



**HAL**  
open science

# Vers une approche non orientée pour l'évaluation de la qualité des odeurs

Massissilia Medjkoune

► **To cite this version:**

Massissilia Medjkoune. Vers une approche non orientée pour l'évaluation de la qualité des odeurs. Autre [cs.OH]. Université Montpellier, 2018. Français. NNT : 2018MONT010 . tel-01866751

**HAL Id: tel-01866751**

**<https://theses.hal.science/tel-01866751>**

Submitted on 3 Sep 2018

**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.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En informatique

École doctorale I2S

Unité de recherche LGI2P et LGEI

Vers une approche non-orientée pour l'évaluation de la  
qualité des odeurs

Présentée par Massissilia MEDJKOUNE

Le 30 mars 2018

Sous la direction de Jacky MONTMAIN et Jean-Louis FANLO

## Devant le jury composé de

M. Eric Gaussier, Professeur, Université Grenoble Alpes  
M. Patrice Bellot, Professeur, Université Aix-Marseille  
M. Serge Garlatti, Professeur, IMT Atlantique  
M. Michel Beigbeder, Maître-Assistant, Mines Saint-Etienne  
M. Jacky Montmain, Professeur au LGI2P, IMT Mines Alès  
M. Jean-Louis Fanlo, Professeur au LGEI, IMT Mines Alès  
M. Sébastien Harispe, Maître-assistant, IMT Mines Alès  
M. Stéphane Cariou, Maître-assistant, IMT Mines Alès  
Mme Cécile Coulon-Leroy, Enseignant- chercheur, ESA, Université d'Angers

Rapporteur  
Rapporteur  
Examineur  
Examineur  
Directeur de thèse  
Co-Directeur de thèse  
Encadrement  
Encadrement  
Invitée



UNIVERSITÉ  
DE MONTPELLIER



# Table des matières

---

|   |    |
|---|----|
| <i>Table des matières</i> .....   | 3  |
| <i>Liste des figures</i> .....  | 6  |
| <i>Liste des tables</i> .....   | 7  |
| <i>Chapitre 1 Introduction Générale</i> .....   | 8  |
| <i>Chapitre 2 Problématique de l'évaluation des odeurs</i> .....                                | 12 |
| <i>2.1 Analyse sensorielle et évaluation de l'odeur</i> .....                                   | 12 |
| <i>2.2 L'odeur et l'odorat</i> .....  | 14 |
| 2.2.1 Anatomie du système olfactif.....   | 14 |
| 2.2.2 Fonctionnement de l'odorat.....   | 16 |
| 2.2.3 Aspects psychologiques de l'olfaction .....   | 17 |
| <i>2.3 Caractéristiques de l'odeur</i> .....  | 17 |
| 2.3.1 Intensité .....   | 17 |
| 2.3.2 Persistence.....  | 18 |
| 2.3.3 Acceptabilité.....  | 18 |
| 2.3.4 Qualité .....   | 19 |
| <i>2.4 Evaluation de l'odeur</i> .....  | 19 |
| 2.4.1 Évaluation de la qualité d'une odeur .....  | 20 |
| 2.4.2 Apprentissage des odeurs .....  | 21 |
| <i>2.5 Méthodes pour l'évaluation de la qualité de l'odeur</i> .....                            | 24 |
| 2.5.1 Méthodes expertes .....   | 24 |
| 2.5.1.1 Roues des odeurs .....  | 24 |
| 2.5.1.2 Champ des odeurs® .....   | 25 |
| 2.5.2 Approche classique dite orientée.....   | 26 |
| 2.5.2.1 Avantages et inconvénients .....  | 27 |
| 2.5.2.2 Conclusion.....   | 28 |
| 2.5.3 Approche à commentaires libres dite non orientée .....                                    | 28 |
| 2.5.4 Vers des approches hybrides de l'évaluation de la qualité .....                           | 29 |
| <i>2.6 Objectif et plan de la thèse</i> .....   | 36 |
| 2.6.1 Processus cognitifs et perception de l'odeur .....  | 36 |
| 2.6.2 Contexte : de la chimie de l'odeur à l'automatisation de l'évaluation de sa qualité ..... | 37 |

|   |           |
|---|-----------|
| 2.6.3 Objectif.....   | 38        |
| 2.6.4 Vers une automatisation cognitive de l'évaluation de la qualité de l'odeur.....                   | 39        |
| <b>Chapitre 3 Conceptualisation des Descriptions Terminologiques .....</b>                              | <b>41</b> |
| <b>3.1 Etat de l'art : techniques de représentation et d'analyse des connaissances de domaine .....</b> | <b>43</b> |
| 3.1.1 Modèles et représentations des connaissances considérés .....                                     | 43        |
| 3.1.2 Mesurer la similarité entre concepts.....   | 45        |
| 3.1.3 Plongement sémantique et comparaison de termes.....   | 49        |
| <b>3.2 Contributions.....</b>   | <b>57</b> |
| <b>3.2.1 Du terme au concept : conceptualisation via transcription explicite.....</b>                   | <b>58</b> |
| 3.2.1.1 Formalisation du problème et proposition.....   | 58        |
| 3.2.1.2 Expérimentation sur la conceptualisation des descripteurs terminologiques.....                  | 60        |
| 3.2.1.3 Conclusion.....   | 69        |
| <b>3.2.2 Evocations complexes : conceptualisation via transcription implicite.....</b>                  | <b>71</b> |
| 3.2.2.1 Stratégies globales et formalisation du problème .....  | 71        |
| 3.2.2.2 Modélisation de la fonction score.....  | 71        |
| 3.2.2.3 Evaluation et discussion .....  | 76        |
| <b>3.3 Conclusion.....</b>  | <b>77</b> |
| <b>Chapitre 4 Synthèse de Descriptions Conceptualisées.....</b>   | <b>78</b> |
| <b>4.1 Etat de l'art sur le résumé automatique et problématiques connexes.....</b>                      | <b>79</b> |
| 4.1.1 Génération automatique de résumé.....   | 79        |
| 4.1.1.1 Résumé par extraction .....   | 79        |
| 4.1.1.2 Résumé par abstraction.....   | 81        |
| 4.1.2 Annotation et indexation.....   | 82        |
| <b>4.2 Approche automatisée de résumé d'un ensemble de descriptions conceptualisées.....</b>            | <b>85</b> |
| 4.2.1 Notations.....  | 85        |
| 4.2.2 Définition formelle du problème .....   | 86        |
| 4.2.2.1 Définition générale et espaces considérés .....   | 86        |
| 4.2.2.2 Formalisation de l'appréciation qualitative .....   | 87        |
| 4.2.2.3 Quantité d'information abstraite $\Psi(Y, \hat{X})$ .....                                       | 88        |
| 4.2.2.4 Modélisation des pénalités .....  | 89        |
| 4.2.3 Génération de résumés conceptuels.....  | 92        |
| 4.2.3.1 Construction de l'ensemble de résumés $S^X$ .....   | 92        |
| 4.2.3.2 Réduction de la complexité .....  | 94        |
| <b>4.3 Expérimentations.....</b>  | <b>94</b> |
| 4.3.1 Analyse du consensus sur la notion de résumé .....  | 94        |
| 4.3.2 Evaluation de la performance de l'approche automatique de résumé .....                            | 99        |
| 4.3.3 Application à l'analyse psychosensorielle : illustration .....                                    | 101       |

|   |            |
|---|------------|
| <i>4.5 Conclusion</i> .....                 | <i>105</i> |
| <i>Chapitre 5 Conclusion Générale</i> ..... | <i>107</i> |
| <i>Références</i> .....                     | <i>110</i> |

## Liste des figures

---

|  |     |
|--|-----|
| Figure 1.1 : Fréquences des termes fournis par les 15 juges pour la caractérisation de la qualité de l'odeur d'un vin (Lawrence et al., 2013). | 10  |
| Figure 2.1 : Anatomie du système olfactif (Poncelet, 2010)   | 14  |
| Figure 2.2 : Structure de la muqueuse olfactive (Poncelet, 2010)   | 15  |
| Figure 2.3 : Bulbe olfactif – Vue en coupe (Balez, 2001)   | 15  |
| Figure 2.4 : Courbe de Stevens (AFNOR, 1996)   | 18  |
| Figure 2.5 : Exemple d'une roue des odeurs (Bruneau <i>et al.</i> , 2000)  | 25  |
| Figure 2.6 : Le champ des odeurs® (Jaubert <i>et al.</i> , 1987a; Jaubert <i>et al.</i> , 1987b)   | 26  |
| Figure 2.7 : Principales étapes de la méthode classique (d'après ISO 11035, 1994)  | 27  |
| Figure 2.8 : Construction d'une roue des arômes généralisée (d'après (Capone et al., 2013))  | 35  |
| Figure 2.9 : Vue radar des OAV des 13 descripteurs sensoriels pour les cépages Negroamaro des Pouilles (d'après (Capone et al., 2013))         | 36  |
| Figure 2.10 : Décomposition de l'évaluation de l'odeur   | 37  |
| Figure 2.11 : Evaluation de la qualité d'une odeur (Méthode Olentica/LGEI)   | 38  |
| Figure 2.12 : Processus d'évaluation de la qualité des odeurs  | 40  |
| Figure 3.1 : Structuration des descripteurs conceptuels en différents niveaux  | 61  |
| Figure 3.2 : Paires termes/concepts utilisées pour l'évaluation  | 64  |
| Figure 3.3 : Evaluation des résultats pour le corpus de pages de Wikipédia   | 65  |
| Figure 3.4 : Evaluation des différents modèles pour le corpus WikiOdor   | 65  |
| Figure 3.5 : Performance d'évaluation terme/concept dans le cas du corpus de pages de Wikipédia  | 68  |
| Figure 4.1 : Représentation graphique de l'espace étudié   | 87  |
| Figure 4.2 : Illustration de la quantité d'information abstraite $\Psi(Y, \hat{X})$  | 88  |
| Figure 4.3 : Illustration des différents facteurs de pénalité  | 91  |
| Figure 4.4 : Répartition des évaluateurs en fonction du type d'exercice  | 96  |
| Figure 4.5 : Analyse des Correspondances Multiples (ACM) des questions en fonctions des évaluateurs qui leurs sont associées                   | 97  |
| Figure 4.6 : Analyse des Correspondances Multiples (ACM) des évaluateurs en fonctions des réponses aux questions proposées                     | 98  |
| Figure 4.7 : Carte de chaleur des similarités d'utilisateurs calculée à partir d'une mesure booléenne  | 98  |
| Figure 4.8 : Performance de l'approche en fonction de différents seuils de sélection du nombre de résumés retournés                            | 101 |
| Figure 4.9 : Roue des arômes de whisky   | 103 |

## Liste des tables

---

|   |     |
|---|-----|
| Tableau 3.1 : Méthodes de calcul des poids.....   | 51  |
| Tableau 3.2 : Mesures de similarité basées sur l’approche géométrique.....  | 51  |
| Tableau 3.3 : Mesures de similarité basées sur des approches ensemblistes.....  | 53  |
| Tableau 3.4 : Mesures de similarité basées sur l’approche probabiliste.....   | 54  |
| Tableau 3.5 : Taille de vocabulaire pour chaque modèle avec le corpus Wikipédia dump.....   | 63  |
| Tableau 3.6 : Taille de vocabulaire pour chaque modèle avec le corpus WikiOdor.....   | 63  |
| Tableau 3.7 : Exemple d’un échantillon d’apprentissage pour le niveau 1.....  | 67  |
| Tableau 3.8 : Ratio de prédiction dans le cas non pondéré.....  | 67  |
| Tableau 3.9 : Ratio de prédiction dans le cas pondéré.....  | 67  |
| Tableau 3.10 : Matrice des poids des différents modèles.....  | 68  |
| Tableau 3.11 : Evaluation terme/concept avec le corpus Wikipédia.....   | 69  |
| Tableau 3.12 : Poids des 6 modèles.....   | 69  |
| Tableau 3.13 : Exemple des entrées d’évaluation.....  | 76  |
| Tableau 3.14 : Résultats d’évaluation.....  | 76  |
| Tableau 4.1 : Choix de résumé pour synthétiser quatre concepts de même intensité.....   | 95  |
| Tableau 4.2 : Nombres d’évaluateurs associés aux résumés proposés au choix, classés par fréquence de sélection.....               | 97  |
| Tableau 4.3 : Performance de l’approche globale en fonction de différents seuils de sélection du nombre de résumés retournés..... | 100 |
| Tableau 4.4 : Descriptions expertes de la qualité de 6 whiskies.....  | 104 |
| Tableau 4.5 : Descriptions libres de la qualité de 6 whiskies.....  | 104 |

## Chapitre 1 Introduction Générale

---

L'analyse sensorielle, dévolue au départ à l'agroalimentaire et à la cosmétique, est aujourd'hui très largement utilisée dans l'Industrie. Relative à la perception des produits par les utilisateurs, elle consiste à saisir et à exploiter les informations issues de nos cinq sens (vue, ouïe, odorat, toucher, goût). L'analyse sensorielle exploite ainsi de manière privilégiée comme instrument de mesure « l'homme ». Incontournable dans le développement de nouveaux produits, elle trouve son origine dans la nécessité de méthodes de contrôle de la qualité, notamment gustative, de produits alimentaires, en l'absence de méthodes de mesures instrumentales appropriées. Son champ d'application s'est ensuite élargi à l'industrie des cosmétiques et des parfums avant de gagner par la suite des domaines variés comme celui de l'industrie automobile.

L'analyse sensorielle est devenue un passage obligé pour le développement et l'amélioration de nouveaux produits. Pour évaluer par exemple les fragrances d'un parfum, l'arôme d'un café ou le bouquet d'un vin avant de le commercialiser, des panels d'évaluateurs sont réunis à grands frais par l'industriel pour caractériser et restituer ce qu'ils « mesurent » par l'intermédiaire de leurs sens. Ces informations sont précieusement recueillies par l'industriel qui devra alors relier ces perceptions aux paramètres physicochimiques de son procédé de fabrication et proposer une reformulation plus à même de satisfaire les consommateurs le cas échéant.

L'enjeu industriel du marketing sensoriel est donc de taille : il s'agit de stimuler et séduire les sens des consommateurs. Seulement pour cette « mesure sensorielle », il n'existe pas d'autres capteurs que l'homme. Les industriels, en particulier dans l'alimentaire, emploient ainsi des professionnels exercés pour tester, qualifier et évaluer les produits nouveaux qu'ils souhaitent lancer sur le marché ou bien des produits dont certains paramètres physicochimiques ont été modifiés pour des raisons de coût par exemple. Il faut donc imaginer que l'homme, avec toute l'imprécision et la subjectivité qui caractérisent ses évaluations sensorielles, se trouve au cœur d'une nouvelle chaîne de mesure et de fiabilité. L'analyse psychosensorielle se doit donc d'être gérée de façon méthodique et rigoureuse pour que l'industriel puisse en tirer les profits attendus.

Améliorer la fiabilité de la mesure descriptive de la qualité sensorielle est un enjeu majeur puisque cela augmentera la probabilité d'un accueil favorable du produit par le public. Cependant, alors que dans l'industrie, l'analyse sensorielle est utilisée comme un instrument de mesure à part entière, les données sensorielles sont entachées d'un degré d'incertitude et d'imprécision que l'on n'admettrait pas en caractérisation physico-chimique ; même si, dans le cas par exemple d'un panel expert, il peut être réduit par l'entraînement des individus et des précautions opératoires strictes. Ceci est lié au fait que la mesure est réalisée par un panel de sujets dont les perceptions sont différentes et variables suivant le contexte et qu'ils évaluent des produits pouvant présenter eux-mêmes une variabilité naturelle.

Les méthodes classiquement utilisées en analyse sensorielle (analyses de la variance et analyses multidimensionnelles) sont efficaces pour interpréter et analyser ces données incertaines, mais dès lors que l'on reconnaît que les données manipulées—les mots—sont de plus de nature imprécise et subjective, qu'il existe des corrélations statistiques et des relations sémantiques entre ceux-ci, alors les outils mathématiques en vigueur deviennent questionnables. Ces traitements inadéquats peuvent fragiliser les décisions associées aux conclusions d'une étude sensorielle. Toutefois, les avancées en traitement de l'information pourraient permettre aujourd'hui de faire évoluer les pratiques. Il s'agit désormais de prendre en compte de façon réaliste l'imprécision et l'incertitude tout au long de la chaîne de traitement des données sensorielles, depuis leur acquisition jusqu'au profil produit, du juge entraîné jusqu'au consommateur naïf. L'objectif est de fiabiliser les données sensorielles et leur interprétation en vue de mieux comprendre les préférences et l'acceptabilité du public (citoyens, consommateurs, utilisateurs) dans un contexte économique et environnemental changeant, nécessitant de se différencier par le biais de l'innovation, de la qualité et de l'image.

Bien que le monde de l'industrie soit de plus en plus numérique, l'homme garde une place nécessaire dans le processus d'évaluation des produits. Il est donc essentiel de s'intéresser à l'interaction de l'homme avec le monde numérique, en dotant la machine de capacités de traitement d'information compatibles avec les modes cognitifs de l'humain pour traiter de façon pertinente ces informations, les transformer, les interpréter, les combiner de façon à créer de nouvelles informations, à les élaborer en connaissances et autoriser la prise de décisions. Notre projet doit permettre le déploiement de nouvelles méthodes d'analyse sensorielle à grande échelle pour l'entreprise qui positionnent l'homme de la rue–l'évaluateur néophyte–comme acteur de la production via son évaluation sensorielle.

D'un point de vue méthodologique, l'évaluation par un panel d'experts de produits, qui pourront ensuite être testés auprès de cohortes d'utilisateurs, est un processus qui met en jeu de multiples formes d'incertitude. Nous pouvons citer tout d'abord la variabilité des produits comparés (ou à l'inverse leur similarité) qui joue sur la difficulté de l'évaluation, sur sa précision, et par conséquent sur l'incertitude des valeurs fournies par les sujets ; mais aussi l'imprécision d'une évaluation, inhérente à la notion de jugement et qui définit la granularité de l'évaluation qui, si elle n'est pas bien choisie, devient source d'incertitude. L'incertitude épistémique (*i.e.*, relative à l'état de la connaissance de l'instant) doit être prise en compte car l'expérience et la culture de l'évaluateur détermine la précision et la vraisemblance de l'évaluation. Elle est *a priori* contrôlée dans un panel d'experts entraînés, elle a un impact plus perturbateur dans l'évaluation par des utilisateurs néophytes. Par ailleurs, les référentiels utilisés par un expert et par un utilisateur lambda sont significativement différents (pas de représentation des connaissances partagée chez les évaluateurs naïfs). En particulier, leurs évaluations respectives ne sauraient se référer à une même échelle : la granularité de l'évaluation attendue d'un utilisateur non entraîné ne peut être celle attendue d'un expert chevronné. Ce biais engendre une incertitude quant à la mise en correspondance entre les évaluations expertes et celles des utilisateurs. Par ailleurs, la dispersion ou la variabilité des avis des experts d'un panel, qui bien qu'entraînés n'ont pas forcément le même ressenti sur un même produit (seuil de détection différent d'un individu à l'autre par exemple), induit une incertitude stochastique dans l'évaluation collective (interprétation classiquement probabiliste). Elle est aujourd'hui souvent occultée en pratique par l'utilisation de grandeurs statistiques pour gérer la disparité d'opinions. En analyse sensorielle, la pratique consiste en effet à calculer des moyennes pour synthétiser l'information multicritère et multi acteurs à différents niveaux de l'évaluation. Ces synthèses constituent une perte informationnelle non réversible dans le traitement des données d'évaluation initiales. L'incertitude des données brutes doit au contraire être préservée tout au long du processus d'évaluation pour être synthétisée *a posteriori* si nécessaire.

Nous nous intéressons principalement à la synthèse d'analyses sensorielles olfactives compte tenu du domaine d'expertise d'une partie de notre encadrement au Laboratoire de Génie de l'Environnement Industriel (LGEI de IMT Mines Alès). Les problèmes liés à la qualité de l'air en général, et aux émissions de particules, de COV (Composés Organiques Volatils) et aux nuisances odorantes associées à certains d'entre eux, font maintenant partie des préoccupations sociétales de premier plan, et sont de fait devenus structurants pour assurer la durabilité de la production industrielle. Les populations sont en effet aujourd'hui de plus en plus souvent alertées sur l'impact que la qualité de l'air qu'elles respirent est susceptible d'avoir sur leur santé. Très souvent, la première perception qu'elles ont de cette qualité est l'odeur qu'elles perçoivent. Une odeur à caractère désagréable est en effet toujours perçue comme un risque pour la santé de l'individu. On comprend ainsi l'enjeu majeur qu'a l'industriel de procéder à la qualification de ses rejets gazeux.

Plusieurs méthodes dites orientées existent pour qualifier et classer les odeurs selon le ressenti. Elles associent toutes une perception à un vocabulaire (champ lexical). L'utilisation efficace de ces approches nécessite un apprentissage en plusieurs étapes : initiation à la compréhension de l'odorat et de son fonctionnement, mémorisation olfactive d'une collection de référents odorants et de leur organisation relative à un espace odorant, entraînement à la reconnaissance des référents odorants dans un mélange complexe et apprentissage d'un vocabulaire structuré de domaine pour une caractérisation précise et non ambiguë. Le principe consiste à associer, pour un domaine d'activité particulier, les perceptions olfactives avec les composés odorants trouvés dans le domaine d'activité. Pour cela, plusieurs niveaux de référents peuvent être spécifiés pour un domaine donné afin d'orienter la perception vers plus de précision. Ces méthodes visent à utiliser un référentiel commun pour qualifier les odeurs. Une étape

d'appropriation de ce vocabulaire ou d'apprentissage des référents est donc nécessaire. Il s'agit là d'un préalable indispensable à toute exploitation de cette dimension fondamentale de l'odeur, ce qui explique le coût élevé des sessions d'analyse sensorielle pour l'industriel. En effet, dans une population non formée, l'appréciation de la qualité d'une odeur se limite généralement à une assimilation de l'odeur perçue avec une autre odeur, fréquemment rencontrée, et donc connue (« ça sent comme ... »). Cela rend impossible l'exploitation de cette information, plusieurs descripteurs pouvant renvoyer en réalité à la même qualité d'odeur.

Pour fiabiliser et automatiser l'obtention des informations concernant la qualité d'une odeur, nous proposons de mettre en place une méthode ne nécessitant ni formation préalable, ni apprentissage d'un champ lexical, mais basée sur des techniques de traitement automatique de la langue et de représentation des connaissances. Il s'agit là d'une démarche diamétralement opposée à celles employées à ce jour. L'objectif est d'identifier et d'extraire, dans un discours de non spécialiste décrivant une odeur, les mots ou locutions pouvant être mis en relation avec des descripteurs communément admis dans le monde des odeurs. En inversant la logique de recherche des données descriptives de la qualité de l'odeur—analyses distributionnelles et sémantiques plutôt que formation des observateurs, l'idée est de pouvoir assoir la détermination de cette grandeur sur des échantillons de population de taille significative et de s'affranchir de l'utilisation de jurys spécialisés.

Si l'utilisation de panels d'évaluateurs non experts permet un changement d'échelle dans l'analyse sensorielle, elle représente aussi une économie substantielle pour l'industriel. Malheureusement, elle souffre de biais liés à la naïveté sensorielle, au manque de vocabulaire métier précis et à l'hétérogénéité du panel de néophytes. A titre illustratif, considérons la description suivante fournie par 15 évaluateurs pour l'évaluation des arômes d'un vin donné. Si les évaluateurs ne sont pas des professionnels de l'analyse sensorielle ou des métiers du vin et qu'aucune contrainte n'est imposée sur leur restitution, alors un grand nombre de termes sera généralement proposé. Ainsi, si l'on associe une fréquence à chacun des termes d'évocation, on risque fort d'obtenir un histogramme comme celui de la Figure 1.1 qui représente le nombre de fois qu'un terme a été utilisé par les 15 juges. Sur les 24 termes utilisés, seulement 3 ont été utilisés au moins quatre fois.

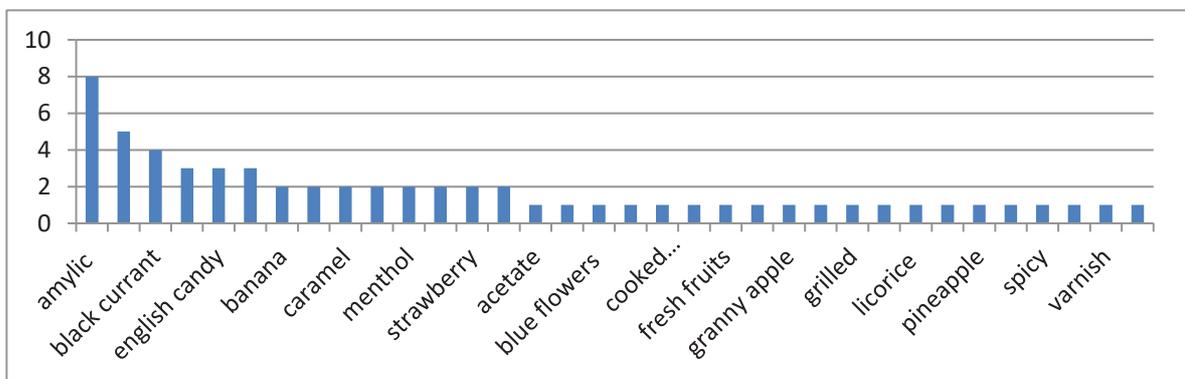


Figure 1.1 : Fréquences des termes fournis par les 15 juges pour la caractérisation de la qualité de l'odeur d'un vin (Lawrence et al., 2013).

Comme les réponses des évaluateurs ne sont pas limitées à une liste de descripteurs prédéfinis, l'ensemble de termes employés peut être important. L'histogramme représentatif des fréquences des termes est alors peu exploitable. On peut remarquer aussi que les descriptions sont hétérogènes, partiellement redondantes (un grand nombre de mots évoquent une même perception) et ne sont pas forcément comprises par un expert du vin. Pour tirer parti de cet histogramme, l'analyste sensoriel doit être capable de rapprocher les évocations similaires et de séparer celles qui expriment vraiment des sensations distinctes. Pour cela, il doit être capable d'interpréter les évocations en langage naturel, c'est-à-dire d'identifier à quels descripteurs utilisés pour caractériser les arômes d'un vin, chaque terme pourrait se référer. Une fois cette projection des termes communs dans un vocabulaire qui fait sens pour un professionnel du vin réalisée, l'analyste doit les regrouper selon la similarité des évocations ce qui lui permet alors d'obtenir un histogramme plus informatif que celui de la Figure 1.1. L'analyste doit en effet pouvoir proposer une évaluation collective qui soit expressive tout en étant synthétique en gérant

l'abstraction et la précision des descripteurs. Si ces tâches de projection dans un vocabulaire structuré et partagé par le monde professionnel et de résumé sémantique des évocations reposent d'abord sur l'expérience et la connaissance du domaine de l'analyste, elles restent néanmoins sujettes à la subjectivité de ce dernier. Leur automatiser ne saurait être réalisée sans admettre qu'elles ne sont pas complètement déterministes et dépendent nécessairement d'une interprétation qu'une subjectivité culturelle ou contextuelle pourra en partie affecter. Cette thèse se propose d'étudier l'automatisation de ces processus cognitifs caractéristiques du métier d'analyste sensoriel.

La contribution majeure de la thèse est la transcription des termes (descripteurs linguistiques) fournis par les évaluateurs en langage naturel en concepts dans le but d'exprimer un terme sous forme d'un descripteur conceptuel, précis, non ambigu et compris par le monde professionnel d'un domaine donné. Comme le nombre de termes est souvent important, il est ensuite nécessaire de synthétiser l'ensemble des descripteurs linguistiques pour obtenir un résumé sémantique qui restitue l'ensemble des descripteurs linguistiques de façon fidèle et synthétique.

Notre chaîne de traitement consiste ainsi à mettre en correspondance les termes fournis en langage naturel par les évaluateurs avec des descripteurs structurés, puis à élaborer un résumé sémantique pour synthétiser de façon pertinente l'information véhiculée par l'ensemble des descripteurs associés aux termes évoqués. Ce traitement est basé sur des techniques de représentations de connaissances et sur des techniques de traitement automatique de la langue (TAL). Les techniques de représentation des connaissances sont utilisées pour modéliser l'information des descripteurs structurés (concepts) ; par la suite, différentes techniques nous permettront de tirer parti de ces représentations pour comparer les descripteurs au travers de la notion de similarité sémantique. Les techniques de TAL permettent d'analyser un corpus de textes afin de construire des représentations sémantiques de termes évoqués par les évaluateurs pour évaluer la proximité entre les termes—étape nécessaire pour la conceptualisation des évocations proposées par les évaluateurs avant l'application d'un résumé sémantique.

Le plan du manuscrit est le suivant. La première partie de la thèse est consacrée à un état de l'art sur l'évaluation de la qualité des odeurs et à la définition de la problématique. Les définitions et les caractéristiques des odeurs y sont présentées. Puis, les méthodes de description sensorielle utilisées aujourd'hui par l'industriel sont décrites. Les contraintes et les enjeux d'une approche non orientée pour la qualité des odeurs sont recensés. Enfin, l'objectif détaillé de la thèse et une présentation générale des contributions concluent ce chapitre. La seconde partie est consacrée à la mise en correspondance entre les termes et les concepts associés à une taxonomie qui exprime les concepts d'intérêt dans un contexte d'analyse sensorielle. Cette phase est basée sur la comparaison de représentations vectorielles de termes et initie le processus d'association des termes aux concepts. Sur la base de modèles distributionnels (ou plus généralement de plongement sémantique), nous proposons de calculer le degré d'évocation avec lequel un groupe de termes évoque un concept de la représentation des connaissances associée à un domaine. L'objectif est de représenter chaque description fournie par un évaluateur non expert par une description conceptuelle. La troisième partie est dédiée au résumé sémantique des annotations conceptuelles individuelles des évaluateurs—ensembles de concepts—en une liste synthétique et fidèle de concepts. Cette étape est d'autant plus cruciale que les évaluateurs sont nombreux. Nous concluons sur les résultats et perspectives de cette automatisation des tâches cognitives associée au métier d'analyste sensoriel.

## Chapitre 2 Problématique de l'évaluation des odeurs

---

De nos jours, l'odorat est généralement considéré comme un sens mineur, archaïque (Lledo & Vincent, 1999), rappelant les plus bas instincts animaux (McKenzie, 2013). Un premier paradoxe réside dans le fait qu'il est aussi reconnu comme le plus subtil et le plus méconnu, par cette défiance vis-à-vis de notre animalité. Le second paradoxe est que les nuisances odorantes sont parmi les plus immédiatement ressenties par un individu avec les nuisances sonores et les poussières. Les mauvaises odeurs sont en effet aujourd'hui considérées comme une agression et sont régulièrement présentées comme étant le deuxième motif de plaintes dans la population après celles liées au bruit (Chemin, 1996).

Les révolutions industrielles successives ont engendré une pollution de l'air de plus en plus perceptible au cours des dernières décennies. Selon plusieurs enquêtes, les citoyens sont devenus de plus en plus sensibles aux problèmes liés à la pollution et à l'environnement (Etudes ISAAC (INSERM-2007) (Asher *et al.*, 2006), APHEKOM (INVS-2011) (Declercq *et al.*, 2012)). Or, la première perception qu'a souvent la population de la qualité de l'air qu'elle respire est l'odeur qu'elle perçoit. Une odeur à caractère désagréable est toujours perçue comme un risque pour la santé. Cette hypersensibilisation des populations aux nuisances odorantes fait que la problématique des odeurs est devenue très structurante pour les industriels et les collectivités locales.

Les entreprises sont aujourd'hui confrontées à la recherche de solutions pour améliorer leur image et leur relation avec les populations riveraines de sites industriels. Il est donc essentiel pour l'industriel d'être capable de caractériser les odeurs liées aux rejets de ses usines afin de comprendre et d'anticiper la réaction des populations. Cette caractérisation est appelée qualité de l'odeur.

Si la première réaction d'un individu confronté à une odeur est la plupart du temps de fournir une information de type hédonique (ça sent bon /ça sent mauvais), faisant ainsi référence à l'acceptabilité de l'odeur, l'information sur la qualité de l'odeur, qui nécessite l'utilisation de descripteurs linguistiques précis, est, elle, beaucoup plus difficile à obtenir. L'individu questionné procède en effet le plus souvent par assimilation de l'odeur perçue avec une odeur connue, perçue antérieurement (*ça sent comme...*). La qualité de cette information est pourtant d'un intérêt stratégique pour l'industriel. La connaissance du niveau d'acceptabilité de l'odeur permet de prévoir l'impact potentiel sur les riverains ou la réaction des utilisateurs à un produit. Une connaissance détaillée des descripteurs qualitatifs de l'odeur permet, elle, de déterminer l'origine des sources de pollution odorante sur un site, ou de comprendre quelles molécules peuvent être à l'origine d'une caractéristique rendant désagréable l'odeur d'un produit. Des mesures correctives sont alors envisageables, permettant aux industriels d'augmenter la probabilité d'une perception plus favorable de l'effluent par les riverains ou du produit par les consommateurs, accroissant ainsi la compétitivité de l'entreprise.

### 2.1 Analyse sensorielle et évaluation de l'odeur

L'analyse sensorielle est un ensemble de méthodes permettant de mesurer les perceptions sensorielles (vue, ouïe, odorat, goût, toucher). Elle est fondée sur trois étapes: percevoir, identifier, discerner. À l'aide de descripteurs utilisés pour qualifier le produit, elle s'attache à identifier un point de vue objectif sur le ressenti. Le principal problème de ces méthodes est que la réaction sensorielle ne se limite pas à

une simple réaction physiologique à un stimulus, mais est conditionnée par le vécu de l'individu, son expérience, son état d'esprit et son environnement. Depuis les années 90, l'évaluation sensorielle a connu un remarquable développement. Cet essor est en particulier dû au besoin de connaissance des produits par les industriels pour déterminer si leurs caractéristiques sont différentes ou similaires de celles des produits lancés par leurs concurrents. On cherche donc à qualifier les performances des produits par rapport à un produit référent dont les caractéristiques sont connues. L'objectif de l'industriel est alors de standardiser la qualité de ses produits.

Bien que l'analyse sensorielle soit née dans l'industrie agro-alimentaire, aujourd'hui, son domaine d'application s'est étendu à bien d'autres domaines : la cosmétique, l'emballage, les produits d'hygiène, l'environnement, etc. L'évaluation sensorielle nécessite l'intervention d'un groupe d'individus représentatif de la population ciblée. Ces sujets peuvent être naïfs (consommateurs, riverains, etc.), ou experts, *i.e.*, des individus ayant suivi des formations afin d'apprendre un vocabulaire contrôlé.

Dans de nombreux secteurs industriels, des opérateurs ont encore aujourd'hui recours à leurs sens pour réaliser leurs tâches : outre les nez des parfumeurs ou de l'œnologie, on peut citer la détection à l'oreille du bon (ou mauvais) fonctionnement d'une machine, la détection de la présence d'une odeur anormale, la détection visuelle d'un défaut de fabrication, etc. Force est donc de constater que l'homme, malgré la subjectivité et l'incertitude qui caractérisent sa perception sensorielle, se positionne comme un outil de mesure et de prise de décision (Depledt & Sauvageot, 2002).

Généralement, l'analyse sensorielle est en effet un outil d'aide à la décision. Elle est chargée d'apporter des réponses objectives aux questions que se pose la production. Elle aide aussi à mieux communiquer, grâce à un vocabulaire partagé pour qualifier les produits et leurs caractéristiques dans un domaine donné.

Deux phases complémentaires peuvent ainsi être distinguées dans le processus de l'évaluation sensorielle :

- La caractérisation de la qualité du produit par le biais de descripteurs linguistiques, cette étape sémantique, appelée étude terminologique, permet de générer l'ensemble des descripteurs linguistiques qui caractérisent le produit (Lefebvre & Bassereau, 2003). Ensuite, l'intensité de ces descripteurs est évaluée sur des échelles de mesure afin d'établir un profil sensoriel, c'est-à-dire un ensemble de descripteurs quantifiés ;
- La caractérisation de la satisfaction ou de l'acceptabilité du produit par les consommateurs s'intéresse, elle, à la manière dont un produit est accepté par le public.

Nous reviendrons plus en détails sur ces notions de qualité et d'acceptabilité dans le domaine de l'odeur dans la section suivante. L'odeur est une dimension clé de l'analyse sensorielle. L'odorat nous sert de système informatif et trouve à ce titre son utilité dans notre vie de tous les jours. Les odeurs nous influencent dans nos choix, nous permettent de juger de la qualité d'un aliment, d'une fleur, d'un produit. C'est l'odorat, et non simplement le goût, qui fait que nous percevons la saveur des aliments. L'odeur joue donc un rôle de premier plan chez l'homme dans l'appréciation du confort et du bien-être.

Les sections suivantes sont consacrées à la description du fonctionnement du système olfactif, des caractéristiques de l'odeur, afin d'introduire un état de l'art concernant l'évaluation de l'odeur, en particulier sa caractérisation verbale, *i.e.*, sa qualité. Enfin, l'objectif global de la thèse est présenté.

## 2.2 L'odeur et l'odorat

### 2.2.1 Anatomie du système olfactif

Le système olfactif humain comporte un dispositif permettant la circulation de l'air et un autre la détection des molécules odorantes (Figure 2.1).

Le dispositif de ventilation qui va des narines aux poumons permet aux molécules odorantes contenues dans l'air que nous respirons de rentrer en contact avec le dispositif de détection (la muqueuse olfactive) situé au sommet et au fond des fosses nasales.

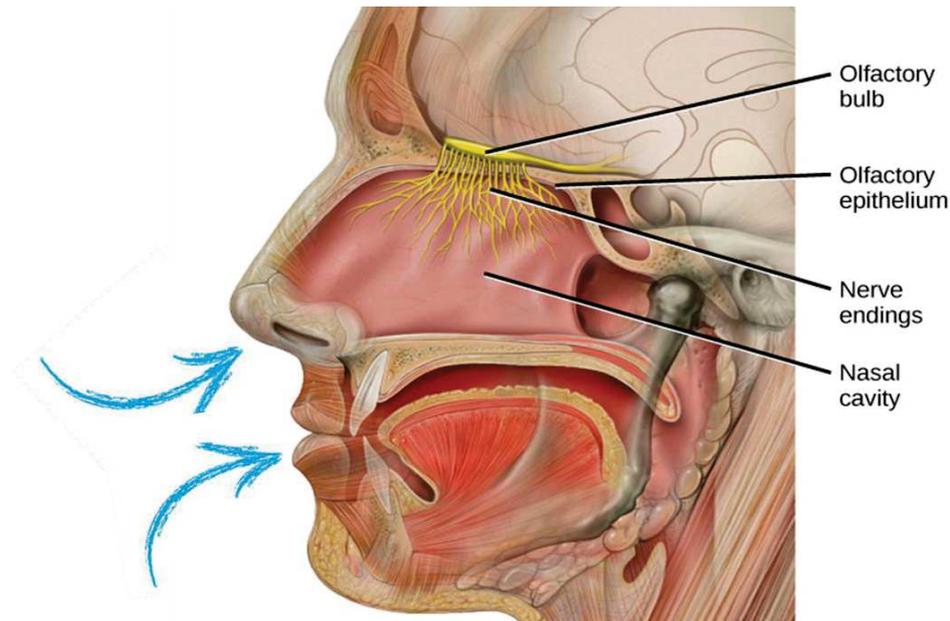


Figure 2.1 : Anatomie du système olfactif (Poncelet, 2010)

Notons qu'il existe deux voies d'introduction des substances odorantes vers la muqueuse olfactive : la voie directe, qui correspond à l'inspiration par le nez, et la voie rétro-nasale, qui correspond à l'expiration par le nez qui entraîne vers la muqueuse l'air contenu dans la bouche.

#### 2.2.1.1 La muqueuse olfactive

Elle occupe, chez l'homme, une surface moyenne de l'ordre de 2 à 3 cm<sup>2</sup> (contre 150 cm<sup>2</sup> chez le berger allemand) (Morrison & Costanzo, 1990). On y distingue trois types de cellules (Figure 2.2):

- Les neurorécepteurs olfactifs (ou cellules réceptrices neurosensorielles) qui sont des neurones bipolaires dont l'une des extrémités se dirige vers l'épithélium et l'autre (l'axone) va vers le bulbe olfactif. Ils portent des bouquets de cils chimiosensibles qui baignent dans la couche de mucus. La membrane de ces cils est tapissée de récepteurs moléculaires au sein desquels s'effectue la transduction du message chimique en message électrique. Les neurones olfactifs via leurs cils sont capables de reconnaître et de se lier aux molécules odorantes et par conséquent de créer un signal vers le cerveau (Buck & Axel, 1991).
- Les cellules de soutien sont des cellules épithéliales non neuronales qui s'étendent sur toute l'épaisseur de l'épithélium olfactif. Elles ne semblent pas intervenir directement dans le traitement de l'information olfactive, mais jouent un rôle dans la facilitation du développement des neurones.

- Les cellules basales qui sont des cellules souches qui ont pour fonction de remplacer soit les cellules de soutien, soit les cellules réceptrices.

La sous-muqueuse, via les glandes de Bowman, sécrète en permanence un mucus qui remplit plusieurs fonctions, dont celle de protéger les cellules réceptrices, mais surtout d'assurer la captation et le transport des molécules odorantes.

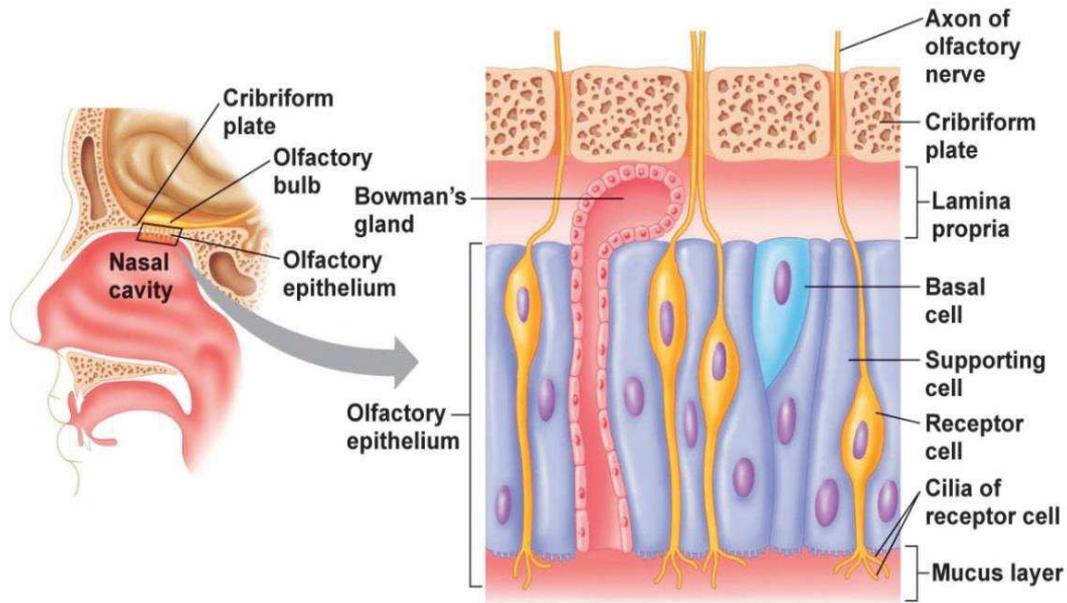


Figure 2.2 : Structure de la muqueuse olfactive (Poncelet, 2010)

### 2.2.1.2 Le bulbe olfactif

Il s'agit du premier relais de l'information olfactive en recevant en entrée les messages provenant de la muqueuse et en envoyant les informations aux différents territoires du cerveau antérieur. Chez l'homme, il est de forme cylindrique allongée (Figure 2.3).

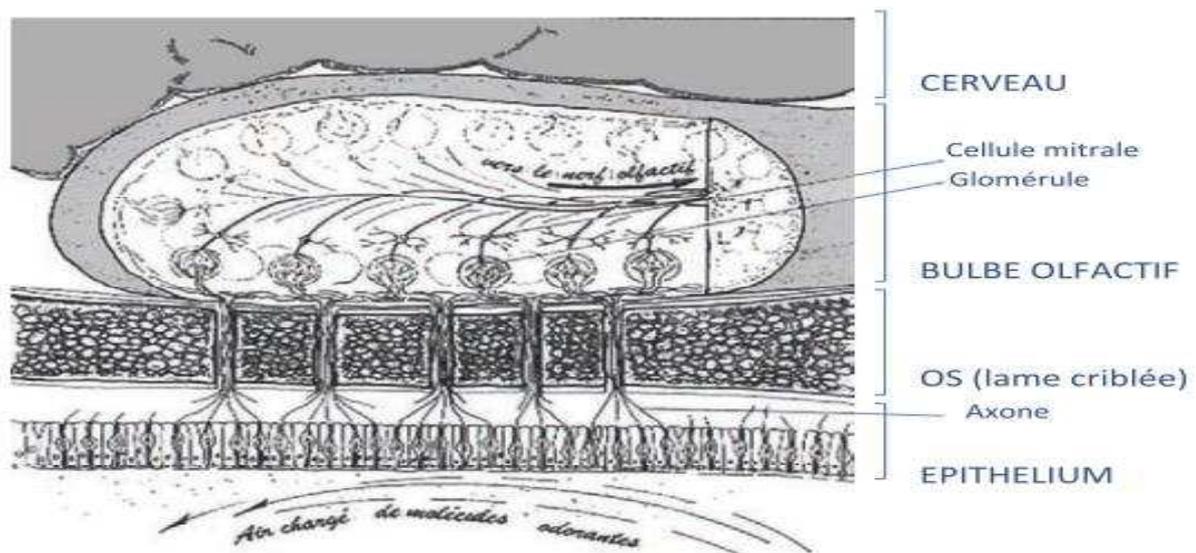


Figure 2.3 : Bulbe olfactif – Vue en coupe (Balez, 2001)

### 2.2.1.3 Autres systèmes participant à la perception olfactive

Il faut noter que la perception olfactive ne se limite pas au strict appareil olfactif tel qu'il a été décrit voméronasal. Le système trigéminal traite des sensations somesthésiques de frais, chaud, piquant, brûlant, irritant, etc., provoquées par la majorité des molécules odorantes, et qui font que ces odeurs peuvent même être perçues par les patients anosmiques (Doty *et al.*, 1978). En premier lieu, ces informations, bien qu'étant peu diversifiées, participent à la reconnaissance d'une odeur. L'odeur de menthe, par exemple, sans son caractère trigéminé frais, n'est plus reconnaissable. En second lieu, ces informations peuvent garantir un rôle de protection pour l'individu. Ainsi, quand les sensations trigéminales sont trop fortes, l'organisme réagit fortement en assurant un ensemble de réflexes physiologiques tels que bradycardie, diminution du rythme respiratoire et fermeture de la glotte. Le nerf terminal dont une partie des fibres se terminent dans la cavité nasale et une autre rejoint le bulbe olfactif et l'hypothalamus participe aussi à la perception.

Enfin, l'organe voméronasal, dont l'existence chez l'homme n'a clairement été établie qu'en 1991, peut court-circuiter les centres cognitifs du cerveau et provoquer des comportements innés et des réponses émotionnelles.

## 2.2.2 Fonctionnement de l'odorat

L'odeur correspond à la détection par le système olfactif de molécules volatiles (les odorants) émis dans l'air ambiant par une substance ou une activité. On peut distinguer plusieurs étapes dans le fonctionnement – complexe – du système olfactif. La première consiste en l'arrivée de la molécule et en sa captation par le système olfactif (réception). La seconde est la transformation du stimulus chimique en un signal électrique (transduction) dans la cellule sensorielle (le neurone). La troisième étape est l'étape de codage de l'information, qui correspond au traitement du signal électrique dans le bulbe olfactif et dans le cerveau.

### 2.2.2.1 Captage des molécules et transduction du signal

Les récepteurs protéiques sont des protéines qui traversent sept fois la membrane des cils. L'interaction avec la molécule odorante est une interaction de faible énergie, réversible, de type liaison de Van Der Waals. Le mécanisme de transduction est la conversion des stimuli en signal électrique. Il s'agit de la transformation de l'interaction des molécules odorantes absorbées présentes dans le mucus avec les récepteurs glycoprotéiques de la membrane des cils olfactifs, en message olfactif. Lorsque l'interaction a lieu, une modification se produit à l'intérieur du neurone qui provoque l'activation d'un type de protéines auquel les protéines réceptrices sont capables de se coupler (protéines G). C'est la première amplification du signal. La libération d'une des sous-unités de ces protéines G va déclencher le deuxième niveau d'amplification du signal en provoquant un transport actif d'ions au travers de la membrane du cil, modifiant le potentiel électrique de la cellule. Le signal ainsi créé se propage alors le long de l'axone, puis de la dendrite, vers le cerveau.

Le nombre très largement supérieur de molécules par rapport aux nombre de récepteurs olfactifs implique une non-spécificité de ceux-ci vis-à-vis des molécules. Des études ont également montré qu'un odorant peut stimuler plusieurs récepteurs et inversement une cellule peut être sensible à plusieurs composés odorants (Firestein *et al.*, 1993). De même, une molécule peut avoir soit un effet d'excitation soit un effet inhibiteur sur les récepteurs (Price, 1987).

### 2.2.2.2 Traitement du signal olfactif

Dans le bulbe olfactif, le signal se transforme en une « image » que le cerveau traite comme une image visuelle. Il semble que chaque neurone ne porte qu'un seul type de protéine réceptrice (il en existe 350

types chez l'homme). Par ailleurs, si les protéines réceptrices apparaissent aléatoirement distribuées à la surface de l'épithélium, les neurones qui les portent sont eux reliés à des glomérules spécifiques du bulbe olfactif, en fonction du type de protéine qu'ils portent.

Une odeur donnée provoque donc l'activation d'un ensemble précis de glomérules, et c'est la répartition spatiale, et sans doute temporelle, de cette activation qui permet au cerveau de reconnaître les odeurs.

### 2.2.3 Aspects psychologiques de l'olfaction

Tout ce qui précède a montré que la perception et l'évaluation des odeurs est un phénomène complexe. Cependant, cette perception n'est pas uniquement régie par des phénomènes physiologiques. Elle est également façonnée par le filtre de la cognition et de la mémoire.

Pour (Hudson & Distel, 2003), « *les odeurs sont des percepts individuels subjectifs, façonnés par la rencontre d'odorants particuliers, conduisant à des réponses aux odeurs organisées à partir du contexte individuel d'une personne ou de la diversité des environnements qu'elle a connus au cours de sa vie* ». (Ferdenzi *et al.*, 2013) montrent que la variabilité de la réponse aux odeurs est fonction de la culture, du filtre de la cognition et de la mémoire (aspect culturel, éducation,...). (Chrea *et al.*, 2005; Hudson, 1999; Rabin, 1988) montrent l'importance de l'expérience et de la culture dans l'évaluation et la reconnaissance et la discrimination des odeurs.

De plus, la place tenue par les images dans l'encodage des odeurs est fondamentale. Lorsque nous percevons un signal olfactif, nous consultons notre banque d'images en procédant par comparaison et association d'idées pour comprendre et reconnaître l'odeur. Celle-ci sera ainsi liée à un ou plusieurs contextes si c'est la première fois que nous la rencontrons, sinon, nous additionnons ce ou ces nouveaux contextes à ceux précédemment enregistrés. Toute odeur perçue est ainsi comparée à celles auparavant mises en mémoire et consolidera ou modifiera notre référentiel olfactif. En somme, de nos rencontres olfactives nous créons un album mental composé d'histoires et de moments vécus intimement relié au temps et aux lieux (Lawless & Engen, 1977). Et, faute de n'avoir aucun vocabulaire universel défini pour communiquer sur les perceptions ressenties, comme il en existe pour les couleurs par exemple, cet album constitue notre lexique olfactif.

## 2.3 Caractéristiques de l'odeur

La perception olfactive possède quatre caractéristiques mesurables qui sont :

- Son intensité
- Sa qualité
- Son acceptabilité
- Sa persistance (ou concentration)

### 2.3.1 Intensité

L'intensité d'une odeur caractérise sa force ou puissance. Cette perception est traduite par les termes « fort », « faible » ou encore « je perçois ». Ce paramètre se mesure, par comparaison entre une échelle de référence basée sur différentes concentrations de butanol dans de l'eau et l'échantillon à évaluer (norme X 43-103).

L'odeur étant une perception, la relation entre le stimulus odorant et la réponse faite par les sujets suit des lois psychophysiques semblables à celles connues dans d'autres modalités sensorielles telles que la vision ou l'audition (Wright, 1978).

