

## Introduction

La majeure partie des recherches que nous allons présenter ont été conduites dans le cadre d'un projet de recherche pluridisciplinaire, nommé « Perception intelligente et traitement cognitif des odeurs ». Ce projet implique notre laboratoire, le LIP, ainsi que plusieurs partenaires : le Laboratoire de Physiologie Neurosensorielle (LPN) de l'Université Claude Bernard Lyon 1 ; l'Unité de Neuropsychologie Cognitive (UNPC) de l'Université Lumière Lyon 2 ; l'Équipe de Recherche en Ingénierie des Connaissances (ÉRIC) de l'Université Lumière Lyon 2 ; et le Laboratoire des Méthodes d'Analyse des Systèmes et des Structures (MASS) de l'Université Claude Bernard Lyon 1. La section 7.1 présente une version simplifiée du modèle des processus de mémorisation à la base de ce projet.

Nous avons conduit une réflexion et réalisé des expérimentations informatiques sur les deux axes du projet : « les mécanismes cognitifs », notamment l'amorçage de répétition ; et le « traitement cognitif des odeurs ».

L'essentiel du projet a porté sur le deuxième axe, c'est-à-dire la problématique de la mémorisation des odeurs que nous présentons dans ce chapitre. Nous allons utiliser le classifieur comme outil d'analyse, sur un problème « difficile à apprendre ». Nous commençons par décrire des expérimentations réalisées auprès de 174 sujets humains par Royet et son équipe du LPN (section 6.1), et notamment le protocole mis en œuvre. Les résultats du classifieur sur une tâche de classification d'odeurs sont présentés dans les sections 6.2 et 6.3. La section 6.4 analyse la base d'exemples afin d'en mesurer la complexité, ce travail conduira à la présentation d'un nouvel algorithme d'apprentissage pour le classifieur (section 6.6) adapté aux problèmes difficiles. Enfin, à titre de comparaison, nous présentons les résultats obtenus par un autre modèle connexionniste : « réseau d'ondelettes » (section 6.7).

Le chapitre 7 est entièrement consacré au premier axe, notamment la simulation d'un amorçage de répétition et une proposition d'implémentation d'une « mémoire associative hétérogène ».

*Travail  
effectué au  
LIP*

## 6.1 Expérimentations auprès des sujets

### *Stimulateur olfactif*

Lors de toutes les expérimentations que nous présentons dans ce chapitre, les odeurs sont présentées aux sujets grâce à un « stimulateur olfactif » mis au point par Vigouroux et Farget, du LPN. Le système contrôle la respiration du sujet grâce à une sonde thermique et envoie une stimulation durant la phase respiratoire. Ce système est complexe mais d'une grande précision, il est par exemple possible de faire des mesures très précises des temps de réponses des sujets.

De nombreuses réunions des équipes impliquées dans le projet de modélisation des processus de mémorisation olfactifs ont permis de discuter et d'établir le protocole des expériences olfactives.

### *Interface d'acquisition des données*

Lors de ces tâches préparatoires, notre part du travail a notamment consisté en la mise au point d'une application (sur micro-ordinateur) pour contrôler les expériences et faire l'acquisition des données : minutage des expériences, informations données aux sujets, pilotage du stimulateur olfactif, acquisition des réponses directement sur un écran tactile, *etc.* Ce travail a été long car les études prévues par le LPN et l'UNPC demandaient une mesure extrêmement précise des temps de réponse des sujets (études non présentées dans cette thèse). Les contraintes de temps réel, ainsi que la nécessité d'un accès de « bas niveau » à l'horloge de l'ordinateur pour une mesure précise du temps, nous ont obligé à écrire notre propre interface graphique.

### *Acquisition des bases d'exemples auprès des sujets*

Une fois les expériences réalisées, nous avons utilisé le classifieur afin d'analyser les données recueillies. Avant de détailler ce travail, il est nécessaire de présenter les expériences réalisées et notamment les protocoles suivis. Les 174 sujets sont répartis en cinq groupes selon la nature des « descripteurs » qui sont à leur disposition pour qualifier les odeurs qu'on leur fait sentir. Ces descripteurs sont des mots du langage courant permettant de décrire l'odeur. Pour chaque sujet de chaque groupe, et pour chacune des 30 odeurs proposées, l'expérience est constituée de deux phases.

### *Première phase*

Durant la **première phase**, le sujet choisit les descripteurs qui lui conviennent pour l'odeur qu'on lui présente. Le choix est limité à une liste proposée à l'écran. Pour ne pas que l'expérience informatique du sujet influe sur les temps de réponse, la sélection s'effectue naturellement en pointant l'écran avec le doigt, sur une zone rectangulaire en regard du nom de l'odeur afin de préciser l'intensité du descripteur. La valeur numérique correspondante, un réel compris entre 0 et 1, n'est pas affichée pour ne pas perturber le sujet. En revanche l'espace situé à la gauche de la position choisie change de couleur pour confirmer l'acquisition. Ainsi, le sujet peut par exemple décrire une odeur « un peu sucrée » en pointant le début de la zone, l'ordinateur enregistrant que le descripteur « sucrée » a été choisi avec une intensité correspondant à la position pointée (par exemple 0.2125).

L'odeur ayant été décrite par les descripteurs à la première phase, au cours de la **seconde phase** le sujet donne à haute voix un nom à l'odeur sentie. La réponse est totalement libre, sans liste préalable. Cependant, il est très difficile pour un non-spécialiste de nommer une odeur. Aussi, nous n'attendons pas des sujets le nom exact de l'odeur. Royer et son équipe ont étudié chaque cas pour juger la validité de la réponse. Les réponses sont séparées en trois catégories selon que le sujet a proposé : une réponse jugée correcte (notée *c*) ; une réponse jugée fautive (noté *w*) ; ou pas de réponse du tout (notée *n*).

