basée sur l'utilisation de techniques statistiques initialement dédiées à l'analyse de l'accord inter-évaluateurs en psychologie. Malheureusement, nous avons démontré que ces approches se basant sur une analyse trop simpliste de l'accord ne permettent pas d'étudier la structure de désaccord entre évaluateurs de manière optimale. Cette conclusion nous a amenés à proposer, comme solution alternative, une approche basée sur un modèle logistique à classes latentes. Celui-ci présente la particularité d'être ajusté sur l'intégralité des profils d'évaluations binaires fournis par le panel. Au-delà de fournir une analyse du désaccord plus cohérente, nous avons vu que cette technique permet d'obtenir des outils graphiques de visualisation de la structure de désaccord intéressants. Ensuite, nous avons illustré comment cette procédure peut être utilisée dans un cadre plus global de classification non supervisée des évaluateurs. Dans ce contexte, une nouvelle distance inter-évaluateurs a été proposée, cette dernière étant basée sur les profils de résidus de déviance d'un modèle d'accord assumant l'absence de classes latentes de désaccord parmi les évaluateurs. Nous avons enfin proposé une procédure originale séquentielle de test du rapport de vraisemblance pour déterminer le nombre approprié de classes latentes de désaccord au sein du panel. Une étude basée sur des simulations a démontré la supériorité de cette procédure de classification par rapport à d'autres techniques plus classiques. Pour finir, nous avons vu comment elle pouvait être utilisée dans un cadre flexible en étant associée à des processus complémentaires de consolidation de partition et d'interprétation automatique des clusters d'évaluateurs obtenus.

Ces différentes propositions méthodologiques ont donné naissance à une nouvelle procédure statistique *ad hoc* pour l'analyse orientée évaluateurs des données de Q-méthodologie binaire. Au-delà de son applicabilité à ces données en particulier, nous pensons que cette approche peut également être utilisée dans d'autres domaines scientifiques. En effet, l'analyse de l'accord inter-évaluateurs représente un intérêt majeur dans de nombreux domaines de recherche tels que la psychologie, la médecine générale, les enquêtes psychologiques et sociales, ou les enquêtes publiques (Maxwell & Pilliner, 1968).

L'intérêt de cette procédure a par ailleurs été illustré à travers son application à des données réelles.

5 Annexe

Impact de la prise en compte des propensions dans le modèle sur le processus de classification

Notre procédure basée sur une approche modèle offre un contexte flexible à l'utilisateur. Celui-ci peut en effet décider de prendre en compte ou non certains scores de propension à être associé à une évaluation positive dans les modèles de désaccord. Cette prise en compte des propensions dans le modèle est notamment essentielle dans la définition de la matrice de désaccord. Dans cette annexe, nous illustrons cet impact sur la classification à travers une étude comparative. Les mesures d'impact des scores de propension sont réalisées sur des données binaires créées par simulations dans lesquels les scores de propension sont connus. Dans cette étude, nous nous focalisons sur les scores de propension associés aux évaluateurs.

Deux procédures de classification sont comparées : celle basée sur le modèle de désaccord dans lequel l'effet *évaluateur* est pris en compte, et celle basée sur le modèle de désaccord dans lequel l'effet *évaluateur* n'est pas pris en compte. Notons ici que les paramètres de propension des stimuli sont supposés nuls. L'objectif est de démontrer que de grandes différences observées sur les scores de propension des évaluateurs peuvent avoir un impact important sur le processus de classification lorsque l'effet *évaluateur* n'est pas pris en compte dans le modèle.

L'idée derrière cette démarche est de prouver que la prise en compte des scores de propension dans le modèle de désaccord est une problématique importante à laquelle l'expérimentateur doit porter attention. Imaginons un cas de figure où l'expérimentateur considère que les différences de propension des évaluateurs ne soulignent pas une différence de perception mais seulement des différences de comportement pendant la réalisation de la tâche. S'il omet de prendre en compte l'effet *évaluateur* dans le modèle logistique alors les résultats de la procédure de classification non supervisée qu'il obtiendra fourniront une interprétation erronée de la structure de désaccord.

Dans cette étude comparative, un modèle de classes latentes est utilisé pour générer des données binaires 15×30 représentant différentes structures de désaccord impliquant 2 clusters d'évaluateurs. Ces structures présentent des différences en termes de propensions des évaluateurs. Dans le but de ne pas conduire à de grandes différences dans les scores de propension des stimuli, les proportions d'évaluateurs dans chacun des clusters sont choisis de manière à être similaires : 60 % des évaluateurs appartiennent au cluster 1 et les 40 % restants appartiennent au cluster 2. Le plan de simulation est essentiellement défini par les paramètres suivants :

- *prop*, la proportion d'évaluations positives fournies par un évaluateur d'un cluster. Ce ratio reflète le score de propension associé à l'évaluateur.

- 2 vecteurs de S probabilités $P_k = \{P_{sk}, s = 1, \dots, S\}$ d'observer une évaluation positive pour chaque stimulus dans chaque cluster k .

Trois configurations de $prop$ sont testées : $prop = 8/15$ pour les évaluateurs du cluster 1 et $prop = 7/15$ pour les évaluateurs du cluster 2 (scores de propension homogènes d'un cluster à l'autre) ; $prop = 10/15$ pour les évaluateurs du cluster 1 et $prop = 5/15$ pour les évaluateurs du cluster 2 (scores de propension modérément hétérogènes d'un cluster à l'autre) ; et $prop = 12/15$ pour les évaluateurs du cluster 1 et $prop = 3/15$ pour les évaluateurs du cluster 2 (scores de propension sévèrement hétérogènes d'un cluster à l'autre). Les 30 vecteurs de 15 probabilités d'observer une évaluation positive sont créés selon l'un des schémas suivants :

1. scores de propension homogènes :
 - Cluster 1 : les huit premiers stimuli sont positivement évalués avec une probabilité $1 - p$ alors que les sept derniers sont positivement évalués avec une probabilité p .
 - Cluster 2 : les huit premiers stimuli sont positivement évalués avec une probabilité p alors que les sept derniers sont positivement évalués avec une probabilité $1 - p$.
2. scores de propension modérément hétérogènes :
 - Cluster 1 : les dix premiers stimuli sont positivement évalués avec une probabilité $1 - p$ alors que les cinq derniers sont positivement évalués avec une probabilité p .
 - Cluster 2 : les dix premiers stimuli sont positivement évalués avec une probabilité p alors que les cinq derniers sont positivement évalués avec une probabilité $1 - p$.
3. scores de propension sévèrement hétérogènes :
 - Cluster 1 : les douze premiers stimuli sont positivement évalués avec une probabilité $1 - p$ alors que les trois derniers sont positivement évalués avec une probabilité p .
 - Cluster 2 : les douze premiers stimuli sont positivement évalués avec une probabilité p alors que les trois derniers sont positivement évalués avec une probabilité $1 - p$.

Dans les paramètres de simulation ci-dessus, p peut être considéré comme une mesure de l'évidence de la structure de désaccord. La valeur $p = 1$ correspond à une situation idéale, où les deux clusters d'évaluateurs sont parfaitement séparés. Dans le but d'évaluer le comportement des deux procédures de classification dans une situation plus réaliste avec un niveau de bruit plus élevé, cette valeur est fixée à $p = 0.8$.

Pour chaque configuration de *prop*, 200 jeux de données sont simulés grâce au modèle de classes latentes présenté plus haut. La structure de désaccord de chaque jeu de données est extraite en utilisant les deux procédures concurrentes.

Les caractéristiques des résultats de chaque procédure sont évaluées en mesurant l'impact des scores de propension des évaluateurs sur la partition des évaluateurs extraite. Cette mesure est réalisée par l'intermédiaire de l'ajustement d'un modèle linéaire généralisé multinomial. Celui-ci permet de prendre en compte le fait que les procédures peuvent potentiellement identifier plus de deux clusters d'évaluateurs. L'impact des scores de propension des évaluateurs sur les résultats du processus de classification est mesuré sur la base de deux types de critères. Le premier critère est la mesure de l'importance de la variable explicative dans le modèle linéaire généralisé multinomial, noté *PropImp*. Cette mesure est réalisée grâce à la fonction `varImp` du package `caret` (Kuhn, 2017). Le deuxième critère est la mesure de la concordance entre la partition des évaluateurs extraite grâce à la procédure et la partition des évaluateurs prédite avec le modèle linéaire généralisé multinomial. Cette mesure est réalisée par l'intermédiaire de l'indice de Rand ajusté, noté *ARI*. Cet indice est calculé en utilisant la fonction `RandIndex` du package `FreeSortR` (Courcoux, 2014). Plus ces deux critères (*PropImp* et *ARI*) sont élevés, plus l'impact des scores de propension des évaluateurs sur les résultats de la procédure de classification est important.

Les résultats de cette étude sont présentés dans la Table 24. Ils montrent que plus les scores de propension des évaluateurs sont hétérogènes, plus leur impact sur le processus de classification est important (augmentation de *PropImp* et de *ARI*). Cette caractéristique est observée quelque soit le modèle utilisé dans la procédure de classification. En d'autres termes, lorsqu'il existe des différences importantes parmi les scores de propension des évaluateurs, celles-ci peuvent affecter les résultats du processus de classification. D'autre part, les résultats montrent que la procédure basée sur le modèle ajusté pour les scores des évaluateurs tend à atténuer la force de cet impact. Cette propriété est observée pour tous les types de scores de proportion. En d'autres termes, la considération de l'effet *évaluateur* dans le modèle de désaccord tend à atténuer l'impact des scores de propension des évaluateurs sur le processus de classification.

Choisir d'atténuer l'impact des scores de propension des évaluateurs dépend de l'expertise de l'expérimentateur. Si celui-ci pense que de grandes différences parmi les scores de propension des évaluateurs ne reflètent pas nécessairement des différences de perception, alors il doit considérer l'effet *évaluateur* dans le modèle de désaccord. Sans cela, les résultats du processus de classification pourront fournir une vision erronée de la structure de désaccord observée au sein du panel.

inputChapitre6-Concep-Gestes/Chapitre6-Concep-Gestes

Table 24: Étude de l'impact des scores de propension des évaluateurs : Valeurs des différents critères (impact des scores de propension sur la partition des évaluateurs extraite *PropImp* et concordance entre la partition d'évaluateurs extraite et la partition prédite *ARI*) en fonction des différentes configurations simulées et du type de modèle de désaccord considéré.