La loi de Stevens permet d'établir une relation entre l'échelle d'intensité du stimulus odorant (concentration du mélange odorant) et l'échelle d'intensité de la perception olfactive.

$$\log I = n \log(C - C_0) + Cte$$

$I$  : intensité perçue

$C$  : concentration dans l'air inhalé (ppm)

$Cte$  : constante relative aux unités employées

$C_0$  : concentration au seuil de perception : c'est la concentration donnant une odeur juste décelable.

Le coefficient " $n$ " traduit l'accroissement de l'intensité odorante en fonction de la concentration (généralement compris entre 0,2 et 0,8).

Chaque odeur peut donc se caractériser par une courbe que l'on peut tracer entre l'intensité olfactive et sa concentration (courbe log-log). Cette courbe présente un palier d'inodorité (intensité nulle), une branche montante et un palier de saturation (Figure 2.4).

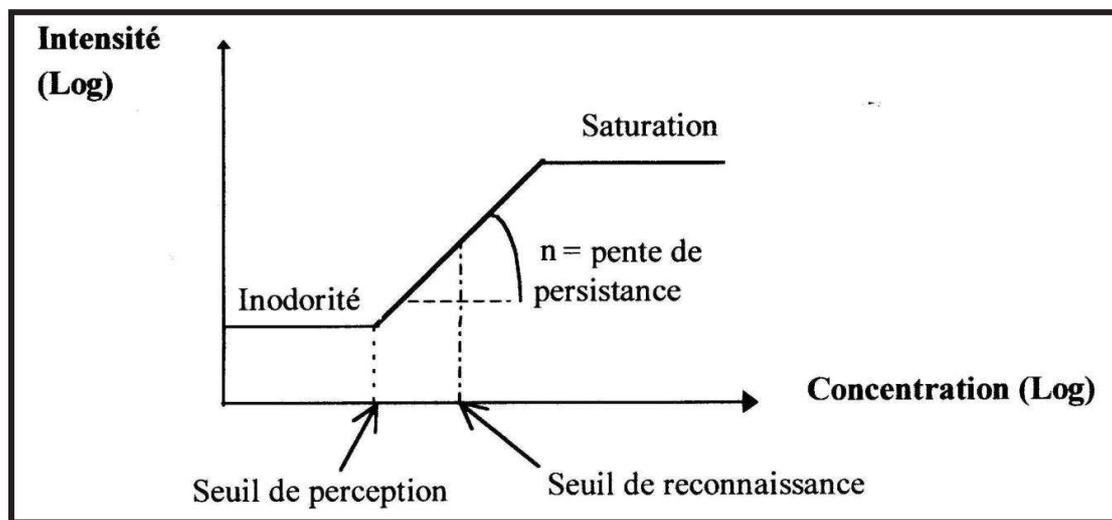


Figure 2.4 : Courbe de Stevens (AFNOR, 1996)

### 2.3.2 Persistance

La pente de la branche montante de la courbe de Stevens traduit la sensibilité de l'odorat à la variation de la concentration du composé odorant. Si  $n$  est faible (0,2), la dilution de la substance a peu d'influence sur sa perception, l'odeur est dite persistante. Si  $n$  est fort (0,8), la dilution réduit largement l'intensité de la perception, l'odeur est dite peu persistante ou fugace.

C'est cette caractéristique qui est à la base de la législation européenne, avec la mesure de la concentration d'odeur (Standard *et al.*, 2003).

### 2.3.3 Acceptabilité

L'acceptabilité représente le caractère hédonique de l'odeur (son côté agréable ou non). Cette notion est mesurée par une valeur de ton hédonique (Chaignaud *et al.*, 2014). Une odeur peut être agréable, acceptable, désagréable, intolérable. Ce classement est subjectif, l'accoutumance pouvant être à l'origine d'un glissement de l'appréciation de l'observateur vers la classe acceptable et le jugement porté dépendant, pour une substance donnée, de la concentration de l'odeur. Généralement, il existe un assez bon accord, dans une population homogène, sur les odeurs considérées comme agréables, désagréables ou intolérables, bien que l'on constate, là encore, des différences d'appréciation individuelles. Par

ailleurs, il faut noter que même si sa qualité ne change pas, une odeur agréable à faible concentration peut s'avérer désagréable à forte concentration, et que certaines substances changent d'odeur en fonction de la concentration.

### 2.3.4 Qualité

La qualité de l'odeur correspond à la description de l'odeur. C'est la première information qui atteint le cerveau de l'individu. Cependant, le premier réflexe des individus est de donner des informations hédonistes plutôt que l'identification ou la dénomination de l'odorat (Benoit & Pannier, 1982; Maille, 2000). L'identification d'une odeur est définie comme la capacité à lui donner un nom. De nombreux travaux se penchent sur la difficulté de la tâche d'identification des odeurs (Engen, 1982). La qualité de l'odeur reste la propriété la plus subjective de la perception olfactive car elle repose sur les souvenirs personnels des individus (variation interindividuelle) (David *et al.*, 1997; Rouby & Sicard, 1997). La qualification des odeurs souffre aussi d'une déficience du vocabulaire à cause du manque de mots pour définir l'odeur (Rouby *et al.*, 2005).

Les caractéristiques de l'odeur qui viennent d'être décrites ne sont pas indépendantes les unes des autres : par exemple, la qualité d'un stimulus olfactif peut changer selon son intensité. De même, le ton hédonique de l'odeur dépend de sa qualité, mais aussi de sa concentration (une odeur agréable à faible concentration peut s'avérer désagréable à très forte concentration).

(Bensafi, 2001) a démontré que ces trois dimensions diffèrent d'un point de vue temporel : les délais de réponses au jugement d'intensité sont moins importants que ceux de la dimension hédonique. Les temps de réponses associés à l'évaluation de la qualité sont les plus longs (Bensafi, 2001). Généralement, une augmentation de l'intensité perceptive d'une odeur s'accompagne d'une diminution de la note hédonique (Distel *et al.*, 1999). Cette relation n'est pas toujours vérifiée et elle reste dépendante de l'odeur étudiée. Par exemple, la vanilline est agréable quelle que soit la concentration à laquelle elle est présentée.

## 2.4 Evaluation de l'odeur

La complexité de l'évaluation de l'odeur revient en grande partie à la difficulté pour des non-spécialistes d'identifier ou de décrire une odeur. L'industrie a longtemps compté sur des procédures bien définies pour l'évaluation de la qualité de l'odeur, comme par exemple dans le cadre des expériences d'analyse sensorielle en agro-alimentaire. Ces expériences sont cependant très coûteuses pour l'industriel.

L'évaluation de la qualité des odeurs est souvent faite en utilisant des descripteurs linguistiques contrôlés fournis par des spécialistes formés capables de qualifier l'odeur en utilisant des descripteurs précis. Parmi les approches utilisées se retrouvent particulièrement la roue des odeurs (Paravisini *et al.*, 2014; Suffet & Rosenfeld, 2007) et le champ des odeurs (Jaubert *et al.*, 1987a; Jaubert *et al.*, 1995; Léger, 2005). Ces méthodes utilisent une approche orientée sur la base d'un référentiel commun pour qualifier les odeurs, ce qui permet de faciliter la caractérisation de l'odeur. En effet, l'obligation faite aux évaluateurs d'utiliser des descripteurs spécifiques facilite la compréhension, l'interprétation et le traitement des résultats. Néanmoins, une phase d'apprentissage des descripteurs est nécessaire pour utiliser ces méthodes, ce qui limite, voire interdit leur utilisation par des non-spécialistes, par exemple par des consommateurs. Ces méthodes présentent quelques inconvénients pour l'industriel :

- des coûts supplémentaires de formation,
- une limitation du nombre de descripteurs grâce auxquels les experts décrivent l'odeur (perte d'information),

- une limitation du nombre d'évaluateurs et d'expériences qui peuvent être utilisés pour évaluer une odeur.

#### 2.4.1 Évaluation de la qualité d'une odeur

L'étude scientifique de l'odorat a été longtemps négligée par rapport à celle de la vue et de l'ouïe. Plusieurs causes expliquent ce retard :

- l'odorat est toujours considéré comme un sens mineur : sa perte n'est en effet toujours pas considérée comme grave par rapport à celle de la vue ou de l'ouïe.
- l'existence d'un très grand nombre de molécules odorantes qui est à l'origine d'un très grand nombre de combinaisons odorantes, comme le souligne André Holley (Holley, 1997) : « *il est compréhensible que la langue n'ait pu individualiser par un terme spécifique chacune des sensations que fournit le monde olfactif, parce qu'il y en a beaucoup trop* ».

Malgré la complexité de l'étude de l'odorat, ces dernières années, de nombreux travaux ont été consacrés à la compréhension du fonctionnement du système olfactif, à la façon dont l'homme perçoit l'odeur et restitue sa perception (Mainland *et al.*, 2014). A la différence du goût, pour lequel il est possible de définir quatre catégories (salé, sucré, amer, acide), il est difficile de définir des catégories d'odeurs fondamentales, même si l'on rencontre quelques travaux sur la classification et la catégorisation d'odeurs qui s'avèrent en réalité n'être qu'une assimilation à une odeur assez fréquemment rencontrée. La catégorisation des odeurs est considérée comme le domaine emblématique du désaccord et de la subjectivité, ce que confirme l'absence de grille universelle de représentation des odeurs (Dubois *et al.*, 1997; Dubois, 2006).

Le langage des odeurs est un domaine de plus en plus exploré par les chercheurs. Les principales études s'intéressent à l'identification des odeurs, à leur catégorisation et à leur dénomination. Ce qui ressort principalement est que le vocabulaire du domaine olfactif est pauvre et que le nombre des termes mis à la disposition des non-spécialistes est réduit (David *et al.*, 1997; Dufour & Barkat-Defradas, 2009). L'olfaction est un sens qui peut être considéré « sans parole » (Howes, 1986). Nous avons souvent l'impression de connaître une odeur sans être capable de la nommer (Jönsson, 2005), Cette impression se verbalise par l'utilisation d'expressions comme « tip of the nose » ou « avoir l'odeur sur le bout du nez ». (Dubois *et al.*, 1997) notent que « *cela s'explique par le fait qu'il n'y a pas (ou peu) de formes lexicales stabilisées qui permettent au locuteur de se référer à un monde des odeurs à partir d'une doxa* ».

La dénomination des odeurs apparaît ainsi comme un processus complexe. L'expression de ces sensations fait appel à l'expérience humaine et à la culture. Cela revient à dire que l'on peut observer une grande variabilité entre les réponses de sujets quand on leur demande de nommer une odeur, contrairement à ce que l'on observe pour les couleurs (Candau, 2004; Howes, 1986). Holley parle de « *vaste réservoir d'expérience humaine des odeurs* » (Holley, 1997).

En résumé, dans la majorité des travaux existants, les résultats obtenus pour qualifier les odeurs révèlent que les formes employées ne renvoient pas à un vocabulaire spécifique et précis mais plutôt soit à la source identifiée être à l'origine de l'odeur (*e.g.* une odeur de rose), soit à une classification des odeurs qui est organisée selon la dimension hédonique distinguant les bonnes odeurs des mauvaises odeurs (*e.g.* une odeur agréable) (Rouby & Sicard, 1997). Cette dimension hédonique de l'odeur constitue la première dimension explicative de l'organisation cognitive des odeurs, les espaces de similitudes entre odeurs varient tant d'une personne à l'autre que la seule dimension commune serait hédonique. L'examen d'autres langues (Boisson, 1997) et d'autres cultures montre que le jugement hédonique est

le mieux représenté dans les langues, et que sa valeur négative (les mauvaises odeurs) qui partitionne l'espace olfactif, est la plus stable entre les sujets et entre les cultures (Schaal *et al.*, 1998).

Le vocabulaire du domaine olfactif est imprécis, le même descripteur peut avoir plusieurs référents et une même odeur peut être associée à plusieurs descripteurs (Candau, 2004). La description de l'odeur est très souvent rapportée à un objet, un lieu, qui sont supposés être à l'origine de l'odeur. De la même manière, les sujets empruntent souvent des termes dans d'autres domaines sensoriels pour qualifier une odeur en utilisant des termes qui n'appartiennent pas strictement à l'univers olfactif. Les odeurs peuvent ainsi être définies comme douces ou piquantes, sucrées ou amères... (Holley, 1997).

Les problèmes du manque du vocabulaire dans le domaine olfactif et de la difficulté à nommer les odeurs font rapidement émerger le besoin d'une démarche dans la dénomination ou la caractérisation. C'est pourquoi un vocabulaire est généralement constitué, soit en utilisant des experts sélectionnés sur la base de leurs capacités à exprimer leurs sensations pour un domaine donné, soit sur la base d'expérimentations menées avec des sujets non spécialistes à qui l'on propose de décrire les odeurs par un choix forcé entre plusieurs descripteurs correspondant le plus souvent à ceux des sources odorantes.

Cette démarche, qualifiée d'**approche orientée**, est la méthode utilisée actuellement pour évaluer la qualité des odeurs. Elle nécessite l'apprentissage du vocabulaire lié aux odeurs d'un domaine donné. Cette démarche force les évaluateurs à utiliser un nombre limité de descripteurs, ce qui peut rendre l'évaluation de la qualité de l'odeur moins fiable et surtout moins représentative de la réalité du ressenti collectif.

## 2.4.2 Apprentissage des odeurs

L'apprentissage des odeurs par un individu est un processus d'acquisition d'informations pour l'identification et la caractérisation de l'odeur. La première étape de l'apprentissage consiste à sentir les odeurs de façon répétitive pour apprendre à les reconnaître et à les nommer, comme le font les « nez » de l'industrie de la parfumerie. Il faut de nombreux mois pour assimiler et devenir capable d'identifier les différentes odeurs. (Engen & Ross, 1973) ont montré que l'identification d'une odeur est plus facile lorsque le descripteur de cette odeur est indiqué au sujet lors de l'apprentissage. Nommer une odeur permet d'améliorer sa reconnaissance (Cain *et al.*, 1998; Rabin, 1988). Cette amélioration reste spécifique à l'odorant sur lequel porte l'entraînement (Rabin, 1988). Cependant, un nom sémantique pauvre peut diminuer la capacité à reconnaître une odeur (par exemple le nom « bubble gum » au lieu de « arôme framboise ») (Engen & Ross, 1973).

Aujourd'hui, les industriels ont développé des classifications spécifiques pour aider à la description comme le « Champ des odeurs » pour la parfumerie (Dratz, 2000; Jaubert *et al.*, 1995) et la « Roue des arômes » pour l'œnologie (Noble *et al.*, 1984).

Le processus d'apprentissage est favorisé par différents facteurs. (Rabin, 1988) démontre que le contact répété avec certaines odeurs améliore la capacité à les discriminer et à les identifier. Dans l'expérience de (Lyman & McDaniel, 1990), les meilleurs résultats de reconnaissance lors d'apprentissage sont atteints lorsque l'odeur est associée en même temps à un descripteur et à une image.

### 2.4.2.1 Définition de l'expertise

L'apprentissage est capable de modifier la mémoire et d'orienter la représentation de l'odeur. Cet apprentissage permet donc de développer une expertise, c'est-à-dire une capacité à décrire, catégoriser et nommer les objets appartenant au domaine d'expertise (Johnson & Mervis, 1997).

Il est possible de distinguer deux niveaux d'expertise :

- « L'expertise produit » définie par (Nicod *et al.*, 2006) comme « *une personne qui, se confrontant très régulièrement au même univers produit a développé des images sensorielles propres aux différents produits de son univers et qui utilise ensuite ces connaissances pour déguster et prendre des décisions liées au produit* » ;
- « L'expertise sensorielle » définie dans les normes NF EN ISO 8586 de Février 2014 (Organisation-Internationale *et al.*, 2005). Selon cette norme, les experts sensoriels, en plus d'une acuité particulière sont entraînés à l'utilisation des méthodes d'analyse sensorielle et sont capables d'effectuer des analyses sensorielles sur différents produits.

Certain travaux montrent que la vitesse d'apprentissage des odeurs agréables est lente, par rapport à l'apprentissage des odeurs désagréables. Ceci montre que l'apprentissage dépend du ton hédonique de l'odeur. L'apprentissage des odeurs est aussi facilité par la formation de classes ou de catégories (Sauvageot, 1990). Lors de la perception, le sujet classe les odeurs dans des structures qui représentent son modèle personnel de la réalité. Dans le domaine sémantique, la catégorisation naturelle a été expliquée par la formation de prototypes, c'est à dire comme l'action de classer les objets par catégorie selon un certain nombre de critères. Selon la théorie des prototypes (Rosch *et al.*, 1976), les catégories naturelles s'organisent autour d'un modèle, un bon représentant de la classe, que l'on appelle prototype. Chaque membre de la catégorie partage quelques caractéristiques avec le prototype. Selon (Herz & Engen, 1996), l'association d'une odeur à un événement peut permettre une meilleure reconnaissance qu'un apprentissage forcé. L'expertise s'acquiert progressivement après de nombreuses années de pratique et elle est limitée à un domaine donné. Un expert pour un domaine d'odeur donné est donc une personne qui a appris à qualifier ces odeurs en utilisant un ensemble de descripteurs bien précis et contrôlés. L'industriel, aujourd'hui, compte sur des experts qui maîtrisent un langage précis pour qualifier les odeurs. Cette dépendance envers les experts entraîne des coûts significatifs.

L'objectif global dans cette thèse est d'évaluer la qualité de l'odeur sans recours à des experts, ce qui permet d'éviter l'étape d'apprentissage du vocabulaire qui est à ce jour une étape obligatoire contraignante et coûteuse pour l'industriel.

#### 2.4.2.2 Langage naturel des odeurs

Les sujets novices ont une capacité limitée à décrire une odeur car ils possèdent peu de termes spécifiques pour exprimer leurs perceptions (Berglund *et al.*, 1973; Rouby *et al.*, 2002). Généralement, le sujet est capable de connaître une odeur sans être capable de la dénommer (Lawless & Engen, 1977). Comme le souligne (Rouby *et al.*, 2005) cette impression « *témoigne d'une mémoire implicite des odeurs, qu'il n'est pas toujours possible de relier aux formes explicites de mémoire, en particulier à la mémoire sémantique, qui relie entre elles nos connaissances* ». Contrairement aux couleurs, le lexique olfactif est imprécis, instable, le même descripteur peut avoir plusieurs référents et, une même odeur, peut être associée à plusieurs descripteurs.

La verbalisation olfactive dépend généralement de l'expérience de chacun, là encore contrairement à l'apprentissage des couleurs. Il en résulte que la dénomination des odeurs sous une seule étiquette est presque toujours impossible : « *l'eugénol composé aromatique associé au clou de girofle est identifié par une partie de la population comme l'odeur "du dentiste" du fait de son emploi comme désinfectant dentaire, et comme une odeur épicée ou chimique par les sujets qui ne peuvent l'identifier. En l'absence d'une norme culturelle ou sémantique, ces réponses sont tout aussi valables* » (Rouby & Sicard, 1997). Le langage naturel des odeurs est imprécis, non contrôlé, variable, émotionnel (un stimulus évoque par exemple un événement), ce qui est significatif de l'absence d'un référentiel commun. Ce référentiel commun est le propre de l'expertise sensorielle.

L'objectif de la thèse sera de construire une projection des termes du langage naturel vers un vocabulaire contrôlé qui permette une communication objective de la qualité de l'odeur, une communication axée sur un profil sensoriel compris par le monde professionnel des odeurs. A partir de descriptions d'une odeur en langage naturel, nous allons donc chercher à définir un ensemble de descripteurs linguistiques compris par le monde professionnel de l'odeur qui permette de caractériser au mieux les odeurs sans faire appel à des experts. Autrement dit, nous cherchons à proposer une approche qui permette de transcrire les descriptions libres données en langage naturel en descripteurs structurés ou descripteurs conceptuels. Le but final sera ensuite de résumer cet ensemble de descripteurs structurés en une caractérisation unique et cohérente pour le monde professionnel.

#### **2.4.2.3 Evaluation de la qualité de l'odeur : novices ou experts ?**

Dans la littérature, plusieurs travaux s'intéressent à la comparaison de l'évaluation de la qualité de l'odeur chez les experts et les novices. (Bende & Nordin, 1997) ont montré que la capacité d'identification des odeurs est plus développée chez les experts que chez des novices, alors que d'autres auteurs (Parr *et al.*, 2002; Parr *et al.*, 2004) n'ont relevé aucune différence de capacité d'identification entre les deux populations... (Gawel, 1997) vérifie que la performance de juges novices est meilleure lorsqu'ils doivent retrouver des vins à partir de la description d'un expert qu'à partir de celle d'un novice. La supériorité des experts a été retrouvée par (Solomon, 1990). Une analyse des termes utilisés par les experts et les novices montre que les experts emploient un vocabulaire plus précis et plus expressif que les novices qui, même s'ils identifient des nuances olfactives, ne sont pas capables de les restituer précisément par manque d'habitude et de vocabulaire descriptif. Dans la même veine de travaux, (Chollet & Valentin, 2000) indiquent que les descriptions réalisées par des experts comportent des termes précis alors que les novices doivent se contenter de descripteurs plus généraux.

(Valentin *et al.*, 2003) ont demandé à un groupe d'experts et un groupe de novices de décrire des vins à l'aide d'une liste de termes organisée en trois niveaux de catégorisation allant du plus général au plus spécifique (arôme des vins). Ensuite, ils ont demandé aux mêmes sujets d'apparier les descriptions réalisées par un autre sujet avec les vins correspondants. Les résultats montrent que les descriptions effectuées par les experts sont plus précises et conduisent à de meilleures performances d'appariement que celles effectuées par les novices. De plus, la distance entre les descriptions d'experts et celles de novices augmente avec le niveau de précision des termes.

(Chollet *et al.*, 2005) ont également évalué la supériorité des juges formés pour de nouveaux stimuli (odeurs). Pour ce faire, un groupe de sujets est formé pour caractériser des saveurs de bière sur une période de deux ans. Une fois que la formation a été accomplie, ils ont comparé les performances de ces évaluateurs formés avec celles de sujets novices pour les tâches de discrimination et d'appariement (une tâche d'appariement consiste pour les sujets à reconnaître les échantillons en fonction de leur description). Les tâches ont été effectuées en utilisant deux catégories de bières (nouvelles bières et bières déjà connues). Les résultats ont montré que les évaluateurs formés ont surclassé les novices dans la tâche de discrimination pour les bières qu'ils avaient appris à reconnaître, mais pas pour les nouvelles bières. Pour le test d'appariement, les évaluateurs formés ont surclassé les novices pour les deux types de bières. Ces résultats suggèrent que l'apprentissage perceptif (tâche de discrimination) est plus difficile à généraliser que l'apprentissage verbal (tâche d'appariement).

D'autres travaux montrent que les experts ont développé des compétences sémantiques pour décrire plus précisément les vins contrairement aux novices (Hughson & Boakes, 2002; Solomon, 1990). (Valentin & Chollet, 2015) constatent encore que le vocabulaire utilisé par les experts offre une plus grande richesse et une meilleure précision que celui des novices pour la description d'une collection de vins et

permet donc une meilleure description des vins. De même, (Tempère, 2010) confirme la meilleure précision des experts par rapport aux novices dans une tâche d'identification .

A contrario, (Lawrence *et al.*, 2013) ont comparé la méthode classique utilisée pour l'analyse sensorielle et la méthode à commentaire libre (qualité de l'odeur exprimée en langage naturel) pour la description de trente-et-un vins de Cabernet Franc de la région Vallée de la Loire. Les propriétés odorantes des vins ont été décrites par deux panels distincts effectuant respectivement un profilage classique (experts utilisant un référentiel de descripteurs) et une méthode à commentaires libres (néophytes s'exprimant en langage naturel). Les deux méthodes ont été mises en œuvre par des professionnels du vin sans formation préalable des évaluateurs aux produits. Les résultats montrent qu'un grand nombre de descripteurs ont été générés par la méthode des commentaires libres, fournissant ainsi des informations riches sur l'ensemble des produits. Les deux méthodes d'évaluation ont néanmoins fait ressortir des caractéristiques similaires des vins de Cabernet Franc. Ce résultat laisse entendre que la méthode des commentaires libres fournit au moins autant d'informations qu'une évaluation classique par des experts alors qu'elle est très avantageuse en coût et temps, et peut représenter une alternative économique à l'analyse descriptive classique. Notre proposition s'inscrit dans cette voie, mais propose des modèles et outils informatiques pour contrôler les transcriptions des descripteurs d'un univers à l'autre.

## 2.5 Méthodes pour l'évaluation de la qualité de l'odeur

Parmi les approches proposées dans la littérature, la roue des odeurs (Suffet & Rosenfeld, 2007) et le champ des odeurs (Jaubert *et al.*, 1987a), sont souvent utilisés pour l'évaluation de la qualité de l'odeur. Ces méthodes sont dites orientées puisqu'elles se basent sur un référentiel de descripteurs partagé pour qualifier les odeurs. Une étape d'appropriation de ce vocabulaire ou d'apprentissage des référents est donc nécessaire. L'approche orientée facilite la caractérisation et la description de l'odeur. En effet, en contraignant les évaluateurs à utiliser des descripteurs spécifiques et précis, la compréhension, l'interprétation et le traitement des résultats sont facilités pour les analyses aval de l'industriel.

A l'inverse, dans une population non formée, l'appréciation de la qualité d'une odeur se limite à des mécanismes de rapprochement entre l'odeur perçue à une odeur déjà rencontrée (« ça sent comme », etc.). Cela rend difficile l'exploitation de cette information, car plusieurs descripteurs peuvent renvoyer en réalité à une même qualité d'odeur.

### 2.5.1 Méthodes expertes

#### 2.5.1.1 Roues des odeurs

Le principe de la roue des odeurs est d'associer une odeur aux descripteurs d'un référentiel. La roue des odeurs rassemble des termes ou descripteurs décrivant un arôme (Suffet & Rosenfeld, 2007). Ces termes sont disposés en couches qui vont du plus général (au centre) au plus spécifique (en bordure, *ex. Chimique >Médicament*). Le but de la présentation de ces niveaux de référents est d'orienter la perception vers plus de précision : les descripteurs du cercle extérieur sont les plus précis ou spécifiques (exemple Figure 2.5).

## RONDE DES AROMES DE MIEL

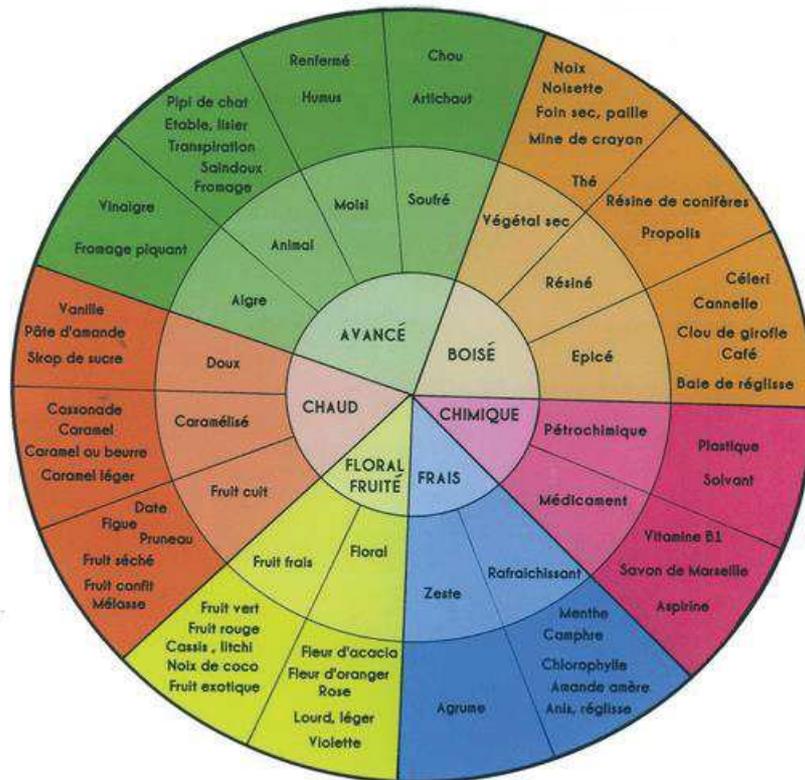


Figure 2.5 : Exemple d'une roue des odeurs (Bruneau *et al.*, 2000)

Généralement, les experts décrivent les odeurs à l'aide de termes de la couche périphérique de la roue des odeurs, puisque ces termes sont plus expressifs et plus précis, contrairement aux non-spécialistes qui utilisent des descriptions plus générales qui correspondent la plupart du temps aux premiers niveaux de la roue. Par exemple dans l'expérience de (Valentin & Chollet, 2015), les experts, formés collectivement à déguster des vins selon une roue des odeurs, se « comprennent mieux » et attribuent donc le même descripteur à une même odeur en utilisant des descripteurs précis et partagés de la roue d'apprentissage.

### 2.5.1.2 Champ des odeurs®

Le champ des odeurs® est une méthode de classification des odeurs. Cette méthode est relativement similaire à la précédente. Elle permet d'obtenir des profils olfactifs objectifs. Elle est mise en place en 1987 (Jaubert *et al.*, 1987a; Jaubert *et al.*, 1987b) à la suite de recherches sur les relations entre la structure chimique et le caractère odorant des molécules. L'approche que propose Jaubert s'appuie sur un travail réalisé avec 1400 molécules odorantes de base. Il retient 45 molécules référentes qui se répartissent dans un espace multidimensionnel structuré. Cette cartographie cherche à positionner à proximité les unes des autres les odeurs qui sont proches, indépendamment de toute proximité en termes de structure chimique. Les 45 référents sont appris progressivement par les juges afin de pouvoir positionner les odeurs à évaluer.

Développée initialement pour la caractérisation des parfums, cette méthodologie a été adaptée ensuite à l'étude des nuisances odorantes, pour la perception olfactive sur les sites émetteurs par les populations riveraines (Dratz, 2000). Les figures suivantes présentent les configurations 2D et 3D qui résultent de ces travaux.

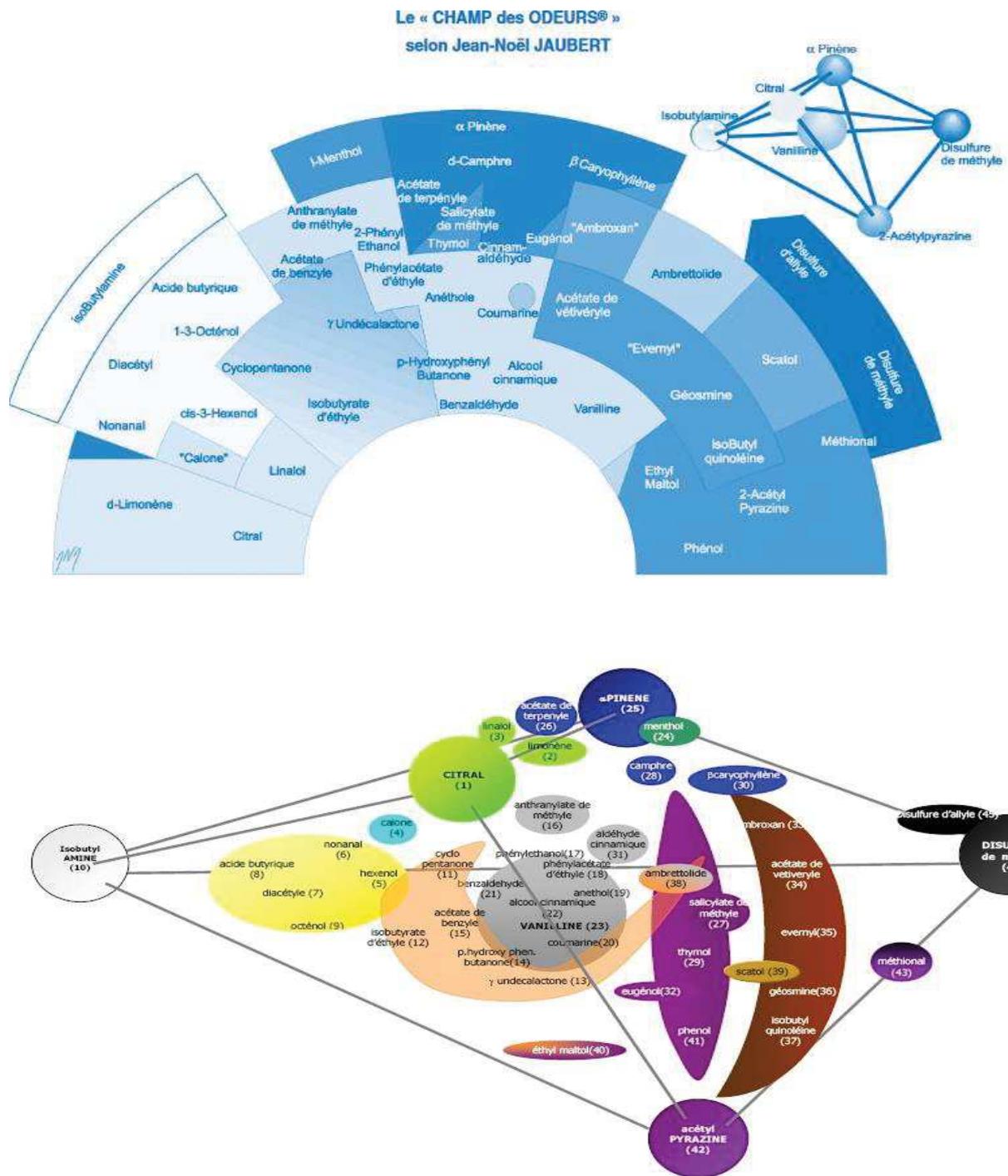


Figure 2.6 : Le champ des odeurs® (Jaubert *et al.*, 1987a; Jaubert *et al.*, 1987b)

## 2.5.2 Approche classique dite orientée

Une approche classique en analyse sensorielle consiste à réunir un panel de jurys ou évaluateurs entraînés pour déterminer la qualité d'un produit (gustative, olfactive, etc.). Ces jurys ont acquis un vocabulaire structuré et partagé et sont entraînés à reconnaître des saveurs, odeurs élémentaires. L'approche dite classique consiste donc dans un premier temps à instruire un ensemble de descripteurs structurés et précis propres à un domaine (ex., vins, miels, yaourts, etc.) qu'il conviendra de faire adopter par les panels de jurys auxquels on aura recours. Ce vocabulaire se veut suffisamment riche et précis pour garantir l'information nécessaire à l'industriel pour que celui-ci bénéficie d'un diagnostic utile pour améliorer la qualité de ces produits. Ainsi, par exemple, des évocations structurées relativement à une odeur permettront à l'industriel de remonter aux molécules susceptibles d'être responsables d'une gêne

olfactive. Les différentes étapes pour l'obtention d'un profil sensoriel sont présentées sur la Figure 2.7 : constitution du panel d'évaluateurs, construction d'une liste de descripteurs contrôlés, entraînement des évaluateurs à l'utilisation de cette liste de descripteurs de façon discriminante et répétable, et enfin évaluation des produits.

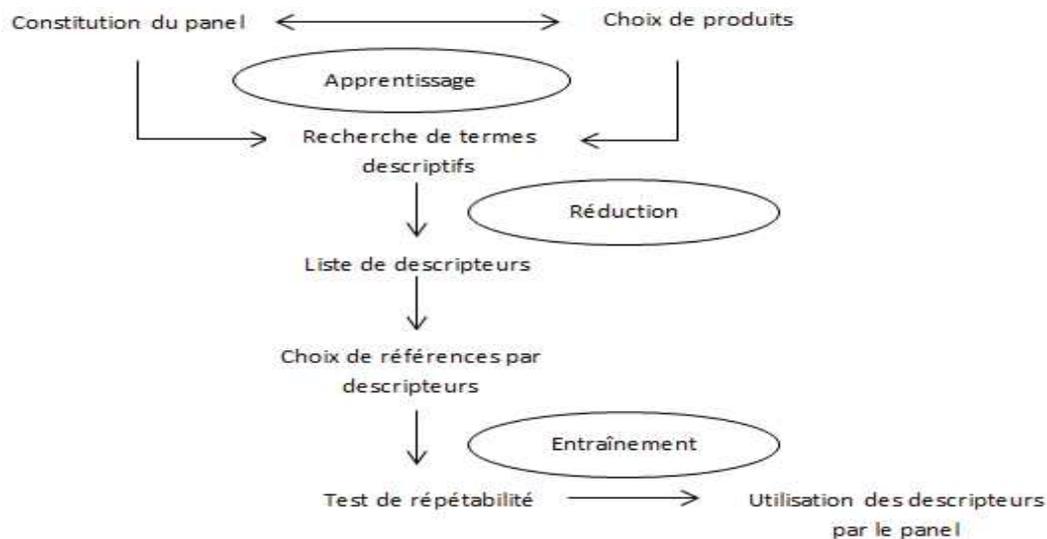


Figure 2.7 : Principales étapes de la méthode classique (d'après ISO 11035, 1994)

Les premières étapes se focalisent sur la constitution d'un panel d'évaluateurs experts qui puissent constituer le vocabulaire du domaine sur un ensemble de produits représentatifs. Notons qu'il convient de s'assurer que les évaluateurs perçoivent normalement les odeurs car la notion de seuil de détection olfactive peut différer significativement d'un individu à un autre (Lawless & Heymann, 1999).

Chaque évaluateur caractérise les échantillons représentatifs du domaine : celui-ci s'exprime librement, mais néanmoins avec un vocabulaire précis que lui confère son statut d'expert de sorte à différencier au mieux les produits du lot. L'union des descripteurs proposés par tous les évaluateurs doit être un ensemble riche de concepts qui couvrent le domaine d'évaluation. La liste est ensuite réduite par le collectif d'experts en supprimant les termes non pertinents, comme les termes qui viseraient à caractériser le caractère hédonique du produit plutôt que sa qualité, à préciser l'intensité d'un arôme, d'une odeur, d'une saveur, etc., les synonymes et antonymes. Cette réduction peut être réalisée par des outils de filtrage et de sélection statistiques (Barthelemy, 1990) et/ou par consensus (Lawless & Heymann, 1999). Finalement, les évaluateurs décident collectivement de la liste des descripteurs. Ils sont ensuite entraînés à l'utilisation de ces descripteurs sur les produits qui ont servi à la constitution du vocabulaire contrôlé afin de vérifier que ce vocabulaire réduit permet d'obtenir des caractérisations individuelles cohérentes. La précision du vocabulaire retenue doit permettre de caractériser le plus unanimement possible les sensations significées par le panel. Une fois la phase d'entraînement terminée, les séances d'évaluation peuvent alors avoir lieu, dans des conditions au protocole expérimental contrôlé (Lawless & Heymann, 1999). Les produits sont présentés un par un, souvent plusieurs fois et dans des ordres différents et les évaluateurs doivent déterminer l'intensité de chacun des descripteurs pour chacun des produits sur des échelles définies au préalable.

### 2.5.2.1 Avantages et inconvénients

Généralement, les données obtenues peuvent être analysées facilement par des techniques statistiques, comme des analyses de variance (*ANOVA*), des tests de comparaisons multiples, des analyses en composantes principales (*ACP*) et des classifications hiérarchiques, etc. Selon (MacLeod *et al.*, 1998), cette approche de la qualité est aujourd'hui encore le meilleur outil disponible pour expliquer le choix des consommateurs *a posteriori* ou prédire l'accueil par le public *a priori*. Cette méthode d'évaluation

de la qualité s'impose comme un instrument de mesure supplémentaire pour l'industriel, largement reconnu comme robuste. Cependant, la formation des évaluateurs est un passage obligé ; elle est longue, fastidieuse et doit être continûment entretenue. La mise en œuvre de cette approche est un processus complexe puisqu'il est nécessaire de réunir les évaluateurs de nombreuses fois : le nombre d'heures d'entraînement peut aller jusqu'à 120h (Meilgaard *et al.*, 2006).

Pour cette raison, une campagne d'évaluation sensorielle peut être écartée voire abandonnée lorsque les résultats de l'étude sont urgents. De plus, malgré la constitution consensuelle du vocabulaire contrôlé, des désaccords ne peuvent être résolus entre les évaluateurs et ce en dépit de la phase d'entraînement (Bárceñas *et al.*, 2003; Barthelemy, 1990; Delarue & Sieffermann, 2004).

Enfin, la recherche du consensus entre les évaluations individuelles peut conduire à produire des synthèses qui correspondent à un petit nombre de descripteurs, peu spécifiques, mais consensuels pour le collectif d'évaluateurs : cette information réduit drastiquement l'information acquise lors de la séance d'analyse sensorielle. Selon (Faye *et al.*, 2004) l'utilisation d'un vocabulaire consensuel, défini par des évaluateurs entraînés, n'est pas une technique représentative des caractéristiques sensorielles telles qu'elles peuvent être perçues par les consommateurs néophytes (contrairement à l'étude de (MacLeod *et al.*, 1998)). Il faut en effet noter qu'un même descripteur peut évoquer des sensations différentes pour les évaluateurs. D'autres limitations proviennent de l'utilisation de ces techniques dans l'environnement industriel, il n'est pas toujours possible pour les analystes sensoriels en entreprise d'analyser une très grande quantité de produits.

#### 2.5.2.2 Conclusion

L'analyse sensorielle est un ensemble de méthodes permettant de mesurer les perceptions sensorielles (vue, ouïe, odorat, goût, toucher). On parle de sensométrie ou de métrologie sensorielle. Pour cette « mesure sensorielle », il n'existe pas d'autres capteurs que l'homme. Les industriels, en particulier dans l'alimentaire, emploient donc des professionnels exercés pour tester, qualifier et évaluer les produits nouveaux qu'ils souhaitent lancer sur le marché ou bien des produits dont certains paramètres physicochimiques ont été modifiés pour des raisons de coût par exemple. Il faut donc imaginer que l'homme, avec toute l'imprécision et la subjectivité qui caractérisent ses évaluations sensorielles, se trouvent au cœur d'une chaîne de fiabilité. Les approches orientées visent à limiter l'impact de cette subjectivité et cette imprécision en contraignant les experts à s'exprimer dans un référentiel métier partagé. En revanche, ce garde-fou peut être pénalisant pour évaluer la variabilité des perceptions (*e.g.*, tout le monde peut s'accorder sur une fragrance d'agrumes si « agrumes » est le seul descripteur disponible dans le vocabulaire contrôlé alors que les évaluateurs auraient été capables de discerner s'il s'agissait de citron, pamplemousse ou de mandarine. Cette précision qui aurait mesuré la variabilité des perceptions est alors perdue.). Enfin, la maîtrise de ce vocabulaire contrôlé représente une sévère contrainte sur la constitution d'un panel de jurys.

#### 2.5.3 Approche à commentaires libres dite non orientée

Des alternatives à cette approche coûteuse et exigeante ont été envisagées pour identifier la qualité de produits. Dans la littérature, on trouve plusieurs désignations pour ces approches où les jurys ne sont plus contraints dans l'expression de leurs évaluations. (Williams & Langron, 1984) appellent leur méthode le « profilage libre » (Free-choice Profiling) et l'appliquent à l'évaluation libre de vins. (Lawrence *et al.*, 2013) ont développé une méthode qu'ils ont baptisée « la méthode à commentaires libres » toujours pour la caractérisation de vins. (Jack & Piggott, 1992) expliquent que ce type de méthodes diminue significativement le coût de la sensométrie en s'affranchissant de la phase d'apprentissage, mais qui plus est, constitue un observateur plus fiable de la réception possible par le

public. Ces approches ont recours à des analyses statistiques relativement simples pour analyser et filtrer l'ensemble du vocabulaire fourni par les consommateurs. Ces méthodes ne contraignent pas le choix des descripteurs (Gerland & Dumont, 2000), elles laissent aux évaluateurs une liberté totale quant au choix des descripteurs. Elles permettent de s'affranchir de la phase d'entraînement des évaluateurs et de mesurer toute la variabilité des évaluations qui est en elle-même une information sensorielle pleine d'enseignements.

Chaque évaluateur est ainsi libre d'utiliser ses propres mots pour décrire les produits. Cette liberté d'expression permet d'avoir accès à des panels d'évaluateurs plus conséquents et par la suite de disposer d'un vocabulaire extrêmement riche pour caractériser une gamme de produits. Si le vocabulaire généré se veut quasi exhaustif, il s'agit ensuite de pouvoir gérer la richesse de cette information, sa complexité, ses redondances, ses corrélations. La possibilité de s'exprimer librement sur le perçu et le ressenti offre davantage de nuances dans la restitution, propriété que bien d'autres domaines comme la psychologie, la sociologie, la publicité et le marketing exploitent déjà (Lebart, 2012). L'industrie alimentaire, en particulier, a pris conscience de l'intérêt de recueillir des avis, certes moins structurés, mais plus représentatifs, directement auprès du consommateur (Martin & Rogeaux, 1994; Symoneaux *et al.*, 2012). Dans la littérature de l'analyse sensorielle, de nombreuses études sont consacrées à l'approche non orientée, en particulier pour les boissons : caractérisation de portos pour (Cristovam, *et al.*, 2000; Williams & Langron, 1984), de cafés pour (Williams & Arnold, 1985) de vins pour (Gerland & Dumont, 2000; Narain *et al.*, 2004), de bières pour (Gains & Thomson, 1990) ou encore de boissons gazeuses pour (Lachnit *et al.*, 2003).

L'avantage majeur de l'approche à commentaires libres est qu'elle ne nécessite pas l'acquisition d'un vocabulaire spécifique par les évaluateurs et qu'elle réduit ainsi les coûts tout en augmentant le nombre d'évaluateurs potentiels. Les résultats des études montrent qu'il est possible de caractériser les plus grandes différences entre les produits étudiés. Cette approche peut ainsi être utilisée comme une alternative rapide et moins coûteuse à l'approche classique. Toutefois, il est difficile d'obtenir une caractérisation aussi précise qu'avec l'approche orientée. De plus, la diversité du vocabulaire utilisé, notamment avec des évaluateurs n'ayant pas de références communes, rend l'interprétation des résultats parfois difficile et pourrait prendre plus de temps (Barthelemy, 1990; Cristovam *et al.*, 2000). (Lachnit *et al.*, 2003) ont montré que des consommateurs fournissent souvent des caractérisations très divergentes dans le cas de l'approche à commentaires libres appliquée à des sodas. Il semble donc assez pertinent d'utiliser des évaluateurs déjà sensibilisés à l'évaluation sensorielle et à la description de produits étudiés. (Lawrence *et al.*, 2013) ont montré que l'approche non orientée est adaptée aux professionnels du vin et à la caractérisation des vins. Elle peut être utilisée comme un outil de cartographie sensorielle pratique pour conduire les phases préliminaires des études sensorielles, par exemple dans l'étape de génération de vocabulaire de l'approche orientée.

## 2.5.4 Vers des approches hybrides de l'évaluation de la qualité

Plusieurs études ont proposé des approches non orientées pour identifier la qualité d'une odeur, d'un arôme, etc. en analyse sensorielle. Les descriptions libres n'y permettent plus de garantir la non ambiguïté des termes utilisés, mais correspondent à de simples sacs de termes extraits du langage naturel. La majorité des travaux cités en littérature introduisent alors une démarche pour extraire un vocabulaire structuré et précis en analysant l'ensemble des termes utilisés par des évaluateurs non nécessairement experts pour caractériser un ensemble de produits représentatifs d'un domaine. L'idée est de proposer une démarche plus ou moins formelle pour transformer les descriptions en langage naturel fournies en descriptions conceptuelles afin qu'elles puissent être comprises par le monde professionnel. Ainsi, peut-on bénéficier conjointement des avantages des approches orientées et non

orientées. Dans ce qui suit, nous citons quelques travaux représentatifs des approches basées sur un choix libre de descripteurs.

Dans (Williams & Langron, 1984), chaque évaluateur utilise son propre vocabulaire pour l'évaluation d'une collection de 8 vins. Cette étude est menée par 10 évaluateurs, experts et non experts. Chaque évaluateur est invité à énumérer des descripteurs sans utiliser de termes hédoniques ou de termes liés à l'intensité qui biaiseraient l'identification objective de la qualité. Pour guider la construction d'un vocabulaire objectif, les évaluateurs sont invités à regrouper les termes en 3 catégories : termes liés à l'apparence, à l'arôme et à la saveur. Chaque évaluateur caractérise la série des 8 vins en utilisant ses propres termes qu'il range dans l'une des trois catégories. L'étude vise à construire un vocabulaire expert pour chaque catégorie, qui permette d'accéder à une communication objective de la caractérisation des 8 vins et qui fasse sens pour des professionnels de la filière viticole. Pour chaque catégorie, apparence, arôme et goût, les termes recueillis sont soumis à une analyse statistique simple. A chaque terme qu'il emploie, un évaluateur associe une intensité. L'union des termes constitue le vocabulaire de l'expérience. Un expert peut réduire ce vocabulaire en introduisant des relations de synonymie ou de similarité sémantique. Une fois le vocabulaire réduit, les termes retenus constituent les dimensions de l'espace d'évaluation. L'évaluation d'un vin par un évaluateur est un point de cet espace dont les coordonnées sont les degrés d'intensité fournis par l'évaluateur (les termes non évoqués ont une intensité de 0). Pour chaque évaluateur, les 8 vins sont ainsi représentés par 8 points dans l'espace défini par le vocabulaire. La distance entre les points modélise les différences et les similitudes qu'il existe entre les caractérisations des vins. Pour chaque vin, l'isobarycentre des 10 points fournis par les 10 évaluateurs est calculé. Cette représentation permet de définir une caractérisation consensuelle dans chacune des 3 catégories. De la même manière, les auteurs proposent de calculer un centre de gravité sur les 8 évaluations d'un évaluateur : un évaluateur est alors vu à son tour comme un point de l'espace d'évaluation et la synthèse du centre de gravité permet une évaluation simple de la variabilité/similarité entre évaluateurs.

Ce modèle vectoriel très simple permet ainsi d'obtenir des caractérisations à partir d'un vocabulaire non contrôlé et d'évaluer la variabilité des évaluateurs. Si l'idée paraît séduisante, la modélisation est plus discutable. Par exemple, deux vins dont quatre des coordonnées seraient (*citron : 1, curcuma : 0, orange : 0, safran : 1*) et (*citron : 0, curcuma : 1, orange : 1, safran : 0*) (toutes les autres dimensions étant égales par ailleurs) seront considérés tout aussi éloignés que deux vins dont quatre des coordonnées seraient (*citron : 1, curcuma 0, orange : 1, safran : 0*) et (*citron : 0, curcuma 1, orange : 0, safran : 1*).

Les deux premiers semblent pourtant relativement proches : pour chacun d'eux, un descripteur évoque un épice et un autre un agrume ; les évaluateurs non contraints dans leur choix des descripteurs peuvent avoir des évocations proches, mais non exactement semblables. L'imprécision intrinsèque au langage naturel peut expliquer cette variabilité des descripteurs non contrôlés. En revanche, les deux derniers vins semblent ne partager aucun arôme, l'un évoquant des agrumes, l'autre des épices. Une distance euclidienne ne saurait rendre compte de la ressemblance/différence entre deux vins (puisque les vins 1 et 2 sont tout aussi dissemblables que les vins 3 et 4). En effet, les termes qui représentent les dimensions de l'espace peuvent évoquer de mêmes arômes et rendre caduque l'utilisation d'une distance classique pour mesurer la similarité/dissimilarité entre les vins (les dimensions non contrôlées sont « liées »). Enfin, l'utilisation des centres de gravité (de moyennes) constitue une perte d'information qui ne permet pas d'apprécier la diversité des évaluations.

Pour rendre compte de la variabilité des évaluateurs, (Martin & Rogeaux, 1994) proposent une méthode pour développer un vocabulaire objectif à partir d'une description libre de six boissons aromatisées. Cette caractérisation libre est effectuée par un panel de 111 consommateurs. La fusion des descripteurs

et l'élimination des termes non pertinents sont faites à l'aide d'outils statistiques. Une analyse qualitative est utilisée pour permettre d'examiner la variabilité des réponses des consommateurs afin de regrouper les termes en différentes catégories (classes). Cette analyse est complétée par une caractérisation des produits par les termes les plus significatifs (synthèse sensorielle).

Devant l'écueil que constitue la connaissance d'un vocabulaire précis et technique pour des évaluateurs non aguerris à l'analyse sensorielle, (Heymann, 1994) essaie lui de montrer que la sensibilité d'évaluateurs néophytes permettrait de distinguer au moins qualitativement des nuances dans les produits testés et que seul le manque de vocabulaire nuit à la qualification des produits. Pour cela, il propose deux expériences : l'une basée sur un tri libre et l'autre sur une qualification non orientée à l'instar des approches précédentes. L'évaluation porte sur 13 extraits de vanille. Deux groupes d'évaluateurs ont évalué les produits. Le premier groupe se compose de 10 personnes ayant reçu une formation en évaluation sensorielle. Le deuxième groupe comporte 10 évaluateurs néophytes. Pour le tri libre, l'ensemble des produits est présenté simultanément aux juges et il leur est simplement demandé de regrouper les extraits de vanille en fonction de leurs similitudes. Pour la méthode de qualification non orientée, il s'agit comme précédemment de qualifier chacun des extraits en langage naturel sans contrainte, puis une étape de filtrage statistique du vocabulaire permet une représentation graphique dans un espace de représentation réduit des 13 extraits. Sur la base de cette représentation vectorielle, un clustering simple permet de regrouper les produits similaires. Les résultats ont montré que les cartes sensorielles obtenues avec les deux approches sont sensiblement similaires sans corrélation marquée avec l'expérience des évaluateurs. En revanche, la dimension de l'espace de caractérisation de l'approche non orientée est trop conséquente pour espérer exploiter une qualification linguistique pour le monde professionnel. Ainsi, si la sensibilité des évaluateurs semble être « innée » (du moins pour des produits simples), la maîtrise d'un vocabulaire technique est en revanche le fruit d'un apprentissage.

Pour faciliter la caractérisation verbale de la qualité d'un produit, l'étude de (Byrne *et al.*, 1999) se donne pour objet d'identifier un vocabulaire précis pour identifier la qualité de pâtés de porc. L'idée est de proposer un résumé de descripteurs (synthèse) à partir d'une liste de descripteurs fournie initialement aux évaluateurs. La fusion des descripteurs et l'élimination des termes redondants reposent sur des traitements statistiques. Dans cette étude, une liste initiale de 45 descripteurs a été développée par des experts à partir de la littérature pour caractériser la viande porcine. Cette liste est modifiée au fur et à mesure des 7 sessions à 16 descripteurs qui sont organisées. Les critères de sélection sont les suivants : 1) les descripteurs doivent être pertinents pour les produits, c'est-à-dire utiles aux évaluateurs, 2) les descripteurs doivent permettre de distinguer les échantillons de produits, c'est-à-dire qu'ils doivent être différenciants, 3) ils ne doivent pas être redondants et enfin 4) avoir une sémantique claire pour l'ensemble des évaluateurs. Les évaluateurs sont également supposés donner une intensité à chacun des descripteurs qu'ils utilisent. Lors de la première session, les évaluateurs sont invités à se familiariser avec l'utilisation des descripteurs et l'échelle d'intensité. Ensuite, les descripteurs retenus lors d'une évaluation sont soumis à une analyse en composante principale (ACP) : les descripteurs redondants (coefficient de corrélation proche de 1) sont rejetés et fusionnés, les descripteurs les moins explicatifs selon l'ACP sont les descripteurs non pertinents et sont éliminés. La liste des descripteurs filtrés est utilisée pour la session suivante. Le processus de sélection est itéré jusqu'à ce que la liste se stabilise. À la fin de chaque session, le groupe est invité à améliorer sa description en proposant des descripteurs qui auraient pu faire défaut lors de l'évaluation. Ces descripteurs peuvent être intégrés au vocabulaire initial. Les capacités discriminatives des évaluateurs peuvent en effet augmenter au cours des sessions et le besoin d'un vocabulaire plus fourni peut être ressenti. Cette démarche itérative basée sur une ACP qui évalue les quatre critères de sélection des descripteurs permet de converger vers un vocabulaire restreint,

partagé et expressif. L'étude de (McDonnell *et al.*, 2001) propose une démarche similaire pour qualifier l'odeur de boissons à l'aide d'une liste initiale de 30 descripteurs et 12 évaluateurs.

(Narain *et al.*, 2004) se sont penchés sur l'analyse des caractères sensoriels de 12 cafés sucrés et non sucrés. L'analyse sensorielle des deux types de café est réalisée avec une approche non orientée et une approche classique. La confrontation des deux approches a pour objectif d'identifier un vocabulaire référent qui fasse sens auprès des consommateurs comme des professionnels. Ce vocabulaire consolidé fournira les descripteurs nécessaires à l'analyse sensorielle par une approche classique plus expressive d'autres gammes de café. Cette étude insiste sur la nécessité d'être à même d'appréhender les relations entre des descripteurs techniques et des évocations en langage naturel. Sous la supervision de trois dégustateurs experts, 13 évaluateurs naïfs ont tenté de caractériser des cafés sucrés et non sucrés, ils ont ainsi proposé des descripteurs dans différentes catégories: *arôme*, *goût*, *arrière-goût* et *sensation en bouche*. Les discussions entre les individus et les expérimentateurs ont clarifié l'utilisation des descripteurs afin d'éviter redondances et ambiguïtés. 26 descripteurs sont ressortis de cette phase de délibération (6 pour l'arôme, 13 pour le goût, 4 pour l'arrière-goût et 3 pour la sensation en bouche). Les 3 experts ont ainsi contrôlé la phase de délibération afin de ne conserver que des descripteurs pertinents, unanimement approuvés et évoquant sans ambiguïté les descripteurs normalisés de l'organisation internationale du café. Les vocabulaires experts et néophytes s'enrichissent mutuellement, s'expliquent l'un l'autre. Relier la capacité à discerner et apprécier différents arômes chez les consommateurs lambda avec les descripteurs contrôlés du domaine est la clé de la réussite du marketing sensoriel. La réussite de cette expérimentation réside dans la capacité des experts à discerner les évocations contenues dans les restitutions des évaluateurs en langage naturel.

(Delarue & Sieffermann, 2004) comparent un profil flash réalisé par des évaluateurs non formés avec une approche orientée réalisée par un panel entraîné. Le profil flash combine une approche descriptive non orientée des produits et un mode d'évaluation comparative des produits à caractériser. Le but est d'obtenir un positionnement sensoriel relatif des produits. Comme dans le cas du profil libre, les évaluateurs utilisent leurs propres termes, ce qui permet de s'affranchir de la phase de mise en commun du vocabulaire et de recherche de consensus. Par contre, l'évaluation des produits est comparative ce qui oblige les évaluateurs à se concentrer sur les différences perceptibles et ainsi à n'utiliser que des descripteurs discriminants.

(Lachnit *et al.*, 2003) ont utilisé l'approche à commentaires libres pour se livrer à une étude de sensibilité chez des évaluateurs néophytes. L'étude portait sur la caractérisation de l'arôme et de l'odeur de limonades à base d'orange. L'expérience est réalisée avec huit échantillons et répétée 4 fois. L'analyse statistique des descriptions libres a montré que les évaluations étaient peu reproductibles : un même évaluateur peut proposer une caractérisation différente pour un même produit d'un test à l'autre. Ce biais s'explique en partie par l'inexpérience des évaluateurs à discerner les nuances olfactives et gustatives, mais également par la difficulté à caractériser verbalement une sensation même commune. Un vocabulaire non contraint laisse place à une grande variabilité des descripteurs dont l'usage n'est pas forcément maîtrisé dans le domaine de l'analyse sensorielle par les évaluateurs, ce qui fait la part belle à l'ambiguïté des évocations. L'interprétation des descripteurs libres par un expert sera au moins en partie affectée par la subjectivité de l'expert. Cette étude montre qu'une analyse sensorielle menée sur des évaluateurs inexpérimentés ne permet pas pour les raisons évoquées précédemment de discerner les différences subtiles qu'il peut exister entre les produits, mais de caractériser les différences majeures et ainsi d'identifier les descripteurs qui expliquent principalement ces différences. Notons pour ce qui nous concerne plus précisément que la variabilité des descripteurs libres ainsi que la non reproductibilité

des expériences mériteraient d'autres traitements que de simples statistiques : par exemple, si un évaluateur propose « madeleine » et un autre « fleur d'oranger », il y a fort à parier qu'ils se font une même idée du produit, mais que leurs expériences et souvenirs personnels donnent lieu à des descriptions différentes *a priori* ; il est nécessaire d'avoir recours à des références culturelles communes pour apprécier la similarité des descriptions proposées, la madeleine étant un biscuit que l'on parfume bien souvent à la fleur d'oranger. Cette remarque vaut pour un même évaluateur évaluant une énième fois un même arôme ou une même odeur : il faut être à même de faire le lien entre ses caractérisations, fonctions de l'humeur du moment, mais évoquant peut-être toutes au final une même impression.

(Perrin *et al.*, 2007) s'intéressent à évaluer l'intérêt d'une approche non orientée réalisée par des professionnels par rapport à l'approche classique réalisée par des professionnels entraînés sur 10 vins blancs du Val de Loire. L'approche orientée est réalisée par 17 évaluateurs entraînés : la sélection du vocabulaire contrôlé est le résultat consensuel de la délibération entre les 17 évaluateurs entraînés sur la base des 10 vins. Les descripteurs retenus doivent être plébiscités par la majorité des évaluateurs. Une fois la liste établie, la phase d'évaluation à proprement parler est engagée : chaque juge attribue une intensité pour chaque descripteur de la liste établie pour chaque vin. L'approche non orientée est, elle, réalisée par 12 professionnels de la région Val de Loire, mais sans entraînement sur les échantillons étudiés. Chaque évaluateur est invité à goûter chacun des vins et propose ses propres descripteurs. Les évaluateurs attribuent ensuite une intensité à chacun de ses descripteurs pour chaque produit. Les données de l'approche orientée sont traitées par une analyse en composantes principales (ACP) et les données de l'approche non orientée par une analyse factorielle multiple (AFM). Ces outils statistiques conduisent à des cartes sensorielles sur lesquelles vins et descripteurs sont projetés. Les cartes sensorielles, pour chacune des deux approches, permettent de voir les crus les plus proches et les plus distincts. L'étude conclut que ces regroupements sont sensiblement les mêmes dans les deux approches. Ainsi, la « mise en correspondance bijective » des clusters entre les deux approches permet d'induire une mise en correspondance entre les descripteurs caractéristiques des clusters. Ces outils permettent de synthétiser les principales similitudes et différences entre les vins, l'approche non orientée offrant davantage de nuances quand elle est réalisée par des professionnels comme c'est le cas dans cette étude. Cette mise en correspondance entre descripteurs, riche d'enseignement, est possible ici parce que les vocabulaires des deux approches ont été proposés par des professionnels du secteur viticole. Cette mise en correspondance nécessite non seulement des outils statistiques, mais relève également de techniques de traitement automatique de la langue (TAL) et doit s'appuyer sur des représentations de connaissances lorsque les descripteurs sont fournis par des évaluateurs naïfs comme nous le montrerons dans nos propositions.

Dans le même esprit, (Lawrence *et al.*, 2013) comparent une approche classique de l'analyse sensorielle et une approche à commentaires libres pour la description olfactive de 31 vins. La qualité olfactive des vins est décrite en utilisant une approche classique à l'aide d'experts formés sur un ensemble de descripteurs caractérisant les vins et une approche à commentaires libres où un ensemble de professionnels caractérisent les vins sans apprentissage préalable d'un vocabulaire partagé. Comme précédemment, des techniques d'analyse statistique sont utilisées pour comparer les profils obtenus. Les professionnels du vin qui ont effectué les évaluations sont divisés en 2 groupes pour participer à l'une des deux méthodes d'évaluation : 17 pour l'approche à commentaires libres et 15 pour l'approche classique. Dans cette étude, tous les évaluateurs sont des experts des Cabernet Franc et n'ont par conséquent pas reçu de formation préalable. Conformément à la définition de la norme ISO Norme 8586-2 (1994), un évaluateur est considéré expert s'il est capable d'identifier les principaux descripteurs d'un domaine. Pour l'approche orientée, une liste de descripteurs communs est élaborée sur la base d'une classification répertoriée des descripteurs habituellement utilisés pour les vins rouges de la vallée de la

Loire. Les descripteurs sont organisés en classes - *e.g.*, Grapefruit, Lemon, Orange sont des sous-classes de Citrus fruits et ce dernier est une sous-classe de Fruits (Fruit > *Citrus fruits* > *Grapefruit, Lemon, Orange, etc.*). 21 descripteurs ont été retenus pour l'étude. Ensuite, les 31 vins sont évalués par les 17 évaluateurs. Chaque évaluateur évalue l'intensité d'évocation sensorielle des 21 descripteurs sur une échelle de 0 à 6 pour chacun des vins. Un test d'analyse de la variance est effectué et 6 des 21 descripteurs jugés non discriminants sont éliminés. Une méthode d'analyse en composantes principales (ACP) est ensuite utilisée pour visualiser les relations entre les 31 vins et les descripteurs (cartes sensorielles). Cette analyse permet d'annoter chaque vin par un ensemble synthétique de descripteurs caractéristiques.

Pour la méthode à commentaires libres, les 15 évaluateurs décrivent les odeurs perçues des 31 vins en utilisant autant de mots ou d'expressions qu'ils jugent nécessaires. Les évaluateurs n'ont reçu aucune instruction quant à la façon de décrire les vins. Par conséquent, le nombre de descripteurs générés est important et 3 experts sont intervenus pour regrouper ou fusionner des descripteurs libres similaires, les transcrire en modalités précises et structurées (Symoneaux *et al.*, 2012). Autrement dit, le travail des experts consiste à associer à chaque descripteur libre un descripteur conceptuel (une classe) compris par l'ensemble des évaluateurs. Chacun des 3 experts réalise sa projection sur son propre ensemble de descripteurs conceptuels. Une phase de délibération entre les experts est ensuite organisée pour constituer un vocabulaire unique commun communicable à l'ensemble des évaluateurs. Les 330 descripteurs ont été regroupés en 34 classes ; 23 classes sont utilisées : « au moins 3 évaluateurs pour au moins 1 vin » constitue le critère de sélection des classes pour les analyses statistiques.

A titre d'exemple, la phrase suivante est citée par un évaluateur pour caractériser l'odeur d'un vin donné: "intensité moyenne, végétal, tabac froid, fruits rouges légers, grenadine". La première étape consiste à supprimer les termes non pertinent par exemple les termes liés aux intensités. La seconde étape consiste à transcrire les termes fournis en langage naturel en modalités précises: {végétal, tabac froid, fruits rouges et grenadine}. Ensuite, les trois spécialistes ont associé à chaque modalité à une catégorie finale : {*Végétal, Végétal sec, Fruits rouge*}.

Les termes obtenus lors des commentaires libres sont analysés quantitativement. Un tableau de contingence (vins  $\times$  classes) fournit les fréquences des classes pour chaque vin. Une analyse en composantes principales (ACP) est ensuite appliquée au tableau de contingence pour visualiser la relation entre vins et classes. Enfin, une analyse factorielle multiple (AMF) est réalisée pour comparer les caractérisations des vins par l'approche orientée et par l'approche à commentaire à libre. Dans cette étude, les résultats des deux approches sont sensiblement similaires. Le traitement de l'information de l'approche non orientée est plus lourd par rapport à la méthode classique puisqu'il nécessite l'intervention d'experts de domaine pour sélectionner, filtrer ou regrouper les descripteurs libres. Cette tâche de filtrage-projection est sujette à interprétation, même si une phase de délibération est organisée entre les experts. Bien que l'approche à commentaire libre laisse une grande liberté de choix des descripteurs par les évaluateurs et ne nécessite aucune formation préalable au vocabulaire du domaine, les étapes de définition des catégories, de regroupement des termes similaires et de projection des descripteurs vers les classes nécessitent des experts de domaine. Cette technique a été mise au point par (Williams & Langron, 1984) qui l'ont appliquée à l'analyse sensorielle du Porto ; elle a ensuite été utilisée par (Gerland & Dumont, 2000) en œnologie pour sélectionner la souche de levure la plus adaptée et (Perrin *et al.*, 2007) pour la description de vins blancs du Val de Loire. L'un des objectifs de notre proposition sera d'objectiver et d'automatiser le processus cognitif de l'expert qui consiste à gérer la proximité des descripteurs et à les projeter dans un espace conceptuel partagé qui permette une description synthétique des produits sans ambiguïté.

D'autres travaux sur la caractérisation d'arômes de vins se sont penchés sur l'utilisation conjointe de méthodes analytiques chimiques et de méthodes descriptives sensorielles. (Capone *et al.*, 2013) s'intéressent ainsi à la définition d'un profil aromatique quantitatif des vins de cépage Negroamaro de la région des Pouilles en Italie du Sud, qui se base à la fois sur une analyse sensorielle et une analyse chimique des composants volatils. Les auteurs proposent une restitution simple du profil aromatique potentiel d'un vin du Negroamaro en traçant une roue des arômes généralisée. L'approche chimique permet d'extraire, de séparer, d'identifier et de quantifier l'ensemble des composés volatils du vin susceptibles de déterminer la perception de son arôme ; elle fournit un profil chimique du vin qui peut être restitué par une vue radar « Volatiles' Aroma Wheel » (Figure 2.8 en haut à gauche). Pour l'approche chimique, un paramètre couramment utilisé pour évaluer la contribution de chaque composé à l'arôme du vin est l'indice d'arôme (*I*), connu également sous le nom « Odour Activity Values » (*OAV*). Cet indice est défini comme le rapport entre la concentration du composé et le seuil olfactif associé à ce composé. Le concept d'*OAV* est une mesure utile pour évaluer l'importance relative des composants chimiques présents dans un échantillon. Parmi toutes les molécules odorantes présentes dans un vin, seul un nombre limité de matières volatiles peuvent être trouvées dans les vins à des concentrations suffisamment élevées pour être perçues.

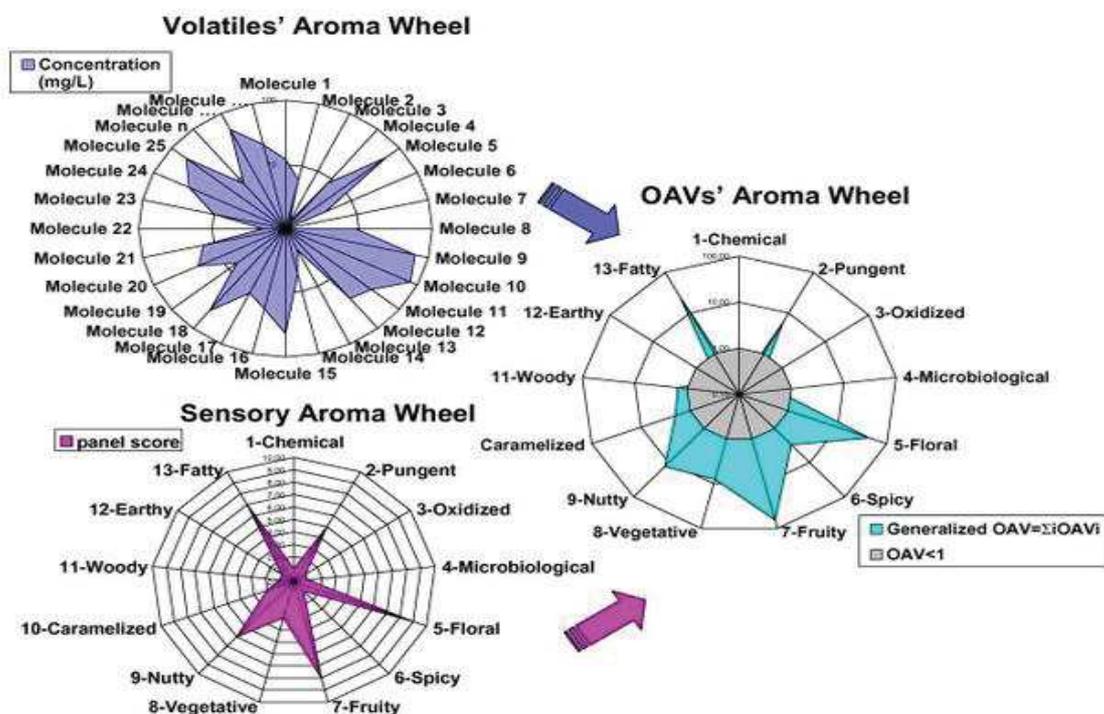


Figure 2.8 : Construction d'une roue des arômes généralisée (Capone et al., 2013)

L'approche descriptive sensorielle est basée sur l'utilisation d'un panel d'experts en vin et débouche sur un profil sensoriel du vin. Ce profil peut également être restitué par une vue radar « Sensory Aroma Wheel » (Figure 2.8 en bas à gauche). Le panel utilise des tests sensoriels descriptifs spécifiques. La roue sensorielle des arômes peut avoir un caractère quantitatif si les descripteurs sensoriels sont associés à des intensités. A chaque molécule odorante, on peut associer un ou plusieurs descripteurs sensoriels. Inversement, deux molécules peuvent être caractérisées par un même descripteur. Un descripteur peut donc être vu comme une classe de composés aromatiques. Une valeur d'activité de l'odeur généralisée (*OAV* généralisée) pour chaque descripteur peut alors être calculée en ajoutant l'*OAV* (valeur de l'activité d'odeur) de chacune des molécules de la série aromatique qui sont associées à ce descripteur.

Les molécules avec un  $OAV \geq 1$  sont généralement considérées comme les composés les plus susceptibles de contribuer à l'arôme du vin.

De cette façon, le profil d'un vin peut être établi en traçant une vue radar  $OAV$  généralisée de tous les descripteurs (Figure 2.8 à droite). Le graphe obtenu permet de relier à la fois des informations quantitatives obtenues par l'analyse chimique avec les perceptions sensorielles et propose ainsi une restitution plus fiable pour comparer les profils d'arôme de vins par fusion des informations chimiques et sensorielles. L'intérêt de la méthode est que le profil généralisé est obtenu d'une manière plus objective en rapportant des informations quantitatives obtenues par une analyse chimique. L'approche chimique a pour rôle de déterminer l'ensemble de molécules susceptibles de contribuer à l'arôme d'un vin, pondérées par les indices  $OAV$  qui sont récupérés pour calculer une intensité sur les classes olfactives. La nouveauté de cette étude est de relier les données quantitatives de la chimie analytique et les descripteurs de l'analyse sensorielle. Cette démarche apporte plus d'information sur l'intensité ou l'importance des descripteurs obtenus avec l'approche descriptive classique. Cette approche permet aussi de nous donner une idée sur les descripteurs dominants et les descripteurs ayant un impact sensoriel insignifiant restent à l'intérieur du  $OAV < 1$  (le descripteur *Fruit* est dominant) (Figure 2.9).

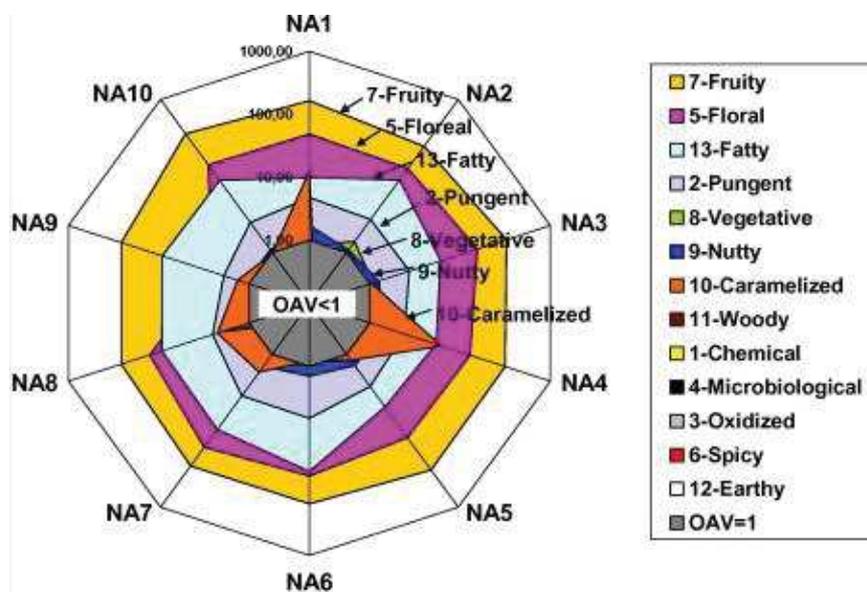


Figure 2.9 : Vue radar des OAV des 13 descripteurs sensoriels pour les cépages Negroamaro des Pouilles (d'après (Capone et al., 2013))

Cette démarche nécessite l'utilisation d'une roue des arômes descriptive obtenue par des experts du domaine. La représentation est considérée comme le résultat de la fusion d'une approche chimique et d'une analyse sensorielle, fusion qui repose sur l'association experte de classes de descripteurs et de molécules. On peut là encore imaginer que le traitement de l'information et l'exploitation de bases de connaissances puissent au moins partiellement remplacer le travail cognitif de l'expert.

## 2.6 Objectif et plan de la thèse

### 2.6.1 Processus cognitifs et perception de l'odeur

Une fois que les molécules chimiques (les odorants) qui constituent l'odeur sont captées par des récepteurs (neurones), c'est un processus cognitif qui va permettre de caractériser et d'évaluer l'odeur. Si l'on décompose le processus cognitif d'évaluation, on peut en particulier distinguer l'évaluation de la qualité de celle de l'acceptabilité. L'évaluation de la qualité consiste à fournir un ensemble de

descripteurs linguistiques qui caractérisent au mieux l'odeur perçue. Ces descripteurs peuvent avoir pour origine une relation olfactive entre l'odeur et une situation vécue, heureuse ou malheureuse. Ce schéma cognitif d'association est à l'origine de l'acceptabilité de l'odeur. Il faut noter que la qualité et l'acceptabilité dépendent de l'intensité de l'odeur, elle-même liée à la fréquence des stimuli reçus par la membrane olfactive. Une odeur agréable à faible concentration peut s'avérer désagréable à forte concentration.

Le schéma de décomposition de l'évaluation est illustré par la figure ci-dessous. L'indice des fonctions  $f$  et  $g$  qui représentent respectivement la qualité et l'acceptabilité, fait référence à un niveau d'intensité donné de l'odeur. Dans cette thèse, nous nous focaliserons sur les modèles mathématiques nécessaires à une aide informatisée pour l'identification des fonctions  $f_i$  associées à l'évaluation de la qualité (Figure 2.10).

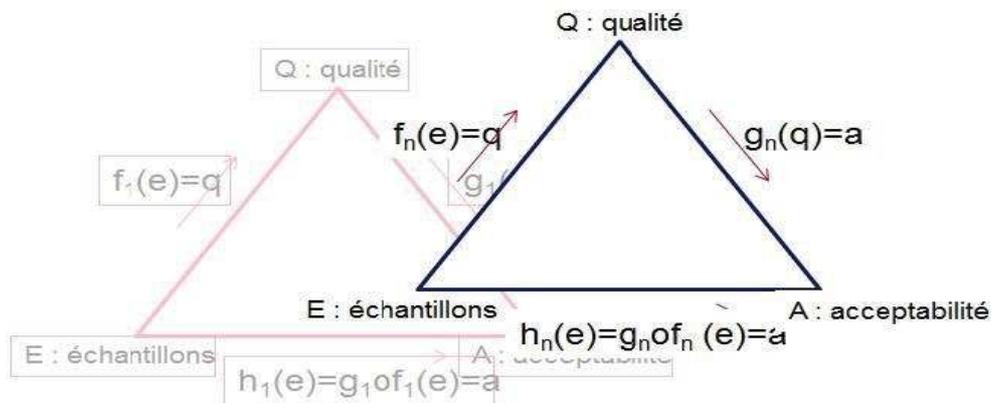


Figure 2.10 : Décomposition de l'évaluation de l'odeur

L'acceptabilité ( $h$ ) est vue comme la combinaison de l'évaluation de la qualité ( $f$ ) et de l'évaluation hédonique ( $g$ ).

### 2.6.2 Contexte : de la chimie de l'odeur à l'automatisation de l'évaluation de sa qualité

Comme expliqué dans ce chapitre, l'évaluation de la qualité d'une odeur se fait aujourd'hui en utilisant des méthodes orientées basées sur l'apprentissage d'un vocabulaire souvent spécifique du domaine concerné et présentent des limites à la fois en termes technique et de coûts de mise en œuvre. La société OLENTICA a développé avec l'équipe « Odeurs et COV » du laboratoire LGEI (Laboratoire de Génie de l'Environnement Industriel) de IMT Mines Alès une approche non orientée, n'exigeant pas l'apprentissage d'un vocabulaire spécifique par les évaluateurs. Cette méthode est décrite ci-après et illustrée par la Figure 2.11.

L'évaluation de la qualité de l'odeur est réalisée par un nombre d'évaluateurs non experts variable (6 à 20). Suite à la présentation d'une odeur donnée à chacun d'entre eux, il est demandé aux évaluateurs :

- 1) D'identifier ou de caractériser l'odeur par un ensemble le plus large possible de descripteurs linguistiques (les « descripteurs bruts ») ;
- 2) D'associer à chaque descripteur un poids entre 1 et 5. Le poids 5 est attribué au(x) descripteur(s) dont la caractéristique ressort le plus clairement dans l'odeur pour l'évaluateur ;

- 3) Cette évaluation est conduite par un analyste sensoriel qui récupère tous les descripteurs fournis par les évaluateurs et opère ensuite des regroupements cohérents de ces descripteurs en fonction des tonalités olfactives auxquelles ils sont associés. Le poids respectif des nouveaux descripteurs associés à ces familles est calculé en sommant celui des descripteurs bruts regroupés. Au final, l'expert obtient un profil olfactif qui se compose d'un ensemble de descripteurs linguistiques pondérés.

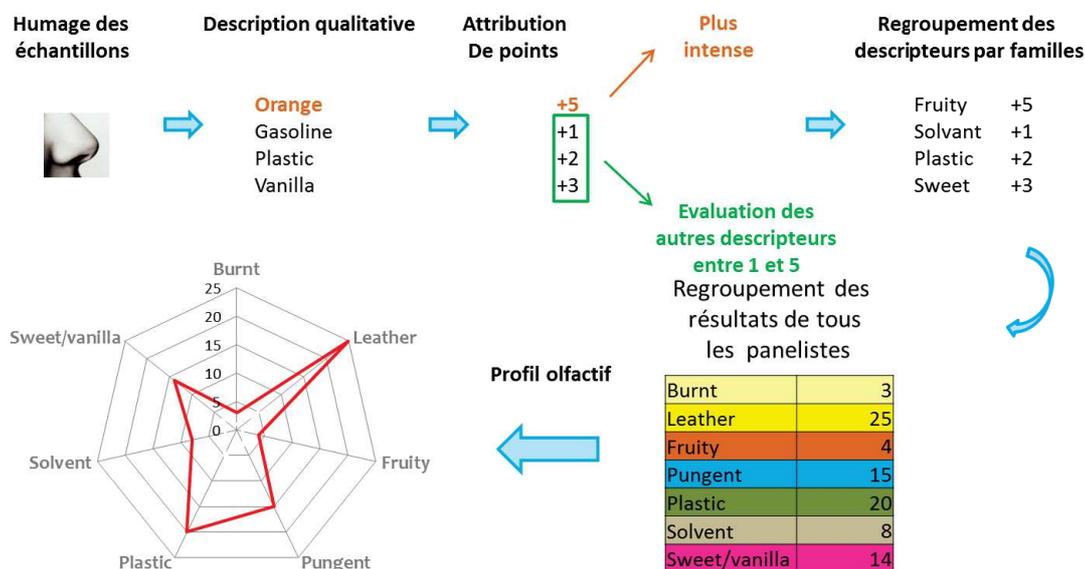


Figure 2.11 : Evaluation de la qualité d'une odeur (Méthode Olentica/LGEI)

L'avantage de cette méthode non orientée est que le vocabulaire utilisé par les juges n'est pas limité et qu'elle évite la coûteuse phase d'apprentissage propre aux méthodes orientées. Son inconvénient principal est qu'elle nécessite un traitement de descripteurs bruts par l'expert conduisant l'analyse, ce qui induit nécessairement une part de subjectivité dans la construction des profils olfactifs.

Notre proposition consiste à automatiser les processus cognitifs mis en œuvre par l'analyste sensoriel afin de limiter la part de subjectivité dans son interprétation des descriptions en langage naturel, formaliser et systématiser les traitements de l'information associés. Ce projet explique la collaboration des laboratoires LGI2P (Laboratoire d'Informatique et d'Ingénierie de la Production) et LGEI de IMT Mines Alès.

### 2.6.3 Objectif

Caractériser la qualité des odeurs consiste à identifier un ensemble de descripteurs qui synthétise au mieux la sensation olfactive. Généralement, cette caractérisation est donnée par un ensemble de descripteurs conceptuels fournis par des professionnels de l'analyse sensorielle. Comme nous l'avons expliqué précédemment, les campagnes d'analyse sensorielles représentent des coûts substantiels pour l'industriel. En revanche, si cette caractérisation de la qualité de l'odeur est confiée à des évaluateurs non experts, par exemple des consommateurs, chaque description en langage naturel sera la plupart du temps chargée d'ambiguïtés. En effet, dès lors que les évaluateurs ne sont plus contraints d'utiliser un vocabulaire contrôlé, surviennent les problèmes d'ambiguïté, d'imprécision langagière et de subjectivité des évocations. Le langage naturel s'oppose aux vocabulaires contrôlés formés intentionnellement par l'entremise de l'homme pour remplir un besoin précis, ici la caractérisation homologuée de vins, miels et autres arômes, etc. Le langage d'experts s'oppose donc à l'expression libre d'évaluateurs néophytes.

L'intérêt d'un vocabulaire contrôlé est de construire un espace qui permette une communication « objective » de la qualité de l'odeur. Cette distinction entre ces deux types de description est fondamentale pour la suite de notre exposé. Une communication objective de la qualité de l'odeur est une communication axée sur un profil sensoriel contraint et compris par le monde professionnel de l'odeur. La caractérisation de la qualité des odeurs par des évaluateurs non-experts, parce que moins contrainte, soulève des problèmes nouveaux liés en particulier à l'ambiguïté inhérente au langage naturel. S'il devient possible de multiplier les évaluations tout en réduisant les coûts de l'analyse sensorielle, faire une synthèse de ces caractérisations libres devient un processus complexe que nous développerons dans les chapitres suivants.

Définir une approche non orientée pour l'évaluation de la qualité de l'odeur nous conduit à proposer un modèle qui permette de transcrire les descriptions libres en langage naturel en descripteurs structurés ou descripteurs conceptuels propres à un domaine métier. Nous parlerons de projection de l'espace terminologique sur l'espace conceptuel. Cette thèse propose une approche basée sur le traitement automatique du langage (TAL) par des modèles distributionnels et des techniques de représentation des connaissances pour formaliser et automatiser la transcription des termes en concepts pour une communication objective de la caractérisation des produits.

#### 2.6.4 Vers une automatisation cognitive de l'évaluation de la qualité de l'odeur

C'est cette interprétation du terme au concept que nous proposons d'automatiser et d'exposer dans le Chapitre 3. Nous nous attacherons en effet à définir une correspondance entre des descripteurs du langage commun, imprécis et ambigus, et des descripteurs techniques, catégoriques et analytiques. (Bergson, 1889) écrivait : « ...*Nous nous exprimons nécessairement par des mots, et nous pensons le plus souvent dans l'espace. En d'autres termes, le langage exige que nous établissions entre nos idées les mêmes distinctions nettes et précises, la même discontinuité qu'entre les objets matériels...* ». Cette faculté à exprimer nuances et gradations avec des mots exige tout de même un vocabulaire de 25 000 à 40 000 mots pour un adulte français cultivé. Cependant si ces milliers de mots sont suffisants pour servir les objectifs premiers du langage, communiquer ou exprimer une idée, identifier la qualité d'un produit ne souffre pas l'ambiguïté et l'imprécision intrinsèque du langage commun. Les approches orientées ont rapidement séduit en analyse sensorielle parce qu'un vocabulaire contrôlé permet justement de « coder » sans équivoque la qualité d'un produit. Si l'on envisage une analyse sensorielle ouverte à des panels d'évaluateurs sans contrainte de nombre, d'acquisition d'un vocabulaire métier, alors il faut être capable de déchiffrer l'évocation des concepts par les termes que proposent les évaluateurs.

Notre proposition relève de l'analyse distributionnelle de corpus de documents et s'appuie sur des représentations vectorielles des mots. La similarité sémantique entre deux mots est alors définie comme une proximité dans un espace de dimension finie où chaque dimension correspond à des contextes partagés possibles. Aux concepts liés à une représentation conceptuelle sont associés des labels qui sont les termes sur lesquels est établie la proximité entre termes et concepts, cette proximité mesure le degré d'évocation d'un concept par un terme. Nous proposerons plusieurs modèles pour la mise en relation des termes et des concepts. Le premier est naïf : il associe de façon unique un concept à un terme. Plusieurs modèles (embeddings) seront évalués pour supporter cette projection. Évaluer automatiquement le degré avec lequel un terme évoque un concept est une tâche difficile. L'évocation est un processus hautement subjectif puisqu'alimenté par nos expériences et notre culture. Pour faire face à ce problème complexe de l'évocation, nous proposons un second modèle. Ce dernier s'intéresse au cas où différents termes proposés par un même évaluateur méritent d'être considérés conjointement pour établir un lien pertinent vers l'espace professionnel (conceptuel). L'objectif visé par ce modèle est

la caractérisation des évocations conceptuelles implicites identifiables par l'utilisation de proximité sémantique. Pour cela, on va chercher à définir des proximités sémantiques entre les termes, ce qui nous permet de regrouper les termes qui partagent la même évocation de l'odeur dans la même classe. Ces deux classes de modèles seront détaillées dans le chapitre 3.

La seconde étape de notre approche est de synthétiser des annotations conceptuelles individuelles en un ensemble limité de concepts. Cette étape est nécessaire en particulier quand plusieurs évaluateurs fournissent leurs annotations pour rendre l'évaluation collective plus synthétique, plus représentative et plus précise. Cette phase de résumé sémantique sera détaillée dans le chapitre 4, elle cherchera à identifier une réduction conceptuelle de l'ensemble des descriptions conceptualisées en se basant sur une analyse taxonomique. Nous proposerons une définition axiomatique du résumé sémantique et discuterons différents aspects techniques relatifs aux algorithmes qui en permettent la génération.

La Figure 2.12 illustre les différentes étapes qui constituent notre approche pour une évaluation automatisée de la qualité des odeurs et structurent notre manuscrit.

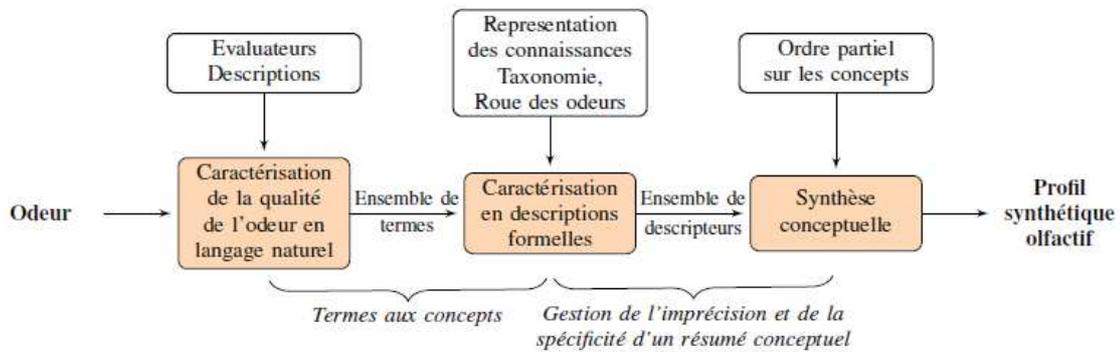


Figure 2.12 : Processus d'évaluation de la qualité des odeurs

## Chapitre 3 Conceptualisation des Descriptions Terminologiques

---

La démarche proposée dans cette thèse qui doit permettre une automatisation de l'évaluation non orientée de la qualité des odeurs a été présentée dans le chapitre précédent. Elle repose sur deux étapes distinctes : (i) la mise en correspondance des descriptions terminologiques non contrôlées avec un espace conceptuel d'intérêt pour le domaine de l'analyse olfactive, et (ii) la synthèse des descriptions conceptualisées obtenues par la phase précédente, en vue d'amener une restitution globale de la qualité de l'odeur – afin notamment d'identifier les tendances majeures jusque-là potentiellement implicites qui se dégagent de l'impression collective recueillie.

Ce chapitre est consacré à la présentation des travaux réalisés pour mener à bien la première phase du traitement. Nous nous concentrerons ici sur les traitements proposés afin de conceptualiser les descriptions terminologiques non contrôlées exprimées par les évaluateurs. Nous présentons en particulier la démarche globale que nous proposons pour mettre en correspondance les termes exprimés en langue naturelle par les évaluateurs avec des descripteurs conceptuels utiles en considération d'une analyse olfactive contextuelle, *i.e.* servant l'analyse d'une gamme identifiée de produits.

En effet, si les approches non orientées pour déterminer la qualité de produits présentent de nombreux avantages en particulier au niveau du coût des sessions, de la dispense d'apprentissage d'un vocabulaire et de la taille des panels, les études précédentes synthétisées dans le Chapitre 2, mettent néanmoins en évidence la difficulté majeure que constitue l'incontournable tâche de projection de termes du langage commun dans l'univers conceptuel des experts métier. C'est cette interprétation du terme au concept, ou plus généralement de l'espace terminologique à l'espace conceptuel, que nous proposons d'automatiser dans ce chapitre. Nous allons nous attacher à définir une correspondance entre des descripteurs du langage commun, intrinsèquement imprécis (et ambigus), et des descripteurs techniques, catégoriques et analytiques.

Notre proposition, à la croisée de différents domaines de l'Intelligence Artificielle, repose sur l'utilisation conjointe de techniques et modèles proposés par les domaines du Traitement Automatique du Langage (TAL), de l'Ingénierie des Connaissances et du Raisonnement approché. Nous exploiterons notamment (i) des modèles distributionnels – aussi appelés modèles de plongement sémantique (*embeddings*) – utilisés en TAL pour caractériser et manipuler la sémantique de termes, et (ii) des représentations des connaissances définissant les concepts et relations taxonomiques utiles pour réaliser la projection visée de l'espace terminologique vers l'espace conceptuel. Plusieurs modèles par ordre de complexité croissante seront proposés, évalués et discutés. Sans prétention d'exhaustivité, nous considérerons en particulier deux classes de modèles permettant de répondre à deux approches de conceptualisation alternatives.

La première classe de modèles s'intéresse à la mise en correspondance d'un terme unique avec un concept de la représentation taxonomique. Autrement dit, si l'on se concentre sur la conceptualisation de l'ensemble des termes proposés par un unique évaluateur, nous chercherons à mettre en correspondance chaque terme proposé avec un seul concept d'une représentation des connaissances. Nous considérons que, dans l'absolu, et en admettant une potentielle perte d'information, chaque terme proposé par un évaluateur peut être associé à une référence conceptuelle unique explicitée dans la

représentation des connaissances. Dans sa forme la plus simple, ce traitement fait référence à une tâche de désambiguïsation triviale : l'évaluateur mentionne le terme *citron* ou *citrus*, une correspondance sera établie avec le concept Citron – la tâche de désambiguïsation est ici qualifiée de triviale car en pratique les cas d'ambiguïté relevant de polysémies restent très largement marginaux et ne méritent pas d'être considérés ; en effet, en pratique, l'étude olfactive porte généralement sur une gamme de produits identifiée pour laquelle sera utilisée une représentation des connaissances spécifique, ce qui réduit *de facto* la taille du vocabulaire d'intérêt pour la caractérisation de la qualité d'une odeur, et vient réduire d'autant le risque induit par d'éventuelles polysémies parmi les termes communément mentionnés pour qualifier la qualité d'une odeur. Le traitement de mise en correspondance de chaque terme avec un concept unique peut aussi faire référence à une tâche d'abstraction ou, plus généralement, de mise en correspondance basée sur une proximité sémantique dans le cas où un terme précisé par l'évaluateur ne dispose pas de correspondance exacte dans la représentation des connaissances. A titre d'exemple, si un évaluateur fait référence à une odeur de *citron* et que la représentation des connaissances considérée n'exprime pas ce concept spécifique, le traitement aura pour objectif d'établir une correspondance avec un concept plus abstrait, *e.g.* Agrumes. De façon plus générale, une mise en correspondance basée sur une proximité cherchera elle à identifier un concept pertinent ne correspondant pas nécessairement à une abstraction, *e.g.* en considérant que la représentation des connaissances utilisée n'exprime pas le concept Limonade tout en définissant le concept Citron, il est pertinent que l'évocation du terme *limonade* soit associée à une évocation du concept Citron.<sup>1</sup> L'étude des modèles permettant de réaliser cette tâche nous donnera alors accès à une analyse plus fine de l'impact du choix de certains paramètres critiques pour la tâche de conceptualisation de descriptions terminologiques ; nous étudierons en particulier de façon détaillée l'impact du choix du modèle de plongement sémantique des termes, et nous explorerons des stratégies couplant différentes approches de plongement.

La seconde classe de modèles de mise en correspondance conceptuelle que nous étudierons s'intéresse au cas où différents descripteurs terminologiques proposés par un même évaluateur méritent d'être considérés pour établir un lien pertinent vers l'espace conceptuel. Nous considérerons donc le cas où un évaluateur n'est pas en mesure d'exprimer un concept identifié, bien que celui-ci, s'il avait été exprimé (au niveau terminologique), aurait pu être associé à un concept d'intérêt pour mener à bien l'analyse olfactive. Ainsi, un évaluateur précisant les termes *cacahuète* et *gras* évoque potentiellement de façon implicite le concept Beurre de cacahuète identifiable par une forme d'agrégation des termes évoqués. Nous le détaillerons par la suite, cette classe de modèles généralise la première classe. Il est cependant important de noter que l'objectif visé par cette classe de modèles ne correspond pas à une phase de synthèse telle que celle visée par la seconde étape de résumé sémantique des descriptions conceptualisées des différents évaluateurs. L'objectif visé par cette classe de modèles est la caractérisation de potentielles évocations conceptuelles implicites identifiables par l'utilisation de proximité sémantique. La phase de résumé sémantique détaillée dans le prochain chapitre cherchera elle à identifier une réduction conceptuelle de l'ensemble des descriptions conceptualisées en se basant sur une analyse purement taxonomique. Cette dernière s'avère utile lorsque le nombre d'évaluateurs et par suite de descriptions terminologiques devient conséquent ; elle diffère radicalement de la réduction proposée par la projection d'un ensemble de termes en un concept.

Le reste du chapitre est organisé comme suit : la section 3.1 introduit l'état de l'art sur les techniques de TAL, d'Ingénierie des Connaissances, et de raisonnement approché sur lesquelles reposent nos travaux ; la section 3.2.1 présente la première classe de modèles étudiée permettant la mise en correspondance entre un terme et un concept ; cette section traite aussi des travaux relatifs à l'analyse des techniques de plongement sémantique ; la section 3.2.2 est consacrée à l'étude menée sur la seconde classe de modèles

---

<sup>1</sup> L'odeur de limonade faisant, de façon consensuelle, souvent référence à une odeur de citron.

s'intéressant à l'identification d'évocations conceptuelles implicites. Pour finir, une discussion des différentes contributions détaillées dans le chapitre est proposée.

### **3.1 Etat de l'art : techniques de représentation et d'analyse des connaissances de domaine**

Nous défendons dans cette thèse l'idée que la définition d'une approche automatisée et non-orientée de l'analyse des odeurs mérite d'être envisagée au travers de la définition d'une chaîne de traitement complexe capable de (i) mettre en correspondance les termes fournis par les évaluateurs avec des descripteurs structurés, (ii) auxquels sera par la suite appliquée une phase de résumé sémantique consistant à synthétiser de façon pertinente l'information portée par l'ensemble des descripteurs associés aux termes évoqués.

Notre proposition pour la réalisation de ce traitement complexe repose sur un ensemble de notions relatives aux domaines du TAL et de la Représentation des Connaissances. Nous proposons alors dans ce chapitre de faire référence aussi largement que possible aux différents travaux relatifs aux techniques et notions sur lesquelles reposent nos travaux. Nous présentons notamment dans un premier temps les techniques de représentation des connaissances utilisées pour modéliser l'information relative aux descripteurs structurés. Différentes techniques qui nous permettront de tirer parti de ces représentations pour comparer les descripteurs au travers de la notion de similarité conceptuelle sont ensuite proposées. Nous exposerons dans un second temps les techniques de TAL utilisées pour caractériser la sémantique des termes évoqués par les évaluateurs, prérequis nécessaire à la conceptualisation des évocations proposées par les évaluateurs avant l'application de la phase de résumé conceptuel. Nous traiterons par la suite des contributions amenées dans cette thèse sur cette partie technique de la chaîne de traitement, à savoir : la proposition et l'étude de différentes approches permettant de mettre en correspondance les termes non-contrôlés exprimés en langage naturel avec les descripteurs d'intérêt du domaine.

Afin de faciliter la lecture, les contributions de l'état de l'art dont l'objet est proche du dessein de nos propositions sont discutées avant la présentation de nos propres contributions.

#### **3.1.1 Modèles et représentations des connaissances considérés**

Nos travaux reposent sur l'utilisation de représentations formelles des connaissances (ontologies). Nous considérons en particulier ces modèles de connaissance afin d'exprimer la connaissance de domaine associée aux descripteurs que nous considérerons lors de la phase de résumé sémantique – nous considérons dans nos travaux que cette connaissance experte, fournissant notamment la structure entre les descripteurs, est connue *a priori* (e.g., une roue des odeurs). Ces modèles seront aussi dans une moindre mesure utilisés dans la phase de caractérisation des descripteurs associés aux termes évoqués par les évaluateurs.

Nous présentons dans un premier temps les modèles taxonomiques retrouvés extensivement dans nos travaux. Le langage de représentation des connaissances OWL ainsi que certains modèles de connaissance existants que nous utiliserons par la suite sont rapidement introduits. Nous présentons dans un second temps l'état de l'art relatif aux mesures sémantiques permettant d'exploiter les ontologies afin de comparer des concepts.

##### **3.1.1.1 Taxonomies**

Les taxonomies permettent de structurer les éléments d'un domaine qui présentent des caractéristiques similaires en classes potentiellement ordonnées. Très intuitives, elles ont été initialement proposées en Biologie pour définir des taxons (classes) en catégorisant des organismes partageant des propriétés

communes. Un modèle de connaissance de type taxonomique permet ainsi une classification systématique et hiérarchisée de taxons dans diverses catégories selon les caractères qu'ils ont en commun, des plus généraux aux plus particuliers, *e.g.*, la classe *Fruit* sera définie comme plus générale que la classe *Fruit rouge*, elle-même définie comme plus générale que la classe *Fraise*.

La sémantique portée par une taxonomie de classes est non-ambigüe, l'interprétation de la relation taxonomique est formellement exprimée par des propriétés particulières / axiomes, et correspond à la relation d'inclusion considérée en logique formelle. La relation taxonomique a la caractéristique d'être i) antisymétrique : cette propriété signifie que l'inclusion d'une classe d'individus  $x$  dans une classe d'individus  $y$  implique que  $y$  n'est pas incluse dans  $x$  ; ii) transitive : si une classe  $x$  subsume une classe  $y$ , qui elle-même subsume  $z$ , alors  $x$  subsume  $z$  ; iii) réflexive : une classe est subsumée par elle-même.

Les notions de concepts et de classes font référence à une notion commune regroupant des objets, parfois appelés instances, de même nature. Nous préférons dans ce manuscrit l'utilisation du terme concept plus communément utilisé en représentation des connaissances.

Formellement, nous considérons qu'une taxonomie définit un ordre partiel sur un ensemble de concepts  $C$ . Elle peut alors être exprimée par  $\leq_c$ , une relation binaire sur les concepts qui est : i) réflexive  $\forall c \in C: c \leq c$ , ii) antisymétrique  $\forall u, v \in C: (u \leq v \wedge v \leq u) \Rightarrow u = v$  et iii) transitive  $\forall u, v, w \in C: (u \leq v \wedge v \leq w) \Rightarrow u \leq w$ . Une taxonomie peut aussi être représentée par un graphe orienté qui ne possède pas de circuit et que l'on considère généralement raciné par un concept abstrait unique (RDAG – *Rooted Directed Acyclic Graph*).

### 3.1.1.2 OWL

*Web Ontology Language* (OWL) est une famille de langages de représentation des connaissances permettant d'exprimer de façon formelle de la connaissance de domaine plus complexe que celle modélisable au travers d'une simple taxonomie. OWL repose pour sa formalisation sur l'utilisation de certaines logiques de description qui correspondent à des fragments décidables de la logique du premier ordre. OWL permet ainsi de décrire des domaines spécifiques au travers de l'expression formelle de définitions axiomatiques propres aux domaines modélisés. Une ontologie exprimée en OWL décrit des concepts (classes), des instances (réalisations de classes/concepts), ainsi que des propriétés reliant les instances. Des étiquettes syntaxiques (labels) seront aussi associées aux concepts, aux instances, et aux propriétés. Proposé par le W3C, OWL s'est aujourd'hui imposé comme un standard informatique et forme une des pierres angulaires du paradigme du Web sémantique. Il est naturellement supporté par de nombreux éditeurs d'ontologie, notamment l'éditeur Protégé qui sera utilisé dans cette thèse (<http://protege.stanford.edu/>).