Propensions	Modèle de désaccord	<i>PropImp</i>		<i>ARI</i>	
		mean	se	mean	se
homogènes	sans effet <i>évaluateur</i>	1.228	0.113	0.260	0.019
homogènes	avec effet <i>évaluateur</i>	0.911	0.079	0.232	0.017
modérément hétérogènes	sans effet <i>évaluateur</i>	2.632	0.206	0.432	0.030
modérément hétérogènes	avec effet <i>évaluateur</i>	1.585	0.162	0.368	0.026
sévèrement hétérogènes	sans effet <i>évaluateur</i>	3.591	0.220	0.657	0.017
sévèrement hétérogènes	avec effet <i>évaluateur</i>	2.209	0.167	0.586	0.017



Annexe

IMPLÉMENTATION DANS R DE L'ANALYSE STATISTIQUE DES DONNÉES DE Q-MÉTHODOLOGIE BINAIRE

RÉSUMÉ : Une procédure *ad hoc* d'analyse statistique des données de Q-méthodologie binaire a été introduite dans cette thèse. Dans cette annexe, son implémentation dans le logiciel libre **R** est présentée. Celle-ci a été réalisée à travers le développement du package **BinQmet**. Ce package repose sur une fonction principale qui fournit à l'utilisateur tous les résultats numériques et graphiques utiles à la compréhension des données de Q-méthodologie binaire. Des fonctions complémentaires permettent à l'utilisateur de personnaliser les sorties graphiques, voire de les rendre interactives pour faciliter leur interprétation et leur diffusion.

Sommaire

1	Fonction AnalyseBinQmet	197
1.1	Arguments en entrée	197
1.2	Sorties	199
2	Fonction plot.AnalyseBinQmet	199
2.1	Arguments en entrée	200
2.2	Sorties	201

Le package `BinQmet`, développé dans le cadre de ce travail de recherche, est dédié à l'analyse statistique des données de Q-méthodologie binaire. Il est disponible gratuitement sur la plateforme GitHub : <https://github.com/MargotBr/BinQmet>. À partir de cette plateforme, l'utilisateur peut télécharger le package. Il peut également visualiser une vignette en anglais présentant un tutoriel de son utilisation pratique.

À travers les fonctions R qu'il contient, le package implémente les procédures statistiques orientée stimuli (voir Chapitre 4) et orientée évaluateurs (voir Chapitre 5). L'implémentation de l'approche orientée stimuli a été construite autour de la fonction MFA du package `FactoMineR` (S. Lê et al., 2008). L'implémentation de l'approche orientée évaluateurs est, elle, réalisée à travers la fonction `AgreeClustBin` du package `AgreeClust` (Brard, Lê, & Causeur, 2017). Ce dernier a également été développé dans le cadre de cette thèse. Il est dédié à l'approche modèle pour l'analyse de la structure de désaccord et pour la classification non supervisée de profils d'évaluations. Dans cette annexe, nous ne détaillerons pas ce package. Il présente en effet des propriétés qui vont au-delà du cadre de cette thèse. Il permet notamment d'implémenter une approche modèle qui n'a pas été décrite dans ce manuscrit. Celle-ci est similaire à celle présentée dans le Chapitre 5 mais est spécifiquement dédiée à l'analyse de profils d'évaluations continues.

Le package `BinQmet` contient des données fictives pédagogiques. Celles-ci permettent à l'utilisateur de comprendre rapidement le fonctionnement des différentes fonctions du package. Des données réelles sont également disponibles. Elles correspondent aux données 'concept d'innovation et parfums' introduites dans les Chapitres 4 & 5. L'utilisateur peut également utiliser les données 'concept de bon geste et ouverture d'huîtres' présentées dans le Chapitre 6. Celles-ci sont disponibles dans le package `AgreeClust`.

Dans cette annexe, nous détaillons succinctement les fonctions de ce package R.

1 Fonction AnalyseBinQmet

La fonction `AnalyseBinQmet` est la fonction principale du package. C'est elle qui permet d'implémenter la procédure statistique des données. Son appel dans R est réalisé grâce aux lignes de commande suivantes :

```
AnalyseBinQmet(dta, id.info.stim = NULL, type.info.stim = NULL,  
               id.info.part = NULL, type.info.part = NULL, graph = TRUE,  
               ext.dev.Rstudio = FALSE, ...)
```

1.1 Arguments en entrée

La fonction `AnalyseBinQmet` admet huit arguments majeurs en entrée :

- `dta` : Jeu de données de Q-méthodologie binaire de dimensions $S \times P$. Il peut être complété par des covariables sur les stimuli (en colonnes) ou sur les évaluateurs (en lignes). Un exemple schématique est présenté sur la Figure 91.
- `id.info.stim` : Identifiants des colonnes de `dta` correspondant aux covariables sur les stimuli. Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les stimuli.
- `type.info.stim` : Nature des covariables sur les stimuli. 'cont' est utilisé pour indiquer une covariable continue et 'cat' est utilisé pour indiquer une covariable catégorielle. Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les stimuli.
- `id.info.part` : Identifiants des lignes de `dta` correspondant aux covariables sur les évaluateurs (= participants). Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les évaluateurs.
- `type.info.part` : Nature des covariables sur les évaluateurs. 'cont' est utilisé pour indiquer une covariable continue et 'cat' est utilisé pour indiquer une covariable catégorielle. Par défaut, cet argument est fixé à `NULL`, signifiant que le jeu de données ne contient pas de covariables sur les évaluateurs.
- `graph` : Choix d'afficher ou non les sorties graphiques produites par la fonction (booléen). Si cette valeur est égale à `TRUE` alors les graphiques sont affichés. Par défaut, cet argument est fixé à `TRUE`.
- `ext.dev.Rstudio` : Choix d'afficher ou non les sorties graphiques à l'extérieur de la fenêtre de RStudio (booléen). Si cette valeur est égale à `TRUE` alors les graphiques sont affichés à l'extérieur. Par défaut, cet argument est fixé à `FALSE`.

- ... : Ensemble des arguments de la fonction `AgreeClustBin` du package `AgreeClust`. Il est par exemple possible de choisir le format du modèle de désaccord en spécifiant les scores de propensions (des stimuli et/ou des évaluateurs) à considérer. Il est également possible de paralléliser le calcul de la distribution sous l'hypothèse nulle pour la stratégie de coupure du dendrogramme. En outre, il est possible de consolider ou non la partition d'évaluateurs à l'aide de l'algorithme de partitionnement. L'utilisateur peut accéder à l'intégralité de ces arguments en affichant l'aide de la fonction `AgreeClustBin`.

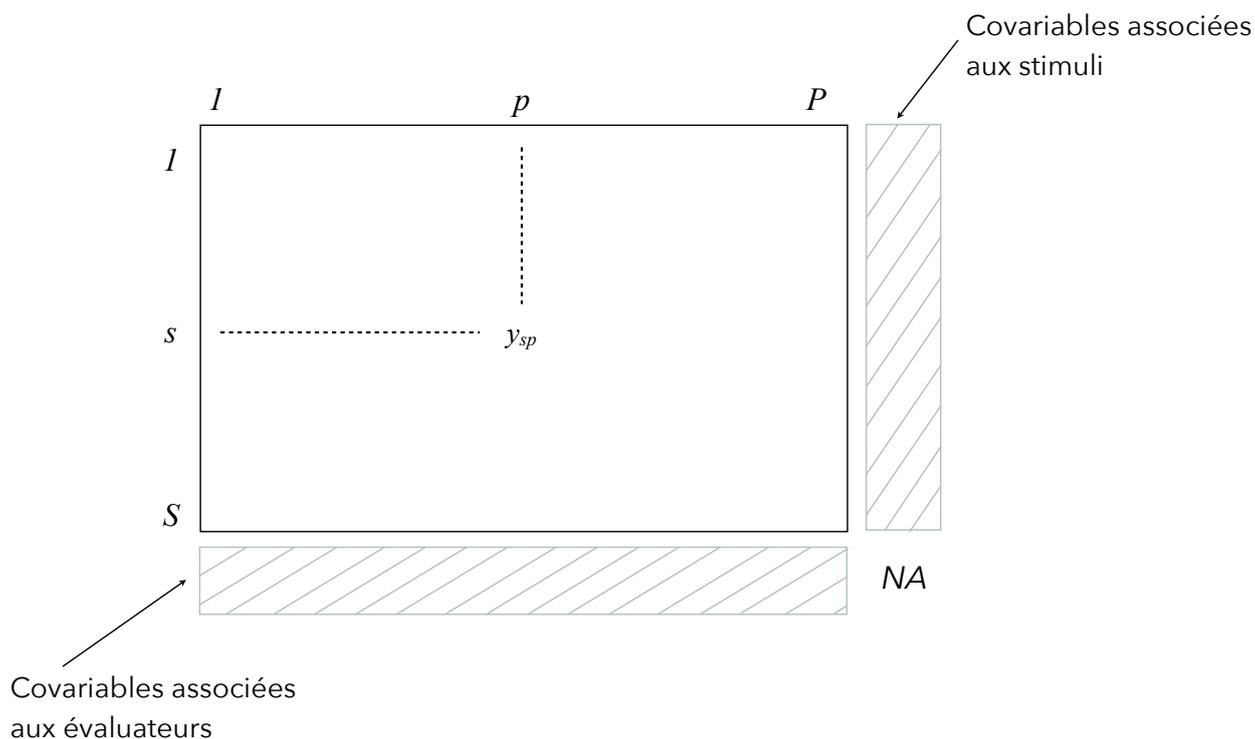


Figure 91: Format du jeu de données `dta` utilisé dans la fonction `AnalyseBinQmet`.

Les sept derniers arguments de la fonction `AnalyseBinQmet` présentent des valeurs par défaut. Si celles-ci conviennent à l'utilisateur, il suffit alors d'entrer uniquement le nom du jeu de données - sans covariables sur les stimuli ou les évaluateurs - dans la fonction pour lancer l'analyse statistique.

Exemple

Le package `BinQmet` contient un petit exemple fictif pédagogique : `pedagdata`. Ce jeu de données correspond à des données de Q-méthodologie binaire relatives à un ensemble de 8 stimuli caractérisés par un panel de 20 évaluateurs. Ces données de Q-méthodologie binaire sont supplémentées par des covariables sur les évaluateurs

et sur les stimuli. Trois covariables sont disponibles sur les évaluateurs : genre, tranche d'âge, fréquence d'utilisation. Les deux premières sont catégorielles et la dernière est continue. Cinq covariables sont disponibles sur les stimuli : intensité en agrumes, intensité en vanille, intensité boisée, intensité en lotus, type de packaging ('packaging actuel' ou 'packaging prototype'). Les quatre premières sont continues et la dernière est catégorielle.