Du fait de la sémantique formelle exprimée par toute ontologie exprimée en OWL<sup>2</sup>, des raisonnements déductifs peuvent être effectués sur (le modèle du) domaine en considération de la connaissance modélisée. Ces raisonnements permettront de tester la consistance du modèle (sa cohérence), mais aussi d'inférer de nouvelles connaissances capables d'enrichir les modèles valides. Parmi les connaissances qui peuvent être inférées par ces phases de raisonnement, la construction de la taxonomie des concepts (classes) est d'intérêt particulier pour nous. Les descripteurs que nous considérerons seront en effet essentiellement représentés par des concepts de l'ontologie, et nos analyses exploiteront largement leur structuration taxonomique. Il est alors important de garder à l'esprit que toute ontologie définit une

---

<sup>2</sup> qui repose, nous le rappelons, sur l'utilisation d'une logique descriptive sélectionnée en fonction de l'expressivité de la modélisation.

structuration plus ou moins expressive des concepts qu'elle exprime par la définition implicite d'une taxonomie. Toutes les ontologies modélisées dans le cadre de nos travaux, bien qu'exprimant la plupart du temps de simples taxonomies (explicites), seront définies à l'aide d'OWL. Celles-ci seront généralement utilisées afin d'exprimer des connaissances propres au domaine olfactif qu'il sera difficile de retrouver dans des bases de connaissances générales, e.g. la structuration entre le concept de menthe (*mint*) et d'odeur fraîche.

### 3.1.1.3 WordNet

WordNet est une base de données lexicale proposant, entre autres, une structuration de *synsets* auxquels sont associés un ensemble de termes – un *synset* fait ici référence à un ensemble de synonymes (*synonyme set*). De manière simplifiée, nous considérerons dans nos travaux que WordNet exprime une taxonomie de concepts, i.e. l'ordre partiel sur les *synsets*, ainsi qu'un ensemble de labels associées aux concepts (<http://wordnetweb.princeton.edu/perl/webwn>). Cette base de données lexicale nous permettra d'obtenir une modélisation générale et consensuelle de la structuration taxonomique de concepts généraux qui pourraient être évoqués par les évaluateurs. L'utilisation de cette ressource nous permettra notamment, dans une certaine mesure, de réduire le travail fastidieux de modélisation de connaissance ; WordNet exprime par exemple la connaissance précisant que les termes menthe poivrée (*peppermint* ou *Mentha piperita*) et menthe des champs (*corn mint* ou *Mentha arvensis*) font référence à deux concepts distincts spécialisant le concept plus général de menthe (*mint*).

### 3.1.2 Mesurer la similarité entre concepts

Les représentations des connaissances permettent d'exprimer une appréciation consensuelle de la connaissance d'un domaine. Ainsi, les modèles ontologiques, même dans leur forme taxonomique la plus simple, fournissent—en permettant par exemple d'apprécier que l'évocation de la menthe fait référence à l'évocation de fraîcheur—la possibilité d'apprécier des subtilités de domaine importantes pour l'automatisation cognitive visée dans nos travaux. Ces modèles peuvent être perçus comme des vecteurs de connaissance permettant la définition d'approches automatisées capables de tirer parti de connaissances expertes nécessaires à l'étude de domaines complexes. Ces modèles seuls ne sont cependant pas suffisants. Ils méritent, pour être pleinement exploités dans des traitements reposant sur des raisonnements approchés automatisés, d'être associés à l'utilisation de *mesures* permettant de comparer, d'apprécier la similarité, des concepts qu'ils décrivent – et alors permettre d'apprécier que les couples d'évocation (menthe poivrée, menthe des champs) et (menthe poivrée, herbe coupée), bien que tous les deux associés à la notion de fraîcheur, ne sont pas composés de termes reliés sémantiquement avec le même degré d'intensité ; le couple d'évocation (menthe poivrée, menthe des champs) faisant référence en effet à deux concepts plus proches, tout le monde s'en accordera. Les mesures sémantiques qui peuvent être utilisées sur les modèles de connaissance pourront alors être considérées comme notre capacité à tirer parti de notre connaissance pour comparer des choses, apprécier la similarité de concepts – traitement essentiel pour la phase d'analyse que nous souhaitons amener sur les descripteurs conceptuels.

De nombreuses mesures de similarité sémantique basées sur la structure d'une représentation des connaissances ont été définies dans la littérature (Harispe *et al.*, 2015; Harispe, 2014). Basées sur des fondements ensemblistes, sur la théorie des graphes, ou encore sur la théorie de l'information, ces mesures permettent d'apprécier la similarité de deux concepts ou de deux groupes de concepts. La section suivante propose d'introduire quelques-unes des mesures topologiques définies dans la littérature.

### 3.1.2.1 Mesure de similarité sémantique entre deux concepts

Trois grandes familles d'approches peuvent être distinguées parmi la large diversité de mesures proposées dans la littérature pour apprécier la similarité de deux concepts. La première famille d'approches repose sur la théorie des graphes (Rada *et al.*, 1989; Wu & Palmer, 1994) : le calcul de la similarité entre deux concepts se base sur une représentation de la taxonomie sous la forme d'un DAG afin d'exploiter les longueurs des chemins qui relient les concepts pour les comparer. Une seconde famille d'approches, inspirée de la théorie de l'information, exploite des aspects topologiques de la taxonomie et la notion d'informativité d'un concept pour définir la notion de similarité (Jiang & Conrath, 1997; Lin *et al.*, 1998; Resnik, 1995). Une troisième famille d'approches propose une vision ensembliste de l'évaluation de la similarité en s'attachant à analyser certaines caractéristiques topologiques associées aux concepts pour estimer la similitude en fonction de leurs caractéristiques partagées et différenciantes. Pour finir, un point de vue hybride consistant à combiner ces différentes approches a aussi été étudié, *e.g.* (Leacock & Chodorow, 1998). Nous présentons certains aspects techniques de chacune des approches évoquées.

#### 3.1.2.1.1 Mesures fondées sur la structure de graphes

Elles estiment la similitude en fonction du degré d'interconnexion entre les concepts. Elles sont généralement considérées comme des mesures qui sont formulées dans le modèle spatial c'est-à-dire que la similarité de deux concepts est estimée en fonction de leur distance dans le graphe. Sur la base de l'analyse des longueurs des chemins qui relient les concepts, (Rada *et al.*, 1989) ont par exemple exprimé la distance taxonomique entre deux concepts  $u$  et  $v$  définie dans un graphe taxonomique comme une fonction du plus court chemin qui les relie  $sp(u, v)$ .

$$Dist_{Rada}(u, v) = sp(u, v)$$

(Leacock & Chodorow, 1998) ont proposé une adaptation non linéaire de la distance de Rada pour définir la mesure de similarité notée  $Sim_{LC}$ . Ils choisissent une normalisation basée sur la profondeur totale de l'ontologie  $max\_depth$  et normalisent avec le logarithme.

$$Sim_{LC}(u, v) = -\log\left(\frac{sp(u, v)}{2 \cdot max\_depth}\right)$$

(Wu & Palmer, 1994) supposent que les termes qui se trouvent plus profondément dans la taxonomie sont toujours plus proches (*menthe poivrée* est perçu plus proche de *menthe des champs* que *plante* semble l'être de *animal*). Ils proposent de prendre en compte la distance entre l'ancêtre commun de  $u$  et  $v$  qui a une profondeur maximale ( $LCA_{u,v}$ ) et la racine.

$$Sim_{WP}(u, v) = \frac{2 \cdot depth(LCA_{u,v})}{depth(u) + depth(v)}$$

avec  $depth(x)$  le nombre d'arcs qui séparent  $x$  de la racine du graphe.

#### 3.1.2.1.2 Mesures fondées sur la théorie de l'information

Ces approches reposent sur la théorie de l'information de Shannon (Shannon, 1984). Les mesures qui en sont issues reposent sur la comparaison des contenus informationnels des concepts, l'information partagée et l'information discriminante. Pour évaluer la pertinence d'un concept, il faut calculer son contenu informationnel ( $IC$ ). L' $IC$  est obtenu en calculant la fréquence du concept dans un corpus (ou une base d'observation des concepts). Resnik calcule la notion d' $IC$  selon la théorie de l'information de Shannon en fonction de son utilisation dans un corpus (Resnik, 1995) :

$$IC(u) = -\log(p(u))$$

où  $p(u)$  est la probabilité d'occurrence de  $u$ . Plus un concept est fréquemment rencontré, moins il est informatif. Par définition, l' $IC$  augmente à chaque fois que l'on se rapproche des feuilles de l'ontologie, *i.e.*,  $u \preceq v \Rightarrow IC(u) > IC(v)$ . Autrement dit, l' $IC$  d'un concept est toujours supérieur à celui de ses ancêtres puisque les occurrences des hyponymes d'un concept sont ajoutées à ses propres occurrences (Sánchez & Batet, 2011; Schickel-Zuber & Faltings, 2007; Seco *et al.*, 2004; Zhou *et al.*, 2008).

(Resnik, 1995) définit la similarité entre deux concepts en fonction de l'ancêtre commun qui maximise l' $IC$ , *i.e.*, leur ancêtre commun le plus informatif ( $MICA$ ).

$$Sim_{Resnik}(u, v) = IC(MICA_{u,v})$$

La mesure de Resnik ne prend pas explicitement en compte les spécificités (la profondeur dans la taxonomie) des concepts comparés. En effet, des paires de concepts avec un  $IC$  de  $MICA$  équivalent auront la même similarité sémantique, indépendamment de leur profondeur dans la taxonomie. Pour remédier à cet écueil, plusieurs auteurs ont amélioré la mesure proposée par Resnik pour y intégrer la spécificité des concepts comparés. Nous proposons ici quelques-unes des plus connues (Lin & Others, 1998) - $Sim_{Lin}$ , (Jiang & Conrath, 1997)-  $Dist_{JC}$ , (Mazandu & Mulder, 2013) - $Sim_{N\ univers}$ , (Pirró, 2009)-  $Sim_{Psec}$ , (Pirró & Euzenat, 2010)  $Sim_{Faith}$  :

$$Sim_{Lin}(u, v) = \frac{2 \cdot IC(MICA_{u,v})}{IC(u) + IC(v)}$$

$$Dist_{JC}(u, v) = IC(u) + IC(v) - 2 \cdot IC(MICA_{u,v})$$

$$Sim_{N\ univers}(u, v) = \frac{IC(MICA_{u,v})}{\max(IC(u), IC(v))}$$

$$Sim_{Psec}(u, v) = 3 \cdot IC(MICA_{u,v}) - IC(u) - IC(v)$$

$$Sim_{Faith}(u, v) = \frac{IC(MICA_{u,v})}{IC(u) + IC(v) - IC(MICA_{u,v})}$$

D'autres auteurs ont proposé de mesurer la similarité entre deux concepts en analysant l'information portée par les ancêtres (Cross & Yu, 2011; Mazandu & Mulder, 2011). On note par  $A(u) = \{v | u \preceq v\}$  l'ensemble des ancêtres du concept  $u$  :

$$Sim_{Mazando}(u, v) = \frac{2 \cdot \sum_{c \in A(u) \cap A(v)} IC(c)}{\sum_{c \in A(u)} IC(c) + \sum_{c \in A(v)} IC(c)}$$

$$Sim_{JacAnc}(u, v) = \frac{\sum_{c \in A(u) \cap A(v)} IC(c)}{\sum_{c \in A(u) \cup A(v)} IC(c)}$$

Ces mesures peuvent aussi être considérées comme des stratégies hybrides entre l'approche ensembliste basée sur les caractéristiques et l'approche basée sur la théorie d'information.

### 3.1.2.1.3 Mesures ensemblistes fondées sur l'analyse des caractéristiques des concepts

Cette approche extrait les caractéristiques des concepts à partir du graphe. Ces caractéristiques sont ensuite analysées pour estimer la similitude en fonction des caractéristiques communes et discriminantes des concepts comparés. L'idée directrice est de modéliser les concepts comme des ensembles de caractéristiques (ancêtres, descendants, etc.). L'approche la plus couramment utilisée pour caractériser un concept est l'association à ses ancêtres. Par exemple, une mesure de similarité sémantique exprimée à partir de l'indice de Jaccard a été proposée par (Maedche & Staab, 2001).

$$Sim_{cMatch}(u, v) = \frac{|A(u) \cap A(v)|}{|A(u) \cup A(v)|}$$

Une autre mesure basée sur les caractéristiques donnée par (Rodríguez & Egenhofer, 2003) :

$$Sim_{RE}(u, v) = \frac{|A(u) \cap A(v)|}{\lambda \cdot |A(u) \setminus A(v)| + (1 - \lambda) \cdot |A(v) \setminus A(u)| + |A(u) \cap A(v)|}$$

avec  $\lambda \in [0, 1]$  un paramètre qui permet de définir la symétrie de la mesure. (Batet *et al.*, 2010) évaluent aussi la distance taxonomique comme le rapport entre des caractéristiques distinctes et communes.

$$Dist_{Batet}(u, v) = \text{Log}_2 \left( 1 + \frac{|A(u) \setminus A(v)| + |A(v) \setminus A(u)|}{|A(u) \setminus A(v)| + |A(v) \setminus A(u)| + |A(u) \cap A(v)|} \right)$$

### 3.1.2.1.4 Mesures hybrides

Plusieurs approches combinent les propriétés des modèles précédents pour produire une mesure de similarité qui considère plusieurs aspects des représentations associées aux concepts. Une méthode présentée par (Leacock & Chodorow, 1998) combine le comptage des arcs et la méthode du contenu informationnel.  $Sim_{Mazando}(u, v)$  et  $Sim_{JacAnc}(u, v)$  proposent des stratégies hybrides qui combinent l'approche basée sur les caractéristiques et l'approche basée sur la théorie de l'information (Cross & Yu, 2011; Mazandu & Mulder, 2011). (Rodríguez & Egenhofer, 2003) ont également proposé une approche hybride exploitant la profondeur des concepts. Enfin, (Paul *et al.*, 2012) ont proposé plusieurs mesures basées sur l'agrégation de mesures existantes.

### 3.1.2.2 Mesures de similarité entre groupes de concepts

Deux principales approches sont généralement distinguées pour la comparaison de deux ensembles de concepts : les approches directe et indirecte.

#### 3.1.2.2.1 Approche directe

L'approche directe correspond à une généralisation des approches définies pour la comparaison de paires de concepts.

- **Approche basée sur la structure de graphe**

On note  $G_T^+(X)$  le graphe induit par l'union des ancêtres des concepts qui composent l'ensemble  $X$ . (Gentleman, 2005) définit la similarité entre deux groupes de concepts ( $U, V$ ) en fonction de la longueur du plus long plus court chemin qui lie un concept  $c \in G_T^+(u) \cap G_T^+(v)$  à la racine.

- **Approche basée sur la théorie d'information**

(Pesquita *et al.*, 2007) ont proposé une mesure qui se base sur le contenu informationnel du concept pour comparer deux groupes de concepts  $U$  et  $V$ .

$$Sim_{GIC}(U, V) = \frac{\sum_{c \in G_T^+(U) \cap G_T^+(V)} IC(c)}{\sum_{c \in G_T^+(U) \cup G_T^+(V)} IC(c)}$$

- **Approche basée sur les caractéristiques**

Les mesures basées sur les caractéristiques des concepts sont caractérisées par l'approche utilisée pour exprimer les caractéristiques d'un ensemble de concepts. Plusieurs mesures ont été proposées, par exemple la mesure  $Sim_{UI}$  de (Gentleman, 2007) : si  $C_T^+(X)$  est l'ensemble des concepts contenus dans  $G_T^+(X)$  :

$$Sim_{UI}(U, V) = \frac{|C_T^+(U) \cap C_T^+(V)|}{|C_T^+(U) \cup C_T^+(V)|}$$

### 3.1.2.2 Approche indirecte

Elles sont généralement de simples agrégations des scores de similitude associés à des paires de concepts définies sur le produit cartésien des deux ensembles à comparer. Parmi les plus couramment utilisées, on peut citer : la mesure max average (*Max - Avg*), la mesure *BMM* (Best Match Max) (Schlicker *et al.*, 2006) et la mesure *BMA* (Best Match Average) (Pesquita *et al.*, 2007) données par les équations suivantes :

$$Sim_{Avg}(U, V) = \frac{\sum_{u \in U} \sum_{v \in V} sim(u, v)}{|U| \times |V|}$$

$$Sim_{Max-Avg}(U, V) = \frac{1}{|U|} \sum_{u \in U} \max_{v \in V} sim(u, v)$$

$$Sim_{BMM}(U, V) = \max(Sim_{Max-Avg}(U, V), Sim_{Max-Avg}(V, U))$$

Nous avons présenté différentes notions relatives à la représentation des connaissances et à l'utilisation de ces représentations au travers de l'utilisation de mesures sémantiques conceptuelles. Nous utiliserons par la suite ces notions afin de représenter de façon non ambiguë les descripteurs, les structurer, et les analyser en tenant compte de la connaissance propre au domaine de l'analyse sensorielle et olfactive. Nous présentons maintenant différentes techniques permettant d'apprécier et d'exploiter la sémantique portée cette fois non pas par des concepts définis dans une ontologie, mais plutôt par des descriptions exprimées en langage naturelle au travers d'ensemble de termes.

### 3.1.3 Plongement sémantique et comparaison de termes

Peu importe le sens donné aux nombres, il est facile de les manipuler afin d'amener la résolution de problèmes complexes. Cette capacité à manipuler les nombres découle notamment du fait que des opérations élémentaires, exprimées par le système axiomatique les définissant, ou encore des propriétés de ces nombres – pouvant être pairs/impairs/premiers –, permettent de les comparer. Si l'on considère par exemple trois entiers naturels, et cela indépendamment de la sémantique qu'ils partagent (taille, âge, salaire), il nous est facile de définir des modalités de comparaison afin de distinguer la paire de nombres regroupant les tailles (respectivement âges, salaires) les plus proches/similaires – intuitivement, on considèrera la valeur absolue minimale de la différence de deux nombres du type taille (respectivement âges, salaires). Comparer des nombres est ainsi intuitif ; encore faut-il s'accorder sur la sémantique portée par la notion de similarité. Plus généralement, l'ajout d'une dimension supplémentaire complexifie légèrement le traitement. Différentes distances permettent ainsi la comparaison de deux

points dans un espace vectoriel, notamment dans un espace euclidien (*e.g.* distance euclidienne de deux points sur une carte). Cette capacité dont nous disposons à comparer ces objets mathématiques, qu'ils soient des nombres, des vecteurs, ou des objets plus complexes tels que des distributions de probabilités, ou, nous l'avons vu, des concepts définis dans une taxonomie, permet alors de manipuler des grandeurs nécessaires pour la résolution de larges classes de problèmes modélisables au travers de ces objets.

Qu'en est-il cependant des mots ? Comment comparer des descriptifs proposés au travers de termes porteurs de sémantique ? Considérez pour l'illustration les trois descriptifs suivants: (i) *citron*, (ii) *orange*, (iii) *citrouille*. Une comparaison des descriptions qui ne se baserait que sur une distance d'édition évaluant les différences syntaxiques, *i.e.* leur forme dans l'espace des caractères – à l'aide d'une distance de Levenshtein par exemple –, définirait alors les descriptions *citron* et *citrouille* comme les plus proches, et ne permettrait pas alors d'apprécier la similarité sémantique, intuitive pour nous, entre les descriptions faisant référence au citron et à l'orange – similarité ici nourrie par l'appartenance de ces deux fruits à la classe des Agrumes.<sup>3</sup> Nous nous concentrerons dans nos travaux sur la comparaison de termes, ce qui nécessite de les représenter de façon à pouvoir apprécier leur sémantique et ainsi pouvoir définir des modalités de comparaison.

La notion de *word embedding* – terme parfois francisé par plongement sémantique ou représentation de mot – fait référence à la représentation de la sémantique portée par un terme dans un espace latent. L'objectif d'un plongement sémantique est de permettre de manipuler la sémantique des termes au travers de leur représentation dans un espace, généralement vectoriel, sur lequel des opérations classiques, notamment de comparaison, pourront être utilisées. Des opérateurs ou métriques applicables sur cet espace permettront par exemple de traduire les notions intuitives de similarité, proximité ou de distance, que l'on considèrera ici *de facto* comme sémantique – car, par construction, l'espace de plongement est défini dans l'objectif de permettre une caractérisation fidèle de la sémantique des mots (au regard, comme nous le verrons par la suite, d'une appréciation consensuelle d'un modèle définissant la sémantique des termes).

Les premières utilisations de la notion de plongement sémantique d'unités lexicales ont été proposées dans le contexte de la recherche d'information visant à distinguer la pertinence d'un document textuel au regard d'une requête exprimée en langage naturel. Ainsi, le modèle vectoriel (*Vector Space Model*), défini par Salton, proposait dès les années 70 de considérer un document comme un vecteur dans l'espace des termes. Chaque dimension vectorielle correspondait ainsi à un terme d'un vocabulaire, et la représentation d'un document était fournie par un vecteur binaire dont la valeur de chacune des dimensions renseignait la présence du terme associé à la dimension dans le document. Une large littérature s'est par la suite attachée à complexifier ce modèle en proposant notamment d'intégrer la pertinence du terme pour la caractérisation du document, ou encore la fréquence d'observation des termes dans les documents. Nous proposons ci-dessous une rapide illustration des techniques de plongement classiquement utilisées en recherche d'information – la compréhension de ces techniques permettra d'apprécier plus facilement les approches proposées pour le plongement sémantique de termes.

---

<sup>3</sup> Il est à souligner que bien que dans certains cas une similarité syntaxique traduira une forme de similarité sémantique, *e.g.*, dans le cas de la comparaison des termes *courge* et *courgette* par exemple, le propos reste entièrement valide car une forte similarité syntaxique ne traduit pas nécessairement une forte similarité sémantique.

### 3.1.3.1 Introduction au modèle vectoriel pour l'analyse de documents

Un modèle vectoriel calculé sur un corpus permet une représentation algébrique des documents de ce corpus. Ce modèle est notamment utilisé en recherche d'information pour la recherche documentaire et la classification de données. Afin de faciliter la manipulation des documents, ceux-ci sont représentés par des vecteurs dont les composantes sont communément appelées *poids* des termes dans le document. Le modèle vectoriel, en considération d'un vocabulaire  $V$ , permet ainsi de représenter un document dans  $\mathbb{R}^{|V|}$ . Les poids qui renseignent les dimensions vectorielles peuvent être définis selon plusieurs métriques dont les plus classiques sont précisées ci-dessous :

| Calcul du poids             | Définitions   |
|-----------------------------|---|
| <i>Approche booléenne</i>   | <i>Si un terme <math>t</math> apparaît dans un document <math>d</math> alors la composante de <math>\vec{d}</math> selon <math>t</math> vaut 1, sinon 0. L'approche booléenne est utilisée lorsque tous les termes et leurs fréquences sont considérés d'égale importance.</i>  |
| <i>Fréquence des termes</i> | <i>Le poids d'un terme est obtenu en comptant les occurrences du terme dans le document : <math>tf_{ij}</math> représente la fréquence du terme <math>i</math> dans le document <math>j</math> (Salton &amp; McGill, 1986).</i>   |
| <i>Tf – Idf</i>             | <i>Le <math>tf</math> permet d'évaluer l'importance d'un terme contenu dans un document relativement à une collection de documents. Le poids augmente avec le nombre d'occurrences du mot dans le document. Il varie également en fonction de la fréquence du mot dans la collection. Ainsi, la fréquence inverse du document (<math>idf</math>) est une mesure de l'importance du terme dans l'ensemble des documents. Elle vise à donner un poids plus important aux termes les moins fréquents, considérés comme les plus discriminants. Il s'agit de calculer le logarithme de l'inverse de la proportion de documents qui contiennent le terme : <math>idf_i = \log \frac{m}{D(i)}</math> avec <math>m</math> le nombre total de documents de la collection et <math>D(i)</math> le nombre de documents de la collection où le mot <math>i</math> apparaît. Le poids <math>w_{ij}</math> du mot <math>i</math> dans un document <math>j</math> est alors : <math>d_{ij} = tf_{ij} \times idf_i</math>.</i> |

Tableau 3.1 : Méthodes de calcul des poids

Une fois la représentation vectorielle des documents calculée (à noter que cette procédure est applicable à tout passage de texte, e.g. une requête exprimée en langage naturel), il est possible de comparer deux représentations à l'aide de différentes approches :

| Mesures de similarité                        | Définitions  |
|--|--|
| <i>Similarité Cosinus</i>                    | <i>La similarité cosinus permet de mesurer la ressemblance entre deux vecteurs <math>u</math> et <math>v</math> en calculant le cosinus de l'angle les séparant :</i><br>$sim_{cos}(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\ \vec{u}\  \ \vec{v}\ } \in [0,1]$   |
| <i>Coefficient de corrélation de Pearson</i> | <i>Le coefficient de corrélation de Pearson peut être utilisé pour calculer la similarité de deux vecteurs. La similarité obtenue <math>sim_{pear}(u, v) \in [-1,1]</math>.</i><br>$sim_{pear}(u, v) = sim_{cos} \sum_i (\mathbf{u}_i - \bar{u})(\mathbf{v}_i - \bar{v})$ <i>où <math>\bar{u}</math> (resp. <math>\bar{v}</math>) représente la moyenne de <math>u</math> (resp. <math>v</math>)</i> |
| <i>Distance euclidienne</i>                  | <i><math>sim_{euc}(\vec{u}, \vec{v}) = \ \vec{u} - \vec{v}\ </math>. Plus généralement, selon la dimension de l'espace, toute norme <math>p</math>-ième ou de Minkovski est envisageable.</i>  |

Tableau 3.2 : Mesures de similarité basées sur l'approche géométrique

La caractérisation sémantique des termes, qui nous intéresse ici tout particulièrement, repose sur une approche, dans l'essence, assez semblable à celle adoptée pour le plongement sémantique de documents. L'objectif de cette section n'est naturellement pas de s'attarder sur les différentes contributions

techniques portant sur la caractérisation d'unités à caractère sémantique aussi variées que les phrases, les paragraphes ou les documents ; nous nous concentrons ici sur le cas particulier du plongement sémantique des termes, plongement d'un intérêt tout particulier pour nos travaux.<sup>4</sup>

La caractérisation de la sémantique des termes s'intéresse au plongement des termes dans un espace latent permettant la caractérisation de la diversité de la sémantique pouvant être associée aux termes. Cette tâche, particulièrement difficile, car reposant sur un objectif qui n'est que faiblement défini – qu'entendons-nous par sémantique ? – a été largement étudiée en Intelligence Artificielle et plus particulièrement par le domaine de la linguistique computationnelle, et par les différentes communautés de recherche intéressées par le traitement automatique du langage naturel. Très tôt, dès les années 60, Firth propose de considérer que la sémantique d'un mot est définie par son voisinage : '*You shall know a word by the company it keeps*' (Firth, 1957). Cette proposition, tout aussi simple que séduisante, résume à elle seule le fondement de la très grande majorité, sinon l'ensemble, des approches permettant le plongement sémantique de termes. L'objet de cette thèse n'est pas de reconsidérer cette proposition aujourd'hui encore discutée, bien que largement plébiscitée notamment du fait de la performance des différents systèmes l'exploitant –des systèmes de reconnaissance vocale présents dans nos smartphones, aux techniques d'analyse automatique d'opinions exprimées en langage naturel, en passant par des systèmes aussi variés que performants permettant aujourd'hui la traduction automatique et l'implémentation de l'intelligence artificielle, capables d'apprécier le langage naturel, tels que ceux développés et largement adoptés par les GAFAs, ou encore IBM (cf. Watson). Il est aussi à noter que bien que séduisante, la proposition exprimée par Firth, ne permettait pas dans les années 60 d'imaginer une quelconque caractérisation automatique de la sémantique des mots ; il faudra attendre pour cela l'avènement de l'ère numérique, la disponibilité de larges corpus de textes (e.g. Wikipédia ou autres corpus de textes numérisés) ainsi que la disponibilité de puissance de calcul informatique permettant leur traitement.

Les premières implémentations du paradigme consistant à caractériser la sémantique d'un mot par l'analyse de son voisinage, ou, autrement dit, par l'analyse de ses contextes d'observation, ont proposé d'adopter une approche semblable à celle considérée pour la projection sémantique de documents (présentée ci-avant). Ces approches ont alors proposé un plongement sémantique dans un espace multidimensionnel pour lequel les dimensions font référence aux termes composant un vocabulaire d'intérêt.

Une grande partie de la littérature est liée à des mesures fondées sur des corpus pour la comparaison des termes (Curran, 2004; Dinu, 2011; Mohammad & Hirst, 2012; Panchenko *et al.*, 2013; Sahlgren, 2006). Les différentes étapes utilisées pour définir des mesures de similarité fondées sur un corpus sont les suivantes :

- 1) La première étape consiste à extraire le vocabulaire à partir d'un corpus de textes non structurés (pages web). Des filtres sont appliqués pour éliminer des documents : les mots creux, e.g. les pronoms, les déterminants, etc., la ponctuation... D'autres mots peuvent être éliminés selon l'application : lorsqu'on définit la qualité d'une odeur, par exemple, on supprimera superlatifs et adverbes qui trahissent une évaluation hédonique. Les mots sont également souvent lemmatisés ce qui permet de ramener les mots d'une même famille à leur racine et de gérer, par exemple, les formes conjuguées des verbes. Des filtres statistiques peuvent également être appliqués selon un critère de discrimination.

---

<sup>4</sup> Plusieurs contributions ont également été proposées pour comparer de plus grandes unités comme les phrases ou des textes. Le lecteur intéressé peut notamment se référer à (Buscaldi *et al.*, 2013; Corley & Mihalcea, 2005; Hughes & Ramage, 2007; Hughson & Boakes, 2002; Ramage *et al.*, 2009; Yu *et al.*, 2006).

2) Le modèle vectoriel est construit selon le vocabulaire extrait du corpus. De manière similaire à l'approche adoptée pour représenter des documents, chaque terme est représenté comme un vecteur dont les mots du lexique constituent la base. Les composantes des représentations vectorielles peuvent correspondre à différentes métriques (fréquences normalisés, cooccurrences, information mutuelle, etc.) : 1) un terme  $t \in T$  peut être représenté comme un vecteur  $\vec{t} = [coocc(t, t_i)]_{i=0,|T|}$ , avec  $coocc: T \times T \rightarrow \mathbb{N}$  la fonction qui compte le nombre de fois où deux mots apparaissent conjointement dans un même contexte ; 2) un mot peut aussi être modélisé par le vecteur  $\vec{t} = [pmi(t, t_i)]_{i=0,|T|}$  où  $pmi(x, y) = \frac{p(x/y)}{p(x)}$  est l'information mutuelle de  $(x, y)$  (pointwise mutual information) et  $p$  désigne une probabilité. Pour calculer ces métriques, il est nécessaire de définir un paramètre, « une longueur de fenêtre », qui définit le contexte pour calculer une cooccurrence, une probabilité conditionnelle, etc. Le contexte utilisé pour caractériser un mot peut être un nombre de mots, une phrase, un paragraphe, un document, etc. A ce stade, un modèle brut est construit, il prend généralement la forme d'une matrice dont les lignes sont les représentations vectorielles de chacun des mots du vocabulaire. Certaines transformations peuvent encore être appliquées. On distingue les phases de raffinement et de réduction :

- Le raffinement se réfère à des traitements qui sont utilisés pour améliorer le modèle en intégrant des informations complémentaires extraites du texte ou de l'application. A titre d'exemple, certaines approches prennent en compte le caractère informatif de mots spécifiques ;
- La réduction se réfère à l'utilisation de techniques spécifiques pour la réduction de la dimension des représentations vectorielles. Cette étape implique généralement des techniques spécifiques de traitement et de réduction de matrices comme la décomposition en valeurs singulières (SVD) (Berry *et al.*, 1995).

Une fois les représentations des termes construites, différentes mesures peuvent être utilisées pour comparer deux représentations – cf. mesures associées au modèle vectoriel (Tableau 3.2). La mesure la plus utilisée pour comparer deux vecteurs de termes est le cosinus.

D'autres approches semblables permettent de comparer des termes relativement à l'observation des occurrences dans des contextes larges qu'ils partagent ou bien qui les différencient (Curran, 2004) – la notion de contexte fait ici référence à des unités de langage relativement grandes telles que des phrases, des paragraphes ou des documents. La comparaison peut être faite en utilisant des mesures classiques (*e.g.*, l'indice de Dice, coefficient de Jaccard).

| Mesures de similarité         | Définitions  |
|-------------------------------|--|
| <i>Coefficient de Jaccard</i> | <p><i>La mesure de Jaccard compare le nombre de contextes (e.g, documents) communs et le nombre de contextes qui différencient une paire de mots. Elle est définie comme le rapport entre la cardinalité de l'intersection des ensembles considérés et la cardinalité de l'union des ensembles. La similarité obtenue entre deux mots <math>u</math> et <math>v</math> est notée <math>sim_{jacc}(u, v) \in [0,1]</math> :</i></p> $sim_{jaccard}(u, v) = \frac{ contextes(u) \cap contextes(v) }{ contextes(u) \cup contextes(v) }$ |
| <i>Indice de Dice</i>         | <p><i>L'indice de Dice mesure la similarité entre deux mots <math>u</math> et <math>v</math> en se basant sur le nombre de contextes communs et différents où ils sont observés :</i></p> $sim_{dice}(u, v) = \frac{2 contextes(u) \cap contextes(v) }{ contextes(u)  +  contextes(v) }$ <p><i>(version « semi-métrique » de l'indice de Jaccard puisque <math>1 - sim_{dice}(\cdot, \cdot)</math> n'est pas une distance, elle ne vérifie pas l'inégalité triangulaire.</i></p>   |

Tableau 3.3 : Mesures de similarité basées sur des approches ensemblistes

Il a aussi été proposé de comparer deux mots en considérant directement leur probabilité d'occurrence, de cooccurrence, ou de façon plus générale les distributions de probabilités associées à chacun des termes calculées par l'analyse de grands corpus de textes.

| Type d'approche       | Mesures de similarité                                     | Définitions   |
|-----------------------|---|---|
| Approche probabiliste | Pointwise Mutual Information (PMI) (Fano & Hawkins, 1961) | Le PMI a été adapté pour la comparaison des mots par (Église et Hanks, 1990). Il est basé sur l'analyse des occurrences et cooccurrences des mots : $pmi(u, v) = \log \frac{p(u, v)}{p(u)p(v)}$ où $p(u, v)$ est la probabilité que $u$ et $v$ apparaissent dans le même contexte et $p(w)$ la probabilité que $w$ apparaisse dans un contexte. |
|                       | Divergence de Kullback-Leibler                            | La Divergence de Kullback-Leibler (entropie relative) est une mesure de dissimilarité entre deux distributions de probabilité. La distance entre deux mots $u$ et $v$ est estimée par l'entropie relative entre leurs distributions respectives $u$ et $v$ : $dist_{KL}(u, v) = \sum_{k=1}^n u_k \log \frac{u_k}{v_k}$                          |
|                       | Distance du $\chi^2$                                      | La distance entre deux mots $u$ et $v$ est :<br>$dist_{\chi^2}(u, v) = \frac{1}{2} \sum_i \frac{(u_i - v_i)^2}{u_i + v_i}$  |

Tableau 3.4 : Mesures de similarité basées sur l'approche probabiliste

Ces dernières années, différentes reformulations des approches classiques de plongement sémantique de termes classiques ont été proposées. Ces reformulations, aujourd'hui préférées et très largement populaires, reposent elles aussi sur la même intuition consistant à caractériser la sémantique d'un terme par une analyse contextuelle de ses observations. En se basant sur des approches d'apprentissage automatique, et notamment d'apprentissage non supervisé, (Mikolov *et al.*, 2013) proposent par exemple de construire les représentations vectorielles de termes de façon indirecte, en entraînant un réseau de neurones formels à inférer les mots observés au voisinage d'un mot donné, ou de façon analogue, en entraînant le système à inférer correctement un mot en considération d'un voisinage donné en entrée du système. L'idée principale de ces modèles de langues neuronales est de construire un espace de projection sémantique latent dans lequel les termes qui partagent des contextes similaires seront proches. Nous présentons rapidement le modèle Skip-Gram qui fait partie des modèles de ce type les plus largement utilisés aujourd'hui – nous nous permettrons cependant certaines simplifications en vue d'alléger l'illustration de l'approche ; l'objectif de la présentation proposée ci-dessous est essentiellement d'introduire l'idée de base sur laquelle repose le modèle - notre travail exploitant l'utilisation de ces modèles sans chercher à traiter directement la problématique de plongement sémantique de termes.

### 3.1.3.2 Aperçu du modèle Skip-Gram

Le modèle Skip-Gram adopte une approche indirecte afin de réaliser le plongement sémantique des termes. Il repose pour cela sur la définition d'une stratégie d'apprentissage automatique non supervisée consistant à entraîner un réseau de neurones artificiels pour une tâche de prédiction ; cette tâche de prédiction repose elle sur la modélisation de l'hypothèse de la caractérisation contextuelle de la sémantique des termes. Le plongement sémantique sera alors réalisé indirectement du fait de la modélisation du problème et de la stratégie de résolution adoptée. Comme nous allons l'illustrer, les représentations vectorielles des termes correspondront en effet à un ensemble de paramètres du réseau

de neurones, obtenus donc par optimisation des paramètres du réseau en considération de la tâche de prédiction pour laquelle il sera entraîné.

Skip-Gram considère en effet une tâche de prédiction consistant à prédire les contextes d'observation d'un terme – la notion de contexte faisant ici référence à un ensemble de mots situés à proximité d'une observation particulière d'un terme. Le modèle considère pour cela un ensemble connu  $K$  de couples (terme, contexte) extraits d'un corpus de textes – couples par exemple extraits de Wikipédia. Comme cela est classiquement le cas en apprentissage automatique, on recherche alors un modèle probabiliste maximisant la vraisemblance des données considérées, *i.e.* de l'ensemble de couples  $K$ , en se basant sur un estimateur de maximum de vraisemblance. Autrement dit, on recherche le modèle rendant l'observation de  $K$  la plus probable. Avec  $p(c|t; \theta)$  la probabilité du contexte  $c$  en considération du terme  $t$  et d'un modèle paramétré par  $\theta$ , Skip-Gram recherche le jeu de paramètres maximisant la vraisemblance au travers de la définition d'un estimateur de vraisemblance :

$$\hat{\theta} = \operatorname{argmax}_{\theta} \prod_{(t,c) \in K} p(c|t; \theta)$$

Dans sa version simplifiée, lors du traitement d'une paire  $(t, c) \in K$ , le modèle ne considère qu'un terme  $t'$  du contexte  $c$  sélectionné aléatoirement parmi l'ensemble des mots de  $c$ . Dans ce cas le modèle cherche alors à maximiser la probabilité d'observer  $t'$  en considérant l'observation de  $t$ . Par optimisation paramétrique, le modèle cherchera par exemple à obtenir  $p(\text{fruit}|\text{citron}) > p(\text{fumier}|\text{citron})$ , *i.e.* à prédire qu'il est plus probable d'observer le terme fruit que le terme fumier dans le voisinage du terme citron. Le réseau produit alors en sortie un vecteur traduisant les probabilités d'observation de chacun des termes d'un vocabulaire considéré et fourni au système.

Techniquement, sans tout de même aborder les subtilités complexes du modèle, Skip-Gram considère un vocabulaire  $V$  contenant l'ensemble des termes qui seront considérés par le modèle. Les termes fournis en entrée au système sont modélisés sous la forme de *one-hot vectors*, *i.e.* tous les mots sont représentés par un vecteur binaire distinct de taille  $|V|$ . Le  $i$ -ème terme du vocabulaire sera ainsi représenté par le vecteur ligne  $[x_1, \dots, x_i, \dots, x_{|V|}]$  avec  $x_i = 1$  et  $x_{j, j \neq i} = 0$ . L'architecture du réseau est définie classiquement (au regard des approches adoptées dans l'ère pré apprentissage profond) ; elle repose sur l'utilisation d'une unique couche cachée composée de  $k$  neurones complètement connectées aux couches d'entrée et de sortie. La matrice de poids reliant la couche d'entrée à la couche cachée est donc de la forme  $WI \in \mathbb{R}^{|V| \times k}$ . Le produit d'un vecteur ligne  $\mathbb{R}^{1 \times |V|}$  par  $WI$  donne en sortie de la couche cachée un vecteur ligne de  $\mathbb{R}^{1 \times k}$ . La probabilité conditionnelle  $p(c|t; \theta)$  est alors calculée par la couche de sortie. Celle-ci repose sur l'application d'une procédure de type *soft-max* afin que le résultat de la multiplication de la sortie de la couche cachée avec la matrice de poids reliant la couche cachée et la couche de sortie ( $WO \in \mathbb{R}^{k \times |V|}$ ) ait une interprétation probabiliste. Grâce à cette procédure, le vecteur de sortie du réseau, défini dans  $\mathbb{R}^{|V|}$ , correspond à la probabilité que le contexte contienne chacun des termes du vocabulaire sachant le terme fourni en entrée.<sup>5</sup> La phase d'apprentissage cherchera alors à définir les paramètres du modèle de façon à maximiser, dans l'intuition, la vraisemblance de couples de  $K$  en considération des valeurs produites par la procédure *soft-max*. On note tout de même que la sortie de la couche cachée lors du traitement d'une paire  $(t, c)$  correspond simplement au produit du *one-hot-vector* associé au terme  $t$  avec la matrice  $WI$  - on ne considère pas ici de fonction d'activation

---

<sup>5</sup> A noter que, comme précisé ci-avant, nous présentons ici une version simplifiée du modèle original qui considère qu'un contexte ne contient qu'un seul mot. Dans sa version originale, le modèle considère des contextes plus larges et la sortie de la couche de sortie ne correspond donc pas seulement à une distribution multinomiale, mais plutôt à plusieurs distributions de ce type paramétrées par une même matrice  $WO$ .

particulière (e.g., classiquement une sigmoïde), mais seulement le produit matriciel. Le résultat de ce produit est alors interprété comme le plongement sémantique dans  $\mathbb{R}^k$  du terme  $t$  fourni en entrée du réseau. Autrement dit, suite à la procédure d’optimisation du réseau en considération de l’estimateur de maximum de vraisemblance, la matrice  $WI \in \mathbb{R}^{|V| \times k}$  contient l’ensemble des plongements sémantiques des termes définis dans  $V$ , i.e. chaque ligne  $i \in [1, 2, \dots, |V|]$  de  $WI$  correspond au plongement sémantique du  $i$ -ème terme de  $V$ .

Le modèle repose sur un ensemble d’hyper-paramètres définis avant la phase d’entraînement qui préciseront notamment la définition de la notion de contexte considérée (e.g. nombre de mots du voisinage) ou encore la taille de la couche cachée qui définit des caractéristiques propres au plongement sémantique qui sera obtenu, en particulier la taille des représentations vectorielles qui seront construites ( $k = 300$  dans la publication initiale).

### 3.1.3.3 Autres modèles de plongement sémantique de termes

Nous présentons succinctement différents modèles alternatifs permettant le calcul de plongements sémantiques au travers du calcul de représentations vectorielles. Nous nous concentrons essentiellement sur les différents modèles utilisés dans le cadre de nos travaux.

#### 3.1.3.3.1 Modèles CBOW et C&W

Nous venons d’introduire un aperçu de l’approche considérée par le modèle Skip-Gram. De manière semblable, le modèle de langue neuronal CBOW repose lui aussi sur une stratégie d’apprentissage non supervisée d’une tâche de prédiction – celle-ci consiste cette fois à prédire un terme en considération de mots caractérisant un contexte. Ce modèle repose ainsi sur une stratégie inverse à celle implémentée dans le modèle Skip-Gram ; il permet lui aussi de modéliser la construction des plongements sémantiques des termes au travers d’une phase d’optimisation des paramètres (i.e. poids) du réseau utilisé pour réaliser la tâche de prédiction (Mikolov *et al.*, 2013).

Le modèle C&W repose lui aussi une approche basée sur un modèle neuronal (Collobert *et al.*, 2011). Proposé dans un contexte plus large que celui associé à la *simple* construction de plongement sémantique, ce modèle propose un cadre générique pour traiter différents problèmes de TAL (détection de fonctions grammaticales, reconnaissance d’entités nommées, etc.). C&W propose ainsi une architecture générique basée sur un réseau de neurones multicouche profond et partiellement convolutionnel adaptée, modulo quelques modifications, à la résolution de différents problèmes classiquement traités en TAL. Ce sont les premières couches de ce réseau qui permettent la construction des plongements sémantiques – l’approche est semblable à celle réalisée par les modèles Skip-Gram et CBOW – ; le plongement calculé sera cependant fonction de la tâche pour laquelle le réseau a été entraîné.

#### 3.1.3.3.2 Modèle GloVe

Le modèle GloVe propose la construction de plongements sémantiques en couplant des informations issues de la factorisation d’une matrice de cooccurrences (telle que celle utilisée par les modèles de langue historiques) et les informations extraites d’une analyse contextuelle par modèle de prédiction non supervisé. Ce modèle repose pour cela sur l’utilisation d’un modèle de régression log-bilinéaire global (type spécifique de modèle de langue exponentiel classiquement basé sur une maximisation de l’entropie) (Pennington *et al.*, 2014).

### 3.1.3.3.3 Modèle HPCA

Ce modèle propose de construire des plongements sémantiques de termes en effectuant une réduction de type *Hellinger PCA* de la matrice de cooccurrences des termes. La motivation de la réduction proposée repose sur l'observation que les statistiques encapsulées par la matrice de cooccurrence sont par nature discrètes, et qu'une réduction basée sur une *ACP* classique (*i.e.* basée sur une distance euclidienne) ne semble alors pas adaptée. Les auteurs proposent d'appliquer une technique de réduction adaptée aux espaces de distributions de probabilité discrètes (ici généralement des distributions multinomiales de  $|V|$  classes); ils utilisent pour cela une *ACP* basée sur la distance de Hellinger : il calcule alors les plongements sémantiques en réalisant une *ACP* de la matrice de cooccurrences en minimisant l'erreur de reconstruction estimée à l'aide d'une distance de Hellinger (Lebret & Collobert, 2014).

### 3.1.3.3.4 Modèle TSCCA

Ce modèle se base sur une technique de réduction spectrale ; technique qui propose différents algorithmes permettant de détecter des structures par l'analyse des vecteurs propres d'une matrice d'affinité étudiée. *TSCCA* se base alors sur l'utilisation de techniques de Transformation Spectrale (TS) couplées à des Analyses Canoniques de Corrélations (*CCA*) – sorte d'*ACP* considérant deux matrices. Les techniques de *CCA* ont l'intéressante propriété d'être invariantes aux échelles (contrairement à une *ACP*), et sont particulièrement intéressantes en traitement automatique des langues car elles permettent de modéliser des contextes de cooccurrence. Elles sont ici appliquées sur des matrices de cooccurrences capturant la fréquence de cooccurrence entre les termes en distinguant celles observées avant et après chacun des mots considérés (Dhillon *et al.*, 2012) .

### 3.1.3.3.4 Modèle Random Projection

Ce modèle repose lui aussi sur approche de reconstruction de la matrice de cooccurrences. Il propose de construire les plongements sémantiques en factorisant la matrice de cooccurrences en utilisant des techniques de projection aléatoire qui permettent de réduire le nombre de dimensions dans des espaces euclidiens tout en préservant les distances entre les points considérés lors de la réduction (Halko *et al.*, 2011).

### 3.1.3.4 Modèles considérés dans notre étude

Dans notre cadre d'application, nous avons sélectionné les 6 modèles présentés dans la section précédente pour avoir une appréciation globale de la performance des différentes méthodes de plongement sémantique. Le choix de ces modèles a été effectué en considération de l'étude proposée par (Schnabel *et al.*, 2015) qui compare les différentes méthodes de plongement sémantique.

## 3.2 Contributions

Nous présentons dans cette partie les deux classes de modèles proposées afin de réaliser la transcription des termes exprimés en langage naturel par les évaluateurs non experts en descripteurs conceptuels utiles au monde professionnel. La première classe de modèles propose une approche directe en s'intéressant au problème de la transcription de chaque terme fourni par un évaluateur par un concept unique – nous dénommons cette classe de modèles par modèles de *transcription explicite* car nous considérons ici que l'évocation conceptuelle mentionnée est explicite (dans l'intention). Il s'agit pour cette classe de modèles de définir une application de l'ensemble des termes vers l'ensemble des concepts. La seconde classe de modèles s'intéresse au cas où plusieurs termes exprimés par un évaluateur méritent d'être considérés afin d'apprécier l'évocation conceptuelle ressentie par l'évaluateur – nous parlerons alors de

modèles de *transcription implicite*. Cette classe de modèles s'intéresse alors à définir une application des parties de l'ensemble des termes vers l'ensemble des concepts. Les différentes propositions évoquées dans cette section reposent sur l'utilisation conjointe de modèles de plongement sémantique de termes et de relations taxonomiques pour réaliser la projection de l'espace des termes vers celui des concepts.

### 3.2.1 Du terme au concept : conceptualisation via transcription explicite

#### 3.2.1.1 Formalisation du problème et proposition

L'objectif est de transcrire chaque description liée à un ensemble de termes exprimés en langage naturel par un ensemble de descripteurs conceptuels. L'objectif consiste alors à définir une correspondance entre les termes et les concepts d'une taxonomie. Pour cela, nous proposons une approche qui exploite la caractérisation sémantique des termes des descriptions et des labels (termes) liés aux concepts. Nous proposons ainsi de calculer le degré avec lequel un terme évoque un concept de la taxonomie en comparant les représentations vectorielles des termes évoqués par l'évaluateur avec les représentations vectorielles des labels associés aux concepts définis dans la représentation des connaissances.

Afin de faciliter la lisibilité de cette section, les principales notations qui seront utilisées pour introduire les modèles proposés sont énumérées ci-dessous :

- $\mathcal{O} = (C, \preceq)$  : la taxonomie qui définit un ensemble de concepts ordonnés ;
- $C$  : l'ensemble des concepts de la taxonomie  $\mathcal{O}$ ,  $C = \{c_1, c_2, \dots, c_n\}$ ,  $|C| = n$  ;
- Un concept  $c \in C$  est subsumé par un concept  $c' \in C$ , *i.e.*,  $c \preceq c'$  si  $c$  est plus spécifique que  $c'$  et inversement  $c'$  est plus imprécis (abstrait) que  $c$ .
- $T$  : l'ensemble des termes qui constituent le vocabulaire ;
- $T_e \subset T$  : l'ensemble des termes fournis par l'évaluateur  $e$  pour qualifier l'odeur,  
 $T_e = \{t_1^e, t_2^e, \dots, t_{k_e}^e\}$
- $L: C \rightarrow \mathcal{P}(T)$ : la fonction permettant de distinguer l'ensemble des labels associés aux concepts définis dans la taxonomie. A noter que nous considérons que tous les concepts définis dans les représentations des connaissances exploitées dans nos travaux disposent d'au moins un label.

La correspondance ou la mesure de proximité sémantique entre les termes et les concepts ne peut pas être établie d'une manière directe ; en effet, les espaces terminologique et conceptuel considérés sont par construction artificiellement indépendants, *i.e.*  $T \cap C = \emptyset$ . Afin d'établir une correspondance entre l'espace conceptuel et l'espace terminologique, il faut alors tenir compte des labels associés aux concepts (ceux-ci étant définis dans  $T$ , *i.e.*  $\forall c \in C, L(c) \subseteq T$ ). Les plongements sémantiques des termes (dont les labels de concepts) sont donc les seuls vecteurs de mise en correspondance permettant d'estimer dans quelle mesure un terme semble évoquer un concept de la taxonomie. Dans le modèle que nous considérerons, l'intensité d'évocation d'un terme à l'égard d'un concept sera évaluée en fonction de la proximité sémantique entre le terme et le concept (*i.e.*, les labels associés au concept).

Nous considérons donc dans un premier temps une mesure  $\sigma_{TT}: T \times T \rightarrow [0,1]$  permettant d'évaluer la proximité entre deux termes. Nous l'avons vu dans la partie introduisant l'état de l'art (Harispe, Ranwez *et al.*, 2015), de nombreuses mesures et techniques peuvent être utilisées afin d'effectuer la comparaison de deux termes. Nous nous intéresserons ici plus particulièrement aux mesures exploitant les techniques de projection sémantiques introduites dans la section état de l'art – celles-ci apparaissent aujourd'hui dans la littérature comme les mesures sémantiques les plus performantes dans de nombreuses tâches de TAL. Nous considérerons qu'une mesure sémantique permettant de comparer des termes est définie par (i) la méthode de plongement sémantique qui permet de calculer les représentations vectorielles associées aux termes, (ii) ses hyper-paramètres, et (iii) la mesure utilisée pour comparer les

représentations vectorielles. Nous considérerons ici que les représentations vectorielles sont systématiquement comparées à l'aide de la mesure cosinus classiquement utilisée dans la littérature à cet effet.

Nous définissons par la suite la correspondance entre les termes du vocabulaire et les concepts de la taxonomie via  $\sigma_{TT}$  et la fonction de labelling  $L$ . Le degré de correspondance, autrement dit la proximité, entre un terme et un concept est alors défini comme une fonction de l'appréciation de la proximité entre le terme et les labels associés aux concepts. Nous faisons référence à une mesure de proximité entre un terme et un concept par la fonction  $\sigma_{TC}$ , avec :

$$\sigma_{TC}: T \times C \rightarrow [0,1]$$

Nous considérons alors que la proximité sémantique entre un terme  $t$  et un concept  $c$  peut être estimée par définition de la fonction  $\sigma_{TC}$ . Nous considérerons par exemple dans nos travaux une formulation intuitive en définissant  $\sigma_{TC}$  comme le maximum des valeurs de proximité entre le terme  $t$  et les labels associés au concept  $c$  :

$$\sigma_{TC}(t, c) = \max_{l \in L_c} \sigma_{TT}(t, l)$$

avec  $L_c$  l'ensemble des labels associés au concept  $c$ . Différentes formulations alternatives peuvent naturellement être explorées. Dans un souci de simplicité et de traçabilité des résultats produits par l'approche globale d'analyse des descriptifs proposés par les évaluateurs, nous avons fait le choix de considérer une fonction de mise en correspondance associée à une complexité maîtrisée.

Nous avons détaillé la modalité d'obtention de l'évocation conceptuelle la plus adéquate en considération d'un terme donné. Une dernière étape consiste alors à synthétiser l'information exprimée par l'ensemble des termes de la description  $T_e$  d'un évaluateur  $e$  pour caractériser le produit testé. Il convient alors de regrouper l'information véhiculée par les représentations vectorielles des termes de la description  $T_e$ . Cette annotation conceptuelle, notée  $annot(T_e)$ , avec  $annot: \mathcal{P}(T) \rightarrow \mathcal{P}(C)$  peut être calculée comme suit :

$$annot(T_e) = \bigcup_{t \in T_e} \left\{ \operatorname{argmax}_{c \in C} \sigma_{TC}(t, c) \right\} \quad (3.1)$$

Ce modèle de mise en correspondance dît de transcription explicite, repose sur le postulat intuitif consistant à considérer (i) qu'un terme évoque de façon préférentielle les concepts associés à des labels qui lui sont proches, et (ii) qu'un ensemble de termes est conceptualisable en traitant de façon indépendante les différents termes. En considérant une approche de transcription explicite de ce type, il semble évident que la performance de l'approche globale de transcription repose sur la qualité de la mise en correspondance réalisée par la fonction  $\sigma_{TT}$ . Du fait de son importance critique afin d'assurer la performance de la mise en correspondance entre l'espace terminologique et l'espace conceptuel, différentes analyses permettant de critiquer le choix de la définition de  $\sigma_{TT}$  ont été réalisées. Nous étudierons ainsi par la suite l'impact du choix de la stratégie de plongement sémantique des termes et de certains hyper-paramètres, en particulier l'impact associé au choix du corpus de textes à partir duquel le plongement sémantique est réalisé. Nous explorerons par la suite l'utilisation de différentes techniques innovantes de couplage de modèles de plongement sémantique.

### 3.2.1.2 Expérimentation sur la conceptualisation des descripteurs terminologiques

#### 3.2.1.2.1 Protocole expérimental

Le protocole expérimental considéré afin d'évaluer les performances de l'approche de projection des termes dans un espace conceptuel repose sur l'analyse d'expériences olfactives menées sur 31 vins de Cabernet. L'approche de projection est évaluée en comparant la mise en correspondance proposée par l'approche automatisée (paramétrée par un modèle de plongement sémantique, *i.e.*, une définition détaillée de  $\sigma_{TT}$ ) avec celle réalisée par un expert en analyse sensorielle suite à la mise en application d'une approche orientée. Ce travail d'évaluation permettra alors d'apprécier les différences entre les mises en correspondance proposées par l'approche non orientée avec celles réalisées par l'implémentation automatisée d'une forme d'approche non orientée. Ce travail d'évaluation a été réalisé en collaboration étroite avec un stagiaire de IMT Mines d'Alès, Jie He (He, 2016).

Les données exploitées dans le cadre de cette analyse sont tirées d'une base de données d'évaluation de 31 vins par 17 professionnels du domaine. La base de données propose les descriptions des évaluateurs sur les odeurs perçues des 31 vins. Les descriptions sont exprimées au travers de l'utilisation d'autant de mots ou d'expressions jugées nécessaires par les évaluateurs. A savoir que les évaluateurs n'ont reçu au préalable aucune instruction quant à la façon de décrire les vins et que les descriptions ont été réalisées sans apprentissage préalable d'un vocabulaire partagé.

Pour l'approche orientée, une liste de descripteurs (concepts) communs est élaborée sur la base d'une classification (taxonomie). Cette classification comporte 4 niveaux de descripteurs désambiguïsés et d'intérêt (concepts) habituellement utilisés pour la caractérisation des vins rouges de la vallée de la Loire. Les descripteurs sont organisés en classes (*e.g.*: Fruits>Fruits rouges>Fruits rouges frais>Framboise, Fraise, etc.). Le travail de l'expert consiste à associer à chaque descripteur libre (terme) fourni par les 17 évaluateurs un descripteur conceptuel pour chaque niveau de la classification. Le but de cette tâche est de rapprocher chaque terme de concepts à différents niveaux d'abstraction (*e.g.*, le terme *framboise* représente pour l'expert le concept Fruits au niveau 1, Fruit rouge au niveau 2, Fruits rouge frais au niveau 3 et Framboise au dernier niveau). L'intérêt est de fournir pour chaque vin quatre caractérisations objectives avec différents niveaux d'abstraction. La Figure 3.1 représente un imprimé écran de l'interface PROTEGE de la structuration des descripteurs conceptuels des quatre niveaux utilisés dans le cadre de l'application.

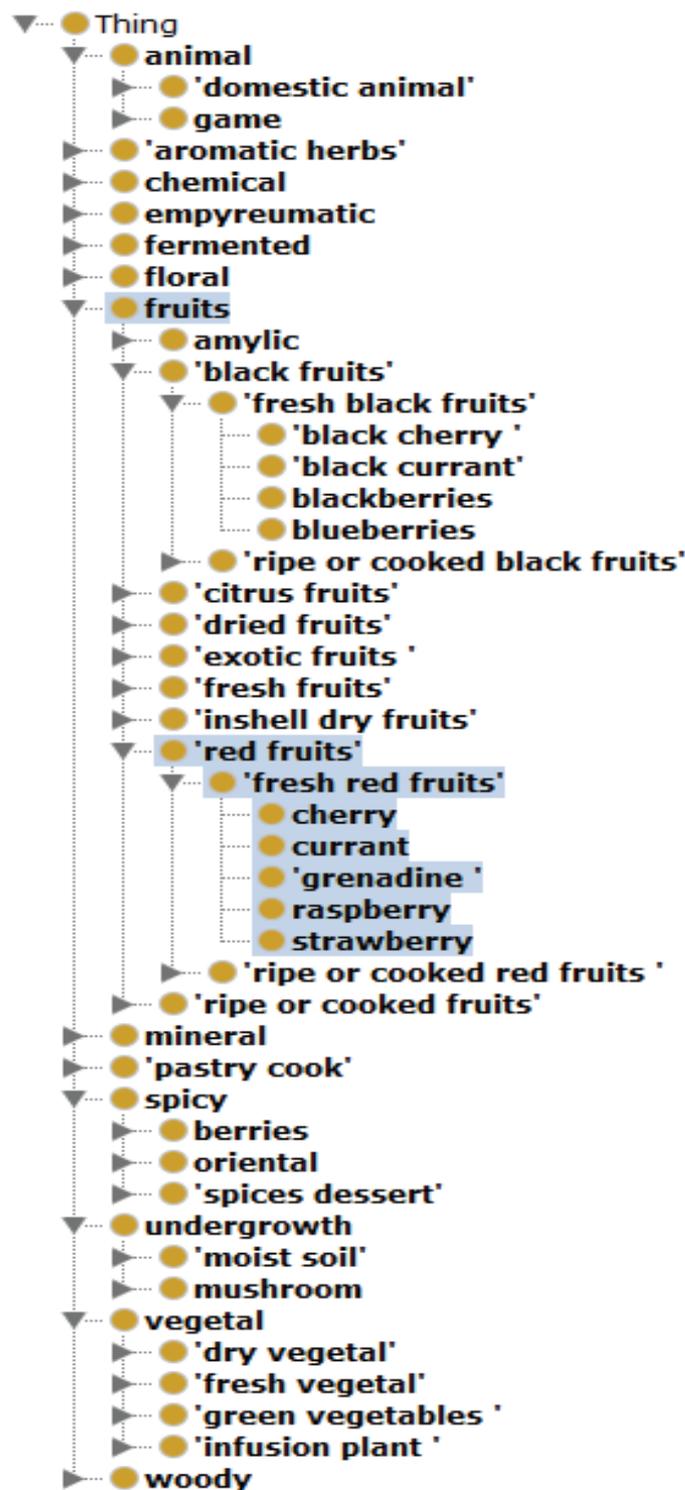


Figure 3.1 : Structuration des descripteurs conceptuels en différents niveaux

Pour la mise en relation des termes et des concepts, nous proposons d'associer à chaque terme un concept de la classification pour les 4 niveaux. Le but est d'analyser les proximités ou le degré d'évocation entre les termes fournis par les évaluateurs et les concepts pour chaque niveau de la taxonomie. Nous disposons ainsi de  $328 \times 281$  couples terme/concept au total, associés à des niveaux d'abstraction numérotés de 1 à 4 du plus abstrait (1) au plus spécifique (4).

Dans une première phase, l'évaluation menée repose à la fois sur (i) l'évaluation du choix de la technique de plongement sémantique, et sur (ii) l'évaluation du choix du corpus considéré pour le calcul des représentations vectorielles construites à partir du plongement sémantique. Sur ce second point, nous souhaitons en particulier tester l'impact de la sélection du corpus pour la définition d'une approche contextuelle et domaine-dépendante en ce qui concerne la construction des représentations vectorielles des termes.

### 3.2.1.2.2 Modèles de plongement sémantique des termes, corpus et détails techniques

Dans notre évaluation, 6 modèles de plongement sémantique sont utilisés : *CBOW*, *C&W*, *HPCA*, *GloVe*, *TSCCA* et *Random Projections*. Ces derniers ont été présentés dans la section introduisant l'état de l'art. Ces modèles de plongement permettent de représenter les termes d'un corpus par des vecteurs afin de faciliter leur analyse sémantique. Le choix spécifique de ces modèles est justifié par la volonté de représenter la diversité des approches de plongement les plus populaires aujourd'hui.

Pour mettre en œuvre l'approche non orientée que nous proposons, deux types de corpus ont été utilisés pour construire les différents plongements sémantiques testés. Le premier est un corpus correspondant à Wikipedia en version anglaise (version du 04-2008). Le second, nommé WikiOdor, est un corpus spécialisé lié au domaine psychosensoriel ; celui-ci a été extrait à partir du premier corpus en se concentrant sur des passages du corpus qui font référence à des termes relatifs au domaine de l'analyse psychosensorielle. L'extraction des paragraphes d'intérêt pour le domaine de l'analyse olfactive a été effectuée en considérant la liste de termes suivants :

*fragrance, smell, aroma, aura, balm, bouquet, essence, incense, odorize, perfume, pheromone, redolence, whiff, stench, stink, deodorize, doggy, lemon, lemony, lilac, lime, mildewed, mint, minty, moldy, pine, plastic, rose, skunky, woody, acid, acrid, airy, biting, clean, crisp, dirty, earthy, faint, feminine, fetid, fishy, fresh, floral, flowery, light, loamy, masculine, moist, musty, nauseating, perfumed, pungent, putrid, rancid, redolent, repulsive, rotten, sharp, sour, spicy, spoiled, stale, stinking, sweaty, sweet, tart, wispy, taste, tastes, odor, odour, flavour, flavor, osmyl.*

L'implémentation des techniques de plongement sémantique évaluées est un travail technique et complexe qui dépasse naturellement le cadre de cette thèse. Bien que différents modèles de plongement sémantique basés sur l'analyse de cooccurrences (éventuellement couplés à des métriques de type PMI) aient été implémentés et testés dans les premières phases de test, les travaux d'évaluation que nous discuterons reposent exclusivement sur les implémentations proposées la plupart du temps par les développeurs des méthodes testées. Or différents traitements sont réalisés dans chacune des implémentations afin, par exemple, de réduire la complexité algorithmique de l'approche. Nous observerons ainsi des variations non négligeables sur le nombre de plongements sémantiques générés par les différentes approches. Ce nombre est fonction de nombreux paramètres, la plupart du temps non modifiables, au sein des différentes implémentations des modèles de plongement. Nous pouvons citer par exemple le nombre minimal d'observations d'un terme afin de réaliser un plongement sémantique (certaines approches requièrent un nombre d'observations minimal important), des stratégies de lemmatisation, d'exclusion des *stop words* et autres techniques de normalisation lexicale définies de façon non modifiables. A titre d'exemple, le Tableau 3.5 présente la taille du vocabulaire considéré par chaque modèle de plongement sur le corpus Wikipedia.

| Model Name               | Vocabulary |
|--------------------------|------------|
| <i>C&amp;W</i>           | 130000     |
| <i>CBOW</i>              | 518914     |
| <i>GloVe</i>             | 840347     |
| <i>HPCA</i>              | 840297     |
| <i>Random Projection</i> | 840297     |
| <i>TSCCA</i>             | 150001     |

Tableau 3.5 : Taille de vocabulaire pour chaque modèle avec le corpus Wikipédia

Ainsi, alors que l'application de l'implémentation du modèle de plongement GloVe ne génère pas moins de 840k représentations vectorielles différentes, l'implémentation du modèle de plongement C&W n'en génère que 130k. Le Tableau 3.6 présente les tailles de vocabulaire obtenues suite à l'analyse du corpus WikiOdor.

| Model Name               | Vocabulary |
|--------------------------|------------|
| <i>CBOW</i>              | 198954     |
| <i>GloVe</i>             | 198954     |
| <i>HPCA</i>              | 21364      |
| <i>Random Projection</i> | 21364      |
| <i>TSCCA</i>             | 43199      |

Tableau 3.6 : Taille de vocabulaire pour chaque modèle avec le corpus WikiOdor

Ces différences de taille des vocabulaires utilisés par les méthodes de plongement induiront nécessairement des variations dans le nombre de paires terme/concept évaluables en fonction du modèle de plongement considéré. En effet, pour évaluer une paire terme/concept en considérant un modèle de projection, *i.e.* afin de pouvoir tester si une correspondance connue entre un terme et un concept peut être retrouvée, il faut nécessairement que le terme de la correspondance et au moins un label du concept associé aient une représentation vectorielle qui ait été obtenue par le modèle de projection considéré. Si cela n'est pas le cas le lien attendu entre le terme proposé et le concept associé ne peut être retrouvé.

La Figure 3.2 présente pour chaque niveau le nombre de paires terme/concept qui peuvent être traitées par les 6 modèles. Le nombre de paires évaluables pour chaque niveau est très faible par rapport au nombre de paires total ; des variations importantes sont en particulier observées pour les niveaux 2, 3 et 4. Cela s'explique par le fait que différents labels de concepts et termes évoqués soient composés (*e.g.*, sirop de framboise). Or, les méthodes de plongement ne construisent que très rarement des représentations vectorielles pour des termes composés. Ainsi, afin d'augmenter le nombre de paires communes, une méthode d'agrégation vectorielle « sum vector » a été appliquée afin de calculer la représentation vectorielle des labels et des descripteurs qui contiennent plusieurs termes. Par la suite, pour une séquence de termes  $s = t_0 t_1 \dots t_n$ , avec  $t_i \in T$  nous considérerons que la représentation vectorielle de  $s$  est définie par :

$$\vec{s} = \sum_{t \in t_0, t_1, \dots, t_n} \vec{t}$$

Cette approche de composition simple, nous permet de ne pas traiter de façon distincte les mots faisant référence à un seul terme de ceux correspondant à une séquence de termes. A titre d'exemple, l'agrégation des mots composés permet d'augmenter le nombre de paires communes (common pairs de la Figure 3.2) de 133 à 247 pour le niveau 1.

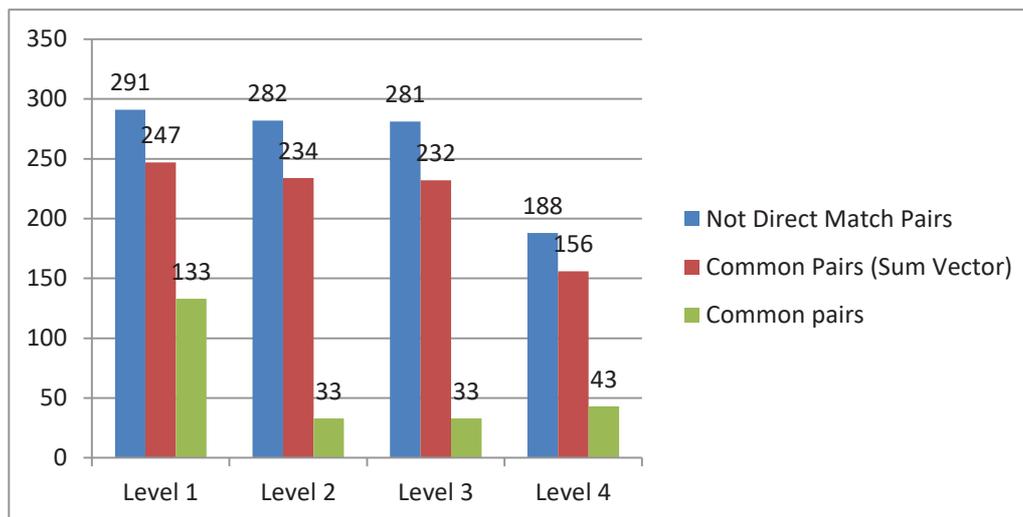


Figure 3.2 : Paires termes/concepts utilisées pour l'évaluation

### 3.2.1.2.3 Evaluation et comparaison

On définit le rang d'un concept  $c \in C$  pour un terme  $t$  en considération d'un modèle  $M_i$  ( $i = 1, \dots, 6$ ), comme l'indice de la position du concept  $c$  dans un ordre total des concepts construit par l'utilisation de  $\sigma_{TT}^{M_i}$ , une fonction de proximité terminologique basée sur le modèle  $M_i$ . L'ordre total est construit comme suit :

$$\forall (c, c') \in C^2, c <_{rank_{M_i}} c' \text{ iff } \sigma_{TC}^{M_i}(t, c) < \sigma_{TC}^{M_i}(t, c')$$

$$\text{avec } \sigma_{TC}^{M_i}(t, c) = \max_{l \in L_c} \sigma_{TT}^{M_i}(t, l)$$

On considère alors  $rank_{M_i}(c, t) \in [0, |C| - 1] \subset \mathbb{N}$  avec  $rank_{M_i}(c', t) < rank_{M_i}(c, t)$  si  $c <_{rank_{M_i}} c'$  et la valeur minimale de rang systématiquement égale à 0.

Ainsi, pour une paire  $P_{tc} \in T \times C$  associée à la mise en correspondance conceptuelle attendue pour le terme  $t$ , le rang  $rank_{M_i}(c, t)$  définit la position de préférence de l'association attendue (*i.e.*, concept  $c$ ). Ainsi, plus le rang du concept  $c$  est petit, plus cela signifie qu'il correspond à une recommandation / correspondance associée à un fort score ; l'erreur d'appariement sera alors considérée comme d'autant plus petite que le rang sera faible.

L'évaluation des différents modèles se base sur deux métriques :

1. L'*erreur relative* pour un modèle donné, en considération d'un niveau d'abstraction fixé, correspond à la moyenne de l'erreur effectuée sur chaque paire associée au niveau d'abstraction considéré. L'erreur relative est naturellement fonction du rang. Ainsi, pour un ensemble de concepts  $C' \subseteq C$  associés aux concepts du niveau d'abstraction considéré, et un rang proposé égale à  $r \in [0, |C'| - 1]$ , l'erreur considérée pour une paire, définie entre 0 et 1, sera égale à :

$$error(r) = \frac{r}{|C'| - 1}$$

2. *Difficulté moyenne* : permet d'apprécier un facteur de dispersion dans l'espace terminologique pour chaque modèle et pour chaque niveau d'abstraction. Cela permet notamment d'illustrer le niveau de difficulté d'association d'un concept et d'un terme. En effet, tous les termes n'ont pas, par nature, le même degré de centralité dans l'espace terminologique, *i.e.* certains termes tels que *orange* seront, du fait de leur usage fréquent, proches de nombreux termes ; d'autres

termes plus spécifiques tels que *pitaya* (fruit du dragon) devraient afficher des niveaux de proximité plus faibles avec l'ensemble des mots du vocabulaire. La difficulté moyenne associée à une paire permet alors, pour un modèle donné, d'apprécier le degré de proximité maximal avec l'ensemble des termes.

$$difficulty(V, M_i) = \frac{1}{|T|} \sum_{t \in T} \max_{t' \in V \setminus \{t\}} \sigma_{TT}^{M_i}(t, t')$$

Ainsi, plus un modèle sera associé à une valeur de difficulté élevée, plus l'on considérera qu'il sera difficile de distinguer le concept attendu par l'utilisation de ce modèle car un degré de proximité, même élevé, ne sera pas nécessairement discriminant.

Les résultats de l'évaluation de la performance des différents modèles pour le corpus Wikipédia et le corpus WikiOdor sont respectivement illustrés sur les Figure 3.3 et Figure 3.4. La barre colorée en bleu représente la moyenne de l'erreur relative, la barre colorée en rouge correspond à la difficulté moyenne.

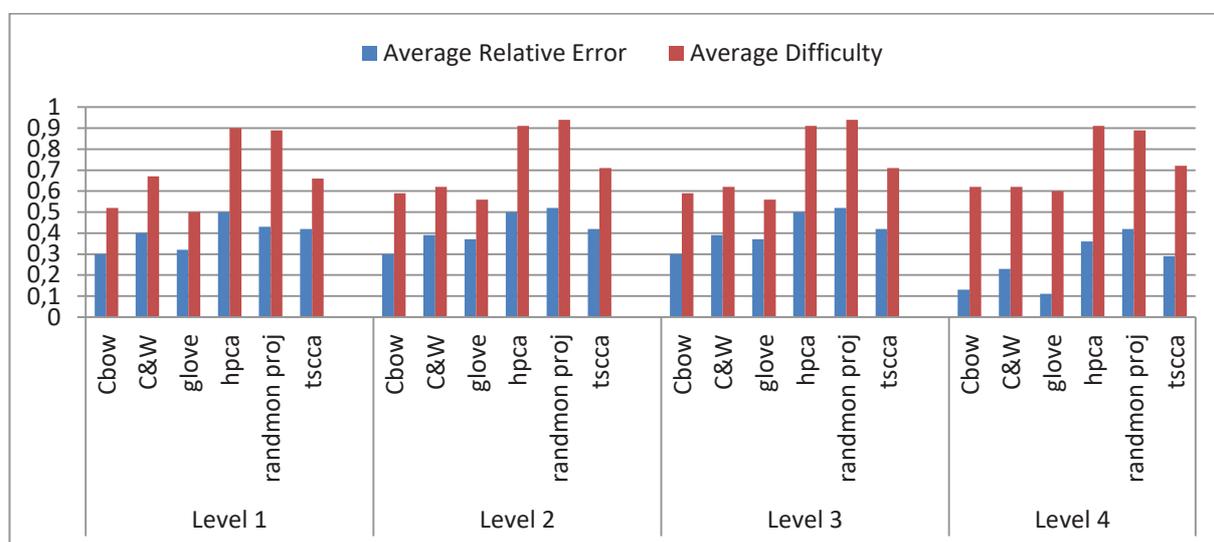


Figure 3.3 : Evaluation des résultats pour le corpus de pages de Wikipédia

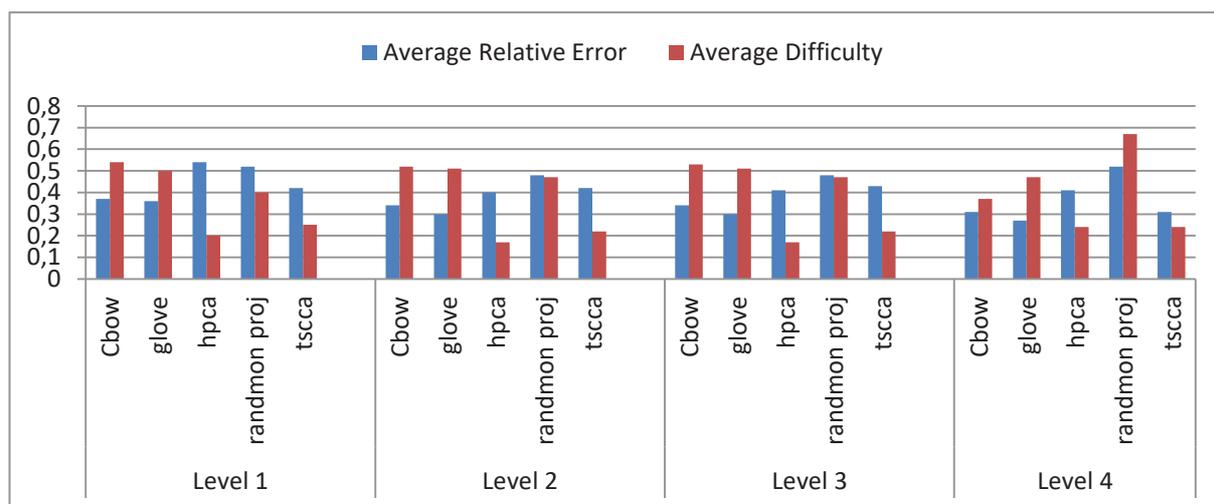


Figure 3.4 : Evaluation des différents modèles pour le corpus WikiOdor

L'analyse des résultats permet d'observer des variations entre les différents modèles analysés. Nous nous concentrons dans un premier temps sur l'analyse des résultats obtenus à partir des modèles

construits sur le corpus complet de Wikipedia Figure 3.3. En ce qui concerne le niveau d'abstraction le plus élevé (Level 1), le modèle Cbow obtient les meilleurs résultats avec une erreur relative égale à 0.3. Plus généralement, de manière consistante avec les résultats obtenus pour de nombreuses tâches de traitement automatique du langage naturel étudiées dans la littérature, les modèles Cbow et Glove généralement très performants, obtiennent dans notre étude les résultats les plus intéressants pour le niveau d'abstraction le plus élevé. Les résultats relatifs à l'erreur relative sont semblablement similaires sur l'ensemble des niveaux d'abstraction. Nous notons aussi que les deux modèles les plus performants sont aussi associés aux valeurs de difficulté les plus faibles : ce qui conforte la pertinence d'utiliser ces modèles dans le cas où les modèles sont utilisés en exploitant un corpus généraliste.

Les résultats obtenus lorsque les modèles exploitent le corpus WikiOdor sont semblables à ceux obtenus avec le corpus généraliste : les modèles Cbow et Glove permettent d'obtenir les erreurs relatives les plus faibles. Nous observons cependant une variation importante du comportement des valeurs de difficulté sur ce corpus. Les modèles Cbow et Glove sont en effet dans ce cas associés aux valeurs de difficulté les plus élevées. Résultat important, nous notons que l'utilisation d'un corpus spécifique ne permet pas dans le cadre expérimental évalué de réduire l'erreur relative des modèles.

Bien que des modèles affichent des performances supérieures en terme d'erreur relative, une analyse détaillée des résultats produits par les différents modèles a cependant permis de mettre en évidence le fait que les erreurs des modèles sont la plupart du temps différentes. Cette observation nous a encouragé à tester une stratégie de sélection d'association basée sur un couplage des différents modèles de plongement sémantique.

#### 3.2.1.2.4 Stratégie de sélection d'association basée sur l'agrégation des différents modèles

Nous proposons un modèle agrégé des 6 modèles testés précédemment. Pour un terme donné  $t$ , et un concept donné  $c$ , nous définissons pour cela une stratégie d'agrégation des rangs des paires  $P_{tc}$  obtenus par les 6 modèles basée sur une sommation des rangs calculés par les différents modèles. Nous considérons alors que le concept le plus pertinent est alors le concept auquel est associée la somme des rangs minimale. On note par  $c(t)$ , le résultat d'évaluation pour un terme donné  $t$ .

$$c(t) = \underset{c}{\operatorname{Argmin}} \sum_{i=1}^6 \operatorname{rank}_{M_i}(c, t) \quad (3.2)$$

Afin de faire varier l'importance de chacun des modèles, et ainsi tenir compte des variations de performance globale observées lors de nos analyses précédentes, un poids est associé à chaque modèle afin d'associer plus ou moins d'importance à un modèle lors de la phase d'agrégation.

$$c(t) = \underset{c}{\operatorname{Argmin}} \sum_{i=1}^6 w_i \cdot \operatorname{rank}_{M_i}(c, t) \quad (3.3)$$

Le paramétrage des poids associés aux 6 modèles a été effectué à l'aide d'une technique d'apprentissage supervisée. Nous avons pour cela cherché à améliorer le ratio de prédiction des paires terme/concept correctes en considération d'une base d'apprentissage composée d'un ensemble de paires terme/concept. L'apprentissage supervisé appliqué se base sur l'utilisation d'un neurone formel.

Pour chaque donnée qui compose l'ensemble d'évaluation, un terme donné est associé à un concept d'un niveau d'abstraction précis. Une matrice de rangs peut alors être obtenue par application de la stratégie de rang minimal détaillée précédemment. Il est alors possible d'obtenir une matrice de rangs pour chaque terme associé à une paire de l'ensemble d'apprentissage – cf. ligne du Tableau 3.7. Une colonne de "sortie" est ajoutée à cette matrice comme étiquette de classification (lorsque le concept est correct, la valeur "0" est précisée, sinon la valeur "1" est renseignée). C'est à partir de cette structure de données

que sera constituée la base d'apprentissage sur laquelle se basera la phase d'apprentissage des poids du neurone formel.

| Term          | Concepts       | Correct concept | Tscca    | Random projection | Glove    | Cbow     | Hpca     | C&W      | Output   |
|---------------|----------------|-----------------|----------|-------------------|----------|----------|----------|----------|----------|
| fennel        | Spicy          | Vegetal         | 2        | 2                 | 7        | 3        | 4        | 3        | 1        |
| fennel        | Undergrowth    | Vegetal         | 6        | 9                 | 6        | 8        | 10       | 7        | 1        |
| fennel        | Fermented      | Vegetal         | 3        | 10                | 5        | 2        | 9        | 1        | 1        |
| fennel        | Chemical       | Vegetal         | 10       | 7                 | 8        | 9        | 6        | 9        | 1        |
| fennel        | Woody          | Vegetal         | 9        | 1                 | 2        | 5        | 1        | 10       | 1        |
| fennel        | Animal         | Vegetal         | 7        | 6                 | 9        | 10       | 8        | 6        | 1        |
| fennel        | Fruits         | Vegetal         | 1        | 4                 | 1        | 1        | 3        | 2        | 1        |
| fennel        | Floral         | Vegetal         | 8        | 3                 | 4        | 4        | 2        | 8        | 1        |
| <b>fennel</b> | <b>Vegetal</b> | <b>Vegetal</b>  | <b>4</b> | <b>8</b>          | <b>3</b> | <b>6</b> | <b>7</b> | <b>4</b> | <b>0</b> |
| fennel        | Mineral        | Vegetal         | 5        | 5                 | 10       | 7        | 5        | 5        | 1        |

Tableau 3.7 : Exemple d'un échantillon d'apprentissage pour le niveau 1

En effet, en considération de cette modélisation, l'objectif est de distinguer un vecteur de poids, paramètres du neurone formel, permettant d'obtenir chaque valeur du vecteur de sortie (*ouput*) en analysant un vecteur de rang renseigné en entrée, *i.e.* les différents rangs renseignés par les différents modèles. Les poids du neurone formel sont au préalable initialisés avec des valeurs aléatoires. Avec  $y^i$  la valeur de sortie et  $g^i$  la valeur de sortie souhaitée pour la donnée d'apprentissage  $i$ , la fonction d'erreur minimisée pour l'obtention des paramètres optimaux est l'erreur quadratique classique ;  $E(w) = \frac{1}{2} \sum_{j=1}^m (y^j - g^j)^2$ ;  $m$  représente le nombre d'échantillons de la base d'apprentissage et  $y^j = f(\sum_{i=1}^6 w_i^j x_i^j)$ , dans notre cas la fonction d'activation  $f$  est la fonction sigmoïde :  $f(u) = \frac{1}{1+e^{-u}}$

Pour l'évaluation de la tâche d'association terme/concept, on s'intéresse dans un premier temps aux valeurs suivantes :

- Ratio de prédiction dans le cas non pondéré : il représente le rapport entre le nombre de paires correctes obtenues avec le modèle agrégé (équation 3.2) et le nombre de paires évaluées.
- Ratio de prédiction dans le cas pondéré : représente le rapport entre le nombre de paires correctes obtenues avec le modèle agrégé pondéré (équation 3.3) et le nombre de paires évaluées.

Les Tableau 3.8 et Tableau 3.9 représentent respectivement le ratio de prédiction des paires correctes pour chaque niveau dans le cas non pondéré et le cas pondéré pour le corpus de Wikipédia.

| Concept Level | Correct prediction | Pairs number | Prediction ratio |
|---------------|--------------------|--------------|------------------|
| Level 1       | 39                 | 133          | 0.293            |
| Level 2       | 8                  | 33           | 0.242            |
| Level 3       | 8                  | 33           | 0.242            |
| Level 4       | 11                 | 43           | 0.256            |
| Average       |                    |              | 0.258            |

Tableau 3.8 : Ratio de prédiction dans le cas non pondéré

| Concept Level | Correct prediction | Pairs number | Prediction ratio |
|---------------|--------------------|--------------|------------------|
| Level 1       | 59                 | 133          | 0.444            |
| Level 2       | 14                 | 33           | 0.424            |
| Level 3       | 14                 | 33           | 0.424            |
| Level 4       | 23                 | 43           | 0.535            |
| Average       |                    |              | 0.457            |

Tableau 3.9 : Ratio de prédiction dans le cas pondéré

Le Tableau 3.10 représente les poids obtenus avec un neurone formel à partir d'un échantillon d'apprentissage pour chaque niveau.

| Model Name | Cbow    | C&W     | Glove   | Hpca    | Random projection | Tessa   |
|------------|---------|---------|---------|---------|-------------------|---------|
| Level 1    | 6.4315  | -1.4513 | 10.2644 | -3.2122 | 4.3821            | -2.3639 |
| Level 2    | 21.8757 | -2.6348 | 3.0634  | 0.277   | -1.5334           | -0.09   |
| Level 3    | 21.8757 | -2.6348 | 3.0634  | 0.277   | -1.5334           | -0.09   |
| Level 4    | 14.5531 | -1.4853 | 12.2572 | -0.6429 | -0.903            | -0.6594 |

Tableau 3.10 : Matrice des poids des différents modèles

Nous constatons que le modèle agrégé pondéré (équation 3.3) semble bien se comporter lorsqu'il est utilisé afin d'agréger les modèles exploitant le corpus de pages de Wikipédia. Une amélioration nette par rapport au modèle agrégé sans poids (équation 3.2) est observée dans les évaluations effectuées. Le ratio moyen de prédiction de paires correctes termes/concepts a augmenté de 0,258 à 0,457.

En complément des résultats proposés précédemment nous présentons les résultats obtenus après application d'une approche de validation croisée ( $k$ -fold cross-validation). Le principe de la méthode  $k$ -fold cross-validation est de répartir l'échantillon initial de manière aléatoire en  $k$  sous-échantillons de même taille. Parmi les  $k$  sous-échantillons, un seul sous-échantillon est conservé comme données de test pour tester le modèle et les  $k - 1$  sous-échantillons restants sont utilisés comme données pour l'apprentissage. Le processus de validation croisée est répété  $k$  fois. Les  $k$  sous-échantillons sont utilisés exactement une fois comme données de test.

La Figure 3.5 compare les précisions (ratio) de prédiction (abrégée par PA) des paires correctes (terme/concept) obtenues avec (i) le modèle agrégé sans poids, (ii) le modèle pondéré et (iii) la moyenne des ratios de prédiction des 6 modèles.

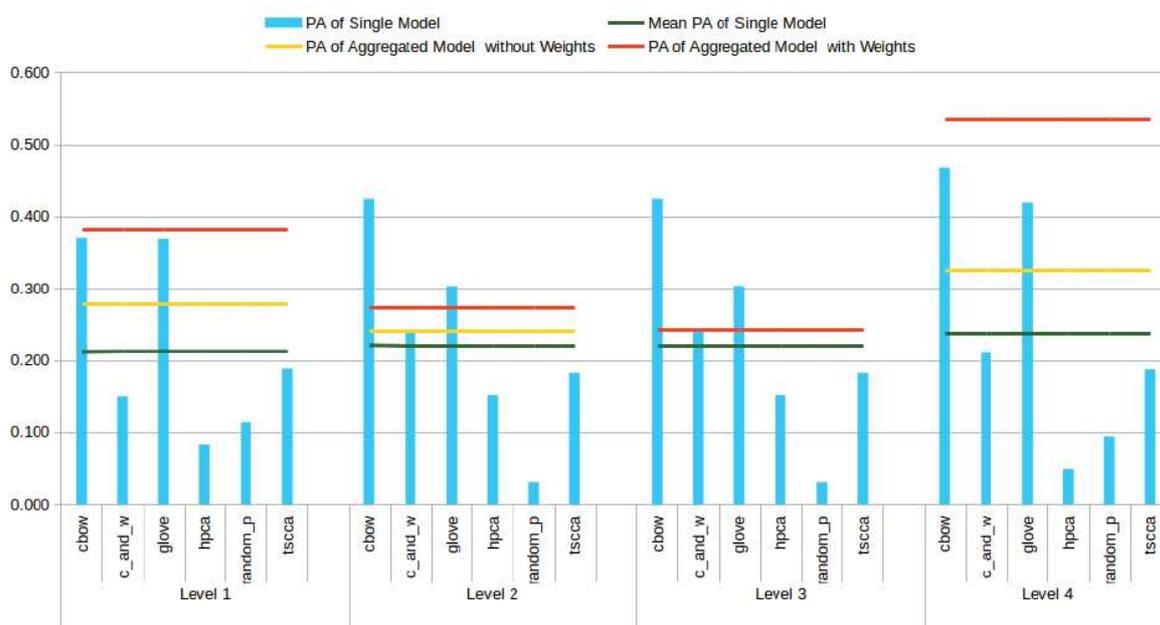


Figure 3.5 : Performance d'évaluation terme/concept dans le cas du corpus de pages de Wikipédia

Nous constatons que le modèle agrégé pondéré (PA of aggregated model with weights - équation 3.3) est le modèle agrégé le plus performant pour la tâche d'évaluation de paires correctes pour les 4 niveaux.

Le Tableau 3.11 fournit les performances des modèles agrégés ainsi que les pourcentages d'amélioration de ratio de prédiction du modèle agrégé appris par rapport au modèle agrégé sans poids ; une

comparaison par rapport à la moyenne des ratios de prédiction des 6 modèles *classiques* (*i.e.* non agrégés) est aussi proposée. Ces résultats portent sur les modèles basés sur l'utilisation du corpus Wikipédia. A titre informatif, le Tableau 3.12 présente les poids obtenus lors des différentes évaluations menées.

| Concept Level  | Pairs Number | Mean PA Single | PA Aggregated | PA Trained   | Imp.from PA Aggregated | Imp.from Mean PA |
|----------------|--------------|----------------|---------------|--------------|------------------------|------------------|
| Level 1        | 44.3         | 0.212          | 0.278         | 0.383        | 38%                    | 81%              |
| Level 2        | 11           | 0.222          | 0.242         | 0.273        | 13%                    | 23%              |
| Level 3        | 11           | 0.222          | 0.242         | 0.242        | 0%                     | 9%               |
| Level 4        | 14.3         | 0.235          | 0.325         | 0.535        | 64%                    | 128%             |
| <b>Average</b> | <b>20.0</b>  | <b>0.223</b>   | <b>0.272</b>  | <b>0.358</b> | <b>32%</b>             | <b>61%</b>       |

Tableau 3.11 : Evaluation terme/concept avec le corpus Wikipédia

| Concept Level | Fold | Cbow    | C&W      | Glove   | Hpca     | Random projection | Tscca    |
|---------------|------|---------|----------|---------|----------|-------------------|----------|
| Level 1       | K=1  | 0.32062 | -0.05606 | 0.41549 | -0.76115 | 0.34954           | -0.14104 |
|               | K=2  | 0.39499 | -0.28613 | 0.35704 | -0.72868 | 0.32172           | 0.01214  |
|               | K=3  | 0.39179 | -0.21911 | 0.35737 | -0.75704 | 0.31202           | -0.01776 |
| Level 2       | K=1  | 0.7177  | -0.45601 | 0.1968  | -0.48403 | -0.03765          | -0.0503  |
|               | K=2  | 0.71349 | -0.45236 | 0.14504 | -0.50457 | -0.0344           | 0.09742  |
|               | K=3  | 0.65431 | -0.37285 | 0.3219  | -0.56849 | 0.03006           | -0.07188 |
| Level 3       | K=1  | 0.57867 | -0.51985 | 0.36174 | -0.51177 | 0.04573           | -0.00703 |
|               | K=2  | 0.72755 | -0.3941  | 0.20004 | -0.52417 | -0.02197          | 0.01002  |
|               | K=3  | 0.80724 | -0.28478 | 0.05906 | -0.47057 | 0.03102           | -0.23043 |
| Level 4       | K=1  | 0.50982 | -0.04045 | 0.29244 | -0.74307 | 0.26265           | -0.17827 |
|               | K=2  | 0.40793 | -0.16472 | 0.45305 | -0.70632 | 0.28486           | -0.14552 |
|               | K=3  | 0.50956 | -0.35011 | 0.34521 | -0.64958 | 0.25722           | -0.10241 |

Tableau 3.12 : Poids des 6 modèles

Les résultats indiquent une amélioration de 32% pour l'évaluation des paires correctes (terme/concept) avec le modèle agrégé pondéré par rapport au modèle agrégé sans poids. Une amélioration plus élevée de 61% par rapport à la moyenne des prédictions de paires correctes des 6 modèles est aussi obtenue. Nous soulignons tout de même que l'augmentation des performances amenée par les modèles agrégés, par rapport aux meilleurs modèles classiques (*e.g.* Cbow, Glove), n'est pas systématique. La performance prédictive globale est en effet dégradée ou similaire lorsque des niveaux abstraits sont considérés (*cf.* résultats Figure 3.5 pour les niveaux 1 à 3). Une amélioration des performances intéressante est cependant observée pour le niveau le moins abstrait (niveau 4) ; ce résultat est particulièrement important dans notre contexte d'utilisation des modèles car le niveau 4 correspond aux mises en correspondance généralement souhaitées lors de l'analyse des descriptions non contrôlées fournies par les évaluateurs.

### 3.2.1.3 Conclusion

Dans cette partie nous avons présenté une stratégie de conceptualisation des descriptions des évaluateurs qui peut être utilisée afin de transcrire les descriptions terminologiques non contrôlées en un ensemble de concepts définis dans une taxonomie. Cette stratégie de conceptualisation se base sur une approche de transcription directe consistant à mettre en correspondance chaque terme proposé par un évaluateur avec un concept de la taxonomie. De manière générale, l'étude menée à l'aide d'expériences portant sur l'analyse olfactive de vins souligne la difficulté de la tâche d'automatisation de la phase de conceptualisation, et cela même si l'on admet au préalable un postulat consistant à chercher à associer chaque terme à un concept, et en le traitant de façon isolée. L'approche proposée permet ainsi d'obtenir des performances à hauteur de 60% si l'on considère une évaluation booléenne stricte, *i.e.* de type

pourcentage de correspondances attendues trouvées. Cette performance, qui peut sembler faible, est tout de même à analyser en tenant compte du fait que certains appariements, bien que ne correspondant pas nécessairement à ceux attendus restent pertinents dans un contexte d'analyse global.

La stratégie de conceptualisation des descriptions non contrôlées proposée est amenée par l'utilisation de modèles de plongement sémantique de termes qui permettent d'évaluer la pertinence d'une mise en correspondance terme/concept en évaluant le degré de proximité sémantique entre un terme proposé par l'évaluateur et les termes correspondants aux labels des concepts d'intérêt pour l'analyse olfactive. Intuitivement, l'approche utilisée se base sur une stratégie de maximisation cherchant à distinguer le concept le plus approprié en maximisant la proximité sémantique entre le terme évalué et la représentation terminologique des concepts (via leurs labels). L'importance du choix de la technique de plongement sémantique – pierre angulaire de l'approche – a été soulignée, et vient justifier nos travaux portant sur l'évaluation et l'analyse des techniques de plongement pour notre contexte applicatif spécifique.

Nous avons ainsi réalisé un travail de comparaison des techniques de plongement dans le cadre de notre applicatif. Cette comparaison nous a permis de distinguer des modèles de plongement sémantique traditionnels qui semblent, de façon globale, plus adaptés à la réalisation des traitements visés en considération de la stratégie de projection proposée. Nous avons notamment souligné que les résultats obtenus sont en accord avec la plupart des évaluations des modèles menées pour d'autres tâches du TAL. Cette étude nous a aussi permis d'évaluer l'intérêt d'appliquer les méthodes de plongement sémantique sur des corpus qui semblent plus spécifiques au domaine. Bien que séduisante théoriquement, l'application d'une phase de restriction sur le corpus ne s'est cependant pas révélée concluante dans nos expérimentations. Les résultats observés soulignent en effet, qu'en pratique, il reste difficile de constituer des corpus spécifiques rivalisant avec des corpus généralistes dans le cas où le domaine traité est très spécifique et peu représenté dans le corpus sur lequel la restriction est appliquée. En complément, une étude approfondie des résultats obtenus par l'utilisation des différents modèles de plongement – et l'observation du fait que les erreurs observées diffèrent en fonction des modèles – nous a permis de mettre en évidence l'intérêt potentiel d'un couplage des différents modèles. En réponse à cela, nous avons proposé une étude de différentes stratégies de couplage agrégeant les résultats fournis par différents modèles traditionnels. Les évaluations empiriques que nous avons menées ont montré l'intérêt de coupler différents modèles en modulant l'importance de chaque modèle, lors de la mise en correspondance entre un terme spécifique et un concept spécifique.

La stratégie de conceptualisation des descriptions des évaluateurs étudiée dans cette section s'est voulue simple afin d'assurer un degré d'interopérabilité maximale en limitant tant que possible l'introduction de biais. Des approches plus complexes tenant compte notamment de contextes d'énonciation du terme peuvent être envisagées. Il serait en effet possible d'intégrer, par exemple, la prise en compte d'informations issues de l'analyse des descriptions intra et inter évaluateurs. Dans l'approche étudiée, les termes sont considérés indépendamment les uns des autres et de façon isolée entre les évaluateurs. Des liens d'évocation intra ou inter évaluateurs pourraient ainsi être considérés, à l'image de la prise en compte d'une sorte de probabilité conditionnelle – *e.g.*, si un évaluateur mentionne une odeur de miel, il y a de fortes chances qu'il mentionne par la suite une odeur de fleur – l'odeur de miel renvoie souvent à une odeur de fleur ; ou encore, si un évaluateur mentionne une odeur spécifique (*e.g.*, d'agrumes), il y a de fortes chances qu'un second évaluateur mentionne une odeur semblable, par exemple plus spécifique (*e.g.* citron). La prise en compte de ces informations supplémentaires nécessite de revisiter le modèle proposé. Conscient de ces limitations, nous avons exclu l'étude de modèles incluant ce type d'informations car nous considérons ici qu'un biais serait introduit avant la phase de résumé, en particulier si les évocations inter évaluateurs sont considérées.

La stratégie de conceptualisation introduite dans cette section se base sur le postulat que l'évaluateur a, dans l'absolu, utilisé un terme unique afin d'exprimer une évocation conceptuelle. Le prochain chapitre s'intéresse au problème relatif à la conceptualisation d'un ensemble de termes, *i.e.* à l'identification de l'évocation implicite mentionnée par l'ensemble de termes proposés par un même évaluateur.

### 3.2.2 Evocations complexes : conceptualisation via transcription implicite

La première contribution présentée précédemment propose une stratégie de conceptualisation d'une description non contrôlée qui se base sur la mise en correspondance d'un terme avec un concept unique de la taxonomie. Nous traitons maintenant le cas de la résolution d'évocations conceptuelles complexes exprimées au travers de l'énonciation de différents termes – nous ferons ici référence à la notion d'évocation implicite. L'idée est de rechercher une évocation conceptuelle unique exprimée par une association de termes afin d'envisager la prise en compte du cas où un évaluateur n'a pas été en mesure d'exprimer une évocation conceptuelle identifiable par l'énonciation d'un terme unique – à titre d'exemple, un évaluateur incapable d'identifier une odeur de *beurre de cacahuète* pourrait mentionner les termes *gras* et *cacahuète*. Cette section présente l'étude de différents modèles permettant de distinguer des évocations conceptuelles implicites à partir d'un ensemble de termes donné. Ce travail est détaillé dans l'étude (Harispe *et al.*, 2017).

#### 3.2.2.1 Stratégies globales et formalisation du problème

##### 3.2.2.1.1 Définition de la fonction de projection

La proposition étudiée dans cette section repose sur l'utilisation conjointe de modèles de plongement sémantique et de relations taxonomiques pour réaliser la projection de l'espace des termes vers celui des concepts. Par rapport à l'objectif considéré dans la section précédente, *i.e.* la mise en correspondance terme/concept, nous cherchons dans ce cas à définir une fonction de projection  $f$  qui permet de projeter un sac de mots  $T' \subset T$  dans un espace conceptuel défini par la taxonomie  $\mathcal{O} = (\mathcal{C}, \preceq)$  ; formellement :

$$f : \mathcal{P}(T) \rightarrow \mathcal{C}$$

Afin de définir cette fonction de projection, nous considérons une fonction score  $s$  qui permet d'évaluer la pertinence d'association d'un concept (évocation conceptuelle) à un ensemble de termes. Cette fonction score permet d'évaluer la proximité entre un ensemble de termes  $T' \subset T$  et chaque concept  $c \in \mathcal{C}$ .

$$s : \mathcal{P}(T) \times \mathcal{C} \rightarrow \mathbb{R}$$

En considération de cette fonction score, nous proposons la définition suivante de la fonction de projection :

$$f(T') = \underset{c \in \mathcal{C}}{\operatorname{argmax}} s(T', c)$$

Nous proposons et détaillons par la suite plusieurs démarches de modélisation de cette fonction.

##### 3.2.2.2 Modélisation de la fonction score

Cette section présente les différentes modélisations que nous avons explorées pour distinguer une liste classifiée d'évocations conceptuelles (classement sur les scores des concepts de la taxonomie) pour un ensemble de termes fournis.

Ces modèles se basent sur les représentations vectorielles des termes, une fonction d'étiquetage reliant les termes aux concepts étant fournie (*i.e.* labels). La fonction d'étiquetage définit l'ensemble des labels (termes) associés à un concept unique, *i.e.*,  $L: C \rightarrow \mathcal{P}(T)$ . Nous avons par exemple,  $L(\text{Fraise}) = \{\text{fraise}, \text{fruit rouge doux}, \dots\}$ .

Deux types de modèles sont présentés :

1. Modèle d'agrégation vectorielle (*VAM*) : l'objectif de ce modèle est de fusionner la *sémantique* d'un ensemble de termes en regroupant les représentations vectorielles associées aux termes ;
2. Modèle basé sur une représentation graphique (*GM*) : l'objectif est de détecter le concept le plus pertinent en utilisant une structure graphique définissant les liens d'interaction/association entre les termes, entre les concepts et entre les termes et les concepts.

Les deux modèles sont détaillés et analysés dans les sections suivantes.

### 3.2.2.2.1 Modèle d'agrégation vectorielle (*VAM*)

Le modèle d'agrégation vectorielle (*VAM*) s'appuie sur une stratégie qui se résume en trois étapes pour analyser un ensemble de termes  $T' \subset T$  :

- A. La première étape consiste à calculer le vecteur d'évocations conceptuelles  $\vec{t} \in \mathbb{R}^{|C|}$  pour chaque terme  $t \in T'$ . Le but de cette représentation est d'englober toutes les évocations conceptuelles potentiellement mentionnées par le terme  $t$  ;
- B. La seconde étape consiste à agréger les vecteurs d'évocations conceptuelles des termes de  $T'$ . L'association des termes de  $T'$  sera ainsi représentée sous la forme d'un vecteur de  $\mathbb{R}^{|C|}$  ;
- C. La dernière étape consiste à analyser le résultat d'agrégation afin de calculer une liste ordonnée des évocations conceptuelles.