L'utilisation de ce jeu de données peut être réalisée grâce aux commandes suivantes :

```
> data(pedagdata)
> res.pedag <- AnalyseBinQmet(pedagdata,
  + id.info.stim = 21 : ncol(pedagdata),
  + type.info.stim = c(rep("cont", 4), "cat"),
  + id.info.part = 9 : nrow(pedagdata),
  + type.info.part = c(rep("cat", 2), "cont"))
> res.pedag
```

1.2 Sorties

La fonction `AnalyseBinQmet` retourne à l'utilisateur une liste contenant les éléments suivants :

- `AnalyseBinQmet.data` : Tableau multiple composé des données catégorielles et des données de comptage utilisé dans le cadre de la procédure statistique orientée stimuli.
- `res.mfa` : Résultats de l'Analyse Factorielle Multiple (AFM) réalisée dans le cadre de la procédure statistiques orientée stimuli. Grâce à cet élément, il est possible d'accéder aux valeurs propres de l'AFM, aux coordonnées des stimuli, etc.
- `res.AgreeClust` : Résultats de la classification non supervisée des évaluateurs réalisée dans le cadre de la procédure statistiques orientée évaluateurs. Grâce à cet élément, il est possible d'accéder aux profils de résidus de la déviance des évaluateurs, à la partition finale des évaluateurs, à la caractérisation automatique des clusters, etc.

En parallèle de cette liste, la fonction `AnalyseBinQmet` retourne à l'utilisateur plusieurs sorties graphiques. Celles-ci sont réalisées grâce au package `ggplot2` (Wickham, 2016). Elles sont présentées en Figure 92.

2 Fonction `plot.AnalyseBinQmet`

La fonction `plot.AnalyseBinQmet` est une fonction secondaire du package. Elle permet d'afficher les représentations graphiques issues de la fonction

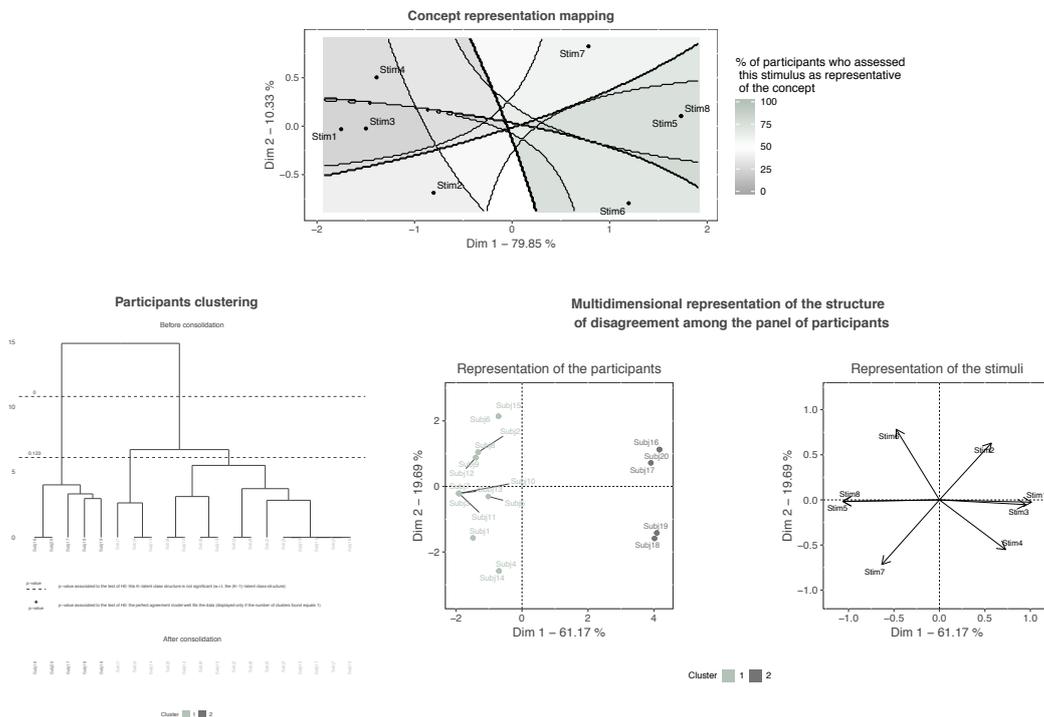


Figure 92: Jeu de données pédagogique du package `BinQmet` : Sorties graphiques de la fonction `AnalyseBinQmet`. Les résultats du haut sont ceux de la procédure statistique orientée stimuli et les résultats du bas sont ceux de la procédure statistique orientée évaluateurs.

`AnalyseBinQmet`. Elle prend donc en entrée l'objet résultant de cette fonction. Elle permet d'afficher de nouveau certains graphiques en les modifiant si besoin. Il est ainsi possible de modifier les couleurs ou de changer les axes des analyses factorielles affichés. Enfin, si l'utilisateur le souhaite, cette fonction peut afficher les graphiques sous un format interactif grâce au package `plotly` (Sievert et al., 2017). L'appel de cette fonction se fait grâce aux lignes de commande suivantes :

```
plot.AnalyseBinQmet(res, choice = "all", interact = FALSE,
  col.pos.ratings = NULL, col.neg.ratings = NULL, col.clust.part =
  NULL, axis = c(1, 2), ext.dev.Rstudio = FALSE)
```

2.1 Arguments en entrée

La fonction `plot.AnalyseBinQmet` admet huit arguments majeurs en entrée :

- `res` : Objet retourné par la fonction `AnalyseBinQmet`.
- `choice` : Type(s) de graphique(s) à créer. `'stim'` indique le souhait d'afficher la représentation multidimensionnelle des stimuli (procédure statistique orientée stimuli). `'part.seg'` indique le souhait d'afficher le dendrogramme et la

partition représentant la structure de désaccord au sein du panel (procédure statistique orientée évaluateurs). `'part.mul'` indique le souhait d'afficher la représentation multidimensionnelle de la structure de désaccord (procédure statistique orientée évaluateurs). `'all'` indique que l'intégralité de ces graphiques doit être affichée. Par défaut, cet argument est fixé à `'all'`.

- `interact` : Choix de créer ou non les graphiques sous un format interactif (booléen). Si cette valeur est égale à `TRUE` alors les graphiques créés sont interactifs. Par défaut, cet argument est fixé à `FALSE`.
- `col.pos.ratings` : Couleur associée aux évaluations positives ('représentatif du concept'). Cet argument peut être spécifié à travers une chaîne de caractères correspondant à la couleur (ex. `'pink'`) ou grâce au code hexadécimal associé à la couleur (ex. `'#B52E57'`).
- `col.neg.ratings` : Couleur associée aux évaluations négatives ('non-représentatif du concept'). Cet argument peut être spécifié à travers une chaîne de caractères correspondant à la couleur (ex. `'pink'`) ou grâce au code hexadécimal associé à la couleur (ex. `'#B52E57'`).
- `col.clust.part` : Couleurs associées aux clusters d'évaluateurs (= de participants) issus de la procédure statistique orientée évaluateurs. Les éléments de cet argument peuvent être spécifiés à travers une chaîne de caractères correspondant à chaque couleur (ex. `'pink'`) ou grâce au code hexadécimal associé à chacune (ex. `'#B52E57'`).
- `axis` : Axes factoriels à afficher sur la représentation multidimensionnelle des stimuli. Par défaut, cet argument est fixé à `axis = c(1, 2)` et les deux premiers axes de l'AFM sont affichés.
- `ext.dev.Rstudio` : Choix d'afficher ou non les sorties graphiques à l'extérieur de la fenêtre de RStudio (booléen). Si cette valeur est égale à `TRUE` alors les graphiques sont affichés à l'extérieur. Par défaut, cet argument est fixé à `FALSE`.

Cette fonction permet de customiser en grande partie les résultats issus de la procédure statistique orientée évaluateurs. Cependant, l'utilisateur peut avoir accès à plus d'options en utilisant la fonction `plot.AgreeClust` du package `AgreeClust` sur l'objet `res$res.AgreeClust`.

2.2 Sorties

Les lignes de commande suivantes permettent de changer les couleurs des graphiques (un exemple de sortie est présenté sur la Figure 93) :

```
> plot.AnalyseBinQmet(res.pedag,
```

```
+ choice = "stim",
+ col.pos.ratings = "#B52E57")
```

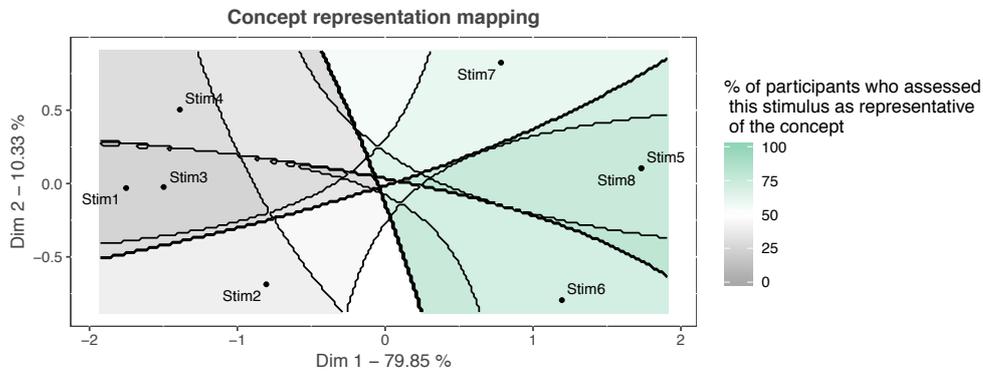


Figure 93: Jeu de données pédagogique du package `BinQmet` : Modification des couleurs de la sortie graphique de la procédure statistique orientée stimuli grâce à la fonction `plot.AnalyseBinQmet`.