**A - Calcul du vecteur d'évocation conceptuelle d'un terme.** Nous considérons que le degré d'évocation d'un terme à l'égard d'un concept est défini par la fonction de proximité  $\sigma_{TC}: T \times C \rightarrow [0,1]$  définie comme dans la section précédente *i.e.*,  $\sigma_{TC}(t, c) = \max_{l \in L(c)} \sigma_{TT}(t, l)$ . La relation envisagée entre un terme et un concept dépend uniquement de la relation sémantique qui peut être distinguée au niveau des termes. Autrement dit, la relation entre un terme et un concept se définit comme la proximité sémantique entre ce terme et les labels associés au concept ; proximité évaluée comme précédemment par la définition d'une fonction  $\sigma_{TT}$ . L'évocation conceptuelle d'un terme  $t \in T$  peut alors être définie par la fonction suivante :

$$\rho_{TC} : T \rightarrow \mathbb{R}^{|C|}$$

$$t \rightarrow \rho_{TC}(t) = [\sigma_{TC}(t, c_1), \dots, \sigma_{TC}(t, c_n)]$$

Ce vecteur représente les évocations conceptuelles potentielles d'un terme. Nous considérons qu'il représente une empreinte faisant référence à toutes les évocations conceptuelles auxquelles un terme pourrait se référer. A noter que nous ne considérons pas de phase de désambiguïsation – qui pourrait par exemple exploiter les fréquences d'usage des concepts – car l'applicatif visé n'est pas directement impacté par des problèmes liés au caractère polysémique des termes évoqués.

**B - Agrégation des vecteurs d'évocation conceptuelle des termes.** Plusieurs approches peuvent alors être considérées afin d'agréger les degrés d'évocation d'un ensemble de termes  $T' \subset T$ . Nous définissons la fonction d'agrégation générale par  $P_C : \mathcal{P}(T) \rightarrow \mathbb{R}^{|C|}$ . Deux définitions de la fonction d'agrégation  $P_C$  seront également prises en considération ; les deux définitions sont basées sur une agrégation élémentaire :

1.  $P_C^{min}(T') = \bigwedge_{t \in T'} \rho_{TC}(t)$  : définissant l'agrégation comme la valeur d'évocation minimale parmi tous les termes pour chaque concept (agrégation de type conjonctif) ;
2.  $P_C^{sum}(T') = \sum_{t \in T'} \rho_{TC}(t)$  : définissant l'agrégation comme la somme des évocations conceptuelles de tous les termes (agrégation de type compromis).

**C - Classement et critère de sélection.** Afin de définir un classement des scores pour une fonction d'agrégation donnée, nous proposons une démarche permettant d'apprécier plus finement l'interprétation des scores d'évocation des concepts. Nous considérons pour cela que la distribution des scores d'évocation des concepts et la taille du vocabulaire sont connues. La distribution des concepts est estimée en calculant les représentations conceptuelles  $P_C$  associées à un ensemble d'échantillons de termes de taille  $|T'|$  tirés de manière aléatoire.

Pour établir le classement des concepts, nous avons introduit une fonction de score qui tient compte de la moyenne  $\mu$  et de l'écart type  $\sigma$  de la distribution de concepts.

$$s_{c_i}(T') = \frac{P_C(T')_i - \mu_{c_i}}{\sigma_{c_i}}$$

$\mu_{c_i}$  et  $\sigma_{c_i}$  désignent respectivement la médiane et l'écart type des valeurs d'évocation pour le concept  $c_i$  calculés au cours du processus d'échantillonnage associé aux échantillons de taille  $|T'|$ . La définition de cette fonction de score permet de définir le modèle complet permettant de distinguer les concepts les plus pertinents en considération d'un ensemble de termes (cf. B).

### 3.2.2.2 Modèle basé sur une représentation graphique (GM)

Le modèle basé sur représentation graphique (GM) est fondé sur une stratégie de propagation dans un graphe visant à modéliser les relations entre les termes, entre les concepts et entre les termes et les concepts. C'est alors l'analyse menée sur ce graphe qui permettra de distinguer les concepts les plus pertinents pour un ensemble de termes donné dans le processus d'évocation.

- **Modèle graphique**

Nous considérons pour ce modèle un graphe orienté pondéré  $G = (V, E)$  avec  $V = T \cup C$  et  $E \subseteq V \times V$ . Trois types de relations peuvent alors être distingués :

1. Relation entre deux termes : la relation entre les termes est pondérée par la mesure  $\sigma_{TT}$  qui permet de modéliser l'intensité de la relation entre les termes. Le poids normalisé du lien entre deux termes  $(t, t') \in E$  est défini par une fonction  $w_{TT} : T \times T \rightarrow [0,1]$  :

$$w_{TT}(t, t') = \frac{\sigma_{TT}(t, t')}{\sum_{t'' \in T} \sigma_{TT}(t, t')}$$

Cette fonction de pondération de la relation entre deux termes vise à normaliser la distribution des score  $\sigma_{TT}$  qui peut différer d'une manière considérable entre les termes – certains termes étant en quelque sorte plus centraux que d'autres.

2. Relation entre deux concepts : les relations entre les concepts sont données par l'ordre partiel qui organise les concepts de la taxonomie  $\mathcal{O}$ . La relation entre les concepts se définit comme suit : 1) avec  $E_{CC} \subseteq C \times C$ , construction d'un graphe  $G' = (C, E_{CC})$  à partir de la taxonomie  $\mathcal{O}$  en considérant que  $(c, c') \in E_{CC}$  si  $c \preccurlyeq c'$ ; 2) appliquer la réduction transitive à  $G'$ ; 3) Le poids des relations entre les concepts est fourni par une mesure de similarité sémantique  $\sigma_{CC}$  (section 3.1.2.1); 4) pondération des relations.

$$w_{CC}(c, c') = \frac{\sigma_{CC}(c, c')}{\sum_{c'' \in C | (c, c'') \in E_{CC}} \sigma_{CC}(c, c')}$$

Cette fonction de pondération des relations définies dans  $G'$  normalise le score  $\sigma_{CC}$ .

3. Relation entre les concepts et les termes : cette relation est donnée par la fonction d'étiquetage des concepts  $L$ . Avec  $c \in C$ ,  $t \in T$ , on considère un lien  $(t, c) \in E$  et  $(c, t) \in E$  si  $t \in L(c)$ , *i.e.*, si le terme  $t$  se réfère au concept  $c$ .

- **Modèle de propagation**

Nous présentons maintenant comment le modèle de propagation permet d'exploiter le graphe pour distinguer les concepts qui semblent être évoqués par un ensemble de termes.

Considérons un ensemble donné de termes  $T' \subset T$ . Le modèle de propagation adopté pour distinguer les concepts pertinents est donné par l'algorithme 1 ; la procédure de propagation est détaillée par l'algorithme 2. La stratégie globale de l'algorithme 1 vise à :

1. Calculer le degré d'évocation de chaque terme de  $T' \subset T$  (lignes 3-9) en propageant une quantité – assimilable à une intensité d'évocation - de chaque nœud de l'ensemble  $T'$  (ligne 7).
2. Agréger ces résultats afin de calculer, pour l'ensemble des termes  $T' \subset T$ , les scores d'évocation pour chaque concept (lignes 10-13).

---

**Algorithme 1** : algorithme de propagation

**Données** : le graphe  $G = (V, E)$ , l'ensemble des termes  $T' \subset T$  avec  $|T'| \ll |T|$ ,  $\varepsilon$  : critère d'arrêt

**Résultat** : une structure de données stockant la pertinence de chaque concept.

---

```

1 query_term_evocation_map ← map();
2 concept_score ← map();
3 pour t ∈ T' faire
4   ev_map ← map();
5   visited_node ← {};
6   score ← 1;
7   propagate(t, visited_node, 1, ev_map, score, ε); // cf. Algorithme 2
8   query_term_evocation_map[t] ← ev_map;
9 fin
10 pour c ∈ C faire
11   concept_score[c] ← aggregate(c, query_term_evocation_map);
12   // fonction d'agrégation : somme, moyenne
13 fin
14 retourner concept_score

```

---

Les structures de données utilisées pour stocker les résultats sont initialisées au début de l'algorithme 1 (lignes 1-2). Le degré d'évocation du terme  $t$  à l'égard des concepts est stocké dans un dictionnaire, cf.

$query\_term\_evocation\_map[t]$ ;  $query\_term\_evocation\_map[t][c]$  correspondant au degré d'évocation du concept  $c$  par le terme  $t$ .

Nous calculons les évocations conceptuelles pour chaque terme de  $T'$  (lignes 3 à 9). Ensuite, ces résultats sont agrégés en utilisant une stratégie d'agrégation (lignes 10 à 13). La fonction *aggregate* peut se définir par exemple comme la somme ou la médiane des représentations conceptuelles de l'ensemble des termes de  $T'$ . Intuitivement, la médiane est utilisée pour exprimer le fait que nous voulons non seulement un score élevé, mais que nous voulons également que le score soit soutenu par une contribution partagée des termes de  $T'$ . Les détails de la propagation sont définis par l'algorithme 2. Le processus de propagation est défini à l'aide d'une procédure récursive visant à propager des valeurs en évitant les nœuds déjà traités. Selon le type de nœud en cours de traitement (terme ou concept), la propagation vise à étendre une quantité fixe d'information à d'autres nœuds (termes ou concepts). Lorsqu'un terme est traité (lignes 2 à 13), une quantité est propagée à tous les concepts qui pourraient être référencés (labélisés) par le terme. Si la quantité propagée est suffisamment importante, l'évocation est propagée à ces concepts (ligne 6). La propagation est également effectuée aux termes voisins en tenant compte de la distance entre les termes  $w_{TT}$  (lignes 8 à 11).

Lorsqu'un nœud concept est traité (ligne 13 à 22), la propagation est effectuée en considérant les étiquettes associées au concept. La propagation est également effectuée au niveau conceptuel, en tenant compte également de la similarité des concepts qui peut être calculée en analysant l'ordre topologique des concepts  $w_{CC}$ .

---

**Algorithme 2 :**

**Données :** un nœud  $v$  du graphe, un ensemble de nœuds visités  $S$ ,  $score$  : une valeur de score (à propager),  $qtem$  : un dictionnaire pour stocker les scores d'évocation conceptuelle

**Résultat :** mise à jour de vecteur d'évocation conceptuel

---

```

1  $S.add(v)$  ;
2 si  $v \in T$  alors
3    $C' = \{c \in C | v \in L(c)\}$  ;
4   pour  $c \in C'$  faire
5      $qtem[c] \leftarrow qtem[c] + score$  ;
6     si  $score \geq \epsilon$  and  $c \notin S$  alors  $propagate(c, score, \epsilon)$  ;
7   fin
8   pour  $t \in T$  faire
9      $p\_value \leftarrow w_{TT}(t, v) \times score$  ;
10    si  $p\_value > \epsilon$  and  $t \notin S$  alors  $propagate(t, p\_value \times synDecFactor, \epsilon)$  ;
11  fin
12 fin
13 sinon
14 //  $v \in C$ 
15 pour  $t \in L(v)$  faire
16   si  $score \geq \epsilon$  and  $t \notin S$  alors  $propagate(t, score, \epsilon)$  ;
17 fin
18 pour  $c \in \{c \in C | (v, c) \in E \vee (c, v) \in E\}$  faire
19    $p\_value \leftarrow w_{CC}(v, c) \times score \times eSmoothingFactor$  ;
20   si  $p\_value \geq \epsilon$  and  $c \notin S$  alors  $propagate(c, p\_value, \epsilon)$  ;
21 fin
22 fin
23  $S.remove(v)$  ;

```

---

### 3.2.2.3 Evaluation et discussion

L'évaluation de l'approche proposée est basée sur un ensemble d'évocations attendues pour des ensembles de mots donnés. Dans notre application, nous avons évalué 220 entrées (ensemble de mots). Nous avons considéré pour cela des entrées générales qui ne font pas nécessairement référence au domaine de l'analyse sensorielle. Le Tableau 3.13 représente quelques exemples considérés extraits du jeu Taboo.

| Sac de mots                | Evocation attendue |
|----------------------------|--------------------|
| place, study, teacher      | School             |
| food, italy, round, tomato | Pizza              |
| yellow, fruits, monkeys    | Banana             |
| city, UK, capital          | London             |

Tableau 3.13 : Exemple des entrées d'évaluation

Pour chaque ensemble de mots, les évocations conceptuelles attendues sont liées à la représentation des connaissances WordNet. WordNet définit un ordre partiel sur des ensembles de synonymes fournissant (i) l'ensemble des *concepts (synsets)* et leur ordre partiel, ainsi que (ii) l'ensemble des labels associés aux concepts (fonction d'étiquetage  $L$ ). La liste de classement des concepts est calculée pour un ensemble  $C' \subset C$  correspondant aux réponses attendues pour toutes les requêtes évaluées  $|C'| = 198$ . Les implémentations de la mesure  $\sigma_{CC}$  ont été réalisées à l'aide de la SML (Semantic Measures Library)<sup>6</sup> (Harispe *et al.*, 2013). La fonction  $\sigma_{TT}$  utilisée dans ces expériences utilise le modèle Glove (Global Vectors for Word Representation) (Pennington *et al.*, 2014). La performance des deux approches est évaluée en calculant le nombre de requêtes pour lesquelles la réponse attendue est fournie parmi les  $k$  premiers résultats ( $k$  premiers classements). Le paramètre  $k$  prend les valeurs 1, 2, 3, 5, 10 ou 20.

Six modèles ont été évalués :

1. Deux modèles d'agrégation vectorielle (*VAM*) : le premier modèle *VAM – MIN* définit l'agrégation comme étant la valeur d'évocation minimale parmi tous les termes pour chaque concept et pour le second modèle *VAM – SUM*, l'agrégation est définie comme étant la somme des évocations conceptuelles de tous les termes.
2. Quatre définitions du modèle graphique (*GM*) :
  - a. deux stratégies utilisent une approche d'agrégation basée sur la médiane. Le modèle *GM – MEDIAN – KB* utilise une propagation au niveau conceptuel contrairement au modèle *GM – MEDIAN*.
  - b. deux stratégies utilisant une approche d'agrégation basée sur la somme. Le modèle *GM – SUM – KB* utilise une propagation au niveau conceptuel contrairement au modèle *GM – SUM*.

| Approche k=  | 1           | 2           | 5           | 10          | 20          | 50          | 100        |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|
| VAM_MIN      | 0.17        | 0.25        | 0.34        | 0.41        | 0.51        | 0.67        | 0.81       |
| VAM_SUM      | <b>0.28</b> | <b>0.39</b> | 0.51        | 0.6         | 0.66        | 0.8         | 0.86       |
| GM_MEDIAN    | 0.20        | 0.29        | 0.46        | 0.54        | 0.59        | 0.66        | 0.74       |
| GM_SUM       | 0.18        | 0.29        | 0.47        | 0.59        | 0.65        | 0.84        | 0.88       |
| GM_MEDIAN_KB | 0.24        | 0.35        | 0.49        | 0.58        | 0.64        | 0.74        | 0.81       |
| GM_SUM_KB    | 0.23        | 0.32        | <b>0.55</b> | <b>0.63</b> | <b>0.72</b> | <b>0.86</b> | <b>0.9</b> |

Tableau 3.14 : Résultats d'évaluation

<sup>6</sup> <http://www.semantic-measures-library.org>

Les résultats montrent que pour  $k = 1$  et  $k = 2$ , la meilleure performance est obtenue en utilisant le modèle d'agrégation vectorielle *VAM – SUM*. Cette approche ne prend pas en compte l'information fournie par l'ordre partiel des concepts. La performance des modèles dépend également de la stratégie d'agrégation utilisée pour le modèle *VAM* ; nous observons notamment qu'en utilisant la stratégie d'agrégation minimale (*VAM – MIN*) la performance du modèle diminue fortement.

Pour le modèle graphique (*GM*), les résultats mettent en évidence l'importance de l'information taxonomique pour chercher l'évocation conceptuelle d'un ensemble de termes. L'incorporation de l'information fournie par la taxonomie entraîne en effet une augmentation significative de la performance. Pour  $k > 2$ , les meilleures performances sont réalisées à l'aide du modèle basé sur une représentation graphique (*GM*) qui profite de l'information taxonomique.

### 3.3 Conclusion

Ce chapitre s'intéresse au problème de conceptualisation d'une description olfactive. Après avoir présenté différentes techniques de représentation des connaissances utilisables pour modéliser de la connaissance de domaine, et ainsi exprimer le type d'espaces conceptuels exploité en analyse sensorielle, nous avons introduit la notion de plongement sémantique terminologique (*embedding*) centrale pour la construction de représentations vectorielles permettant de manipuler des termes au travers de leur sémantique. Nous avons par la suite introduit comment ces représentations des connaissances et techniques de plongement sémantique terminologique peuvent être utilisées pour comparer des paires de termes ou de concepts.

Nous avons par la suite proposé un modèle permettant de conceptualiser un ensemble de termes par maximisation des évocations conceptuelles directes observées en considérant les termes indépendamment les uns des autres. Du fait de l'importance centrale des techniques de plongement afin d'assurer la performance du modèle, une étude approfondie des méthodes de projection a été effectuée pour le contexte d'utilisation relatif à l'analyse sensorielle considéré dans nos travaux. Un modèle général d'agrégation a aussi été proposé et évalué pour répondre à certaines limites rencontrées suite à l'utilisation des approches traditionnelles. Ces travaux ont permis la mise en place d'une approche globale de performance respectable pour la conceptualisation d'une description olfactive non contrôlée proposée par un évaluateur unique.

Nous avons par la suite proposé une étude portant sur la définition et l'analyse de modèles permettant de considérer le cas d'évocations implicites complexes pour lesquelles plusieurs termes méritent d'être considérés afin de détecter une évocation conceptuelle évoquée de façon sous-jacente. Ces modèles peuvent être utilisés afin de complexifier l'approche de conceptualisation d'une description non contrôlée proposée dans ce chapitre. Nous préférons cependant dans le prochain chapitre nous attarder sur la problématique de synthèse de différentes conceptualisations de descriptions non contrôlées proposées par un panel d'évaluateurs.

## Chapitre 4 Synthèse de Descriptions Conceptualisées

---

Le chapitre précédent a présenté une approche permettant la conceptualisation de descriptions exprimées sous la forme d'un ensemble de termes ambigus. Lors de l'analyse non orientée d'une odeur, cette approche permet d'exprimer chaque description à l'aide de descripteurs conceptuels d'intérêt pour le domaine d'étude de l'odeur analysée. Ce chapitre s'intéresse à la synthèse d'un ensemble de descriptions conceptuelles associées à une même odeur en tenant compte de la diversité des descriptions qui lui sont associées. En effet, l'objectif global visé par l'approche non orientée et automatisée que nous étudions est la synthèse conceptuelle de l'ensemble des descriptions proposées par les évaluateurs – *i.e.* l'obtention d'une annotation conceptuelle qui résume les profils olfactifs calculés (comme proposé au chapitre précédent) à partir des évocations proposées par les évaluateurs. Il nous reste alors à définir une approche automatisée permettant de synthétiser l'ensemble des descriptions conceptualisées. Cette approche doit être capable de traiter l'information globale véhiculée par l'ensemble des descriptions, et requiert pour cela une analyse fine de la sémantique des descripteurs conceptuels à partir desquels celles-ci sont exprimées.

La définition, même « manuelle », d'un résumé conceptuel caractérisant correctement une odeur à partir d'un ensemble de descriptions (conceptuelles) individuelles est cependant une tâche difficile. Elle nécessite une forte expertise dans le domaine de l'analyse sensorielle afin (i) d'apprécier les liens sémantiques entre les différents descripteurs – qu'ils soient contrôlés ou non –, (ii) d'agréger l'information véhiculée par les différents descripteurs, et (iii) de proposer une restitution fidèle de cette information qui soit pertinente pour le métier. L'objectif est en effet d'obtenir un résumé synthétique, cohérent, et représentatif de l'ensemble de l'information évoquée par les évaluateurs. Le résumé doit alors correspondre à une annotation conceptuelle collective (résumé conceptuel) qui reprend de façon concise les principales informations fournies par l'ensemble des évaluateurs (l'équivalent conceptuel d'une analyse en composantes principales). Pour cela le résumé doit en particulier reprendre les éléments les plus pertinents et informatifs en éliminant les évocations redondantes. Prérequis à la définition d'une approche automatisée de synthèse conceptuelle, notre étude nous amènera alors à nous attarder sur le problème complexe de la formalisation de la notion de résumé étudiée dans nos travaux. En particulier, il nous faudra gérer le dilemme suivant : à trop abstraire, on perd de l'information ; mais à être trop spécifique, on perd le critère de synthèse.

Ce chapitre est organisé comme suit. Nous proposons dans un premier temps une discussion générale sur la problématique de résumé ainsi qu'un état de l'art présentant différentes approches de TAL proposées pour résumer du contenu à caractère sémantique (*e.g.* textes, ensembles de concepts). Nous présentons et évaluons dans un second temps une nouvelle approche de résumé sémantique permettant de synthétiser un ensemble de descripteurs conceptuels définis dans une taxonomie. Pour finir, une discussion des résultats obtenus, des limites de l'approche, et des perspectives envisageables conclut ce chapitre.

## 4.1 Etat de l'art sur le résumé automatique et problématiques connexes

La capacité d'abstraire des observations spécifiques en distinguant des motifs (*patterns*) est centrale pour la mise en place de nombreux processus cognitifs complexes. C'est en effet une des capacités essentielles de tout agent intelligent qui analyse finement des contextes en vue d'inférer de la connaissance d'intérêt potentiellement générale et transposable. La capacité d'abstraction, essentielle dans toute phase d'apprentissage, est alors étroitement liée au processus d'analyse catégorielle d'observations basé sur une prédisposition à analyser leurs caractéristiques. Ainsi, un enfant qui se brûle avec une source de chaleur spécifique, *e.g.* un feu de bois, apprendra de son erreur en évitant dans le futur de s'exposer aux dangers associés (i) à ce feu spécifique, (ii) à d'autres feux de bois, mais aussi (iii) à d'autres sources de chaleur semblables, *e.g.* briquets, gazinières. Dans une certaine mesure, un lien étroit peut alors être fait entre la capacité d'abstraire des observations, *i.e.* de distinguer des motifs pertinents qui les caractérisent, et la capacité de les résumer, *i.e.* de faire la synthèse des enseignements qui peuvent en être tirés. La tâche de synthèse – ou de résumé des informations, nous utiliserons les deux termes de façon interchangeable – est alors assimilée à un processus cognitif complexe d'importance majeure dans de nombreux domaines relatifs à l'analyse de données et au développement d'intelligences artificielles. Différents travaux, en particulier dans des domaines liés à l'analyse de données, se sont alors intéressés à la définition d'approches automatisées de résumé de l'information, *e.g.* statistiques. Nous nous concentrerons plus particulièrement par la suite sur les approches de résumé adaptées à l'analyse de données à caractère sémantique (*e.g.* langage naturel, descriptions basées sur des représentations des connaissances).

L'avènement de la société de l'information nous obligeant à faire face quotidiennement aux flux massifs d'informations a tout particulièrement contribué à créer un besoin pressant d'approches rapides, fiables et automatisées de résumé. Dans le cas spécifique de données complexes à caractère sémantique (*e.g.* textes, discours) les résumés sont très souvent cependant encore le produit d'équipes d'experts (*e.g.* journalistes, éditorialistes dans le cas de journaux télévisés ou papiers). Cela s'explique notamment par la difficulté de développer des techniques automatisées capables de (i) distinguer des sources d'informations pertinentes et les informations importantes qu'elles communiquent, pour par la suite (ii) les résumer de façon condensée en préservant une cohérence et un style satisfaisant.

En TAL, de nombreux travaux proposent diverses approches de résumé automatique. Le but d'un résumé automatique est dans ce contexte informellement défini comme la production d'une représentation précise et synthétique couvrant l'essentiel des informations originales importantes véhiculées par différentes sources d'information – ici vues comme des (producteurs de) contenus à caractère sémantique. La nature de la source d'information peut en effet être très variée : document multimédia, e-mail, ensemble de mots etc., et faire référence à divers domaines : scientifique, littérature, juridique, etc.

Dans la littérature, les méthodes de création automatique de résumé sont généralement principalement classées en deux types d'approches : les approches par extraction et les approches par abstraction. Ces deux types d'approches sont présentés par la suite ; par souci de concision nous nous concentrerons principalement sur les travaux relatifs à l'analyse de textes et à l'indexation sémantique.

### 4.1.1 Génération automatique de résumé

#### 4.1.1.1 Résumé par extraction

L'approche de résumé par extraction est aujourd'hui la plus répandue. Elle propose de générer un résumé à partir de textes en identifiant les éléments (*e.g.* paragraphes, phrases, mots) les plus informatifs tout

en supprimant les éléments répétitifs et non informatifs. Quelle que soit la technique d'identification utilisée, la phase d'extraction est ainsi un processus de sélection des unités saillantes à partir de l'entrée originale (Mani *et al.*, 1999). L'approche par extraction produit alors généralement une contraction des textes basée sur une extraction de quelques phrases ; le résumé produit respecte la plupart du temps dans sa structure finale l'ordre d'apparition des phrases extraites.

Les approches de résumé par extraction se contentent souvent de repérer les éléments importants en analysant différentes caractéristiques telles que leur position dans le texte, la fréquence d'apparition des mots ou d'autres indicateurs statistiques, en exploitant dans certains cas des techniques d'apprentissage automatique (Turenne, 2000). Les méthodes existantes sont ainsi généralement basées sur des techniques d'extraction des éléments clés basées sur la définition d'une modalité d'évaluation de leur représentativité (Li *et al.*, 2006). Dans le cas d'une sélection de mots-clés (ou termes significatifs), l'évaluation peut par exemple se baser sur l'analyse de leur score  $tf * idf$ ,  $BM25$  ou tout autre mesure classiquement utilisée en théorie de l'information ou en recherche d'information (Robertson & Jones, 1976). Dans le cas de l'analyse de concepts (tels que des événements ou des entités), une analyse statistique exploitant des connaissances linguistiques peut aussi être utilisée.

Lorsqu'une approche par extraction basée sur une sélection de phrases est utilisée, le traitement peut être modélisé comme un problème de classification binaire de textes (phrases) – celui-ci s'intéresse alors à distinguer les phrases pertinentes de celles non pertinentes. En considération d'une base d'apprentissage identifiant les phrases importantes pour un ensemble de textes ; différentes techniques d'apprentissage supervisées applicables à la classification binaire de textes peuvent ainsi être utilisées. Une approche classique en TAL consisterait par exemple à : (i) appliquer une technique de représentation vectorielle des phrases (*e.g.* en considérant des caractéristiques, *features*, pertinentes) pour par la suite (ii) utiliser un classifieur (linéaire, Bayésienne probabiliste, SVM, réseau de neurones) afin de discriminer les phrases pertinentes (à partir de leur représentation) pour la construction du résumé ; des techniques plus récentes basées sur l'apprentissage de représentations (*i.e.* orientées *deep learning*) peuvent aussi être utilisées pour ce traitement. Dans la même veine, une approche alternative consisterait à traiter le problème comme une tâche de segmentation, avec ici aussi la possibilité de mobiliser une large gamme d'approches traditionnellement utilisées en TAL, *e.g.* modèles graphiques probabilistes tels que les Champs aléatoires conditionnels (CRF), ou encore les réseaux de neurones récurrents (LSTM par exemple). Cependant, quelle que soit la modélisation du problème choisie, l'utilisation d'approches semi-supervisées nécessite dans la plupart des cas des volumes importants de données labélisées pour la phase d'entraînement. Le manque de données labélisées et la difficulté de les obtenir explique très certainement que ces approches soient aujourd'hui délaissées pour le traitement de la tâche de résumé. Les approches mentionnées ci-avant, basées sur la définition et sur l'utilisation de fonctions *ad hoc* d'évaluation et de scores sont ainsi largement préférées.

(Lin & Bilmes, 2011) introduisent par exemple une approche de résumé de textes par extraction qui se base sur l'utilisation de fonctions monotones sous-modulaires. A partir d'un ensemble de phrases caractéristiques d'un document, ils utilisent les fonctions sous-modulaires pour distinguer un sous-ensemble de ces phrases qui résume au mieux l'ensemble de l'information évoquée dans le document. Pour cela, ils caractérisent la pertinence d'un résumé par deux propriétés fondamentales : la *fidélité* (la couverture) et la *diversité*. La propriété de fidélité consiste à mesurer la similarité entre le résumé candidat et le contenu du document traité ; la diversité quant à elle a pour objectif de favoriser la pluralité des éléments constitutifs du résumé - elle contribue alors implicitement à éviter les redondances. Le problème de la recherche d'un meilleur résumé pour un document est ainsi dans cette étude exprimé sous la forme d'un problème d'optimisation : la fonction objectif étant définie comme l'agrégation des fonctions sous-modulaires associées aux deux propriétés (fidélité et diversité).

(Tschiatsek *et al.*, 2014) proposent eux-aussi une famille de fonctions sous-modulaires pour résumer une collection d'images ; l'objectif de cette étude est de proposer une approche automatisée capable de sélectionner un sous-ensemble d'images représentatif d'une collection d'images. Ils ont pour cela eux aussi identifié la fidélité et la diversité comme les deux propriétés principales pour l'identification d'une bonne synthèse. La fidélité est une propriété qui modélise une bonne couverture de l'information véhiculée par le résumé ; ce qui signifie que tous les « concepts » distincts dans la collection d'images ont au moins un représentant dans le résumé. Par exemple, un résumé d'une collection de photos de montagnes et de plages devrait contenir des images de ces deux types de scènes. Le score de cette propriété sera alors d'autant plus élevé que la partie des images fournies qui sont couvertes par le résumé est grande. Par la suite, la diversité évalue le caractère redondant du résumé ; ainsi, le résumé ne doit pas contenir des images identiques ou trop proches. Cette propriété est représentée sous la forme d'une pénalité basée sur une analyse de la similarité des éléments qui composent le résumé. Plus la similarité entre les images qui composent le résumé est élevée, plus le score de la pénalité sera grand. La recherche du meilleur résumé d'images est alors ici aussi exprimée au travers de la résolution d'un problème d'optimisation reposant sur une fonction objectif définie comme la somme pondérée de ces deux propriétés. Pour valider leur approche, les auteurs ont comparé les résumés générés automatiquement avec les résumés subjectifs humains. Dans ce travail, tout comme dans l'approche de (Lin & Bilmes, 2011) mentionnée ci-avant et dans d'autres approches similaires (Sinha & Jain, 2011; Sinha *et al.*, 2011), les propriétés de fidélité et de diversité sont centrales. Dans notre approche, nous avons aussi tenu en compte de ces deux propriétés en adoptant une démarche de modélisation alternative.

Nous notons aussi que l'état de l'art relatif au résumé automatique propose certaines méthodes de compression qui étendent l'approche par extraction classique en ajoutant une phase de traitement sur les phrases extraites en vue de supprimer l'information redondante qui peut apparaître dans les contenus pertinents sélectionnés.

L'avantage principal de l'approche de résumé par extraction repose sur sa simplicité de mise en œuvre : les éléments du résumé (*e.g.* phrases) sont *simplement* extraits du contenu initial ce qui ne demande pas *de facto* de s'attaquer aux problèmes complexes de compréhension approfondie des contenus à traiter et de génération de contenus.

#### 4.1.1.2 Résumé par abstraction

Le résumé produit par abstraction est le résultat d'un processus complexe de compréhension du contenu source et de génération d'une forme condensée de ce dernier. L'approche se base en effet sur une première phase de construction d'une représentation de l'information véhiculée par le contenu. Cette représentation sera par la suite analysée et exploitée pour générer un texte traduisant de façon synthétique l'information jugée importante qui y est stockée. *A contrario* de l'approche par extraction, l'approche par abstraction est alors une approche génératrice capable de synthétiser de nouveaux contenus. Elle fait intervenir des processus complexes de compréhension et de génération de textes. Nettement plus ambitieuse que l'approche par extraction, l'approche par abstraction reste cependant, aujourd'hui encore, moins applicable en pratique du fait de ses faibles performances.

Un exemple d'approche de résumé automatique par abstraction a été proposé par (Endres-Niggemeyer & Wansorra, 2004). Cette approche, nommée SummIT-BMT (*Summarize It in Bone Marrow Transplantation*), a été proposée dans un contexte particulier de recherche d'information dans des textes de médecine relatif à la transplantation de moelles épinières. L'approche proposée exploite une ontologie de domaine au cours du traitement. Cette ontologie définit 4500 concepts portant sur la transplantation de la moelle épinière, et a été conçue pour aider à résumer adéquatement des textes liés

à ce domaine spécifique.<sup>7</sup> Le système repose sur le déclenchement d'agents qui exécutent certaines actions sur le texte en tenant compte de la connaissance précisée dans l'ontologie. Le comportement des agents correspond à un ensemble de stratégies humaines qui ont été identifiées lors d'analyses psycholinguistiques de l'activité de résumé.<sup>8</sup>

Ce chapitre s'intéresse à la définition d'une approche de résumé capable de synthétiser des ensembles de descripteurs contrôlés définis dans une taxonomie. L'objectif visé est, sur certains aspects, différent de celui traité par les approches de résumé de textes par extraction et par abstraction. Nous considérerons en effet que le problème traité peut être assimilé à un problème simplifié de résumé par abstraction en considérant (i) qu'une représentation des connaissances fournit une partie de la représentation interne qui mérite d'être considérée pour amener la phase de résumé, et (ii) que l'espace de restitution, *i.e.* l'espace des résumés, est nettement plus contraint dans le sens où le résumé visé prendra la forme – dans un premier temps – d'un ensemble de descripteurs contrôlés. On traitera alors d'une forme d'approche par abstraction pour laquelle les problèmes de représentation interne et de génération sont nettement simplifiés. Un problème similaire est traité dans l'état de l'art portant sur l'indexation et la synthèse d'annotations structurées. Des éléments de l'état de l'art associé à ces deux problématiques sont présentés par la suite.

#### 4.1.2 Annotation et indexation

Différents travaux s'intéressent à la définition de méthodes capables de suggérer automatiquement des annotations afin d'indexer une source d'information, *e.g.* textes, images, ou tout autre entité digitalisée (patients, maladies, gènes, catastrophes, clients, etc.). Dans le cadre d'une tâche d'annotation de corpus de documents, il s'agit par exemple de développer des outils et des méthodes permettant de caractériser de manière explicite ou implicite la sémantique des contenus documentaires en vue de faciliter leur recherche et leur exploitation. Dans ce contexte, plusieurs études portant sur la définition de la notion de résumé conceptuel (synthèse conceptuelle) ont été menées.

L'annotation automatique est une problématique liée aux domaines de l'extraction et de la recherche d'information traitée pour répondre à des problématiques aussi variées que le TAL et le traitement d'images. De nombreux travaux liés à cette thématique ont ainsi été proposés pour l'indexation de textes (Jimeno-Yepes *et al.*, 2012), pour l'annotation d'images (Carson *et al.*, 1999; Tschatschek *et al.*, 2014; Zhang *et al.*, 2014), ou encore pour l'annotation de documents audio (Turnbull *et al.*, 2008), de vidéos (Tseng *et al.*, 2008), ou de gènes (Beissbarth & Speed, 2004).

De manière générale, la tâche d'annotation automatique vise à synthétiser un ensemble d'informations fournies par une source par une annotation synthétique traduisant l'information véhiculée par la source. Tout comme les approches automatiques de résumé, l'annotation automatique cherche alors à identifier des éléments pertinents au regard de différents critères spécifiques et contextuels qui peuvent dans certains cas être définis par un utilisateur. L'annotation enrichit alors la source d'information initiale, *e.g.* les documents, avec des *étiquettes* suivant un besoin spécifique de description. Les étiquettes peuvent être de natures variées selon la tâche d'annotation, *e.g.* simples labels non structurés, ou des concepts définis dans une représentation des connaissances de type taxonomie par exemple, ou de type ontologie. Dans certains cas les labels seront générés lors du processus d'annotation.

---

<sup>7</sup> La notion de concept fait ici référence à une clause de Horn exprimée en Prolog.

<sup>8</sup> (Endres-Niggemeyer & Wansorra, 2004) proposent une approche de synthèse qui applique les résultats d'un travail empirique et expérimental antérieur sur la synthèse humaine. La synthèse humaine se base sur la recherche des passages les plus pertinents dans un texte en faisant référence par exemple à la table des matières.

En TAL, la littérature liée à l'annotation de documents est riche. À titre d'exemple, le modèle *LDA* (allocation latente de Dirichlet) proposé par (Blei *et al.*, 2003) permet d'identifier les thèmes mentionnés dans un corpus de textes, et de manière incidente caractérise les textes par ces thèmes. En effet, *LDA* regroupe les documents et produit simultanément un ensemble de thèmes (*topics*) à partir desquels les documents sont caractérisés. Le modèle *LDA* est un modèle probabiliste : chaque document est vu comme un mélange de thèmes (*mixture*), et chaque thème correspond à une distribution de probabilités sur l'ensemble des mots observés dans le corpus traité. *LDA* peut ainsi être vu comme un système d'indexation de documents au travers de thèmes inférés par l'analyse du corpus. Des extensions prenant en compte une structuration hiérarchique des thèmes ou des corrélations entre les thèmes ont également été proposées (Blei & Lafferty, 2006; Blei *et al.*, 2004). Par exemple, (Bartlett *et al.*, 2011) introduisent le modèle *HSLDA* (allocation latente de Dirichlet supervisée hiérarchiquement) qui permet d'inférer les étiquettes d'un document en considérant une hiérarchie d'étiquettes existante sous la forme d'un arbre. Le modèle infère alors une ou plusieurs étiquettes de sorte que, si une étiquette *l* est déduite comme pertinente pour un document, toutes les étiquettes de *l* à la racine de l'arbre sont également déduites comme étant pertinentes pour le document.

(Medelyan *et al.*, 2008; Ferragina & Scaiella, 2010) proposent eux de détecter les principaux thèmes d'un document en utilisant les noms d'articles Wikipedia. Ces approches permettent une annotation contrôlée en l'absence d'un vocabulaire créé manuellement. L'objectif est d'extraire les informations pertinentes d'un document texte (par exemple des phrases) puis d'identifier les articles Wikipédia qui correspondent aux documents afin d'attribuer ces noms d'articles comme thème d'indexation pour le document. L'évaluation de cette démarche d'indexation se fait en comparant les thèmes assignés automatiquement avec ceux choisis manuellement par des annotateurs humains.

Des approches pour le regroupement et l'annotation d'une collection de documents annotés par des concepts d'une taxonomie ont également été étudiées par (Fiorini, 2015). Dans ces travaux, le processus d'annotation des groupes de documents est considéré comme un problème d'optimisation continue des distances entre documents. La similitude entre les éléments internes d'un cluster est minimisée alors que la distance entre les éléments de deux groupes distincts est maximisée. La recherche d'une synthèse se fait alors par la génération d'un ensemble de concepts correspondant au barycentre conceptuel des différents ensembles à synthétiser en considération d'une mesure de proximité sémantique. Cette approche peut aussi être appliquée dans notre cas. Elle peut cependant être limitante lorsque différents points de vue distincts sont à résumer. De manière illustrée, une approche barycentrique définira que le concept Miel est pertinent pour résumer un groupe d'évaluateurs (A) exprimant une appréciation d'odeur de fruit et un groupe d'évaluateurs (B) exprimant une appréciation d'odeur de fleur, et cela bien que dans ce cas aucun évaluateur de A et de B n'ait jamais rapporté une appréciation de Miel. L'approche numérique adoptée par cette approche induit aussi dans certains cas la génération de résumés pouvant sembler contre-intuitifs car ne respectant pas, par définition, les attentes fréquemment exprimées en termes de résumé. L'approche génère en effet dans certaines conformations des résumés comportant des redondances d'information - cela a été observé empiriquement sur certaines données analysées lors de nos tests et s'explique théoriquement par l'approche gloutonne adoptée pour générer l'ensemble de concepts correspondant au barycentre. Des adaptations peuvent cependant être envisagées pour répondre à ce problème. Une limite supplémentaire de cette approche nous a cependant paru plus contraignante. Celle-ci repose sur le fait que l'approche se base sur une définition implicite de la qualité d'un résumé. En effet, un résumé est ici estimé comme de qualité s'il correspond au barycentre de l'ensemble des descriptions, ce qui, comme nous l'avons vu, peut dans certains cas être inapproprié. Il est alors, par exemple, impossible de critiquer un résumé restitué par l'approche en évaluant un éventuel degré de

performance sur des critères fréquemment d'intérêt pour la définition de résumés, *e.g.* concision, fidélité (couverture), diversité.

(Bairi *et al.*, 2015) traitent eux aussi du problème d'annotation d'un ensemble de documents. Ils supposent que les documents sont déjà associés à un ou plusieurs thèmes et proposent de trouver une synthèse cohérente pour un groupe de documents en utilisant une hiérarchie de thèmes structurés sous la forme d'un *DAG* (graphe orienté acyclique). Un lien direct peut être effectué avec notre problématique – les documents traités dans ces travaux correspondent à nos descriptions conceptualisées. Pour traiter cette problématique, les auteurs introduisent une famille de fonctions sous-modulaires afin d'identifier un ensemble approprié de thèmes en tirant parti de l'information fournie par le *DAG* de thèmes. Ils caractérisent ainsi la pertinence d'un résumé au travers de la définition d'un ensemble de propriétés souhaitables telles que la couverture (fidélité), la diversité, la spécificité, la clarté et la pertinence. Ces différentes propriétés permettent de choisir un ensemble de thèmes qui ne sont ni trop abstraits ni trop spécifiques. Elles permettent ainsi d'assurer que les thèmes sélectionnés sont alors clairs et pertinents pour représenter les documents analysés. L'objectif est en effet ici d'obtenir un résumé – représenté par un ensemble de thèmes – qui soit diversifié, pertinent, précis, spécifique et caractéristique d'un groupe de documents. Les fonctions sous-modulaires définies au regard de ces objectifs sont par la suite agrégées au travers d'une somme pondérée. La somme pondérée définit alors la fonction objectif du problème d'optimisation dont résulte le meilleur résumé conceptuel des documents. La propriété de couverture est centrale dans cette approche. En effet, dans cette étude, contrairement au raisonnement humain, aucun raisonnement approximatif n'est autorisé sur cette propriété. Ainsi l'approche proposée ne permet pas de considérer la fréquence des occurrences lors de la phase de résumé. A titre d'exemple, les thèmes associés aux documents à résumer ne peuvent pas être exclus du résumé final même dans le cas où ces derniers sont peu représentés (poids négligeables) – un aspect extrêmement limitant pour l'application de cette approche dans notre contexte applicatif.

Cette section souligne la richesse de l'état de l'art relatif à la synthèse automatisée d'information à caractère sémantique : une tâche située à la croisée des problématiques associées au résumé et à l'annotation/indexation automatique de contenus, notamment textuels. De nombreux travaux qui étudient (i) des approches de résumé automatique par extraction et par abstraction, et (ii) des approches d'annotation/indexation s'attardent ainsi sur des problématiques semblables ou connexes à la nôtre. De manière intéressante, bien que de nombreuses approches ne se basent ni sur une définition explicite de la notion de résumé, ni sur une évaluation explicite de la qualité des résumés, d'autres approches proposent de répondre à la problématique de synthèse d'information au travers de l'évaluation explicite de la qualité d'un résumé – qualité qui, nous l'avons souligné, est par nature subjective et contextuelle. Ces dernières approches adoptent une stratégie d'évaluation de la qualité des résumés par décomposition. Elles proposent ainsi d'attaquer le problème de résumé au travers de la définition de certaines propriétés souhaitées d'un résumé de qualité. Le problème est alors décomposé en une phase d'évaluation basée sur l'agrégation de l'appréciation de différentes caractéristiques : couverture (fidélité), spécificité, clarté... L'avantage d'une telle stratégie est de permettre une restitution analysable et évaluable par l'opérateur humain. L'opérateur sera ainsi capable, lorsqu'il interagit avec le système, d'explicitier son besoin par l'expression d'un objectif en termes de résumé, *i.e.* il lui sera alors possible de maîtriser d'autant plus la sémantique d'un résumé produit. L'opérateur peut par exemple exprimer qu'il souhaite un résumé plus abstrait, plus fidèle, plus concis ; il est aussi possible d'associer aux résumés restitués une appréciation relative à chacune des caractéristiques étudiées – en précisant par exemple à l'opérateur qu'un résumé spécifique implique une perte d'information, qui pourrait être ici potentiellement quantifiée.

Les travaux proposés par (Bairi *et al.*, 2015) et (Fiorini, 2015) sont très certainement les plus proches des nôtres dans le sens où ils s'intéressent à la génération de résumés d'ensembles de descripteurs (concepts) définis dans une taxonomie – ou structure semblable de type DAG. En réponse aux limites de ces approches en termes d'expressivité et de définition de la notion de qualité d'un résumé en particulier<sup>9</sup>, la section suivante est consacrée à la présentation d'une nouvelle approche de résumé automatique. Nous proposons en particulier un modèle capable de synthétiser l'information évoquée par un ensemble de concepts définis dans une taxonomie. Celle-ci adopte une stratégie par décomposition permettant d'exprimer la qualité d'un résumé afin de synthétiser l'information véhiculée par un ensemble de descripteurs contrôlés – correspondant dans notre cas au résultat de la conceptualisation des descriptions non contrôlées fournies par plusieurs évaluateurs lors de l'évaluation d'une odeur donnée.

## 4.2 Approche automatisée de résumé d'un ensemble de descriptions conceptualisées

Dans cette section, nous nous concentrons sur la définition d'une approche automatisée capable de résumer un ensemble de descriptions conceptuelles exprimées sous la forme d'ensembles de concepts. Même si le résumé est très souvent assimilé à un processus intuitif dans lequel nous souhaitons résumer autant d'informations que possible tout en gardant une représentation abstraite à la fois fidèle et informative, la pertinence d'un résumé est nécessairement contextuelle ce qui implique qu'aucune définition consensuelle *du meilleur résumé* ne puisse être considérée.

Nous proposons par la suite une démarche de modélisation de la notion de qualité d'un résumé conceptuel basée sur une approche de décomposition. Différentes propriétés (critères) d'un résumé d'un ensemble de descriptions conceptuelles sont données. Ces propriétés modélisent principalement, la quantité d'information évoquée par un résumé au regard de la perte d'information qu'il engendre, l'information redondante, l'information potentiellement ajoutée et finalement la concision et la spécificité des descripteurs qui composent le résumé. L'approche proposée s'attachera donc à rendre aussi explicite que possible l'impact de ces différents critères dans la construction automatique du résumé afin que l'utilisateur puisse interpréter le résultat et modifier ces impacts le cas échéant : cette interactivité permet alors de pallier le caractère contextuel *du meilleur résumé*.

Avant de discuter des aspects techniques de notre contribution, nous introduisons dans un premier temps les différentes notations et notions sur lesquelles s'appuie notre approche. Celles-ci seront par la suite utilisées pour définir formellement la notion de résumé que nous considérerons, ainsi que l'approche que nous proposons.

### 4.2.1 Notations

Nous définissons les fonctions suivantes :

---

<sup>9</sup> L'approche proposée par (Bairi *et al.*, 2015) ne permet pas de considérer la fréquence des occurrences lors de la phase de résumé. A titre d'exemple, les thèmes associés aux documents à résumer ne peuvent pas être exclus du résumé final même dans le cas où ces derniers ont un poids négligeables – notre approche intègre explicitement des masses liées aux concepts. L'approche proposée par (Fiorini, 2015) pour la recherche d'un résumé se fait par la génération d'un ensemble de concepts correspondant au barycentre conceptuel des différents ensembles à synthétiser en considération d'une mesure de proximité sémantique – cette approche ne respecte pas les critères de redondance et de couverture proposés par notre approche.

- Deux fonctions permettant de faire référence aux concepts ancêtres et descendants d'un ensemble de concepts :  $\mathcal{A}: \mathcal{P}(C) \rightarrow \mathcal{P}(C)$  et  $\mathcal{D}: \mathcal{P}(C) \rightarrow \mathcal{P}(C)$  respectivement telles que  $\forall C' \subseteq C$  :

$$\mathcal{A}(C') = \bigcup_{c \in C'} \{c' \mid c \preceq c'\},$$

$$\mathcal{D}(C') = \bigcup_{c \in C'} \{c' \mid c' \preceq c\},$$

- Le contenu informationnel ( $IC$ ) d'un concept associé à son degré de spécificité ; plusieurs définitions ont été proposées pour son calcul dans la littérature, *e.g.* (Seco *et al.*, 2004; Harispe *et al.*, 2015). Pour chaque fonction  $IC$ , nous avons :  $\forall (x, y) \in C^2, x \preceq y \Rightarrow IC(x) \geq IC(y)$  (un concept plus spécifique est plus informatif). Certaines formulations d'IC sont basées sur l'analyse des propriétés topologiques de l'ordre partiel des concepts (modèles intrinsèques). Un exemple de contenu d'information intrinsèque est celui proposé par (Seco *et al.*, 2004) :

$$IC(c) = 1 - \frac{\log(|\mathcal{D}(\{c\})|)}{|C|}$$

- Lorsque des occurrences sur des concepts sont disponibles, nous considérons la fonction de masse  $m : C \rightarrow [0,1]$  satisfaisant  $\sum_{c \in C} m(c) = 1$ ;  $m(c)$  correspond au nombre d'observations du concept  $c$ , *i.e.* au nombre d'évaluateurs qui ont explicitement mentionné le concept, divisé par le nombre d'observation total. Tout élément de  $C$  qui a une masse non nulle est appelé un élément focal.
- Les fonctions de croyance  $bel: C \rightarrow [0,1]$  et de plausibilité  $pl: C \rightarrow [0,1]$  sont définies comme suit :<sup>10</sup>

$$bel(c) = \sum_{x \preceq c} m(x)$$

$$pl(c) = \sum_{x \in C, \mathcal{D}(x) \cap \mathcal{D}(c) \neq \emptyset} m(x)$$

- A chaque évaluateur  $e_i$  est associé un ensemble de concepts (annotation conceptuelle) noté  $X_i \in \mathcal{P}(C)$ .
- On note par :  $\hat{X} = (X_1, X_2, \dots, X_n)$  la séquence d'annotations des  $n$  évaluateurs à résumer ; nous notons  $X$  l'ensemble des concepts fournis par les évaluateurs, *i.e.* :

$$X = \bigcup_{i=1}^n X_i$$

## 4.2.2 Définition formelle du problème

### 4.2.2.1 Définition générale et espaces considérés

L'objectif est de définir une fonction de synthèse notée  $f$  qui résume la séquence d'annotations conceptuelles  $\hat{X}$ .

$$f: \hat{X} \rightarrow \mathcal{P}(C)$$

<sup>10</sup> (Harispe, Imoussaten, *et al.*, 2015) proposent une définition motivée et détaillée de ces expressions.

Nous considérons que chaque résumé  $Y \in \mathcal{P}(C)$  doit respecter les propriétés suivantes :

- 1) Synthèse :  $|Y| \leq |X|$  (en pratique, nous aurons généralement  $|Y| \ll |X|$ ),
- 2) Fidélité :  $\forall y \in Y, \exists x \in X$  tel que  $x \preceq y$ ,
- 3) Non-redondance :  $\forall (x, y) \in Y^2, x \not\preceq y \wedge y \not\preceq x$ .

On note par  $\mathcal{S} \subseteq \mathcal{P}(C)$  les sous-ensembles de  $C$  respectant la propriété de non-redondance. On considère  $\mathcal{S}^X \subseteq \mathcal{S}$  le sous-ensemble des résumés de la séquence d'annotations  $\hat{X}$  qui respectent les deux propriétés de synthèse et de fidélité (vis-à-vis de  $X$ ).

On note par  $S$  la fonction de résumé qui synthétise l'information fournie par les  $n$  annotations conceptuelles  $\hat{X}$  par un seul résumé de  $\mathcal{S}^X$  :

$$S: \mathcal{P}(C)^n \rightarrow \mathcal{S}^X$$

En considération d'une séquence d'annotations conceptuelles  $\hat{X}$ , l'objectif visé est alors de définir une approche permettant de distinguer le résumé  $Y \in \mathcal{S}^X$  qui résume au mieux  $\hat{X}$ .

Nous l'avons déjà souligné, la notion de *meilleur résumé* ou de *résumé optimal* dépend du contexte et de la problématique ; on peut toutefois distinguer quelques notions et quantités qui peuvent être étudiées afin d'évaluer la pertinence d'un résumé ; celles-ci seront exploitées pour la définition de la fonction  $S$ . Dans notre contribution, la notion de couverture est centrale pour la définition de meilleur résumé. Cette notion permet de modéliser la quantité d'information couverte par le résumé.