En plus de changer les couleurs des graphiques, les lignes de commande suivantes permettent de les afficher sous format interactif. Un simple déplacement du curseur de la souris de la part de l'utilisateur permet d'afficher des informations complémentaires sur les stimuli ou sur les évaluateurs (un exemple de sortie est présenté sur la Figure 94) :

```
> plot.AnalyseBinQmet(res.pedag,
+ interact = TRUE,
+ choice = "part.mul",
+ col.clust.part = c("#B52E57", "#737575"))
```

Multidimensional representation of the structure of disagreement among the panel of participants

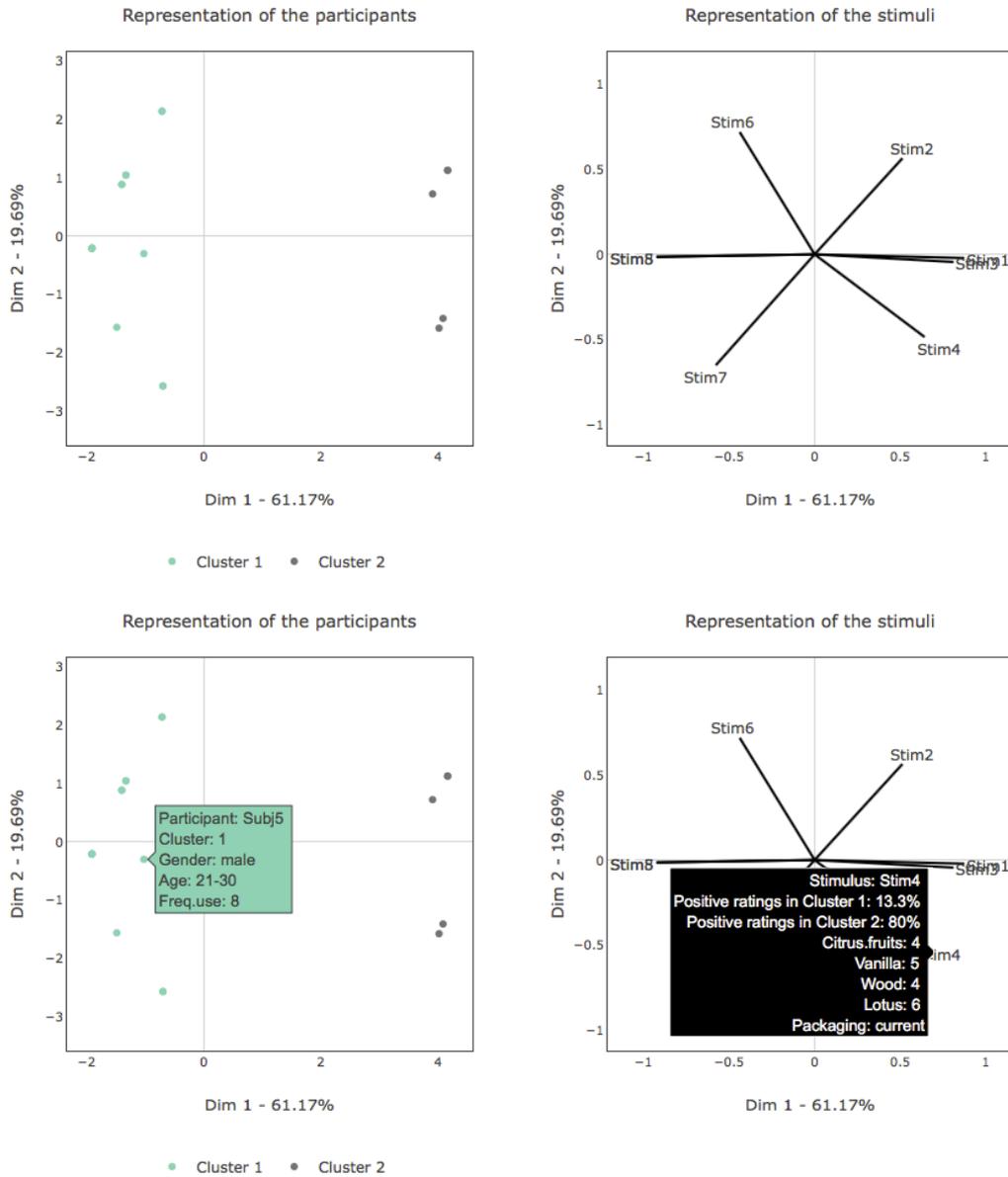
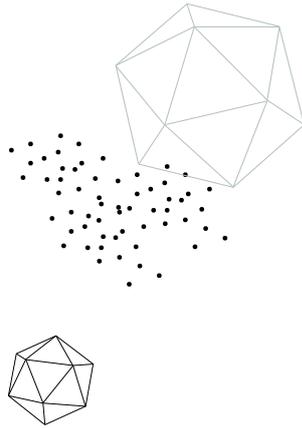


Figure 94: Jeu de données pédagogique du package `BinQmet` : Modification des couleurs des sorties graphiques de la procédure statistique orientée évaluateurs et affichage sous format interactif grâce à la fonction `plot.AnalyseBinQmet`.

Dans l'objectif de rendre les fonctionnalités de ce package accessibles au plus grand nombre d'utilisateurs - y compris des utilisateurs non familiers avec l'environnement

de R - une application Shiny doit également être développée. Cette application web permettra à l'utilisateur, grâce à une interface ergonomique, d'accéder à toutes les sorties du package R de manière dynamique et interactive.



CONCLUSION

L'objectif de ce travail de recherche est de contribuer au développement de nouvelles méthodes d'évaluation. Celles-ci permettent d'obtenir une caractérisation - perceptuelle ou conceptuelle - de grands ensembles de stimuli de la part d'évaluateurs non-experts. Les méthodes développées sont flexibles dans la mesure où elles sont applicables à différents types de stimuli. Appliquées à des produits industriels, elles permettent de dépasser certaines limites rencontrées jusqu'alors dans le domaine de l'analyse sensorielle. Il est en effet désormais possible de caractériser des espaces produits composés de plusieurs dizaines de références. Appliquées à des gestes culinaires, elles contribuent à l'émergence de nouvelles perspectives dans le domaine de l'innovation par les usages culinaires. Elles permettent d'étudier la diversité des gestes culinaires réalisés par les usagers ainsi que d'analyser la subjectivité du jugement apporté par les évaluateurs non-experts. Le développement des deux méthodologies a été axé autour de la définition de protocoles de recueil et d'analyse statistique de données adaptés.

L'introduction de ce manuscrit expose le contexte de ce travail de recherche. La définition de ce dernier a constitué une grande partie du travail entrepris. En effet, une part importante de cette thèse a consisté à établir une analogie entre les problématiques rencontrées en innovation par les usages culinaires - motivations à l'initiative de ce projet - et celles identifiées en analyse sensorielle. Dans ce chapitre, nous proposons donc des rappels et des généralités relevant de ces deux domaines. Cette analogie a ensuite permis de mettre en place une démarche de recherche adaptée. Elle consiste à s'inspirer de deux méthodes de caractérisation sensorielle : le tri libre (dans le cadre de la caractérisation perceptuelle) et la Q-méthodologie binaire (dans le cadre de la caractérisation conceptuelle). Des développements méthodologiques sont alors mis en place pour pallier aux limites qui leur sont associées : incapacité du protocole de recueil de données à gérer un grand nombre de stimuli, inexistence d'un protocole d'analyse des données adapté, etc. La démarche de recherche consiste enfin à étudier l'applicabilité de ces méthodes de caractérisation sensorielle 'améliorées' lorsqu'elles sont mises en œuvre sur un grand nombre de gestes culinaires enregistrés sous format vidéo.

Le Chapitre 1 présente une méthode de caractérisation sensorielle adaptée à la problématique d'évaluation perceptuelle de grands ensembles de stimuli. Dans un contexte de tri libre, un nombre important de stimuli rend difficile la tâche de caractérisation pour les évaluateurs. Les problèmes de saturation de mémoire et de fatigue entraînés par le nombre de stimuli expliquent cette difficulté. Pour s'affranchir de cette saturation cognitive, nous proposons de nous appuyer sur un protocole de recueil de données original tirant profit des solutions mises en place pour gérer des grands espaces produits en tri libre (c.-à-d. la présentation d'un sous-ensemble de stimuli par évaluateur et l'utilisation de prototypes fixés). La méthode proposée consiste à diviser les stimuli en plusieurs sous-ensembles de taille restreinte. Le premier sous-ensemble est caractérisé par l'évaluateur au travers d'une tâche de tri libre verbalisé. À la fin de cette première phase, chaque groupe de stimuli est ainsi synthétisé au travers de ses caractéristiques perceptuelles. Ces dernières constituent des prototypes fictifs. Lors de la deuxième phase, les stimuli des sous-ensembles restants sont agglomérés à ces prototypes. Ce protocole s'apparente à une tâche de tri verbalisé lors de laquelle les stimuli sont présentés séquentiellement et agglomérés les uns aux autres. Pour cette raison, cette nouvelle méthode de caractérisation a été dénommée 'Tri Séquentiel Agglomératif' (TSA). Une étude montre que cette approche donne de bons résultats dans une situation impliquant un espace produit facile à caractériser. Il reste cependant à poursuivre son processus de validation. Une nouvelle étude doit être réalisée pour étudier les performances du TSA dans une situation impliquant un espace produit difficile à évaluer. Par ailleurs, des démarches sont à entreprendre pour déterminer un plan de présentation des stimuli présentant des propriétés optimales. Même si pendant cette thèse nous nous sommes toujours assurés que chaque couple de stimuli était évalué de manière homogène sur la globalité du panel d'évaluateurs à chaque étape du TSA, nous ne nous sommes pas encore penchés sur la définition d'un plan d'expérience spécifiquement dédié à ce type de tâche de caractérisation. Enfin, nous envisageons de tester le TSA sur de nouvelles familles de stimuli, comme par exemple des produits alimentaires.

Le Chapitre 2 est centré autour du développement d'une procédure d'analyse statistique des données de TSA. Au préalable, nous rappelons le format original qu'elles présentent. Nous abordons ensuite la manière avec laquelle il est possible d'analyser ces données pour obtenir des résultats orientés stimuli ainsi que des résultats orientés évaluateurs. Concernant l'approche orientée stimuli, la procédure statistique proposée consiste dans un premier temps à analyser les partitions finales comme des données de tri libre verbalisé usuelles. Pour cela, une Analyse des Correspondances Multiples (ACM) est implémentée de manière à obtenir une représentation superposée des stimuli et des ensembles de caractéristiques utilisés par les évaluateurs. Cette représentation euclidienne est ensuite améliorée grâce à des techniques dédiées à l'analyse de données textuelles. Ces différentes techniques statistiques n'ont pas été développées dans le cadre de cette thèse. Il s'agit

cependant de la première fois que leur combinaison dans un contexte d'analyse de données de tri libre verbalisé est décrite. La procédure statistique orientée stimuli repose dans un second temps sur l'analyse des partitions intermédiaires. Pour cela, une Analyse Factorielle des Correspondances (AFC) est appliquée à chaque ensemble de partitions intermédiaires correspondant à une étape donnée du TSA. La contribution des stimuli à la construction des deux premiers axes factoriels est alors proposée comme un indicateur numérique de leur particularité perceptuelle. L'étude de son évolution dans le temps permet de nuancer certaines conclusions tirées de l'analyse statistique des partitions finales seules. Concernant l'approche orientée participants, la procédure statistique proposée consiste à analyser les partitions finales comme des données de tri libre verbalisé usuelles. Pour cela, une démarche de classification non supervisée des partitions fournies par les évaluateurs est implémentée. Basée sur un indice de dissemblance entre partitions, elle permet d'extraire des clusters homogènes d'évaluateurs ayant une perception similaire des stimuli. L'interprétation des clusters d'évaluateurs est ensuite facilitée grâce à l'intégration de covariables sur les évaluateurs et grâce à la définition d'une partition consensus des stimuli pour chaque cluster d'évaluateurs. Bien que cette dernière démarche ait déjà été décrite, nous proposons de l'approfondir grâce à l'intégration originale des caractéristiques perceptuelles utilisées par les évaluateurs du cluster pour décrire les stimuli. Cette proposition, ici utilisée pour analyser des données de TSA, offre de nouvelles perspectives pour l'analyse des données de tri libre verbalisé. Bien que la procédure d'analyse orientée évaluateurs permettent d'obtenir des résultats opérationnels, certaines interrogations restent cependant en suspens. La plus importante d'entre elles repose sur la possibilité de valoriser les partitions intermédiaires fournies par les évaluateurs. Une telle utilisation des partitions intermédiaires permettrait de prendre en compte la cinétique de perception des stimuli dans le processus de classification. Deux évaluateurs seraient ainsi dans le même cluster s'ils ont globalement perçu les stimuli de la même façon tout au long du TSA, et non uniquement à la fin de la tâche. Des réflexions méthodologiques doivent être mises en place pour répondre à cette problématique, notamment pour pallier aux limites liées à la présence de données manquantes au sein des partitions intermédiaires. Le développement de solutions spécifiques répondra également à d'autres problématiques rencontrées en analyse sensorielle. Ces solutions permettront effectivement de segmenter un panel d'évaluateurs ayant réalisé une tâche de tri libre impliquant un plan de présentation des stimuli incomplet. L'intégralité de la procédure d'analyse statistique décrite dans ce chapitre est implémentée dans le package R `SeqAgSorting`.

Le Chapitre 3 propose une application du TSA pour la caractérisation perceptuelle de gestes culinaires. Dans ce cadre, une étude réalisée sur 27 recettes de tarte au citron meringuée est présentée. Elle montre que le TSA est applicable à un ensemble de gestes culinaires. En plus de fournir des résultats intéressants sur les gestes évalués, elle démontre que des évaluateurs non-experts sont capables de fournir une évaluation sur des gestes d'utilisateurs. Du point de vue des résultats, cette

étude souligne l'importance de l'expertise de l'évaluateur dans la caractérisation fournie sur les gestes. Malgré les résultats encourageants de cette étude, de nouvelles démarches doivent être mises en place. En particulier, il est nécessaire d'appliquer le TSA à d'autres ensembles de gestes culinaires ; et notamment à des gestes culinaires recueillis dans une situation écologique (et non dans une situation proche du laboratoire comme cela est le cas dans l'étude présentée).

Le Chapitre 4 est axé sur la Q-méthodologie binaire. Cette méthode de caractérisation sensorielle avait déjà démontré ses performances pour l'évaluation conceptuelle de grands ensembles de stimuli. Sa mise en œuvre opérationnelle était pourtant limitée par l'inexistence d'une procédure d'analyse statistique adaptée à la nature des données collectées. Pour pallier à cette limite, nous proposons dans ce chapitre une approche statistique fournissant des résultats orientés stimuli. Grâce à des jeux de données pédagogiques, nous illustrons que l'Analyse des Correspondances Multiples (ACM) - technique pouvant être utilisée dans ce cadre - ne prend pas totalement en compte la nature des données. Elle considère en effet les données de Q-méthodologie binaire comme des données de tri libre. En ce sens, elle omet que les stimuli sont associés à des catégories prédéfinies. Nous voyons ensuite comment l'Analyse Factorielle des Correspondances (AFC) prend en compte cette association. Enfin, nous voyons comment l'Analyse Factorielle Multiple (AFM) permet de combiner de manière harmonieuse la distance inter-stimuli de l'ACM et celle de l'AFC à travers une distance globale. Toujours grâce aux jeux de données pédagogiques, nous montrons que cette distance prend en compte la nature des données de Q-méthodologie binaire. Nous proposons également dans ce chapitre une procédure de visualisation facilitant l'interprétation de la représentation euclidienne des stimuli. La représentation graphique obtenue est similaire aux représentations de type *preference mapping* largement utilisées en analyse sensorielle. Notre approche présente cependant la particularité de gérer des données externes de nature binaire. Cet outil original apporte de nouvelles perspectives méthodologiques aux analystes sensoriels. Il permet par exemple de remplacer dans leur *preference mapping* les données hédoniques de nature continue par des données binaires. L'intégralité de la procédure d'analyse statistique décrite dans ce chapitre est implémentée dans le package R `BinQmet`.

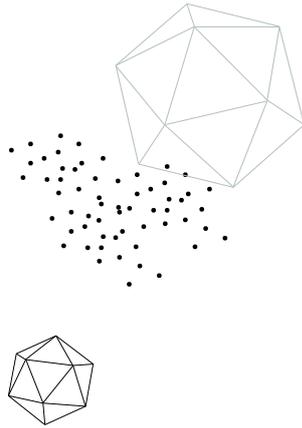
Le Chapitre 5 présente la poursuite du développement d'une procédure d'analyse statistique adaptée à la nature des données de Q-méthodologie binaire. Nous proposons cette fois-ci une approche fournissant des résultats orientés évaluateurs. Après avoir présenté les principaux indices d'accord inter-évaluateurs utilisés en psychologie, nous démontrons qu'ils ne permettent pas d'analyser la structure de désaccord entre les évaluateurs de manière optimale. Pour pallier à ce problème, nous proposons d'utiliser un modèle de régression logistique à classes latentes ajusté sur l'intégralité des profils d'évaluations binaires. Une étude comparative basée sur des simulations démontre les bonnes performances de cette approche. Elle permet en effet de prendre en compte des structures de désaccord plus générales que les

méthodes existantes basées sur des indices synthétiques d'accord inter-évaluateurs. Nous montrons ensuite comment le cadre modèle proposé peut être utilisé dans un contexte plus général de classification non supervisée. Il permet de définir une distance inter-évaluateurs ainsi qu'une procédure de sélection de modèle définissant le nombre de clusters homogènes d'évaluateurs au sein du panel. Malgré les bonnes performances de cette dernière, d'autres alternatives doivent être explorées. Il est par exemple envisagé de comparer cette procédure à d'autres approches permettant la classification non supervisée de profils binaires. C'est par exemple le cas des modèles des blocs latents. Par ailleurs, la stratégie de coupure du dendrogramme que nous proposons pourrait être comparée à des techniques basées sur des critères plus classiques de sélection de modèles tels que le BIC. L'interprétation des clusters d'évaluateurs est ensuite facilitée grâce à l'intégration de covariables sur les évaluateurs et sur les stimuli. L'intégralité de la procédure d'analyse statistique décrite dans ce chapitre est implémentée dans les package R `AgreeClust` et `BinQmet`.

Le Chapitre 6 propose une application de la Q-méthodologie binaire pour la caractérisation conceptuelle de gestes culinaires. Dans ce cadre, une étude réalisée sur 39 gestes d'ouverture d'huîtres est présentée. Elle montre que la Q-méthodologie binaire est applicable à un ensemble de gestes culinaires. En plus de fournir des résultats intéressants sur les gestes évalués, elle démontre que des évaluateurs non-experts sont capables de fournir une évaluation sur des gestes d'usagers. Du point de vue des résultats, cette étude souligne l'importance de l'expertise de l'utilisateur sur la qualité du geste culinaire réalisé. L'expertise de l'évaluateur est également déterminante sur la caractérisation des gestes fournies. Ces résultats sont en concordance avec ceux tirés de l'étude présentée dans le Chapitre 3. De nouvelles études réalisées sur d'autres types de gestes culinaires doivent être mises en œuvre pour confirmer l'intérêt de la Q-méthodologie binaire dans ce cadre.

En conclusion, ce travail de recherche en statistique appliquée à l'agroalimentaire a été l'occasion de se focaliser sur des aspects méthodologiques statistiques (Chapitres 1, 2, 4 et 5) et sur des aspects très appliqués relatifs à la caractérisation des usages culinaires (Chapitres 3 et 6). La part la plus importante de ce doctorat a consisté à développer les propositions méthodologiques, relatives à la fois au recueil et à l'analyse statistique des données. Ceci peut s'expliquer par un objectif de formation assumé en statistique - et plus particulièrement en sensométrie - en accord avec mon projet professionnel. Bien que l'application des méthodes développées aux gestes culinaires ait été réalisée de manière moins poussée, les études menées lors de ce doctorat ont fourni des résultats encourageants. Pour la première fois, elles ont démontré que l'étude des usages pouvait être réalisée grâce à des méthodes de caractérisation relevant de la recherche quantitative. Au-delà de fournir des informations jusqu'alors inaccessibles (p. ex. la prédominance de certains usages au sein de la population), cette nouvelle approche fournit à l'expérimentateur des outils graphiques pour l'interprétation des usages étudiés. La présentation des résultats sous une telle forme est innovante dans ce domaine relevant classiquement

de la recherche qualitative. Une interprétation sociologique ou éthologique poussée nécessaire à la validation des méthodes développées reste cependant à faire. Cette démarche n'a effectivement pas été entreprise dans le cadre de ce doctorat.



LISTE DES TRAVAUX

Publications réalisées dans le cadre de la thèse - Revue à comité de lecture

Brard, M., & Lê, S. The Sequential Agglomerative Sorting Task, a new sensory methodology to characterize large sets of products. *Submitted*.

Brard, M., Lê, S., & Causeur, D. A latent class regression framework for modeling agreement: clustering of multivariate profiles of binary ratings based on a new distance and a model selection strategy. *Submitted*.

Brard, M., & Lê, S. (2018). Adaptation of the Q-methodology for the characterization of a complex concept through a set of products: from the collection of the data to their analysis. *Food Quality and Preference*, 67, 77-86.

Distinctions et bourses obtenues dans le cadre de la thèse

Bourse à la mobilité décernée par la commission scientifique de la Société Française d'Analyse Sensorielle (SFAS) [750€]

Bourse à la mobilité décernée par la commission scientifique d'AGROCAMPUS OUEST [1500€]

Best Young Researcher Award du congrès Sensometrics, décerné par XLSTAT [1000€]

Communications réalisées dans le cadre de la thèse - Conférences

Le nom de l'orateur est indiqué avec un astérisque. Le format de la présentation (sous format oral ou sous format poster) est précisé entre parenthèses.