Par définition, un résumé  $Y \in \mathcal{S}^X$  couvre la séquence d'annotations  $\hat{X}$  si tous les concepts mentionnés dans  $\hat{X}$  sont au moins abstraits par un concept du résumé  $Y$ , c'est-à-dire :  $\forall x \in X, \exists y \in Y$  tel que  $x \preceq y$ . On note les résumés qui couvrent  $\hat{X}$  par  $\mathcal{S}_{cov}^X$ . On note aussi par  $\mathcal{S}_{p-cov}^X = \mathcal{S}^X \setminus \mathcal{S}_{cov}^X$  l'ensemble des résumés couvrant partiellement  $X$ , c'est-à-dire :  $\forall Y \in \mathcal{S}_{p-cov}^X, \exists X' \neq \emptyset$  et  $X' \subset X$  tel que  $Y \in \mathcal{S}_{cov}^{X'}$ . La Figure 4.1 illustre l'espace présenté.

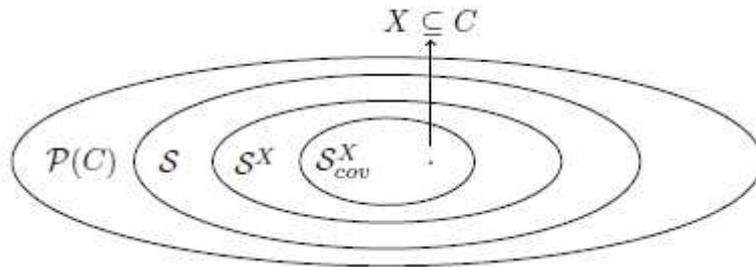


Figure 4.1 : Représentation graphique de l'espace étudié

#### 4.2.2.2 Formalisation de l'appréciation qualitative

Pour étudier le problème de recherche d'une synthèse conceptuelle  $Y \in \mathcal{S}^X$  pertinente pour une séquence de descriptions conceptuelles  $\hat{X} \in \mathcal{P}(C)^n$ , nous modélisons le problème sous la forme d'un problème d'optimisation. La fonction objectif est définie comme suit :

$$S(\hat{X}) = \operatorname{argmax}_{Y \in \mathcal{S}^X} (w_{psi} \cdot \Psi(Y, \hat{X}) - w_{loss} \cdot L(Y, \hat{X})) \quad (4.1)$$

avec,

$$L(Y, \hat{X}) = \Delta(Y, \hat{X}) + \lambda(Y) + \gamma(Y, \hat{X})$$

La fonction  $\Psi(Y, \hat{X})$  modélise la quantité d'information de  $\hat{X}$  résumée par  $Y$  et  $L(Y, \hat{X})$  la pénalité induite par l'abstraction du contenu d'information de  $X$  par  $Y$  avec :

- $\Delta(Y, \hat{X})$  : une fonction de pénalité en cas de perte, d'addition ou de distorsion de l'information véhiculée par  $\hat{X}$ .
- $\lambda(Y)$  : une fonction d'évaluation de la concision du résumé  $Y$ .
- $\gamma(Y, \hat{X})$  : une fonction qui permet d'exprimer des contraintes supplémentaires sur  $Y$  (e.g. permettant par exemple de ne retenir que les résumés couvrants).

#### 4.2.2.3 Quantité d'information abstraite $\Psi(Y, \hat{X})$

L'objectif de la fonction  $\Psi(Y, \hat{X})$  est d'estimer la quantité d'information véhiculée par  $\hat{X}$  qui est résumée par  $Y$ . On s'intéresse pour cela à l'étude de la couverture par  $Y$  des concepts mentionnés par  $\hat{X}$ . La quantité  $\Psi(Y, \hat{X})$  peut intuitivement être définie comme fonction de l'ensemble  $\mathcal{A}(X) \cap \mathcal{A}(Y)$  ; une expression spécifique peut par exemple être :

$$\Psi(Y, \hat{X}) = \sum_{c \in \mathcal{A}(X) \cap \mathcal{A}(Y)} w(c) \times IC(c)$$

Avec  $w(c) = \sum_{c' \in X \cap \mathcal{D}(\{c\})} m(c')$  ; puisque les éléments de  $X$  correspondent aux éléments focaux, nous avons  $w(c) = bel(c)$ .

La fonction  $w: C \rightarrow [0,1]$  est utilisée pour évaluer l'importance de chaque concept. Elle permet de mesurer le degré d'évocation selon les masses spécifiées par  $\hat{X}$  et intègre la prise en compte du caractère informatif des concepts considérés lors du calcul. La fonction  $\Psi(Y, \hat{X})$  est illustrée dans la Figure 4.2. Le cadre en pointillé de la figure représente l'ensemble des concepts évalués lors du calcul de  $\Psi(Y, \hat{X})$ .

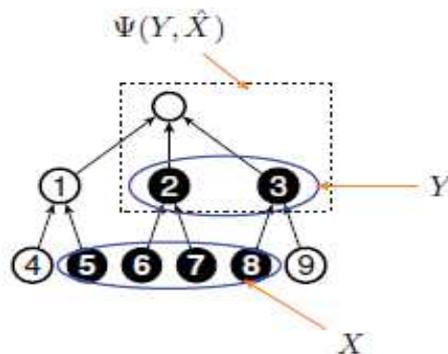


Figure 4.2 : Illustration de la quantité d'information abstraite  $\Psi(Y, \hat{X})$

La formulation de la fonction  $\Psi$  que nous venons de mentionner pose cependant souci. En vous référant à la Figure 4.2, considérez par exemple le cas où nous souhaitons résumer la séquence d'annotations  $\hat{X} = (\{6,7\})$ . En considérant la formulation de  $\Psi$  présentée ci-avant, nous aurons inévitablement  $\Psi(\{2\}, \hat{X}) < \Psi(\{6\}, \hat{X})$ . Ce résultat va à l'encontre du comportement souhaité, et cela indépendamment de la subjectivité associée à l'appréciation de la qualité d'un résumé. Informellement, nous souhaiterions diminuer la contribution de la masse associée au concept numéro 7 dans le calcul de

$\Psi(\{6\}, \hat{X})$ . En effet, bien que le résumé correspondant au seul concept 6 ne traduise pas explicitement l'évocation du concept 7, il bénéficie de la totalité de la masse associée à ce dernier lors du calcul de  $\Psi$  ; cela parce que le résumé fait implicitement référence au concept 2 qui lui généralise (couvre) le concept 7. Nous proposons alors de considérer la formulation alternative suivante :

$$bel_Y(c \in C) = \sum_{c' \in X \cap D(\{c\}) \cap D(Y)} m(c')$$

Et de considérer :

$$\Psi(Y, \hat{X}) = \sum_{c \in \mathcal{A}(X) \cap \mathcal{A}(Y)} bel_Y(c) \times IC(c)$$

#### 4.2.2.4 Modélisation des pénalités

Cette section est consacrée à la modélisation des différents facteurs de la pénalité  $L(Y, \hat{X})$  qui contribue à évaluer la pertinence et la cohérence d'un résumé.

$$L(Y, \hat{X}) = \Delta(Y, \hat{X}) + \lambda(Y) + \gamma(Y, \hat{X})$$

##### 4.2.2.4.1 Pénalité d'abstraction $\Delta(Y, \hat{X})$

Nous avons précédemment proposé un modèle d'estimation de la quantité d'information exacte véhiculée par  $\hat{X}$  qui est transmise par un résumé.<sup>11</sup> Pour critiquer la pertinence d'un résumé, il est également important de discuter des pénalités liées à la perte, l'addition et la distorsion de l'information induites par l'abstraction amenée par la phase de résumé.

Nous définissons ainsi la pénalité associée à la phase d'abstraction par :

$$\Delta(Y, \hat{X}) = H(\Delta^{E-}, \Delta^{E+}, \Delta^{P+}, \Delta^{P-}, \Delta^D)$$

Avec :

1.  $\Delta^{E-}$  Pénalité liée à la suppression de l'information exacte.
2.  $\Delta^{E+}$  Pénalité de l'ajout de l'information exacte.
3.  $\Delta^{P+}$  Pénalité de l'ajout d'informations plausibles.
4.  $\Delta^{P-}$  Pénalité de la suppression d'informations plausibles.
5.  $\Delta^D$  Pénalité de la distorsion de l'information (déformation de l'information).
6.  $H$  un opérateur d'agrégation.

Nous détaillons dans un premier temps la définition des pénalités  $\Delta^{E-}, \Delta^{E+}, \Delta^{P+}, \Delta^{P-}$  et  $\Delta^D$ .

---

<sup>11</sup> La notion d'information exacte fait référence à la notion de croyance (belief), la notion d'information plausible fera alors référence à la notion de plausibilité. Lorsqu'un évaluateur dit sentir une odeur d'agrume, nous pouvons dire (i) de manière exacte qu'il mentionne une odeur de fruit ; (ii) qu'il est aussi plausible qu'il mentionne une odeur d'orange, et (iii) qu'il ne mentionne pas une odeur de brûlé.

- **Modélisation de  $\Delta^{E-}$  et  $\Delta^{E+}$**

La fonction  $\Delta^{E-}$  modélise la quantité d'information exacte de  $\hat{X}$  qui n'est pas véhiculée à  $Y$ , i.e., suppression de l'information exacte ; cette quantité est intuitivement fonction de l'ensemble  $\mathcal{A}(X) \setminus \mathcal{A}(Y)$ . Nous proposons la définition suivante pour son calcul :

$$\Delta^{E-}(Y, \hat{X}) = \sum_{c \in \mathcal{A}(X) \setminus \mathcal{A}(Y)} bel(c) \times IC(c)$$

En raison de la définition d'un résumé et de la propriété de fidélité que nous considérons, tous les résumés correspondent à des ensembles de concepts pour lesquels aucune information exacte n'est ajoutée. Chaque concept d'un résumé évoque alors nécessairement au moins une notion fournie par les évaluateurs ; intuitivement  $\Delta^{E+}$  est associé à l'ensemble  $\mathcal{A}(Y) \setminus \mathcal{A}(X)$  qui du fait de la définition considérée est systématiquement égal à l'ensemble vide.

- **Modélisation de  $\Delta^{P+}$  et  $\Delta^{P-}$**

$\Delta^{P+}$  modélise la quantité d'information plausible véhiculée par  $Y$  qui n'est pas mentionnée par  $\hat{X}$ , i.e., ajout de l'information plausible. Nous définissons cette quantité comme fonction de l'ensemble  $\mathcal{D}(Y) \setminus \{\mathcal{D}(X) \cup \mathcal{A}(X)\}$ . De manière semblable  $\Delta^{P-}$  modélise la quantité d'information plausible véhiculée par  $\hat{X}$  qui n'est pas mentionnée par  $Y$ , i.e., suppression de l'information plausible ; quantité définie comme fonction de l'ensemble  $\mathcal{D}(X) \setminus \mathcal{D}(Y)$ .

Nous choisissons d'exprimer ces fonctions comme suit :

$$\Delta^{P+}(Y, \hat{X}) = \sum_{y \in \mathcal{D}(Y) \setminus \{\mathcal{D}(X) \cup \mathcal{A}(X)\}} pl(y) \times IC(y)$$

$$\Delta^{P-}(Y, \hat{X}) = \sum_{y \in \mathcal{D}(X) \setminus \mathcal{D}(Y)} pl(y) \times IC(y)$$

- **Modélisation de  $\Delta^D$**

L'objectif de  $\Delta^D$  est de pénaliser la *distorsion* d'information qui est faite par un résumé qui couvre partiellement l'ensemble  $X$ . Cette pénalité peut alors être définie comme fonction des éléments de  $X$  qui ne sont pas résumés, i.e. des concepts de  $X \setminus \mathcal{D}(Y)$  (par exemple, les concepts 1 et 5 de la Figure 4.3). On propose le modèle suivant pour estimer la distorsion de l'information :

$$\Delta^D(Y, \hat{X}) = \tau \sum_{x \in X \setminus \mathcal{D}(Y)} \sum_{x' \in \mathcal{A}(\{x\}) \setminus \mathcal{A}(Y)} (bel_{\{x\}}(x') \cdot IC(x'))$$

Le paramètre  $\tau$  est utilisé pour tenir compte de l'importance des concepts de  $X$  qui ne sont pas couverts par le résumé  $Y$  ; son calcul est et détaillé dans (Harispe *et al.*, 2017).

- **Agrégation des pénalités  $\Delta$**

Finalement, la pénalité d'abstraction est définie comme suit :

$$\Delta(Y, \hat{X}) = \delta_{E-} \Delta^{E-} + \delta_{P+} \Delta^{P+} + \delta_{P-} \Delta^{P-} + \delta_D \Delta^D$$

Avec  $\delta_{E-}$ ,  $\delta_{P+}$ ,  $\delta_{P-}$ ,  $\delta_D$  les paramètres utilisés pour définir l'importance de chaque facteur de pénalité d'abstraction ; une somme pondérée est ici utilisée comme opérateur d'agrégation.

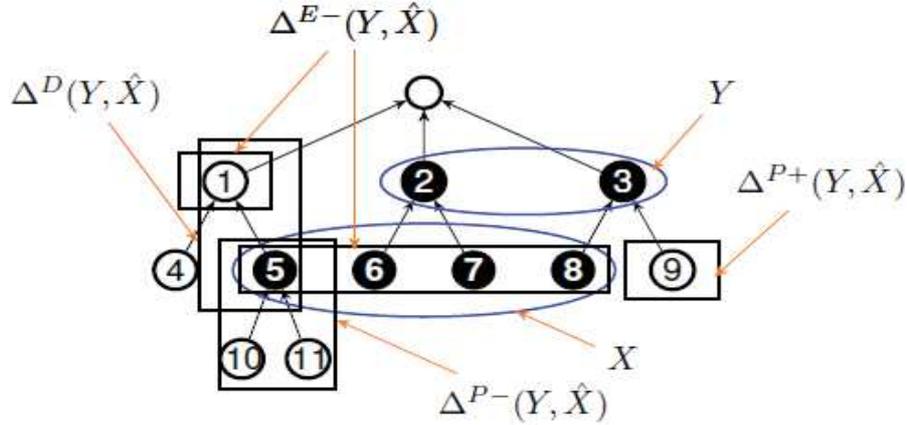


Figure 4.3 : Illustration des différents facteurs de pénalité

#### 4.2.2.4.2 Pénalités supplémentaires pour le résumé

L'objectif de la définition de pénalités supplémentaires est de permettre l'amélioration de la qualité de la synthèse conceptuelle en augmentant sa concision et ou en permettant la définition de contraintes supplémentaires sur les résumés attendus.

- **Modélisation de  $\lambda$  - Concision et redondance des pénalités**

$\lambda(Y) \in \mathbb{R}^+$  est une pénalité utilisée pour évaluer la concision d'un résumé en pénalisant les informations redondantes implicitement véhiculées par un résumé.

$$\lambda(Y) = \epsilon \sum_{y' \in \mathcal{A}(Y)} ((|\{y \in Y | y' \in \mathcal{A}(y)\}| - 1) \times IC(y'))$$

Avec cette expression, nous privilégions l'abstraction et donc la concision, c'est-à-dire les résumés qui ne résumant pas suffisamment l'information portée par  $\hat{X}$  seront automatiquement pénalisés. La pénalisation est définie de telle sorte que chaque notion abstraite qui est implicitement répétée plus d'une fois sera pénalisée le nombre de fois où les informations redondantes (implicites) apparaîtront. Le paramètre  $\epsilon$  peut donc indirectement être utilisé pour contrôler le nombre de descripteurs qui composent le résumé. Plus la valeur de  $\epsilon$  est grande, plus le résumé aura tendance à être abstrait.

- **Modélisation de  $\gamma$  - Contraintes additionnelles**

La fonction  $\gamma(Y, \hat{X})$  est utilisée pour exprimer des contraintes supplémentaires sur le résumé  $Y$ . La définition de cette fonction permet aussi de réduire l'espace des solutions. Les contraintes peuvent en effet être utilisées pour appliquer des restrictions sur la solution  $Y$  souhaitée, e.g., en ne gardant que les résumés qui couvrent tous les éléments de  $X$  ( $\forall x \in X, \exists y \in Y$  tel que  $x \preccurlyeq y$ ). Cette fonction peut être aussi définie pour imposer la taille du résumé ou pour préciser les types de concepts contenus dans le résumé selon les degrés de spécificité.

Un exemple de contrainte est proposé afin de ne considérer que les résumés couvrants :

$$\gamma(Y, \hat{X}) = \begin{cases} 0, & \text{si } \text{valid}(Y, X) \\ +\infty & \end{cases}$$

Avec  $\text{valid}(Y, X)$  définie comme vraie si  $\forall x \in X, \exists y \in Y$  tel que  $x \preceq y$ .

La décomposition de la fonction objectif en fonction de la quantité d'information reproduite par le résumé et la quantité d'information perdue, elle-même décomposée en critères selon la nature de la perte d'information, ainsi que l'agrégation sous la forme d'une somme pondérée doivent permettre à l'utilisateur de mieux comprendre l'impact des différents critères d'évaluation des résumés et ainsi d'agir le cas échéant sur tel ou tel paramètre de pondération afin d'augmenter ou de réduire le poids d'un critère.

### 4.2.3 Génération de résumés conceptuels

Nous avons jusque-là présenté la fonction d'évaluation proposée pour évaluer le score d'un résumé. Nous présentons maintenant les traitements algorithmiques proposés pour la recherche d'un résumé en se concentrant sur la construction de l'espace de recherche  $\mathcal{S}^X$ . Rappelons tout d'abord certaines informations relatives à l'espace de recherche : le nombre de partitions de  $C$  est  $2^{|C|}$ . Ainsi, dans le pire des cas et en considérant que toute représentation de connaissance a un concept abstrait généralisant tous les autres concepts (une racine, Thing), nous avons la règle théorique suivante  $\forall X \subset C, |\mathcal{S}^X| \leq 2^{|C|-1} + 1$ . Cependant, en pratique, la borne théorique est toujours nettement supérieure à la taille des espaces de recherche réels grâce aux contraintes définissant un résumé, et en particulier dans notre cas du fait des propriétés de fidélité et de redondance. Ces deux contraintes permettent en effet de réduire considérablement l'espace de recherche  $\mathcal{S}^X$ . Néanmoins, la taille de  $\mathcal{S}^X$  doit être prise en considération, puisqu'elle affecte largement le temps de calcul de notre approche. En effet, pour toute séquence d'annotations conceptuelles  $\hat{X}$  mentionnant un grand nombre de concepts ( $|X|$  grand) l'application d'une recherche itérative naïve sur  $\mathcal{S}^X$  n'est pas envisageable.

Dans cette section, nous proposons un algorithme pour construire l'espace  $\mathcal{S}^X$ . Nous proposons ensuite des restrictions qui peuvent être appliquées à  $\mathcal{S}^X$  afin de traiter les séquences d'annotations qui mentionnent un grand nombre de concepts.

#### 4.2.3.1 Construction de l'ensemble de résumés $\mathcal{S}^X$

Considérons un ensemble de concepts  $X \subset C$ , l'algorithme 1 définit comment générer l'espace de recherche  $\mathcal{S}^X$ , *i.e.* l'ensemble de résumés de la séquence d'annotations  $\hat{X}$ . Rappelons que  $\mathcal{S}^X = \mathcal{S}_{cov}^X \cup \mathcal{S}_{p-cov}^X$ . On note aussi :  $\mathcal{S}_{p-cov}^X = \bigcup_{X' \in \mathcal{P}(X) \setminus X} \mathcal{S}_{cov}^{X'}$ . Cela permet de reformuler le problème de la construction de  $\mathcal{S}^X$  comme suit :

$$\mathcal{S}^X = \bigcup_{X' \in \mathcal{P}(X)} \mathcal{S}_{cov}^{X'}$$

L'algorithme utilise une représentation de  $\mathcal{O} = (C, \preceq)$  sous la forme d'un graphe orienté acyclique nommé  $G = (C, E)$  avec  $E \subset C \times C$  ;  $(c_1, c_2) \in E$  signifie que  $c_2$  surclasse  $c_1$ , *i.e.*,  $c_1 < c_2$ . Pour des raisons d'optimisation, le graphe  $G$  doit être réduit en tenant compte de la transitivité des relations définissant  $\mathcal{O}$  (ligne 1). Nous appliquons pour cela une réduction transitive afin de supprimer tout arc  $(c_1, c_2) \in E$  s'il existe un chemin alternatif reliant  $c_1$  et  $c_2$ . Ainsi, pour chaque sous-ensemble  $X' \in \mathcal{P}(X)$ , l'ensemble  $\mathcal{S}^{X'}$  utilisé pour calculer  $\mathcal{S}^X$  est étendu en ajoutant tout résumé couvrant  $X'$  (ligne 4). Cette étape correspond à  $\mathcal{S}^{X'} = \mathcal{S}^{X'} \cup \mathcal{S}_{cov}^{X'}$ . A la fin de la procédure, l'algorithme permet d'obtenir  $\mathcal{S}^X = \mathcal{S}^X$ .

---

**Algorithme 1** : Génération de résumés  $\mathcal{S}^X \subseteq \mathcal{S}$  pour  $X \subset C$

---

- 1: réduction transitive de  $G$  – graphe de  $\mathcal{O} = (C, \preceq)$
  - 2:  $\mathcal{S}^X \leftarrow \emptyset$  ;
  - 3: **pour**  $X'$  in  $\mathcal{P}(X)$  **faire**
  - 4:  $\mathcal{S}^X = \mathcal{S}^X \cup \text{extend}_{cov}(G, X', \mathcal{S}^X)$  ;
  - 5: **fin de pour**
  - 6: **retourner**  $\mathcal{S}^X$  (équivalent à  $\mathcal{S}^X$ )
- 

Etant donné un ensemble de concepts  $X$  et un ensemble de résumés  $\mathcal{S}_r \subset \mathcal{S}$  respectant la propriété  $\forall Y \in \mathcal{S}_r, \mathcal{S}_{cov}^Y \subset \mathcal{S}_r$ , l'algorithme 2 calcule l'ensemble des résumés  $\mathcal{S}_{cov}^X$  non présents dans  $\mathcal{S}_r$ . En définissant  $\mathcal{S}_r = \emptyset$ , l'algorithme 2 calcule alors  $\mathcal{S}_{cov}^X$  c'est-à-dire :  $\mathcal{S}_{cov}^X = \text{extend}_{cov}(G, X, \emptyset)$  pour l'ensemble de concepts  $X$ .

L'idée de l'algorithme est simple, étant donné un ensemble de concepts  $X$ , l'algorithme va calculer récursivement les résumés couvrants pour chaque résumé de  $X$ . Ces résumés peuvent être obtenus en substituant un élément  $x \in X$  par une de ses abstractions, *i.e.* un de ses parents, ou un sous-ensemble de ses parents. La ligne 2 itère sur l'ensemble des éléments de  $X$ . La ligne 3 calcule les parents de l'élément  $c$ . Ensuite, l'objectif est de générer un résumé de  $X$  en remplaçant  $c$  par l'une de ses représentations abstraites. La ligne 4 permet d'itérer sur toutes les substitutions possibles de  $c$ . Pour chacune de ces abstractions, un résumé de  $X$  est construit en ajoutant la forme abstraite de  $c$  (*i.e.*  $P'_c$ ) et en supprimant chaque concept de  $X$  qui a été généralisé par un élément de  $P'_c$ . Cela garantit l'obtention d'un ensemble de concepts respectant la propriété de non redondance. La propriété de fidélité est assurée par le fait que chaque concept n'est remplacé que par l'une de ses représentations abstraites. La propriété de synthèse est testée à la ligne 6. Avant de considérer tout résumé, nous vérifions qu'il n'a pas été déjà traité (déjà dans  $\mathcal{S}_{cov}^X$ ) et s'il ne se trouve pas dans l'ensemble des résumés restreints (en  $\mathcal{S}_r$ ). Si ce n'est pas le cas, nous ajoutons le résumé actuel à l'ensemble des résumés et, en appliquant une procédure récursive, nous ajoutons également tous les résumés qui peuvent être générés à partir de ce dernier en excluant cette fois-ci tous les résumés déjà calculés (ligne 7).

---

**Algorithme 2** :  $\text{extend}_{cov}$  : étant donné un ensemble de concepts  $X$ , un ensemble de résumés  $\mathcal{S}_r \subset \mathcal{S}$  tel que  $\forall Y \in \mathcal{S}_r, \mathcal{S}_{cov}^Y \subset \mathcal{S}_r$  l'algorithme génère  $\mathcal{S}_{cov}^X = \mathcal{S}_{cov}^X \setminus \mathcal{S}_r$ , et avec  $\mathcal{S}_{cov}^X \subseteq \mathcal{S}_{cov}^X \subseteq \mathcal{S}^X$ . Par définition si  $\mathcal{S}_r = \emptyset$ , l'algorithme génère  $\mathcal{S}_{cov}^X$ . On considère que l'information définie dans le graphe  $G = (C, E)$  est ici accessible.

---

- 1:  $\mathcal{S}_{cov}^X \leftarrow \emptyset$  ;
  - 2: **pour**  $c$  in  $X$  **faire**
  - 3:  $P_c = \{p \mid \exists (c, p) \in E\}$  ; // parents de  $c$
  - 4: **pour**  $P'_c$  in  $\mathcal{P}(P_c)$  **faire**
  - 5:  $X' = X \setminus \{x \in X \mid P'_c \cap \mathcal{A}(\{x\}) \neq \emptyset\} \cup P'_c$  ;
  - 6: **si**  $|X'| < |X| \wedge X' \notin \mathcal{S}_r \wedge X' \notin \mathcal{S}_{cov}^X$  **alors**<sup>12</sup>
  - 7:  $\mathcal{S}_{cov}^X = \mathcal{S}_{cov}^X \cup \{X'\} \cup \text{extend}_{cov}(G, X', \mathcal{S}_{cov}^X \cup \mathcal{S}_r)$  ;
  - 8: **fin si**
  - 9: **fin de pour**
  - 10: **fin de pour**
  - 11: **retourner**  $\mathcal{S}_{cov}^X$
- 

L'algorithme 2 se base sur l'analyse d'un graphe dirigé acyclique composé d'une racine unique. La procédure récursive se base sur la substitution progressive des concepts par des formes toujours plus abstraites en évitant de traiter à nouveau des ensembles pour lesquels les résumés ont été calculés. En

<sup>12</sup> On note la réflexion associée à la restriction de la cardinalité, afin que pour tout résumé  $Y$  d'un ensemble de descripteurs  $X$  nous ayons  $|Y| \leq |X|$ . Cette restriction peut être contre-intuitive mais est cependant nécessaire dans le cas d'héritages multiples. En effet, dans ce cas, la substitution d'un concept par une de ses formes abstraites peut découler sur une augmentation de la cardinalité de l'ensemble correspondant résumé. Notre définition de la notion de résumé implique  $|Y| \leq |X|$  ; la pertinence de ce choix peut cependant être discutée dans certains cas.

considération de l'acyclicité du fait de l'ordre sur les concepts, nous avons alors la garantie que le traitement se termine, *i.e.* tout appel de la procédure  $extend_{cov}$  se terminera. Cela implique l'existence de la condition d'arrêt pour l'algorithme 1.

#### 4.2.3.2 Réduction de la complexité

Des réductions de temps de calcul peuvent être obtenues lors de la construction de  $\mathcal{S}^X$ . En considérant la procédure proposée, il est en effet possible d'obtenir  $\mathcal{S}^X$  en supprimant les concepts redondants de  $X$  (*i.e.* les concepts de  $X$  spécialisés par un concept de  $X$ ). La complexité globale du traitement étant nécessairement dépendante de la taille de  $\mathcal{S}^X$ , il peut dans certains cas être judicieux de réduire tant que possible sa taille. Différentes approches peuvent ainsi être proposées. Il est par exemple possible de supprimer les valeurs aberrantes, *e.g.* tout concept  $x \in X$  associé à une masse très faible (*e.g.* un concept observé une seule fois) peut être substitué par son concept abstrait (élément de  $(X) \setminus \{x\}$ ) qui a une croyance (bel) supérieure à celle de  $x$ . Cette idée peut être étendue pour réduire le nombre de résumés à évaluer en considérant l'hypothèse que tout résumé intéressant doit nécessairement résumer de l'information pour être intéressant. Conformément à cette hypothèse, nous pouvons exclure tout résumé  $Y \in \mathcal{S}^X$  pour lequel  $\exists Y' \in \mathcal{S}^X$  avec  $Y' \ll_{\mathcal{S}^X} Y$ ,  $|Y'| = |Y|$ . Cette restriction peut facilement être calculée avant l'application de l'Algorithme 1 ; il faut pour cela supprimer les concepts de  $G$  qui n'ont pas une croyance plus importante que celle des concepts qu'ils généralisent (*i.e.*, les concepts qui n'ont pas d'occurrence propre et dont un seul des concepts subsumés a une occurrence non nulle).

### 4.3 Expérimentations

Trois expérimentations sont proposées pour à la fois étendre nos discussions sur la notion de résumé et discuter la performance de l'approche automatique de résumé proposée. La première expérimentation a été définie pour l'étude des attentes en terme de résumé, et en particulier afin d'évaluer si une appréciation consensuelle peut être observée – question ouverte du fait du caractère subjectif associé à la notion de résumé. Nous avons pour cela considéré un jeu de données général à partir duquel différentes propositions de résumés ont été analysées par différents évaluateurs. Nous utiliserons aussi par la suite ces données dans une seconde expérimentation afin d'évaluer la performance de notre approche automatisée pour la génération de résumé. Pour finir, la troisième expérimentation proposée s'intéresse plus particulièrement à l'utilisation de l'approche de résumé définie dans un contexte d'analyse psychosensorielle.

#### 4.3.1 Analyse du consensus sur la notion de résumé

Pour différentes requêtes représentées par des ensembles de descripteurs contrôlés pondérés, la préférence en termes de résumé a été recueillie pour différents évaluateurs. Pour chaque requête, la préférence exprimée par un évaluateur correspond à un choix de résumé parmi un ensemble de résumés éligibles. Les données complètes correspondent aux résumés choisis pour 44 questions, pour lesquelles 4 à 6 choix de résumé étaient proposés. Deux connotations de poids sur les descripteurs étaient parfois proposées pour une même question. En moyenne, 43 choix de résumé, *i.e.* résultats d'évaluateurs, ont été recueillis par question. Un exemple de question est proposé ci-dessous :

| <b>Exemple de question</b> : citron (25%), banane (25%), orange (25%), pamplemousse (25%) |                                |
|---|--------------------------------|
| <b>Résumés proposés</b>   | <b>Fréquence d'évaluateurs</b> |
| Fruit   | 27                             |
| Agrume, Banane  | 15                             |
| Citron, Orange, Banane  | 0                              |
| Citron  | 0                              |
| Agrume  | 1                              |
| <b>Total</b>  | <b>43</b>                      |

Tableau 4.1 : Choix de résumé pour synthétiser quatre concepts de même intensité

Cette question porte sur le choix du résumé le plus pertinent pour résumer quatre évocations de même intensité : citron, banane, orange, pamplemousse ; quatre choix de résumé sont ainsi proposés. Les fréquences de sélection des résumés sélectionnés par les évaluateurs sont présentées. On observe dans cet exemple une préférence nette pour le choix des deux résumés couvrants (« Fruit », et « Agrume, Banane ») – nous rappelons que chaque évaluateur ne peut sélectionner qu'un seul résumé.

Nous souhaitons au travers de cette expérimentation nous intéresser à l'appréciation d'un consensus associé à la notion de résumé, *i.e.* les évaluateurs, ici sujets, se comportent-ils de manière similaire lors d'un processus cognitif consistant à résumer une information exprimable sous la forme d'un ensemble de descripteurs pondérés ?

Nous précisons que bien que les données aient naturellement été acquises au travers de requêtes et de choix de résumé exprimés en langage naturel, *i.e.* à l'aide de termes, les données sélectionnées sont assimilables à des descripteurs contrôlés car aucune liberté dans l'interprétation n'a été donnée aux évaluateurs. La Figure 4.4 présente les répartitions des choix des évaluateurs pour l'ensemble des résumés (cf. barre données complètes), et plus particulièrement pour les résumés pour lesquels les masses associées aux descripteurs étaient égales ou différentes (cf. barres poids égaux et différents respectivement). Le positionnement des évaluateurs est étudié en fonction :

- du fait que le choix sélectionné soit (1) ajoute de l'information, *i.e.* fait référence à un descripteur jamais explicitement ou implicitement mentionné dans l'ensemble à résumer, soit (2) contient une redondance d'information, *i.e.* deux concepts ordonnés apparaissent dans le choix proposé. On souligne que les choix correspondant à ces deux classes ne correspondent pas à des résumés si l'on considère la définition proposée dans ces travaux.
- d'une perte éventuelle d'information induite par le résumé, dans le sens où le résumé sélectionné n'est pas couvrant (3),
- du caractère plus ou moins abstrait d'un résumé couvrant (4-5)

Les classes sélectionnées (1 à 5) sont mutuellement exclusives, de façon à ce que chaque choix proposé par un évaluateur ne soit associé qu'à une seule et unique classe.

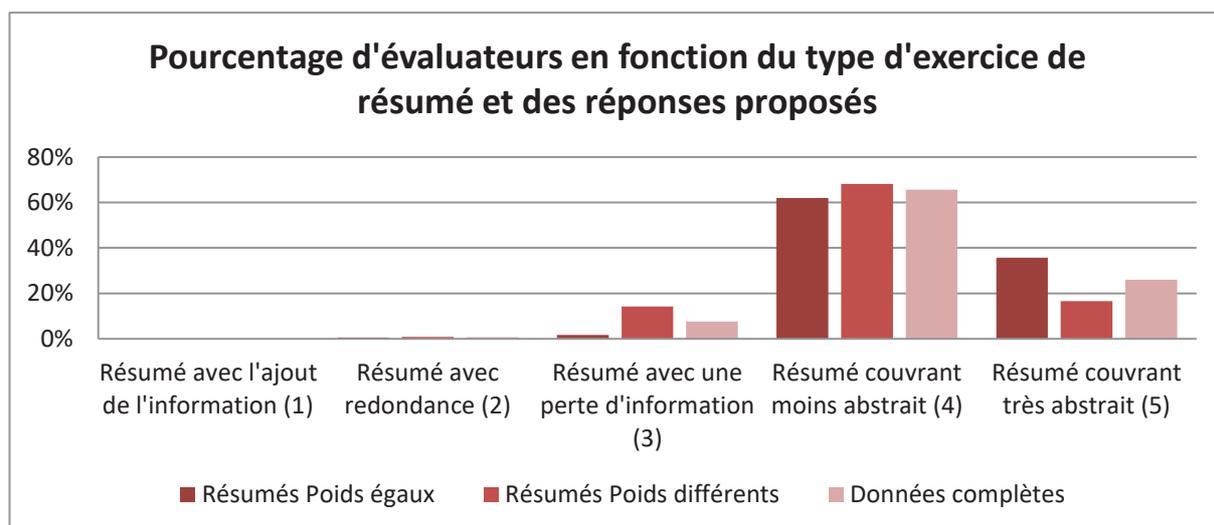


Figure 4.4 : Répartition des évaluateurs en fonction du type d'exercice

Trois conclusions principales peuvent être tirées de l'expérimentation menée :

1. La première conclusion porte sur les attentes relatives à la notion de résumé exprimée par les évaluateurs. Nous observons en effet, de façon intuitive, que la notion de résumé fait référence à la sélection de groupes de descripteurs n'induisant pas (i) d'ajout d'information au regard de l'information à traiter et (ii) de redondances dans la restitution proposée (les classes 1 et 2 sont ainsi marginales). L'observation de ce résultat, bien que sans réelle surprise, conforte naturellement la définition formelle de la notion de résumé considérée dans nos travaux.
2. La seconde conclusion porte sur la préférence des résumés couvrants, *i.e.* la très grande majorité des résumés sélectionnés pour les différentes questions étaient en effet couvrants lorsque les poids traduisant l'importance donnée aux descripteurs étaient répartis de façon uniforme sur les différents descripteurs. Dans ce cas, moins de 2% des évaluateurs choisissent un résumé induisant une perte d'information (dans le sens considéré ici). Nous notons ainsi l'accroche importante associée à l'évocation d'un descripteur. En effet, les choix effectués par les évaluateurs montrent qu'il leur est difficile d'exclure un descripteur lors de la phase de résumé – et cela même lorsque le pourcentage des masses associées au descripteur conservé n'excède pas 5% du total des masses. Autrement dit, des descripteurs qui semblent *a priori* en marge de l'analyse (très souvent assimilés à du bruit en première intention) sont dans de nombreux cas jugés importants afin d'assurer une restitution fidèle des observations sous la forme d'un résumé. En effet, le nombre d'évaluateurs ayant sélectionné un résumé associé à une perte d'information (classe 3) est de manière globale relativement faible, et cela même dans le cas spécifique où des différences de masses entre les descripteurs étaient proposées de manière à favoriser la sélection de tels résumés – dans ce cas seulement 14% des résumés étaient associés à une perte d'information.
3. La troisième conclusion amenée par notre analyse porte sur le niveau d'abstraction souhaité lors du processus de résumé. La plupart des questions contenaient en effet différents choix de résumés couvrants associés à différents niveaux d'abstraction, *e.g.* les choix préférés dans l'exemple mentionné (Tableau 4.1) sont tous les deux couvrants et de niveau d'abstraction différents (*e.g.* le choix « Fruit » abstrait le choix « Agrume, Banane »). Les données analysées montrent que parmi les choix sélectionnés, les résumés couvrants les moins abstraits étaient généralement préférés. Nous soulignons en effet la préférence des résumés appartenant à la classe 4 par rapport à ceux plus abstraits appartenant à la classe 5. Cette observation souligne la volonté de l'évaluateur, le plus souvent, de minimiser la perte d'information en minimisant le caractère abstrait du résumé. Cependant, de façon intéressante, nous observons aussi qu'un nombre non négligeable d'évaluateurs semblent dans certains cas préférer des résumés plus

abstrait ; cela peut notamment s'expliquer par le caractère plus explicite de certains résumés, *e.g.* si 15 fruits peuvent être résumés par Agrumes, Banane et Pomme, de nombreux évaluateurs préféreront considérer l'unique descripteur Fruit – cet aspect n'est pas considéré dans nos travaux car il est selon nous associée à la restitution ; il peut cependant avoir une importance fondamentale notamment dans certains contextes applicatifs (*e.g.* publicité).

En complément, le Tableau 4.2 présente les nombres d'évaluateurs associés aux résumés en fonction du classement des résumés en termes de choix. On observe ainsi que si l'on se concentre sur la sélection du meilleur résumé pour l'ensemble des questions proposées, seuls 48.5% des votes seraient recueillis. En se concentrant sur les 3 premiers choix pour chaque requête, nous observons que 88.62% des votes peuvent être recueillis. Ces résultats soulignent la variabilité non-négligeable sur les choix de résumé.

| N-ième meilleur résumé | Nombre d'évaluateurs associés au n-ième meilleur résumé (somme pour l'ensemble des questions + pourcentage total) | Nombre d'évaluateurs associés à un résumé de rang $x \leq n$ (somme pour l'ensemble des questions + pourcentage total) |
|------------------------|---|--|
| 1                      | 891 (48.5%)   | 891 (48.5%)  |
| 2                      | 468 (25.48%)  | 1359 (73.98%)  |
| 3                      | 269 (14.64%)  | 1628 (88.62%)  |
| 4                      | 153 (8.33%)   | 1781 (96.95%)  |
| 5                      | 55 (3%)   | 1836 (99.95%)  |
| 6                      | 1 (0.05%)   | 1837 (100%)  |
| Total                  | 1837  | 1837 (100%)  |

Tableau 4.2 : Nombres d'évaluateurs associés aux résumés proposés au choix, classés par fréquence de sélection

Les Figure 4.5 et Figure 4.6 permettent de visualiser la répartition des questions et des évaluateurs sur les deux dimensions les plus informatives obtenues suite à une Analyse des Correspondances Multiples – Méthode factorielle applicable sur des données qualitatives. Dans la Figure 4.5, nous observons de fortes variations dans les réponses proposées aux questions ; un groupe de questions pour lesquelles les réponses proposées sont sensiblement similaires peut tout de même être identifié. Il semblerait ainsi que, bien que la stratégie de résumé adoptée par un évaluateur soit naturellement dépendante de la question, une stratégie globale de résumé soit tout de même identifiable.

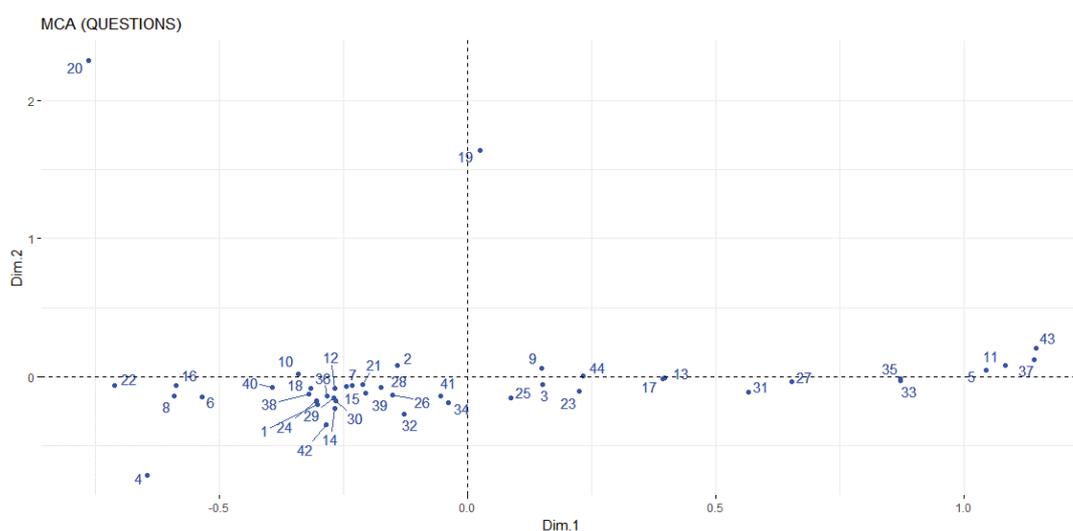


Figure 4.5 : Analyse des Correspondances Multiples (ACM) des questions en fonction des évaluateurs qui leur sont associés



### 4.3.2 Evaluation de la performance de l'approche automatique de résumé

Nous proposons de discuter la performance de l'approche automatisée de résumé en fonction, pour une configuration de paramètres spécifique, du nombre d'évaluateurs associés aux résumés qu'elle propose. Nous utilisons pour cela les données étudiées dans la section précédente.

Pour une configuration donnée, l'algorithme permet d'obtenir un classement des meilleurs résumés, avec pour chaque résumé un score associé (cf. équation 4.1). Ainsi, nous proposons, pour différents seuils de nombre de résumés considérés, d'évaluer le nombre d'évaluateurs ayant choisi un résumé parmi ceux retournés. Plus le nombre d'évaluateurs associés à des résumés sélectionnés par l'approche automatisée est élevé, meilleure devra être considérée l'approche (ou plutôt son paramétrage).

Nous définissons la performance *relative* d'un algorithme de résumé paramétré nommé  $A$  par la fonction d'évaluation suivante :

$$perf_{rel}(A, k) = \frac{\sum_{q \in Q} \sum_{r \in R_{A(q)}^k} \#eval(q, r)}{\sum_{q \in Q} \sum_{r \in R_q^k} \#eval(q, r)} \quad (4.2)$$

Avec  $Q$  l'ensemble des questions proposées aux évaluateurs,  $R_q^k$  l'ensemble des  $k$ -meilleurs résumés associés à une question  $q \in Q$ , *i.e.* les  $k$  résumés les plus sélectionnés par les évaluateurs pour la question  $q$ ;  $R_{A(q)}^k$  correspond à l'ensemble des  $k$ -meilleurs résumés proposés par l'algorithme  $A$  pour la question  $q$  et  $\#eval(q, r)$  correspond au nombre d'évaluateurs ayant sélectionné le résumé  $r$  pour la question  $q \in Q$ .

Pour un nombre donné de résumés retournés par l'approche automatisée, le score maximal pouvant être obtenu est égal à la somme des évaluateurs associés à l'ensemble des  $k$ -meilleurs résumés proposés pour chacune des questions. Ce score fournit ainsi une borne maximale permettant d'apprécier la performance globale de l'approche pour un seuil  $k$  fixé. Il est cependant à noter que certains choix de résumés proposés aux évaluateurs ne sont pas évalués par l'approche automatisée car ils ne correspondent pas à des résumés répondant aux contraintes fixées par la notion de résumé considérée – certains résumés violent par exemple la contrainte de non-redondance et n'étaient alors proposés qu'afin d'évaluer l'attente des évaluateurs en terme de résumé. Ces résumés n'étant en effet pas évalués et ne pouvant donc pas être *retrouvés* par l'approche, pour certains seuils  $k$ , le score maximal correspond alors à une borne maximale ne pouvant être atteinte. Nous précisons cependant que le score obtenu pour un seuil  $k$  très élevé fournit une bonne appréciation de la performance optimale qui peut être obtenue par l'approche automatisée.

Le paramétrage de l'algorithme pour lequel les résultats sont présentés a été défini de manière empirique lors d'une phase de paramétrage menée au préalable sur des jeux de tests plus modestes. Les paramètres utilisés associent une importance plus grande à l'estimation de la partie commune en définissant  $w_{psi} = 300.0$  et  $w_{loss} = 100.0$ . Dans ce paramétrage, la perte d'information exacte  $\delta_{E-} = 400$  y est fortement pénalisée, peu d'importance est accordée à la perte ou à l'ajout d'information plausible  $\delta_{P+} = \delta_{P-} = 1$ , une importance critique pour la distorsion  $\delta_D = 500000$  est considérée, avec ici  $\alpha = 1$  défini pour le calcul de  $\tau$ ; le paramètre contrôlant le degré de synthèse est défini tel que  $\epsilon = 500$ .

La représentation des connaissances utilisée pour le traitement a été définie par un binôme de modélisateurs qui ont pris en compte les liens entre les descripteurs proposés dans les questions soumises et les descripteurs proposés dans les résumés – la modélisation sélectionnée a été validée de manière consensuelle par un troisième modélisateur. Nous précisons que cette représentation des

connaissances n'était pas connue des évaluateurs lors de la phase de résumé. Techniquement la représentation des connaissances a été par la suite traduite en ontologie à l'aide du logiciel Protégé (<http://protege.stanford.edu/>).

Nous reportons ci-dessous, pour différents seuils de  $k$  :

1. les scores de performance relative obtenus (équation 4.2);
2. la performance non normalisée (performance brute), *i.e.* la valeur du numérateur associé au calcul de la performance proposée ci-avant,
3. la performance brute maximale, soit le nombre d'évaluateurs associés à un résumé de rang  $x \leq k$ ,
4. la performance relative rectifiée qui considère que la performance maximale est atteinte lorsqu'un très grand nombre de résumés sont retournés, *i.e.* que tous les évaluateurs associés à des résumés pouvant être retrouvés par l'approche automatisée le sont. La performance rectifiée correspond donc à la performance normalisée par la performance maximale du système lorsque  $k \rightarrow \infty$ .
5. La performance globale (rectifiée) définie comme la performance brute pour la valeur  $k$  considérée normalisée par la performance brute obtenue lorsque  $k \rightarrow \infty$  (dans la table qui suit la performance brute est donc divisée par 1740).

| k   | Performance Relative | Performance Brute | Performance Brute Maximale | Performance Relative Rectifiée | Performance Globale |
|-----|----------------------|-------------------|----------------------------|--------------------------------|---------------------|
| 1   | 0,503                | 448               | 891                        | 0,531                          | 0,257               |
| 2   | 0,483                | 656               | 1359                       | 0,510                          | 0,377               |
| 3   | 0,521                | 848               | 1628                       | 0,550                          | 0,487               |
| 4   | 0,503                | 896               | 1781                       | 0,531                          | 0,515               |
| 5   | 0,498                | 914               | 1836                       | 0,526                          | 0,525               |
| 6   | 0,513                | 942               | 1837                       | 0,541                          | 0,541               |
| 8   | 0,551                | 1013              | 1837                       | 0,582                          | 0,582               |
| 10  | 0,590                | 1084              | 1837                       | 0,623                          | 0,623               |
| 15  | 0,673                | 1236              | 1837                       | 0,710                          | 0,710               |
| 20  | 0,743                | 1365              | 1837                       | 0,784                          | 0,784               |
| 50  | 0,900                | 1654              | 1837                       | 0,951                          | 0,951               |
| inf | 0,947                | 1740              | 1837                       | 1                              | 1                   |

Tableau 4.3 : Performance de l'approche globale en fonction de différents seuils de sélection du nombre de résumés retournés

Les résultats obtenus sont aussi représentés graphiquement dans la Figure 4.8.

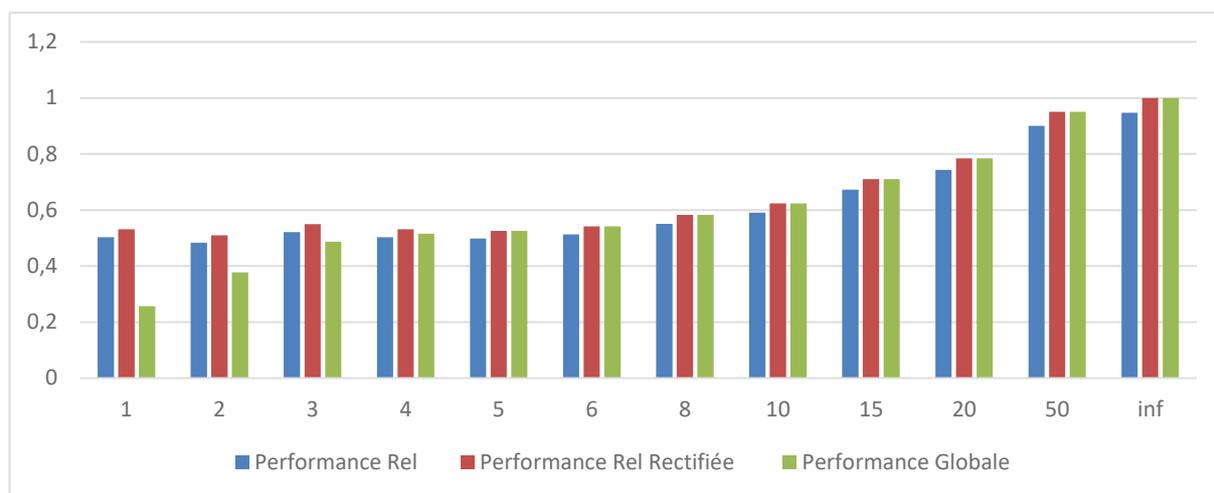


Figure 4.8 : Performance de l'approche en fonction de différents seuils de sélection du nombre de résumés retournés

Les résultats obtenus soulignent la performance de l'algorithme pour l'identification de résumés. Nous observons tout d'abord que la performance de l'approche pour une valeur de  $k$  égale à 5 permet de distinguer les résultats associés à plus de 50% des choix exprimés par les évaluateurs.

Une validation croisée a aussi été effectuée (*leave-one-out cross-validation*) dans le cas d'un paramétrage correspondant à  $k = 3$ . Le jeu de données a pour cela été décomposé en 11 blocs, 4 requêtes par bloc (utilisé dans le  $k$ -fold). L'objectif était alors d'analyser les performances du système sur un bloc de donnée dans le cas où les poids sont calculés sur les 10 blocs restant. Le nombre maximal moyen pouvant être atteint est alors de  $1628/44 \times 4 = 148$  évaluateurs (cf. performance brute maximale pour  $k = 3$ ) (voir Tableau 4.3).

Les performances obtenues sur la validation croisée réalisée sont (dans le cas où nous nous intéressons aux performances pour les 3 meilleurs résumés<sup>13</sup>) ; en moyenne, une performance rectifiée de 0.553 est obtenue. Ces résultats sont sensiblement similaires à ceux obtenus dans l'évaluation précédente.

### 4.3.3 Application à l'analyse psychosensorielle : illustration

Nous proposons dans cette étude une illustration d'utilisation de l'approche automatisée de résumé. L'illustration repose sur l'analyse de données relatives au domaine de l'élaboration de Whiskies. Nous disposons pour 6 marques de whiskies d'une description exprimée sous la forme d'un ensemble de descripteurs. Ces descriptions ont été proposées par des experts du domaine, elles correspondent aux descriptifs qui peuvent être utilisés sur l'emballage associés aux différents whiskies. En supplément de ces descriptions conceptuelles expertes rattachées à des whiskies identifiés, nous disposons d'un ensemble de descriptions conceptuelles. Ces descriptions conceptuelles correspondent à l'union de descriptions proposées par des néophytes pour un même whiskey testé – nous ferons par la suite référence à ces descriptions au travers du terme description d'un panel.