Brard, M.*, & Lê, S. Qu'est-ce qu'un bon geste culinaire ? Caractérisation du concept de bon geste à travers un ensemble de vidéos de gestes d'ouvertures d'huîtres. *Workshop de la Société Française d'Analyse Sensorielle*, Lyon (France), 10 novembre 2017. [Présentation orale]

Brard, M.*, & Lê, S. What makes a good culinary gesture ? Characterization of the concept of good gesture through a set of videos representing openings of oysters gestures. *Pangborn symposium*, Providence (États-Unis d'Amérique), 20 - 24 août 2017. [Présentation poster]

Brard, M.*, & Lê, S. The Sequential Agglomerative Sorting Task, a new methodology for the sensory characterization of big sets of stimuli: Application to the definition of perceptual characteristics of sounds of trams driving noise annoyance in urban areas. *Pangborn symposium*, Providence (USA), 20-24 août 2017. [Présentation poster]

Brard, M.*, Lê, S., & Causeur, D. Modeling agreement for the clustering of multivariate binary profiles. *Journées de Statistique de la Société Française de Statistique (JDS)*. Avignon (France), 29 mai - 2 juin 2017. [Présentation orale]

Brard, M.*, & Lê, S. The Sequential Agglomerative Sorting task, a new methodology for the sensory characterization of big sets of stimuli: Highlighting the main features perceived by consumers in a set of videos of culinary recipes. *Meeting annuel de la European Sensory Science Society (E3S)*. Paris (France), 9 - 10 mai 2017. [Présentation orale flash = Présentation poster]

Brard, M.*, & Lê, S. The Sequential Agglomerative Sorting task, a new methodology for the perceptual characterization of big sets of stimuli: Highlighting the main features perceived by consumers in a big set of culinary recipes. *Food Innovation Asian Conference (FIAC)*. Bangkok (Thaïlande), 15 - 16 juin 2017. [Présentation orale]

Brard, M.*, & Lê, S. The QC-methodology: adapting the Q-methodology for analyzing concepts through categorical data. *Meeting de la Society for Computers in Psychology (SCiP)*. Boston (États-Unis d'Amérique), 17 november 2017. [Présentation orale]

Brard, M.*, & Lê, S. The QC-methodology as a way for understanding some

complex concepts: An innovative case study. *Eurosense*. Dijon (France), 12 - 14 septembre 2016. [Présentation poster]

Brard, M.*, & Lê, S. The QC-methodology: twisting the Q-methodology for understanding some complex concept. *Sensometrics*, Brighton (Grande-Bretagne), 26 - 29 juillet 2016. [Présentation orale]

Brard, M., & Lê, S*. The QC-methodology as a way for understanding some complex concept: an innovative case study. *Summer Program in Sensory Evaluation (SPISE)*, Ho Chi Minh City (Vietnam), 26 - 29 juillet 2016. [Présentation orale]

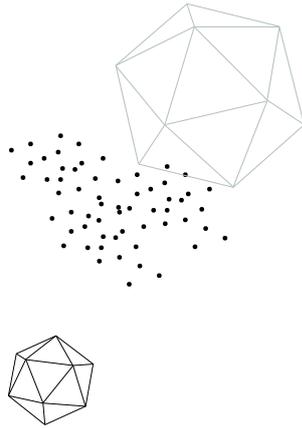
Publications réalisées dans un cadre extérieur à celui de la thèse - Revues à comité de lecture

Lê, T.M., Brard, M., & Lê, S. (2017). Holos: a collaborative environment for similarity-based holistic approaches. *Behavior Research Methods*, 49(5), pages 1597 - 1604.

Brard, M., & Lê, S. (2016). The Ideal Pair Method, an alternative to the Ideal Profile Method based on pairwise comparisons: application to a panel of children. *Journal of Sensory Studies*, 31(4), pages 306 - 313.

Révisions réalisées dans un cadre extérieur à celui de la thèse - Revues à comité de lecture

Réviseur pour *Journal of Sensory Studies* (2016).



BIBLIOGRAPHIE

- Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. In *Encyclopedia of measurement and statistics* (p. 651-657). Thousand Oaks, CA: Sage.
- Abdi, H., Valentin, D., Chollet, S., & Chr ea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*, 18, 627-640.
- Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). *Advantages and uses check-all-that-apply response compared to traditional scaling of attributes for salty snacks*. (7th Pangborn Sensory Science Symposium, Minneapolis, MN, USA, August 2007)
- Adejumo, A., Heumann, C., & Toutenburg, H. (2004). *A review of agreement measure as a subset of association measure between raters*. (SFB386-Discussion Paper 385, Ludwig-Maximilians-Universitat, Munchen, Germany)
- Agresti, A. (1992). Modelling patterns of agreement and disagreement. *Statistical Methods in Medical Research*, 1, 201-218.
- Bailey, J. (2008). First steps in qualitative data analysis: transcribing. *Family Practice*, 25, 127-131.
- Ballester, J., Dacremont, D., Fur, Y. L., & Eti evant, P. (2005). The role of olfaction in the elaboration and use of the chardonnay wine concept. *Food Quality and Preference*, 16, 351-359.
- Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1), 3-23.
- Bayona-Saez, C., Cruz-C azares, C., Garc ia-Marco, T., & Garc ia, M. S. (2017). Open innovation in the food and beverage industry. *Management Decision*, 55, 526-546.

- Bech, S., & Zacharov, N. (2006). *Perceptual audio evaluation: Theory, method and application*. John Wiley & Sons: Chichester.
- Bécue-Bertaut, M., & Pagès, J. (2005). Clustering units from frequency and nominal variables: Definition of a global distance. application to survey data with closed and open-ended questions. In *Applied stochastic models and data analysis* (p. 89-97). Brest.
- Bécue-Bertaut, M., & Pagès, J. (2007). Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics & Data Analysis*, *52*, 3255-3268. Retrieved from doi:10.1016/j.csda.2007.09.023
- Benzécri, J.-P. (1992). *Correspondence analysis handbook*. New York: Marcel Dekker.
- Bibliardi, B., & Galati, F. (2016). Which factors hinder the adoption of open innovation in smes? *Technology Analysis & Strategic Management*, *28*, 869-885.
- Blumenthal, D., & Herberth, N. (2015). Use of rapid sensory methods in the automotive industry. In *Rapid sensory profiling techniques and related methods: Applications in new product development and consumer research*. Cambridge: Woodhead Publishing.
- Boinbaser, L., Parente, M., Castura, J., & Ares, G. (2015). Dynamic sensory characterization of cosmetic creams during application using temporal check-all-that-apply (tcata) questions. *Food Quality and Preference*, *45*, 33-40.
- Bonnin, G. (1998). *L'observation des comportements au point de vente : vers une étude de la valorisation de la visite au magasin*. Disponible sur le site http://leg.u-bourgogne.fr/images/stories/pdf/doc_trav1998/g1998-04.pdf. (Consulté le : 11/03/2016)
- Brard, M., & Lê, S. (2016). The ideal pair method, an alternative to the ideal profile method based on pairwise comparisons: Application to a panel of children. *Journal of Sensory Studies*, *31*, 306-313.
- Brard, M., Lê, S., & Causeur, D. (2017). AgreeClust: agreement-based unsupervised clustering of a set of ratings through a latent class regression modeling framework [Computer software manual].
- Brickenkamp, R., & Zillmer, E. (1998). *d2: Test of attention*. Seattle: Hogrefe and Huber.
- Brown, S. (1993). A primer on Q methodology. *Operant Subjectivity*, *16*, 91-138.

- Brown, S. (2000). Q methodology in assessment and research [Computer software manual]. Retrieved from <http://schmolck.userweb.mwn.de/qmethod/syllabus08.pdf> (Workshop notes. Presented Summer 2000. 22 May-2 June)
- Bruzzese, D., & Vistocco, D. (2015). Despotia : Dendrogram slicing through a permutation test approach. *Journal of Classification*, *32*, 285-304.
- Bruzzone, F., Ares, G., & Giménez, A. (2012). Consumers' texture perception of milk desserts: Comparison with trained assessors' data. *Journal of Texture Studies*, *43*, 214-226.
- Bucher, T., Collins, C., Diem, S., & Siegrist, M. (2016). Adolescents' perception of the healthiness of snacks. *Food Quality and Preference*, *50*, 94-101.
- Buck, D. (2007). Methods to understand consumer attitudes and motivations in food product development. In H. MacFie (Ed.), *Consumer-led food product development*. Boca Raton: CRC Press.
- Busse, M., & Siebert, R. (2017). The role of consumers in food innovation process. *European Journal of Innovation Management*, *21*, 20-43.
- Cadoret, M. (2010). *Analyse factorielle de données de catégorisation, application aux données sensorielles* (Unpublished doctoral dissertation). Université Européenne de Bretagne & AGROCAMPUS OUEST.
- Cadoret, M., Lê, S., & Pagès, J. (2009). A factorial approach for sorting task data (fast). *Food Quality and Preference*, *20*, 410-417.
- Cariou, V., & Qannari, E. (2018). Statistical treatment of free sorting data by means of correspondence and cluster analyses. *Food Quality and Preference*, *In press*.
- Carr, B., Craig-Petsinger, D., & Hadlich, S. (2001). A case study in relating sensory descriptive data to product concept fit and consumer vocabulary. *Food Quality and Preference*, *12*, 407-412.
- Carroll, J., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart Young" decomposition. *Psychometrika*, *35*, 283-319.
- Carvalho, F., & Spence, C. (2018). The shape of the cup influences aroma, taste, and hedonic judgements of specialty coffee. *Food Quality and Preference*, *In Press*.
- Cavazzana, A., Larsson, M., Hoffmann, E., Hummel, T., & Haeh, A. (2017). The vessel's shape influences the smell and taste of cola. *Food Quality and Preference*, *59*, 8-13.

- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*, 1-36.
- Chollet, S., Lelièvre, M., Abdi, H., & Valentin, D. (2011). Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality and Preference*, *22*, 507-520.
- Chollet, S., Valentin, D., & Abdi, H. (2014). Free sorting task. In P. Varela & G. Ares (Eds.), *Novel techniques in sensory characterization and consumer profiling* (p. 207-227). CRC Press.
- Cicchetti, D., & Feinstein, A. (1990). High agreement but low kappa: II. resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 551-558.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological measurement*, *20*(1), 37-46.
- Costa, A., & Jongen, W. (2006). New insights into consumer-led food product development. *Trends in Food Science and Technology*, *17*, 457-465.
- Courcoux, P. (2014). Freesortr : Free sorting data analysis [Computer software manual].
- Courcoux, P., Faye, P., & Qannari, E. (2014). Determination of the consensus partition and cluster analysis of subjects in a free sorting task experiment. *Food Quality and Preference*, *32*, 107-112.
- Courcoux, P., Qannari, E., & Faye, P. (2015). Free sorting as a sensory profiling technique for product development. In *Rapid sensory profiling techniques and related methods: Applications in new product development and consumer research*. Cambridge: Woodhead Publishing.
- Coxon, A. (1999). *Sorting data: Collection and analysis*. California: Sage Publications.
- Danzart, M., Sieffermann, J.-M., & Delarue, J. (2004). *New developments in preference mapping techniques: Finding out a consumer optimal product, its sensory profile and the key sensory attributes*. (7th Sensometrics Meeting, David, CA, USA, August 2004)
- Davis, C., & Michelle, C. (2011). Q methodology in audience research: bridging the qualitative/quantitative 'divide'? *Food of Audience & Reception Studies*, *8*.
- de Jong, J., & von Hippel, E. (1978). Transfers of user process innovations to producers: A study of dutch high tech firms. *Research Policy*, *38*, 1181-1191.

- Delarue, J. (2015). The use of rapid sensory methods in r&d and research: an introduction. In *Rapid sensory profiling techniques and related methods: Applications in new product development and consumer research*. Cambridge: Woodhead Publishing.
- Derndorfer, E., & Baierl, A. (2014). Multidimensional scaling (mds). In *Mathematical and statistical methods in food science and technology*. Chichester: Wiley.
- de Saldamando, L., Delgado, J., Herencia, P., Giménez, A., & Ares, G. (2013). Polarized Sensory Positioning: Do conclusions depend on the poles? *Food Quality and Preference*, *29*, 25-32.
- Di Eugenio, B., & Glass, M. (2004). The kappa statistic: a second look. *Computational linguistics*, *30*(1), 95-101.
- Dimitriadou, E., Dolnicar, S., & Weingessel, A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika*, *67*, 137-160.
- Douthwaite, B., Keatinge, J., & Park, J. (2001). Why promising technologies fail: the neglected role of user innovation during adoption. *Research Policy*, *30*, 819-836.
- Faye, P., Brémaud, D., Daubin, M. D., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mapping. *Food Quality and Preference*, *15*, 781-791.
- Faye, P., Courcoux, P., Giboreau, A., & Qannari, E. (2013). Assessing and taking into account the subjects' experience and knowledge in consumer studies. application to the free sorting of wine glasses. *Food Quality and Preference*, *28*, 317-327.
- Faye, P., Courcoux, P., Qannari, E., & Giboreau, A. (2011). Méthodes de traitement statistique des données issues d'une épreuve de tri libre. *MODULAD*, *43*, 1-24.
- Fleiss, J. (1975). Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, *31*, 651-659.
- Fleiss, J., Levin, B., & Paik, M. (2003). *Statistical methods for rates and proportions* (3rd ed.). John Wiley & Sons, Inc.
- Fleming, E., Ziegler, G., & Hayes, J. (2015). Check-all-that-apply (cata), sorting and polarized sensory positioning (psp) with astringent stimuli. *Food Quality and Preference*, *45*, 41-49.

- Fortuin, T., & Omta, S. (2009). Innovation drivers and barriers in food processing. *British Food Journal*, *111*, 839-851.
- Galanakis, C. (2016). Preface. In *Innovation strategies in the food industry: Tools for implementation*. London: Elsevier.
- Garcez, A., Duarte, R., & Eisenberg, Z. (2011). Production and analysis of video recordings in qualitative research. *Educação e Pesquisa*, *37*, 249-261.
- Giboreau, A., Navarro, S., Faye, P., & Dumortier, J. (2001). 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*, 311-322.
- Gisev, N., Bell, J., & Chen, T. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *9*, 330-338.
- Goodman, L., & Kruskal, W. (1954). Measures of association for cross classifications. *The Journal of the American Statistical Association*, *49*, 732-765.
- Gower, J.-C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*, 33-51.
- Greenacre, M. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
- Grunert, K., Jensen, B., Sonne, A.-M., Brunso, K., Byrne, D., Clausen, C., . . . Scholderer, J. (2008). User-oriented innovation in the food sector: relevant streams of research and an agenda for future work. *Trends in Food Science and Technology*, *19*, 590-602.
- Gwet, K. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*, 29-48.
- Gwet, K. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics.
- Gwet, K. (2017). *R functions for calculating agreement coefficients*. From: http://www.agreestat.com/r_functions.html. (Retrieved : 24/03/2017)
- Hands, S., & Everitt, B. (1987). A monte carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Prevention Science*, *16*, 1007-1016.
- Heide, M., & Olsen, S. (2017). Influence of packaging attributes on consumer evaluation of fresh cod. *Food Quality and Preference*, *60*, 9-18.

- Henry, D., Dymnicki, A., Mohatt, N., Allen, J., & Kelly, J. (2015). Clustering methods with qualitative data: a mixed-methods approach for prevention research with small samples. *Prevention Science, 16*, 1007-1016.
- Heymann, H. (2017). *Pushing the limits*. (12th Pangborn Sensory Science Symposium, Providence, RI, USA, August 2017)
- Homa, D. (1984). On the nature of categories. In *The psychology of learning and motivation*. Academic Press: Orlando.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*, 193-218.
- Husson, F., Lê, S., & Cadoret, M. (2017). Sensominer: Sensory data analysis [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=SensOMineR> (R package version 1.23)
- Husson, F., Lê, S., & Pagès, J. (2011). *Exploratory multivariate analysis by example using r*. CRC Press.
- Husson, F., Le Dien, S., & Pagès, J. (2001). Which value can be granted to sensory profiles given by consumers? methodology and results. *Food Quality and Preference, 12*, 291-296.
- Jaeger, S., Hort, J., Porcherot, C., Ares, G., Pecore, S., & MacFie, H. (2017). Future directions in sensory and consumer science: Four perspectives and audience voting. *Food Quality and Preference, 56*, 301-309.
- Jaffré, J., Valentin, D., Dacremont, C., & Peyron, D. (2009). Burgundy red wines: Representation of potential for aging. *Food Quality and Preference, 20*, 505-513. Retrieved from doi:10.1016/j.foodqual.2009.05.001
- Kergoat, M., Giboreau, A., Nicod, H., Faye, P., Diaz, E., Beetschen, M.-A., & Meyer, T. (2012). Consumer preference for tactile softness: A question of affect intensity? *Journal of Sensory Studies, 27*, 232-246.
- Kim, J., & Mueller, C. (1978). *Factor analysis : statistical methods and practical issues*. Newbury Park: Sage Publications.
- Kostov, B. (2015). *A principal component method to analyse disconnected frequency tables by means of contextual information* (Unpublished doctoral dissertation). Universitat Politècnica de Catalunya.
- Kostov, B., Bécue-Bertaut, M., & Husson, F. (2014). An original methodology for the analysis and interpretation of word-count based methods: Multiple factor analysis for contingency tables complemented by consensual words. *Food Quality and Preference, 32*, 35-40.

- Kruskal, J. (1964). Nonmetric multidimensionnal scaling: A numerical method. *Psychometrika*, *29*, 115-129.
- Kuhn, M. (2017). caret: Classification and regression training [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=caret> (R package version 6.0-77)
- Lacour, J.-P. (2004). Observation éthologique du consommateur dans une grande surface d'articles de sport. *Revue Française du Marketing*, *199*, 39-48.
- Lawless, H. (1989). Exploration of fragrance categories and ambiguous odors using multidimensionnal scaling and cluster analysis. *Chemical Senses*, *14*, 349-360.
- Lawless, H., & Heymann, H. (2010). *Sensory evaluation of food* (2nd ed.). Springer.
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1-18. doi: 10.18637/jss.v025.i01
- Lê, S., & Worch, T. (2015). *Analyzing sensory data with R*. Boca Raton: CRC Press.
- Lê, T. (2015). *Réflexions méthodologiques autour du napping : vers une intégration du comportement du sujet dans l'analyse des données de napping* (Unpublished doctoral dissertation). Université Européenne de Bretagne & AGRO-CAMPUS OUEST.
- Lebart, L. (1975). L'orientation du dépouillement de certaines enquêtes par l'analyse des correspondances multiples. *Consommation*, *2*, 73-96.
- Lee, H.-S., & O'Mahony, M. (2005). Sensory evaluation and marketing: measurement of a consumer concept. *Food Quality and Preference*, *16*, 227-235.
- Lefebvre, A., Bassereau, J., Pensé-Lhéritier, A.-M., Rivière, C., Harris, N., & Duchamp, R. (2010). Recruitment and training of a sensory expert panel to measure the touch of beverage packages: issue and methods employed. *Food Quality and Preference*, *21*, 156-164.
- Leonard, D., & Rayport, J. (1997). Spark innovation through empathic design. *Harvard Business Review*, *75*, 102-113.
- Liu, Y., Hayes, D., Nobel, A., & Marron, J. (2008). Statistical significance of clustering for high-dimension, low-sample size data. *Journal of the American Statistical Association*, *103*, 1281-1293.
- Luthje, C., Herstatt, C., & von Hippel, E. (2005). User-innovators and 'local' information: The case of mountain biking. *Research Policy*, *34*, 951-965.

- MacFie, H. (2007). Preface. In *Consumer-led food product development*. Boca Raton: CRC Press.
- Mammasse, N., & Schlich, P. (2014). Adequate number of consumers in a liking test. insights from resampling in seven studies. *Food Quality and Preference*, *31*, 124-128.
- Maxwell, A., & Pilliner, A. (1968). Deriving coefficients of reliability and agreement for ratings. *The British Journal of Mathematical and Statistical Psychology*, *21*, 105-116.
- McClain, J., & Rao, V. (1975). Clustisz: A program to test for the quality of clustering of a set of objects. *Journal of Marketing Research*, *12*, 456-460.
- Mirabito, A., Oliphant, M., Doorn, G. V., Watson, S., & Spence, C. (2017). Glass shape influences the flavour of beer. *Food Quality and Preference*, *62*, 257-261.
- Moskowitz, H. (1996). Experts versus consumers: A comparison. *Journal of Sensory Studies*, *11*, 19-37.
- Moussour, M., Lavarde, M., Pensé-Lhéritier, A.-M., & Bouton, F. (2016). Sensory analysis of cosmetic powders: personal care ingredients and emulsions. *International Journal of Cosmetic Science*, *39*, 83-89.
- Murray, J., & Delahunty, C. (2000). Mapping consumer preference for the sensory and packaging attributes of cheddar cheese. *Food Quality and Preference*, *11*, 419-435.
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, *31*, 274-295.
- Van der Kloot, W., & Van Herck, H. (1991). Multidimensional scaling of sorting data: A comparison of three procedures. *Multivariate Behavioral Research*, *26*, 563-581.
- Pagès, J. (2003). Recueil direct de distances sensorielles : application à l'évaluation de dix vins blancs du val de loire. *Sciences des Aliments*, *23*, 679-688.
- Pagès, J. (2014). *Multiple factor analysis by example using r*. Boca Raton: CRC Press.
- Pagès, J., & Périnel, E. (2004). Panel performance and number of evaluations in a descriptive sensory study. *Journal of Sensory Studies*, *19*, 273-291.