<sup>13</sup> En considérant une optimisation des paramètres effectuée en considérant un seuil  $k=3$ , les paramètres sont optimisés pour maximiser le nombre d'évaluateurs associés aux 3 meilleurs résumés sélectionnés par l'approche. On note qu'une évaluation de la performance pour une partition précise a été exclue de l'analyse (résultats non représentatif du fait du caractère atypique des questions regroupées dans la partition de test). Les scores bruts sont  $70+138+57+56+61+80+36+44+151+115 = 808$  ; soit en moyenne  $834/10 = 80.8$  ; ce qui normalisé par la performance maximale à  $k=3$  nous donne une performance relative de  $80.8/146 = 0.553$ .

Nous proposons ainsi d'évaluer l'approche automatique de résumé au travers de son utilisation pour l'identification de whiskies décrits au travers de descriptions d'un panel. Nous souhaitons en particulier tester si l'approche automatisée est en mesure d'identifier correctement le whiskey utilisé pour l'obtention d'une description spécifique d'un panel. Le jeu de données utilisé pour cette illustration est de taille réduite : nous disposons seulement de six descriptions expertes de whiskies et tout autant de descriptions d'un panel. L'ensemble des descriptions proposées sont exprimées à l'aide de descripteurs contrôlés définis dans la roue des odeurs propre au domaine d'analyse, cf. Figure 4.9.<sup>14</sup>

---

<sup>14</sup> Cette roue des odeurs a été légèrement augmentée afin de remédier à l'inexistence de certains descripteurs utilisés par les experts.



## *Les chemins secrets d'Irlande et d'Écosse* à travers leurs grands whiskies

Six notes de dégustation (terme commercial pour désigner la qualité du whisky) ont été proposées par six maîtres-distillateurs d'Irlande et d'Écosse : Cragganmore, Caol Ila, Talisker, Knockando, Glenkinchie et Bushmills. Ils ont chacun résumé les impressions olfactives et gustatives de leur whisky préféré dans la table suivante.

|                    | Notes de dégustation (descriptions expertes)      |
|--------------------|---|
| <b>Cragganmore</b> | Fruits frais, Fruits secs, Floral, Réglisse, Fumé |
| <b>Caol Ila</b>    | Fruité, Tourbé, Olive, Amande, Fumé, Iodé         |
| <b>Talisker</b>    | Iodé, Tourbé, Poivré, Épicé                       |
| <b>Knockando</b>   | Fruité, Noisette, Amande                          |
| <b>Glenkinchie</b> | Malté, Herbe coupée, Gingembre, Boisé             |
| <b>Bushmills</b>   | Céréales, Agrume, Vanille, Amande, Malté, Épicé   |

Tableau 4.4 : Descriptions expertes de la qualité de 6 whiskies

Nous nous sommes livrés à une analyse sensorielle pour laquelle un panel de juges amateurs (de whisky) a eu à choisir parmi 127 descripteurs : 6 de niveau 1 (le plus abstrait), 18 de niveau 2 et 103 de niveau 3 (le plus spécifique). Voici les évaluations recueillies :

|  | Description de panels   |
|--|---|
| <b>Whisky 1</b><br><b>Talisker*</b>    | Hareng fumé, brise de mer, terreux, cheminée, algues, iodé, feu de bois, humus, tourbe, tourbé, épicé   |
| <b>Whisky 2</b><br><b>Knockando*</b>   | Noisette, boisé, hydromel, chocolat au lait, huile de lin, dattes, fruits secs, vanille, figues séchées, cire de bougie, beurre, mélasse, amande                  |
| <b>Whisky 3</b><br><b>Cragganmore*</b> | Pêche, saumon fumé, réglisse, pomme, compote de pommes, feu de bois, pruneau, lavande, raisins secs, bois, parfumé, parfumé, sciure, rose, noix de coco, cheminée |
| <b>Whisky 4</b><br><b>Glenkinchie*</b> | Céréales de petit déjeuner, bière ambrée, sève, porridge, légumes verts, herbe coupée, tiges vertes, son, bois, gingembre   |
| <b>Whisky 5</b><br><b>Bushmills*</b>   | Amande, porridge, caramel mou, flocons d'avoine, crème brûlée, son, fruits frais, vanille, noisette, céréales   |
| <b>Whisky 6</b><br><b>Caol Ila*</b>    | Dattes, olive, coquillages, tourbé, beurre, fruité, noisette, amande, jambon fumé, boisé, huileux, fruits de mer  |

Tableau 4.5 : Descriptions libres de la qualité de 6 whiskies

Cette information n'était naturellement pas précisée lorsque le descriptif de la tâche était fourni aux évaluateurs en charge d'effectuer l'association. Ces informations correspondent en effet aux réponses correctes attendues dans le cas où (i) les descripteurs proposés permettent l'identification du bon Whisky, et (ii) la phase de résumé de l'information est effectuée correctement.

L'approche automatisée basée sur les paramètres appris sur le jeu de données utilisé dans la précédente section permet l'obtention des appariements attendus. Lors d'un test rapide réalisé auprès de 5 testeurs seule une experte de l'analyse sensorielle d'odeurs aura elle aussi réussi un sans-faute dans la mise en correspondance ; deux autres, initialement rebuté par la combinatoire des possibilités de regroupement

des termes des descriptions de panels, ont *a posteriori* validé les réponses du premier. Sur un exemple de petite taille comme celui-ci, une roue des arômes comportant à peine plus de 120 descripteurs, des panels d'évaluateurs réduits, ce qui explique que les descriptions de panels soient limitées à quelques termes seulement, il est d'ores et déjà difficile à un testeur humain, même exercé, de gérer à la fois la combinatoire des descriptions de panels (gérer les redondances, les regroupements, la représentativité) et la structure de la connaissance que constitue la roue des arômes (gestion de l'abstraction, de la spécificité). On imagine donc aisément l'intérêt de l'approche automatisée proposée si l'on envisage de collecter des centaines d'évaluations pour caractériser les nuisances olfactives autour d'un site industriel.

## 4.5 Conclusion

Dans ce chapitre, nous nous sommes intéressés à la synthèse d'un ensemble de descripteurs conceptuels, traitement d'importance centrale pour la mise en place de l'analyse automatisée globale visée dans cette thèse. Nous avons en particulier proposé une approche automatisée de résumé procédant par identification, tant que possible, d'un ensemble de descripteurs conceptuels synthétiques de l'information portée par l'ensemble des descripteurs à résumer. La formulation de cette proposition repose sur un travail préliminaire de formalisation de la notion de résumé, et sur l'identification de différentes dimensions caractéristiques qui méritent d'être considérées afin de critiquer la pertinence d'un résumé – notions de couverture et de perte d'information, caractère synthétique d'un résumé, etc. Issus de ces travaux préliminaires, le modèle retenu repose sur une fonction d'évaluation qui tient compte de l'analyse des différentes propriétés attendues du résumé et qui permet d'identifier les résumés les plus pertinents au regard de cette analyse. La qualité d'un résumé est alors calculée par une fonction d'agrégation qui prend en compte l'ensemble des critères d'évaluation : cette décomposition en critères et le choix de l'opérateur d'agrégation (a minima un jeu de poids pour une agrégation linéaire comme proposé ici) permettent l'interprétation des scores et le cas échéant, l'agrégation peut être modifiée par l'utilisateur pour tenir compte d'un contexte spécifique.

Les nombreuses expérimentations menées soulignent la capacité de l'approche proposée à produire des résumés d'intérêt et à obtenir des performances intéressantes dans les contextes d'évaluation étudiés. Ces évaluations illustrent aussi la difficulté toute particulière de la tâche de résumé. Ces dernières ont en effet souligné que l'existence même d'un résumé consensuel idéal mériterait d'être discutée dans certains cadres applicatifs – différents profils d'évaluateurs semblent en effet dans certains cas pouvoir être identifiés, ce qui souligne le caractère subjectif de la tâche, qui dans le cas étudié portait sur des résumés produits par des évaluateurs non entraînés.

Les perspectives d'amélioration et d'extension des travaux proposés sont nombreuses. Des études complémentaires et plus larges méritent notamment d'être menées pour caractériser plus finement, et contextuellement, les attentes des opérateurs humains en terme de résumé. Une étude approfondie et contextuelle du caractère subjectif de la tâche de résumé serait ainsi très intéressante pour la définition de modèles plus performants – en s'autorisant éventuellement à considérer des approches tenant compte des particularités des utilisateurs. Des travaux supplémentaires doivent aussi être proposés pour améliorer l'approche proposée, en particulier dans l'objectif de réduire les temps de calcul. Nous proposons pour cela de nous intéresser à la réduction de l'espace de recherche (l'espace des résumés) qui, de par sa taille importante dans le cas de représentation des connaissances volumineuses, empêche *de facto* une évaluation de l'ensemble des résumés éligibles. L'identification des paramètres optimaux, ou du moins pertinents, mérite elle aussi d'être étudiée. Différentes techniques peuvent ici être utilisées : algorithme génétique, descente du gradient, etc. De manière plus générale, une confrontation étendue de l'approche proposée avec des approches reposant sur des techniques d'apprentissage automatique

supervisé doit être menée – on imagine dans ce cas considérer des fonctions génératrices de jeux de données d'entraînement basées à la fois sur la connaissance *a priori* exprimée par la représentation des connaissances et sur les contraintes définies par la notion de résumé considérée.

## Chapitre 5 Conclusion Générale

---

Dans ce travail, nous avons proposé une approche non orientée pour l'évaluation sensorielle et plus spécifiquement l'évaluation de la qualité des odeurs. La méthode développée dans le cadre de ce travail avait en effet pour objet initial de trouver des applications dans le domaine de la caractérisation des émissions gazeuses industrielles et constituer une approche originale, en rupture avec les approches actuelles d'analyse sensorielle, contraintes par la taille et la formation des panels d'évaluateurs. En inversant la logique de recherche des données descriptives de la qualité de l'odeur—apprentissage automatique plutôt que formation des observateurs, l'idée était d'assoir la détermination de cette grandeur sur des échantillons de population de taille significative et de s'affranchir de l'utilisation de jurys spécialisés. Cependant le recours à des non-spécialistes pour caractériser la qualité d'une odeur nécessite d'être à même de comprendre et d'interpréter des descriptions en langage naturel alors que des panels professionnels maîtrisent, eux, un vocabulaire structuré de descripteurs non ambigus. Le défi à relever consiste alors à apprécier l'évocation implicite d'un descripteur que suggère un mot ou un groupe de mots du langage courant en s'appuyant, d'une part, sur des techniques d'analyse distributionnelle appliquées à des corpus de textes pour disposer d'une représentation vectorielle des mots, sur une gestion de l'imprécision et de l'abstraction du langage guidée par une organisation taxonomique des connaissances d'autre part.

Lorsque l'évaluation de la qualité d'une odeur est confiée à un panel néophyte, chaque description n'est alors plus associée à un ensemble de descripteurs non ambigus, mais à un simple sac de mots en langage naturel. Deux problématiques ont alors été rattachées à la caractérisation d'odeurs dans nos travaux et les réponses que nous y avons apportées dans les chapitres 3 et 4 constituent les deux contributions techniques principales de cette thèse. La première a consisté à transformer des descriptions terminologiques en descripteurs structurés ; la seconde s'est donné pour objet de résumer un ensemble de descripteurs formels proposés par un panel d'évaluateurs conséquent en une caractérisation synthétique et cohérente à des fins industrielles. Ainsi, la première partie de notre travail s'est focalisée sur la définition et l'évaluation de modèles qui peuvent être utilisés pour résumer un ensemble de mots en un unique descripteur désambiguïsé. Parmi les différentes stratégies envisagées dans cette contribution, nous avons proposé de comparer des approches hybrides exploitant à la fois des bases de connaissances et des plongements lexicaux définis à partir de grands corpus de textes. Nos résultats illustrent le bénéfice substantiel à utiliser conjointement représentation symbolique et plongement lexical. Nous avons ensuite défini de manière formelle le processus de synthèse d'un ensemble de concepts et avons proposé un modèle qui s'apparente à une forme d'intelligence humaine pour évaluer les résumés alternatifs au regard d'un objectif de synthèse donné. L'approche non orientée que nous avons proposée dans ce manuscrit apparaît ainsi comme l'automatisation cognitive des tâches confiées aux opérateurs des séances d'analyse sensorielle. Elle ouvre des perspectives intéressantes pour développer des analyses sensorielles à grande échelle sur de grands panels d'évaluateurs lorsque l'on essaie de caractériser les nuisances olfactives autour d'un site industriel.

L'idée de base de ces travaux est née de la similitude entre l'organisation hiérarchique de la connaissance que proposent tout aussi bien une roue des odeurs qu'une taxonomie que l'on utiliserait par exemple pour indexer des documents. Dans chacun des cas, un élément *parent* est plus abstrait qu'un élément *enfant* plus précis. La subsomption désigne une relation d'inclusion entre des concepts dans les logiques de description. Elle correspond ainsi à la relation hyponymie—hyperonymie en linguistique, c'est-à-dire deux termes dont le sens de l'un inclut celui de l'autre. Cependant cette similitude doit être

utilisée avec précaution. Ainsi, dans une roue des arômes du whisky, Bois est un concept qui généralise Gâteau Brûlé, Marc de Café, Réglisse ou encore Gingembre. Il est clair que cette relation d'inclusion est uniquement basée sur le ressenti sensoriel et n'a rien à voir avec la relation d'hyponymie–hyperonymie. Ainsi, dans la phase de résumé basée sur une roue des odeurs (transposée en taxonomie), la factorisation conceptuelle peut produire des résultats contre intuitifs car c'est la relation d'hyponymie–hyperonymie que l'on pense implicitement exploiter pour la synthèse linguistique. Ainsi, résumerait-on naturellement les concepts Bière Ambrée, Porto et Cognac par Alcool dans un exercice linguistique alors que dans notre contexte sensoriel, on a Malté qui subsume Céréales qui subsume Bière Ambrée d'un côté et de l'autre, Boisé généralise Vineux qui généralise Porto et Cognac... L'analyste sensoriel n'a donc aucune raison de « rapprocher » par synthèse les concepts d'évaluation qui appartiennent à deux secteurs angulaires bien distincts de la roue des arômes dans ce cas. Une fois entendu que l'ordre de la représentation des connaissances porte sur le sens olfactif ou gustatif et non pas sur la sémantique, notre algorithme de résumé reste pour autant correct. En revanche, les choses se compliquent lorsque l'on intègre la gestion de la hiérarchie des concepts dans le modèle distributionnel utilisé dans la phase de plongement des termes dans l'espace conceptuel puisque peuvent coexister les deux interprétations.

Après cette mise en garde sur l'interprétation de l'ordre sur les descripteurs, portons notre attention sur les notions de distance ou de similarité qui ont été utilisées dans cette thèse. Des métriques classiques du TAL nous ont permis de définir la proximité entre deux mots par le biais de leurs représentations vectorielles. Des mesures que nous avons appelées similarités sémantiques nous ont permis d'apprécier la similarité entre deux descripteurs. Nous avons ainsi pu évaluer la similitude de profils sensoriels en combinant ces mesures définies dans un cadre linguistique. La similarité sémantique et la proximité syntaxique de deux descriptions impliquent que les perceptions olfactives sont voisines. C'est l'hypothèse que nous avons retenue. En réalité, il peut sembler plus naturel de faire le chemin inverse et penser que deux odeurs voisines doivent impliquer des descriptions similaires. Ainsi, dans le chapitre 2, nous avons introduit le champ des odeurs®, méthode de classification des odeurs proposée par Jaubert en 1987 à la suite de recherches sur les relations entre la structure chimique et le caractère odorant des molécules. Comme nous l'avons expliqué, Jaubert s'est appuyé sur un travail réalisé avec 1400 molécules odorantes de base dont il a retenu 45 molécules référentes qui se répartissaient dans un espace multidimensionnel structuré. Cette cartographie lui a permis de positionner les odeurs les unes par rapport aux autres (voir chapitre 2 section 2.5.1.2). Les 45 référents sont appris progressivement par les panelistes afin de pouvoir positionner les odeurs à évaluer. Ainsi, Jaubert est-il parvenu à associer à chacune de ses molécules référentes une liste de descripteurs qu'il a pu relever au fil des analyses sensorielles. Nous aurions alors pu envisager de construire nos modèles de langue sur la base de ces 45 « petits » vocabulaires en utilisant le principe que les molécules odorantes proches ont des vocabulaires qui se recouvrent en partie (nous n'avons pas vérifié quantitativement cet énoncé). La limite serait venue de la taille limitée des vocabulaires proposés par Jaubert, même augmentés de synsets. Une autre idée pourrait être d'utiliser la distance olfactive au sens de Jaubert comme une mesure de contrôle qui autorise ou non la factorisation conceptuelle lors de la phase de résumé que nous avons proposée : la synthèse reposerait alors bien sur une véritable relation d'hyponymie–hyperonymie qui serait utilisée pour synthétiser seulement quand la proximité olfactive est vérifiée. Cette remarque conclut la seconde mise au point que nous souhaitons apporter cette fois-ci non plus sur les structures de connaissances, mais sur les mesures de similarité manipulées dans nos propositions.

Parmi les autres extensions que nous souhaiterions développer par la suite, citons la généralisation du prolongement de l'espace des termes vers l'espace conceptuel. Nous avons proposé d'associer un terme à un concept, puis un sac de termes à un concept, la contextualisation permettant de réduire l'impact de l'ambiguïté des descripteurs en langage naturel. Il serait intéressant de généraliser le prolongement d'un

groupe de termes vers un groupe de descripteurs. Pour illustrer notre propos, reprenons l'exemple des arômes de whisky et imaginons que l'on ait recueilli les descripteurs « feu de bois, humus, hareng fumé ». A ce groupe de trois termes, le prolongement d'un sac de termes à un concept pourrait associer le concept *Tourbé*, comme « bière ambrée, porridge, tomate verte » pourrait permettre de découvrir le descripteur *Malté*. En conditions expérimentales, il nous faut maintenant imaginer que ces descripteurs recueillis auprès des panelistes ne forment qu'un seul et même ensemble : « feu de bois, bière ambrée, humus, porridge, hareng fumé, tomate verte ». La généralisation du prolongement consiste à trouver la partition de cet ensemble de termes qui maximisera la moyenne (par exemple) des degrés d'évocation attribués à l'appariement de chaque sous-ensemble de la partition à un concept. La combinatoire de cette extension permet d'apprécier la difficulté de l'exercice, même pour un analyste sensoriel confirmé, de sortir les deux concepts *Tourbé* et *Malté* du sac de termes « feu de bois, bière ambrée, humus, porridge, hareng fumé, tomate verte ».

Cette remarque nous permet de conclure sur l'objectif de cette thèse : l'automatisation cognitive des tâches confiées aux opérateurs des séances d'analyse sensorielle. Lorsque le nombre de descripteurs recueillis auprès d'un panel de néophytes devient conséquent, l'analyste doit non seulement avoir une maîtrise parfaite de la représentation des connaissances qui structure les descripteurs du domaine, mais également une remarquable capacité à gérer la combinatoire des descriptions. Généralement, l'analyste connaît au moins qualitativement le résultat attendu et va filtrer et regrouper les termes, implicitement guidé par ce résultat. Ce filtrage dirigé et la connaissance plus ou moins complète de l'agencement des descripteurs du domaine induisent une part de subjectivité dans l'interprétation des descripteurs par l'analyste. L'homme, avec toute l'imprécision et la subjectivité qui caractérisent ses évaluations sensorielles, se trouvent pourtant au cœur d'une nouvelle chaîne de fiabilité de l'industrie. Nous avons vu que les approches orientées visent à limiter l'impact de cette subjectivité et cette imprécision en contraignant les panelistes à s'exprimer dans un référentiel métier partagé. Notre approche s'est donné le même objectif dans le cas d'une approche non orientée pour la détermination de la qualité d'une odeur. L'automatisation cognitive des tâches de l'analyste sensoriel que nous proposons doit permettre de s'abstraire de la connaissance imparfaite de la structure des descripteurs du domaine et de la combinatoire des descriptions ; la subjectivité résiduelle ne sera plus liée qu'au choix des corpus d'apprentissage pour le modèle distributionnel. Cette automatisation laisse espérer de nouvelles perspectives pour développer des analyses sensorielles à grande échelle sur des panels d'évaluateurs néophytes ce qui est nécessaire lorsque l'on souhaite caractériser les nuisances olfactives autour d'un site industriel.

Enfin, les exemples qui ont servi à illustrer nos propos ne sont pas tous issus de l'analyse sensorielle olfactive, ce qui confère à notre approche toute sa genericité et son applicabilité à tous les domaines du sensoriel, mais aussi aux domaines variés intéressés par l'analyse de données à caractère sémantique (e.g., TAL, Recherche d'Information).

---

## Références

---

- Asher, M. I., Montefort, S., Björkstén, B., Lai, C. K., Strachan, D. P., Weiland, S. K., & Williams, H. (2006). Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *The Lancet*, 368(9537), 733–743.
- Bairi, R., Iyer, R. K., Ramakrishnan, G., & Bilmes, J. A. (2015). Summarization of Multi-Document Topic Hierarchies using Submodular Mixtures. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, 553–563.
- Balez, S. (2001). *Ambiances olfactives dans l'espace construit : perception des usagers et dispositifs techniques et architecturaux pour la maîtrise des ambiances olfactives dans des espaces de type tertiaire. Thèse de doctorat.* Université de Nantes, France.
- Bárcenas, P., Elortondo, F. P., & Albisu, M. (2003). Comparison of free choice profiling, direct similarity measurements and hedonic data for ewes' milk cheeses sensory evaluation. *International Dairy Journal*, 13(1), 67–77.
- Barthelemy, J. (1990). Evaluation d'une grandeur sensorielle complexe: description quantifiée in Evaluation sensorielle-manuel méthodologique.
- Bartlett, N., Wood, F., & Perotte, A. (2011). Hierarchically Supervised Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, 1–9.
- Batet, M., Sanchez, D., Valls, A., & Gibert, K. (2010). Exploiting taxonomical knowledge to compute semantic similarity: an evaluation in the biomedical domain. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 274–283.
- Beissbarth, T., Speed, T. P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics (Oxford, England)*, 20(9):1464–5.
- Bende, M., & Nordin, S. (1997). Perceptual learning in olfaction: Professional wine tasters versus controls. *Physiology and Behavior*, 62(5), 1065–1070.
- Benoit, O., & Pannier, G. (1982). *Les polluants atmosphériques et les odeurs-nature, mesures, méthodes d'élimination-synthèse bibliographique.*
- Bensafi, M. (2001). *Le traitement affectif des odeurs : Aspects implicites et explicites. Thèse de doctorat.* Université de Lyon 2, France.
- Berglund, B., Berglund, U., Engen, T., & Ekman, G. (1973). Multidimensional analysis of twenty-one odors. *Scandinavian Journal of Psychology*, 14(1), 131–137.
- Bergson, H. (1889). Essai sur les données immédiates de la conscience. *Alcan, Paris*, 1889–1896.
- Berry, M. W., Dumais, S. T., & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM Review*, 37(4), 573–595.
- Blei, D. M., Jordan, M. I., Griffiths, T. L., & Tenenbaum, J. B. (2004). Hierarchical Topic Models and the Nested Chinese Restaurant Process. *Advances in Neural Information Processing Systems*, 16, 17.

- Blei, D. M., & Lafferty, J. D. (2006). Correlated Topic Models. *Advances in Neural Information Processing Systems*, 18, 147.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Boisson, C. (1997). La dénomination des odeurs : variations et régularités linguistiques. *Intellectica*, 24(1), 29–49.
- Bruneau, E., Barbier, E., Gallez, L., & Guyot, C. (2000). La roue des arômes des miels. *Abeilles & Cie*, 77.
- Buck, L., & Axel, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell*, 65(1), 175–187.
- Buscaldi, D., Le Roux, J., Flores, J. J. G., & Popescu, A. (2013). Lipn-core: Semantic text similarity using n-grams, wordnet, syntactic analysis, esa and information retrieval based features. In *Second Joint Conference on Lexical and Computational Semantics*, 63.
- Byrne, D. V, Bak, L. S., Bredie, W. L. P., Bertelsen, G., & Martens, M. (1999). Development of a sensory vocabulary for warmed-over flavor: part I. in porcine meat. *Journal of Sensory Studies*, 14(1), 47–65.
- Cain, W. S., De wijk, R., Lulejian, C., Schiet, F., & See, L. C. (1998). Odor identification: Perceptual and semantic dimensions. *Chemical Senses*, 23(3), 309–326.
- Candau, J. (2004). L’odeur médiane du couvre-lit de tante Léonie. *Cognition*, 75, 72–87.
- Capone, S., Tufariello, M., & Siciliano, P. (2013). Analytical characterisation of Negroamaro red wines by Aroma Wheels”. *Food Chemistry*, 141(3), 2906–2915.
- Carson, C., Thomas, M., Belongie, S., Hellerstein, J. M., & Malik, J. (1999). Blobworld : A System for Region-Based Image Indexing and Retrieval ( long version )? In *Visual Information and Information Systems*, 509–517.
- Chaignaud, M., Cariou, S., Poette, J., Fages, M., Despres, J. F., & Fanlo, J. L. (2014). A new method to evaluate odour annoyance potential. *Chemical Engineering Transactions*, 40.
- Chemin, P. (1996). Évolution de la législation sur les émissions d’odeurs, In: Proceedings du 1 er Congrès International sur les nez électroniques et les composés odorants. In *Odours and VOC’s Journal* (pp. 57–10). Paris.
- Chollet, S., & Valentin, D. (2000). Le degré d’expertise a-t-il une influence sur la perception olfactive ? Quelques éléments de réponse dans le domaine du vin. *L’année Psychologique*, 100(1), 11–36.
- Chollet, S., Valentin, D., & Abdi, H. (2005). Do trained assessors generalize their knowledge to new stimuli ? *Food Quality and Preference*, 16(1), 13–23.
- Chrea, C., Valentin, D., Sulmont-Rossé, C., Hoang Nguyen, D., & Abdi, H. (2005). Odeurs et catégorisation : notre représentation mentale des odeurs est-elle universelle ou dépendante de notre culture ? *Actes de La 3ème Journée Du Sensolier, Paris (France)*, 3–7.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12, 2493–2537.
- Corley, C., & Mihalcea, R. (2005). Measuring the semantic similarity of texts. In Proceedings of the ACL workshop on empirical modeling of semantic equivalence and entailment. *Association for*

- Computational Linguistics*, 13–18.
- Cristovam, E., Paterson, A., & Piggott, J. (2000). Differentiation of port wines by appearance using a sensory panel: Comparing free choice and conventional profiling. *European Food Research and Technology*, 211(1), 65–71.
- Cross, V., & Yu, X. (2011). Investigating ontological similarity theoretically with fuzzy set theory, information content, and Tversky similarity and empirically with the gene ontology. *Scalable Uncertainty Management*, 387–400.
- Curran, J. R. (2004). From distributional to semantic similarity.
- David, S., Dubois, D., Rouby, C., & Schaal, B. (1997). L'expression des odeurs en français : Analyse lexicale et représentation cognitive. *Intellectica*, 24(1), 51–83.
- Declercq, C., Pascal, M., Chanel, O., Corso, M., Lefranc, A., Medina, S., & others. (2012). Impact sanitaire de la pollution atmosphérique dans neuf villes françaises. Résultats du projet Aphekom. *Revue d'Épidémiologie et de Santé Publique*, 60, 60–61.
- Delarue, J., & Sieffermann, J. (2004). Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15(4), 383–392.
- Depled, F., & Sauvageot, F. (2002). Evaluation sensorielle des produits alimentaires. *Techniques de L'ingénieur. Agroalimentaire*, 3(4000).
- Dhillon, P., Rodu, J., Foster, D., & Ungar, L. (2012). Two step cca: A new spectral method for estimating vector models of words. *arXiv preprint arXiv:1206.6403*.
- Dinu, G. (2011). *Word meaning in context: A probabilistic model and its application to Question Answering*. Thèse de doctorat. Saarländische Universitäts-und Landesbibliothek.
- Distel, H., Ayabe-kanamura, S., Martinez-gomez, M., Schicker, I., Kobayakawa, T., Saito, S., & Hudson, R. (1999). Perception of everyday odors: Correlation between intensity, familiarity and strength of hedonic judgement. *Chemical Senses*, 24(2), 191–199.
- Doty, R. L., Brugger, W. E., Jurs, P. C., Orndorff, M. A., Snyder, P. J., & Lowry, L. D. (1978). Intranasal trigeminal stimulation from odorous volatiles: Psychometric responses from anosmic and normal humans. *Physiology & Behavior*, 20(2), 175–185.
- Dratz, M. T. (2000). Le champ des odeurs de Jean-Noël Jaubert. Intérêt pédagogique et application en analyse sensorielle descriptive. *Annales des Falsifications, de l'Expertise Chimique et Toxicologique*, 93(954), 69–78.
- Dubois, D. (2006). Des catégories d'odorants à la sémantique des odeurs. Une approche cognitive de l'olfaction. *Terrain. Revue d'Ethnologie de l'Europe*, (47), 89–106.
- Dubois, D., Rouby, C., & Sicard, G. (1997). Catégories sémantiques et sensorialités : de l'espace visuel à l'espace olfactif. *Enfance*, 50(1), 141–151.
- Dufour, F., & Barkat-Defradas, M. (2009). Opérations linguistiques de catégorisation : application au domaine olfactif. *ARCO'09, Interprétation et Problématiques Du Sens*, 83–91.
- Endres-Niggemeyer, B., & Wansorra, E. (2004). Making Cognitive Summarization Agents Work In A Real-World Domain. *NLUCS*, 86–96.
- Engen, T. (1982). *The perception of odors*. New York Academic Press.
- Engen, T., & Ross, B. M. (1973). Long-term memory of odors with and without verbal descriptions.

- Journal of Experimental Psychology*, 100(2), 221.
- Fano, R. M., & Hawkins, D. (1961). Transmission of information: A statistical theory of communications. *American Journal of Physics*, 29(11), 793–794.
- Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality and Preference*, 15(7), 781–791.
- Ferdenzi, C., Roberts, S. C., Schirmer, A., Delplanque, S., Cekic, S., Porcherot, C., ... Grandjean, D. (2013). Variability of Affective Responses to Odors: Culture, Gender, and Olfactory Knowledge. *Chemical Senses*, 38(2), 175–186.
- Ferragina, P., & Scaiella, U. (2010). Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*.
- Fiorini, N. (2015). *Semantic similarities at the core of generic indexing and clustering approaches. Thèse de doctorat*. Université de Montpellier, France.
- Firestein, S., Picco, C., & Menini, A. (1993). The relation between stimulus and response in olfactory receptor cells of the tiger salamander. *The Journal of Physiology*, 468, 1–10.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.
- Gains, N., & Thomson, D. (1990). Contextual evaluation of canned lagers using repertory grid method. *International Journal of Food Science and Technology*, 25(6), 699–705.
- Gawel, R. (1997). The use of language by trained and untrained experienced wine tasters. *Journal of Sensory Studies*, 12(4), 267–284.
- Gentleman, R. (2005). Visualizing and distances using GO. URL [Http://www. Bioconductor.Org/docs/vignettes. Html](http://www.Bioconductor.Org/docs/vignettes.Html), 38.
- Gerland, C., & Dumont, A. (2000). Analyse sensorielle descriptive libre choix (FCP): utilisation pratique: De la dégustation à l'analyse sensorielle. *Revue Française D'oenologie*, (182), 36–39.
- Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2), 217-288.
- Harispe, S., Imoussaten, A., François, T., & Jacky, M. (2015). On the consideration of a bring-to-mind model for computing the Information Content of concepts defined into ontologies. *International Conference on Fuzzy Systems (FUZZ IEEE), Istanbul, Turkey*.
- Harispe, S., Medjkoune, M., & Montmain, J. (2017). Eliciting Implicit Evocations using Word Embeddings and Knowledge Representation. *11th International Conference on Scalable Uncertainty Management*, 1–14.
- Harispe, S., Montmain, J., & Medjkoune, M. (2017). Summarizing Conceptual Descriptions using Knowledge Representations. *IEEE Symposium Series on Computational Intelligence*,.
- Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2013). The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. *Bioinformatics*, 30(5), 740–742.
- Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2015). Semantic Similarity from Natural Language and Ontology Analysis. *Synthesis Lectures on Human Language Technologies*, 8(1), 1–254.
- He, J. (2016). *Analysis of Word Vector Representations for the Automatic Procedures Relating to*

*Psychosensorial Analysis.*

- Herz, R. S., & Engen, T. (1996). Odor memory: Review and analysis. *Psychonomic Bulletin and Review*, 3(3), 300–313.
- Heymann, H. (1994). A Comparison of Free Choice Profiling and Multidimensional Scaling of Vanilla Samples. *Journal of Sensory Studies*, 9(4), 445–453.
- Holley, A. (1997). Le physiologiste et la catégorisation des odeurs. *Intellectica*, 24(1), 21–27.
- Howes, D. (1986). Le sens sans parole: vers une anthropologie de l'odorat.
- Hudson, R. (1999). From molecule to mind: the role of experience in shaping olfactory function. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, 185(4), 297–304.
- Hudson, R., & Distel, H. (2003). La nature subjective du monde des odeurs : considérations théoriques et conséquences méthodologiques. *Psychologie Française*, 48(4), 103–115.
- Hughes, T., & Ramage, D. (2007). Lexical Semantic Relatedness with Random Graph Walks. In : *EMNLP-CoNLL*, 581–589.
- Hughson, A. L., & Boakes, R. A. (2002). The knowing nose: The role of knowledge in wine expertise. *Food Quality and Preference*, 13(7-8), 463–472.
- Jack, F.-R., & Piggott, J. (1992). Free choice profiling in consumer research. *Food Quality and Preference*, 3(3), 129–134.
- Jaubert, J. noel, Gordon, G., & Doré, J. C. (1987a). Une organisation du champ des odeurs. Part 1 : Recherche de critères objectifs. *Parfums, Cosmétiques, Arômes*, 77, 53–56.
- Jaubert, J. noel, Gordon, G., & Doré, J. C. (1987b). Une organisation du champ des odeurs. Part 2 : Modèle descriptif de l'organisation de l'espace odorant. *Parfums, Cosmétiques, Arômes*, 78, 71–82.
- Jaubert, J. noel, Tapiero, C., & Dore, J. C. (1995). The field of odors: toward a universal language for odor relationships. *Perfumer & Flavorist*, 20(3), 1–16.
- Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *ArXiv Preprint Cmp-lg/9709008*.
- Jimeno-Yepes, A., Mork, J. G., Wilkowski, B., Demner-Fushman, D., & Aronson, A. R. (2012). MEDLINE MeSH Indexing : Lessons Learned from Machine Learning and Future Directions Categories and Subject Descriptors. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 737–742.
- Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, 126(3), 248.
- Jönsson, F. (2005). *Olfactory Metacognition A Metamemory Perspective on Odor Naming*. Acta Universitatis Upsaliensis.
- Lachnit, M., Busch-Stockfisch, M., Kunert, J., & Krahl, T. (2003). Suitability of free choice profiling for assessment of orange-based carbonated soft-drinks. *Food Quality and Preference*, 14(4), 257–263.
- Lawless, H., & Engen, T. (1977). Associations to odors: interference, mnemonics, and verbal labeling. *Journal of Experimental Psychology: Human Learning and Memory*, 3(1), 52.
- Lawless, H., & Heymann, H. (1999). Descriptive analysis. *Sensory Evaluation of Food*. Springer US,

341–378.

- Lawrence, G., Symoneaux, R., Maitre, I., Brossaud, F., Maestrojuan, M., & Mehinagic, E. (2013). Using the free comments method for sensory characterisation of Cabernet Franc wines : Comparison with classical profiling in a professional context. *Food Quality and Preference*, 30, 145–155.
- Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. *WordNet: An Electronic Lexical Database*, 49(2), 265–283.
- Lebart, L. (2012). Processing texts and open-ended questions in sample surveys. *Sensometrics, Agrocampus Ouest, Rennes, France*.
- Lebret, R., & Collobert, R. (2014). Word embeddings through Hellinger PCA. *In EACL*.
- Lefebvre, A., & Bassereau, J. F. (2003). L'analyse sensorielle , une méthode de mesure au service des acteurs de la conception : ses avantages , ses limites , ses voies d'amélioration. *Application Aux Emballages*, 10, 3–11.
- Léger, C. (2005). Smells supervision setting by AIR NORMAND, air pollution monitoring network Smells, straight perceptible nuisances. *Pollution Atmospherique*, 47(187), 373–384.
- Li, W., Wu, M., Lu, Q., Xu, W., & Yuan, C. (2006). Extractive summarization using inter-and intra-event relevance. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, 369–376.
- Lin, D., & Others. (1998). An Information-Theoretic Definition of Similarity. *Proceedings of ICML*, 98(1998), 296–304.
- Lin, H., & Bilmes, J. (2011). A Class of Submodular Functions for Document Summarization. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 510–520.
- Lledo, P. M., & Vincent, J. D. (1999). Odeurs. *Médecine/sciences*, 15(11), 1211.
- Lyman, B. J., & McDaniel, M. A. (1990). Memory for odors and odor names: modalities of elaboration and imagery. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16(4), 656.
- Mac Leod, P., Sauvageot, F., & Köster, E. (1998). Les caractéristiques d'une réponse sensorielle. *Manuel Méthodologique D'évaluation Sensorielle", SSHA 2ème Édition P*, 7–9.
- Maedche, A., & Staab, S. (2001). Comparing ontologies-similarity measures and a comparison study. AIFB.
- Maille, V. (2000). La réponse du consommateur au produit parfumé: l'incidence de la congruence de l'odeur. *Atti Del L Convegno "Le Tendenze Del Markeing in Europa,"* 24.
- Mainland, J. D., Lundström, J. N., Reisert, J., & Lowe, G. (2014). From molecule to mind: an integrative perspective on odor intensity. *Trends in Neurosciences*, 37(8), 443–54.
- Mani, I., House, D., Klein, G., Hirschman, L., & Sundheim, B. (1999). The TIPSTER SUMMAC Text Summarization Evaluation. *Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics*.
- Martin, N., & Rogeaux, M. (1994). Etude par analyse textuelle de commentaires de consommateurs après dégustation de boissons. *Sciences des aliments*, 14(3), 265–280.
- Mazandu, G. K., & Mulder, N. J. (2011). *IT-GOM: An Integrative Tool for IC-based GO Semantic Similarity Measures. Technical report, University of Cape Town (South Africa)*.

- Mazandu, G. K., & Mulder, N. J. (2013). Information content-based gene ontology semantic similarity approaches: toward a unified framework theory. *BioMed Research International*, 2013.
- Mcdonnell, E., Hulin-Bertaud, S., Sheehan, E. ., & Delahunty, C. (2001). Development and learning process of a sensory vocabulary for the odor evaluation of selected distilled beverages using descriptive analysis. *Journal of Sensory Studies*, 16(4), 425–445.
- McKenzie, D. (2013). *Aromatics and the Soul. A Study of Smells*. Elsevier.
- Medelyan, O., Witten, I. H., & Milne, D. (2008). Topic indexing with Wikipedia. *Proceedings of the AAAI WikiAI Workshop*, 1, 19–24.
- Meilgaard, M. C., Carr, B. T., & Civille, G. V. (2006). Sensory evaluation techniques. *CRC Press*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.
- Mohammad, S. M., & Hirst, G. (2012). Distributional Measures of Semantic Distance: A Survey. *arXiv Preprint arXiv:1203.1858*.
- Morrison, E. E., & Costanzo, R. M. (1990). Morphology of the human olfactory epithelium. *The Journal of Comparative Neurology*, 297(1), 1–13.
- Narain, C., Paterson, A., & Reid, E. (2004). Free choice and conventional profiling of commercial black filter coffees to explore consumer perceptions of character. *Food Quality and Preference*, 15, 31–41.
- Nicod, H., Kieser, S., Bremaud, D., & Nante, A. (2006). L'expertise sensorielle : leurre ou réalité. *Proceedings of the Le Sensolier, Les Expertises Sensorielles: Nature et Acquisition*. Paris. *Le Sensolier*.
- Noble, A. C., Arnold, R., Masuda, B. M., Pecore, S., Schmidt, J., & Stern, P. (1984). Progress Towards a Standardized System of Wine Aroma Terminology. *American Journal of Enology and Viticulture*, 35(2), 107–109.
- Organisation-Internationale, de N., & Others. (2005). Analyse Sensorielle. Méthodologie. Lignes directrices générales. *Recueil Des Normes Françaises, Produits Alimentaire, Analyse Sensorielle, NF ISO 13301 (V09-023)*.
- Panchenko, A., & Others. (2013). *Similarity measures for semantic relation extraction*. Thèse de doctorat. Université catholique de Louvain and Bauman Moscow State Technical University.
- Paravisini, L., Septier, C., Moretton, C., Nigay, H., Arvisenet, G., Guichard, E., & Dacremont, C. (2014). Caramel odor: Contribution of volatile compounds according to their odor qualities to caramel typicality. *Food Research International*, 57, 79–88.
- Parr, W. V, Heatherbell, D., & White, K. G. (2002). Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses*, 27(8), 747–755.
- Parr, W. V, White, K. G., & Heatherbell, D. A. (2004). Exploring the nature of wine expertise: what underlies wine experts' olfactory recognition memory advantage? *Food Quality and Preference*, 15(5), 411–420.
- Paul, R., Groza, T., Zankl, A., & Hunter, J. (2012). Semantic similarity-driven decision support in the skeletal dysplasia domain. In *International Semantic Web Conference*. Springer, Berlin Heidelberg, 164–179.

- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *EMNLP*, 14, 1532–1543.
- Perrin, L., Symoneaux, R., Maitre, I., & Asselin, C. (2007). Apport Des Methodes Spontanees Pour L'Analyse Sensorielle Des Vins, 1–6.
- Pesquita, C., Faria, D., Bastos, H., Ferreira, A. E., Falcão, A. O., & Couto, F. M. (2007). Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC Bioinformatics*, 9(5), 1.
- Pirró, G. (2009). A semantic similarity metric combining features and intrinsic information content. *Data & Knowledge Engineering*, 68(11), 1289–1308.
- Pirró, G., & Euzenat, J. (2010). A semantic similarity framework exploiting multiple parts-of speech. In *OTM Confederated International Conferences On the Move to Meaningful Internet Systems. Springer Berlin Heidelberg*, 1118–1125.
- Poncelet, J. (2010). *Influence de l'expérience précoce et des connaissances sémantiques acquises sur la perception hédonique de l'odeur : études psychophysiques et neurophysiologiques. Thèse de doctorat. Université Claude Bernard Lyon 1, France.*
- Price, S. (1987). Effects of odorant mixtures on olfactory receptor cells. *Annals of the New York Academy of Sciences*, 510, 55–60.
- Qualité de l'air. (1996). Mesurage de l'odeur d'une atmosphère gazeuse. Méthode supraliminaire. *Norme AFNOR NF X 43-103*.
- Rabin, M. D. (1988). Experience facilitates olfactory quality discrimination. *Perception & Psychophysics*, 44(6), 532–540.
- Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(1), 17–30.
- Ramage, D., Rafferty, A. N., & Manning, C. D. (2009). Random walks for text semantic similarity. *Proceedings of the 2009 Workshop on Graph-Based Methods for Natural Language Processing*, 23–31.
- Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *arXiv Preprint Cmp-ig/9511007*.
- Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the Association for Information Science and Technology*, 27(3), 129-146.
- Rodríguez, M. A., & Egenhofer, M. J. (2003). Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2), 442–456.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439.
- Rouby, C., Schaal, B., Dubois, D., Gervais, R., & Holley, R. (2002). *Olfaction, taste, and co.*
- Rouby, C., & Sicard, G. (1997). Des catégories d'odeurs. *Catégorisation et Cognition : De La Perception Au Discours*, 95–81.
- Rouby, C., Thomas-Danguin, T., Sicard, G., Vigouroux, M., Jiang, T., Poitevineau, J., & Issanchou, S. (2005). Influence du contexte sémantique sur la performance d'identification d'odeurs. *Psychologie Française*, 50(2), 225–239.
- Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Disability Studies*, 20(1), 33.–53.

- Salton, G., & McGill, M. J. (1986). Introduction to modern information retrieval.
- Sánchez, D., & Batet, M. (2011). Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. *Journal of Biomedical Informatics*, 44(5), 749–759.
- Sauvageot, F. (1990). Les caractéristiques d'une réponse sensorielle – Intensité. Evaluation sensorielle, Manuel Méthodologique, SSHA., *Lavoisier Tec & Doc, Apria*, 15–31.
- Schaal, B., Rouby, C., Marlier, L., Soussignan, R., Kontar, F., & Tremblay, R. (1998). Variabilité et universaux au sein de l'espace perçu des odeurs : approches inter-culturelles de l'hédonisme olfactif. *Géographie Des Odeurs*, 25–47.
- Schickel-Zuber, V., & Faltings, B. (2007). OSS: A Semantic Similarity Function based on Hierarchical Ontologies. *IJCAI*, 7, 551–556.
- Schlicker, A., Domingues, Francisco Rahnenführer, J., & Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, 7(1), 302.
- Schnabel, T., Labutov, I., Mimno, D. M., & Joachims, T. (2015). Evaluation methods for unsupervised word embeddings. *EMNLP*, 298–307.
- Seco, N., Veale, T., & Hayes, J. (2004). An intrinsic information content metric for semantic similarity in wordnet. In *16th European Conference on Artificial Intelligence (ECAI 2004)*, 16, 1–5.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423.
- Sinha, P., & Jain, R. (2011). Extractive summarization of personal photos from life events. In *2011 IEEE International Conference on Multimedia and Expo* (pp. 1–6). IEEE. <http://doi.org/10.1109/ICME.2011.6011843>
- Sinha, P., Mehrotra, S., & Jain, R. (2011). Summarization of personal photologs using multidimensional content and context. *International Conference on Multimedia Retrieval*, 1–8.
- Solomon, G. E. A. (1990). Psychology of novice and expert wine talk. *The American Journal of Psychology*, 103, 495–517.
- Standard, B., & Others. (2003). Standard, British and others. *BS EN, 13725*, 2003.
- Suffet, I. H., & Rosenfeld, P. (2007). The anatomy of odour wheels for odours of drinking water , wastewater , compost and the urban environment. *Water Science & Technology*, 55(5), 335–344.
- Symoneaux, R., Galmarini, M. V, & Mehinagic, E. (2012). Comment analysis of consumer's likes and dislikes as an alternative tool to preference mapping. A case study on apples. *Food Quality and Preference*, 24(1), 59–66.
- Tempère, S. (2010). *Métrologie sensorielle olfactive et apprentissage olfactif appliqués à l'œnologie*. Université Lumière Lyon 2.
- Tschiatschek, S., Iyer, R. K., Wei, H., & Bilmes, J. A. (2014). Learning Mixtures of Submodular Functions for Image Collection Summarization. *Neural Information Processing Systems (NIPS)*, 1–9.
- Tseng, V. S., Su, J. H., Huang, J. H., & Chen, C. en. (2008). Integrated mining of visual features, speech features, and frequent patterns for semantic video annotation. *IEEE Transactions on Multimedia*, 10(2), 260–267.
- Turenne, N. (2000). *Apprentissage statistique pour l' extraction de concepts à partir de textes . Application au filtrage d'informations textuelles*. Sciences-New York.

- Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), 467–476.
- Valentin, D., & Chollet, S. (2015). Les mots du vin : experts et novices diffèrent-ils quand ils décrivent des vins ?, 2(2003).
- Valentin, D., Chollet, S., & Abdi, H. (2003). Les mots du vin : experts et novices diffèrent-ils quand ils décrivent des vins ? *Corpus*, 2(2), 183–200.
- Williams, A. A., & Arnold, G. M. (1985). A comparison of the aromas of six coffees characterised by conventional profiling, free choice profiling and similarity scaling methods. *Journal of the Science of Food and Agriculture*, 36(3), 204–214.
- Williams, A. A., & Langron, S. P. (1984). The use of free choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture*, 35(5), 558–568.
- Wright, R. (1978). The perception of odor intensity: physics or psychophysics? *Chemical Senses*, 3(1), 73–79.
- Wu, Z., & Palmer, M. (1994). *Verbs semantics and lexical selection. Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics.
- Yu, G., Pei, Y., Zhu, Z., & Chen, H. (2006). Research of text similarity based on word similarity computing. *Computer Engineering and Design*, 2, 67–70.
- Zhang, S., Tian, Q., Hua, G., Huang, Q., & Gao, W. (2014). ObjectPatchNet: Towards scalable and semantic image annotation and retrieval. *Computer Vision and Image Understanding*, 118, 16–29.
- Zhou, Z., Wang, Y., & Gu, J. (2008). A new model of information content for semantic similarity in WordNet. *Future Generation Communication and Networking Symposia, 2008. FGCNS'08. Second International Conference on*, 3, 85–89.

## Résumé

Caractériser la qualité d'une odeur est une tâche complexe qui consiste à identifier un ensemble de descripteurs qui synthétise au mieux la sensation olfactive au cours de séances d'analyse sensorielle. Généralement, cette caractérisation est une liste de descripteurs extraite d'un vocabulaire imposé par les industriels d'un domaine pour leurs analyses sensorielles. Ces analyses représentent un coût significatif pour les industriels chaque année. En effet, ces approches dites orientées reposent sur l'apprentissage de vocabulaires, limitent singulièrement les descripteurs pour un public non initié et nécessitent de coûteuses phases d'apprentissage. Si cette caractérisation devait être confiée à des évaluateurs naïfs, le nombre de participants pourrait être significativement augmenté tout en réduisant le coût des analyses sensorielles. Malheureusement, chaque description libre n'est alors plus associée à un ensemble de descripteurs non ambigus, mais à un simple sac de mots en langage naturel (LN). Deux problématiques sont alors rattachées à la caractérisation d'odeurs. La première consiste à transformer des descriptions en LN en descripteurs structurés ; la seconde se donne pour objet de résumer un ensemble de descriptions formelles proposées par un panel d'évaluateurs en une synthèse unique et cohérente à des fins industrielles. Ainsi, la première partie de notre travail se focalise sur la définition et l'évaluation de modèles qui peuvent être utilisés pour résumer un ensemble de mots en un ensemble de descripteurs désambiguïsés. Parmi les différentes stratégies envisagées dans cette contribution, nous proposons de comparer des approches hybrides exploitant à la fois des bases de connaissances et des plongements lexicaux définis à partir de grands corpus de textes. Nos résultats illustrent le bénéfice substantiel à utiliser conjointement représentation symbolique et plongement lexical. Nous définissons ensuite de manière formelle le processus de synthèse d'un ensemble de concepts et nous proposons un modèle qui s'apparente à une forme d'intelligence humaine pour évaluer les résumés alternatifs au regard d'un objectif de synthèse donné. L'approche non orientée que nous proposons dans ce manuscrit apparaît ainsi comme l'automatisation cognitive des tâches confiées aux opérateurs des séances d'analyse sensorielle. Elle ouvre des perspectives intéressantes pour développer des analyses sensorielles à grande échelle sur de grands panels d'évaluateurs lorsque l'on essaie notamment de caractériser les nuisances olfactives autour d'un site industriel.

**Mots clés :** Représentation des connaissances, Synthèse conceptuelle, Fusion d'information, Traitement Automatique du Langage Naturel, Analyse sensorielle, Taxonomie.

## Abstract

Characterizing the quality of smells is a complex process that consists in identifying a set of descriptors best summarizing the olfactory sensation. Generally, this characterization results in a limited set of descriptors provided by sensorial analysis experts. These sensorial analysis sessions are however very costly for industrials. Indeed, such oriented approaches based on vocabulary learning limit, in a restrictive manner, the possible descriptors available for any uninitiated public, and therefore require a costly vocabulary-learning phase. If we could entrust this characterization to neophytes, the number of participants of a sensorial analysis session would be significantly enlarged while reducing costs. However, in that setting, each individual description is not related to a set of non-ambiguous descriptors anymore, but to a bag of terms expressed in natural language (NL). Two issues are then related to smell characterization implementing this approach. The first one is how to translate such NL descriptions into structured descriptors; the second one being how to summarize a set of individual characterizations into a consistent and synthetic unique characterization meaningful for professional purposes. Hence, this work focuses first on the definition and evaluation of models that can be used to summarize a set of terms into unambiguous entity identifiers selected from a given ontology. Among the several strategies explored in this contribution, we propose to compare hybrid approaches taking advantages of knowledge bases (symbolic representations) and word embeddings defined from large text corpora analysis. The results we obtain highlight the relative benefits of mixing symbolic representations with classic word embeddings for this task. We then formally define the problem of summarizing sets of concepts and we propose a model mimicking Human-like Intelligence for scoring alternative summaries with regard to a specific objective function. Interestingly, this non-oriented approach for identifying the quality of odors appears to be an actual cognitive automation of the task today performed by expert operators in sensorial analysis. It therefore opens interesting perspectives for developing scalable sensorial analyses based on large sets of evaluators when assessing, for instance, olfactory pollution around industrial sites.

**Key words:** Knowledge Representation, Conceptual synthesis, Information fusion, Automatic Natural Language Processing, Sensory analysis, Taxonomy.