- Park, P., Manjourides, J., Bonetti, J., & Pagano, M. (2009). A permutation test for determining significance of clusters with applications to spatial and gene expression data. *Computational Statistics & Data Analysis*, *53*, 4290-4300.
- Peryam, D., Pilgrim, F., & Peterson, M. (1954). *Food acceptance testing methodology*. (A symposium sponsored by the QMFCI, National Academy of Sciences—National Research Council)
- Petersen, J., Larsen, K., & Kreiner, S. (2010). Assessing and quantifying inter-rater variation for dichotomous ratings using a rasch model. *Statistical Methods in Medical Research*, *21*, 635-652.
- Picard, M., Tempere, S., de Revel, G., & Marchand, S. (2015). A sensory study of the ageing bouquet of red bordeaux wines: A three-step approach for exploring a complex olfactory concept. *Food Quality and Preference*, *42*, 110-122.
- R Core Team. (2017). *R: A language and environment for statistical computing. R foundation for statistical computing*. From: <http://www.R-project.org/>.
- Rand, W. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American statistical association*, *66*, 846-850.
- Rasch, G. (1961). *On general laws and the meaning of measurement in psychology*. Wiley, Berkley University Press.
- Ratkowsky, D., & Lance, G. (1978). A criterion for determining the number of groups in a classification. *Australian Computer Journal*, *10*, 115-117.
- Rogeaux, M., Lawlor, B., Punter, P., & Delarue, J. (2011). *Current status and future directions for alternative descriptive sensory methods workshop*. (9th Pangborn Sensory Science Symposium, Toronto, Canada, August 2011)
- Ryota, S., & Shimodaira, H. (2011). pvcust: Hierarchical clustering with p-values via multiscale bootstrap resampling [Computer software manual].
- Sarkar, S., & Costa, A. (2008). Dynamics of open innovation in the food industry. *Trends in Food Science and Technology*, *19*, 574-580.
- Schoen, A. (2017). Openness and collaboration in the food sector: mapping the field. *British Food Journal*, *119*, 2493-2506.
- Schwender, H. (2015). scime: Analysis of high-dimensional categorical data such as snp data [Computer software manual].
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2017). plotly: Create interactive web graphics via 'plotly.js' [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=plotly> (R package version 4.7.1)

- Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.
- Stewart-Knox, B., & Mitchell, P. (2003). What separates the winners from the losers in new food product development? *Trends in Food Science and Technology*, *14*, 58-64.
- Stone, H., & Sidel, J. (2004). *Sensory evaluation practices* (3rd ed.). Academic Press.
- Stone, H., Sidel, J., Oliver, S., Woolsey, A., & Singleton, R. (1974). Sensory evaluation by quantitative descriptive analysis. *Food Technology*, *28*, 24-34.
- Takane, Y. (1980). Analysis of categorizing behavior by a quantification method. *Behaviormetrika*, *8*, 75-86.
- Talavera, M., & Chambers, E. (2017). Using sensory sciences help products succeed. *British Food Journal*, *119*, 2130-2144.
- Teillet, E. (2015). Polarized sensory positioning (psp) as a sensory profiling technique. In *Rapid sensory profiling techniques and related methods: Application in new product development and consumer research*. Woodhead Publishing: Cambridge.
- Teillet, E., Schlich, P., Urbano, C., Cordelle, S., & Guichard, E. (2010). Sensory methodologies and the taste of water. *Food Quality and Preference*, *21*, 967-976.
- ten Klooster, P., Visser, M., & de Jong, M. (2008). Comparing two image research instruments: The q-sort method versus the likert attitude questionnaire. *Food Quality and Preference*, *19*, 511-518.
- Thomson, D. (2015). Expedited procedures for conceptual profiling of brands, products and packaging. In *Rapid sensory profiling techniques and related methods: Applications in new product development and consumer research*. Cambridge: Woodhead Publishing.
- Tijssen, I., Zandstra, E., de Graaf, C., & Jager, G. (2017). Why a 'light' product package should not be light blue: Effects of package colour on perceived healthiness and attractiveness of sugar- and fat-reduced products. *Food Quality and Preference*, *59*, 46-58.
- Trijp, H., & van Kleef, E. (2008). Newness, value and new product performance. *Trends in Food Science and Technology*, *19*, 562-573.

- van Kleef, E., & van Trijp, H. (2007). Opportunity identification in new product development and innovation in food product development. In *Consumer-led food product development*. Boca Raton: CRC Press.
- van Kleef, E., van Trijp, H., & Luning, P. (2005). Consumer research in the early stages of new product development: a critical review of methods and techniques. *Food Quality and Preference*, *16*, 181-201.
- Vezen, L. (1972). Rôle de l'exemple dans l'identification de concepts et l'acquisition des connaissances. *L'année psychologique*, *72*, 463-486.
- von Hippel, E. (1976). The dominant role of users in the scientific instrument innovation process. *Research Policy*, *5*, 212-239.
- von Hippel, E. (1978). Successful industrial products from customer ideas. *Journal of Marketing*, *42*, 39-49.
- von Hippel, E. (1988). *Sources of innovation*. New York: Oxford University Press.
- von Hippel, E. (2005). *Democratizing innovation*. Cambridge: MIT Press.
- von Hippel, E., & Katz, R. (2002). Shifting innovation to users via toolkits. *Management Science*, *48*, 821-833.
- Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.
- Warrens, M. (2008). *Similarity coefficients for binary data* (Unpublished doctoral dissertation). Leiden University.
- Warrens, M. (2015). Five ways to look at cohen's kappa. *Journal of Psychology and Psychotherapy*, *5*.
- Watts, S., & Stenner, P. (2005). Doing Q methodology: theory, method, and interpretation. *Qualitative Research in Psychology*, *2*, 67-91.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>
- Worch, T. (2012). *L'analyse du profil idéal : de la validation à l'analyse statistique des données issues du profil idéal* (Unpublished doctoral dissertation). Université Européenne de Bretagne & AGROCAMPUS OUEST.
- Worch, T., Crine, A., Gruel, A., & Lê, S. (2014). Analysis and validation of the ideal profile method: Application to a skin cream study. *Food Quality and Preference*, *32*, 132-144.

- Worch, T., Lê, S., & Punter, P. (2010). How reliable are consumers? comparison of sensory profiles from consumers and experts data. *Food Quality and Preference*, *21*, 309-318.
- Zhang, T., Lusk, K., Miroso, M., & Oey, I. (2016). Understanding young immigrant chinese consumers' freshness perceptions of orange juices: A study based on concept evaluation. *Food Quality and Preference*, *48*, 156-165.
- Zhu, Y., Zeng, X., Koehl, L., Lageat, T., Charbonneau, A., & Chaigneau, C. (2010). A general methodology for analyzing fashion oriented textile products using sensory evaluation. *Food Quality and Preference*, *21*, 1068-1076.

Titre : Méthodologies statistiques pour le recueil et l'analyse de données de caractérisation de grands ensembles de stimuli – Application à la caractérisation sensorielle de produits & à la caractérisation de gestes culinaires

Mots clés : méthodologies statistiques, grands ensembles de stimuli, caractérisation sensorielle, caractérisation de gestes culinaires

Résumé : Dans le secteur agroalimentaire, la caractérisation des gestes culinaires est considérée de plus en plus comme un levier d'innovation. En observant et en analysant la manière avec laquelle leurs produits sont appréhendés en cuisine par les usagers (consommateurs ou professionnels), les industriels peuvent en effet déceler des pistes d'amélioration ou des idées de nouveaux produits. Dans ce travail de recherche, nous proposons de nouvelles méthodes dédiées à la caractérisation des gestes culinaires. Leur caractère innovant repose sur le fait qu'elles relèvent d'une approche quantitative, et non d'une approche qualitative comme cela est usuellement le cas. Elles s'inspirent en partie de deux méthodes de caractérisation utilisées en analyse sensorielle : le tri libre et la Q-méthodologie binaire.

Nous voyons comment des développements méthodologiques apportés à ces dernières permettent de caractériser des grands ensembles de stimuli. Ces développements sont liés à la fois à la procédure de recueil des données et à la procédure d'analyse statistique des données.

En particulier, nous proposons plusieurs procédures statistiques permettant d'aborder des problématiques variées : l'analyse d'un ensemble de partitions contenant des données manquantes, la classification non supervisée de profils d'évaluations binaires basée sur la notion d'accord inter-évaluateurs, etc.

Nous voyons ensuite que les deux méthodes de caractérisation quantitatives 'améliorées' sont applicables à des gestes culinaires.

Title : Statistical methodologies for the collection and analysis of data characterizing large sets of stimuli – Application to the sensory characterization of products & to the characterization of culinary techniques

Keywords : statistical methodologies, large sets of stimuli, sensory characterization, characterization of culinary techniques

Abstract: In the agri-food sector, the characterization of culinary techniques - through their observation and analysis - is increasingly seen as a lever for innovation. By analyzing how their products are used in the kitchen by the users (consumers or professionals), the manufacturers can detect improvement tracks or ideas for new products.

In this research, we propose new methods dedicated to the characterization of culinary techniques. Their innovative nature is based on the fact that they are based on a quantitative approach, and not on a qualitative approach as is usually the case. They partially draw their inspiration in two methods of characterization used in sensory analysis: the free sorting and the binary Q-methodology.

We see how methodological developments brought to the latter make it possible to characterize large sets of stimuli. These developments are both related to the procedure of data collection and to the procedure of statistical analysis of the data.

In particular, we propose several statistical procedures to address various issues: the statistical analysis of a set of partitions containing missing data, the unsupervised agreement-based clustering of a set of profiles of binary evaluations, etc.

Then, we see that these two 'improved' quantitative methods of characterization can be successfully applicable to culinary techniques.