*Seconde phase*  
*Appréciation des réponses*

Le but de l'expérience est de déterminer si l'aptitude à bien distinguer les odeurs (en trouvant le nom correct) dépend de la profondeur de la réflexion demandée au sujet. La profondeur de réflexion est conditionnée par la liste des descripteurs proposée aux sujets. Les différentes versions de cette liste définissent les cinq groupes : (i) un groupe, noté GS, qui ne se voit proposé que 3 descripteurs qualifiés de « superficiels » (« intense », « agréable », et « familier ») ; (ii) un groupe noté GP, avec une liste de descripteurs plus « profonds » donnée dans la table 6.1 ; (iii) un groupe GSP, répondant avec les descripteurs des groupes GS et GP réunies ; (iv) un autre groupe profond, noté GP2, qui dispose des descripteurs différents de ceux du groupe GP<sup>1</sup> ; (v) et un groupe témoin noté GT ne fait que nommer l'odeur (c'est-à-dire la seconde phase). Les données des groupes GP2 et GT n'ont pas été utilisées dans notre étude informatique, leur présence est motivée par les besoins des psychologues.

*Objectifs de l'expérience*

1. Boisée	5. Fruitée	9. Acide	13. Entêtante
2. Corporelle	6. Fumée	10. Amère	14. Fraîche
3. Epicée	7. Herbacée	11. Salée	15. Graisseuse
4. Florale	8. Médicinale	12. Sucrée	16. Irritante

TAB. 6.1 – Liste des descripteurs proposés au groupe profond GP

La table 6.2 résume le nombre de descripteurs mis à la disposition des sujets de chaque groupe.

*Effectifs*

Nom du groupe	GS	GP	GSP
Nombre de descripteurs	3	16	19

TAB. 6.2 – Nombres de descripteurs pour les groupes utilisés

1. Notre vocabulaire est très pauvre pour décrire une odeur. Les descripteurs de GP2 caractérisent des sensations auditives (une odeur peut par exemple être décrite comme « bourdonnante »), de formes (par exemple « allongée ») et de couleurs (« colorée », « verte », *etc.*).

## 6.2 Utilisation du classifieur

Nous allons essayer d'apprendre au classifieur à discriminer les odeurs à l'aide des descripteurs. La tâche est clairement difficile, le but n'est donc pas de réaliser un classifieur d'odeurs, mais plutôt de juger dans quelles mesure les descripteurs permettent de classifier les odeurs. Le travail est donc plus axé sur l'étude des données que sur la recherche de bons taux de reconnaissance.

*Liste des odeurs*

Pour apprendre, nous avons besoin de données. La table 6.3 donne les noms des 30 odeurs présentées aux sujets. Certaines odeurs, marquées d'une étoile, ont été particulièrement mal reconnues par la plupart des sujets. En raison des très faibles effectifs de la catégorie des bonnes réponses, nous éliminons ces 5 odeurs de notre base d'exemples.

1. Lilas	11. Transpiration	21. Saumon fumé
2. Menthe	12. Citronnelle	22. Lavande
3. Champignon	13. Thym	23. Vernis à ongles
4. Poivre *	14. Orange	24. Anis
5. Camphre	15. Chocolat *	25. Banane
6. Fruit de la passion	16. Ether	26. Goudron
7. Rose	17. Pêche	27. Cannelle *
8. Herbe	18. Fraise	28. Verveine
9. Caramel	19. Pin *	29. Amande amère
10. Clou de girofle	20. Vinaigre *	30. Citron

TAB. 6.3 – *Noms des odeurs*

*Bases d'apprentissage et de généralisation*

Nous ne pouvons pas utiliser toutes les données pour l'apprentissage. En effet, nous avons besoin d'un ensemble de généralisation pour estimer la qualité de l'apprentissage, or nous avons vu que ces deux ensembles doivent être disjoints. Aussi, les réponses des sujets pour les 25 odeurs que nous avons conservées sont réparties à 80 % pour l'ensemble d'apprentissage et à 20 % pour l'ensemble de généralisation. La table 6.4 résume le nombre d'exemples restant disponibles pour l'apprentissage, pour chaque groupe de sujets (GS, GP, et GSP) et pour chaque catégorie de réponses (correcte, fausse, et l'abstention).

Groupe	Nb d'exemples	c (correct)	w (faux)	n (non rép.)
GS	850	435 (51%)	208 (25%)	207 (24%)
GP	825	440 (53%)	304 (37%)	81 (10%)
GSP	825	459 (56%)	240 (29%)	126 (15%)

TAB. 6.4 – *Nombres d'exemples, selon les groupes et la nature de la réponse*

Chaque exemple comprend une liste de descripteurs sélectionnés avec leur intensité. De plus, tous les exemples sont associés à une classe (c'est-à-dire à un nom d'odeur). Cette classe correspond toujours à l'odeur présentée. Aussi, les descripteurs ayant conduit à une mauvaise dénomination (catégorie  $w$ ) ou à pas de dénomination du tout (catégorie  $n$ ) sont étiquetés avec le nom correct.

Les exemples présentés au classifieur sont donc des vecteurs de réels (notés  $X$ ) reprenant comme composantes les intensités des descripteurs (dans un ordre donné). La valeur 0 est attribuée à l'intensité lorsque le descripteur n'a pas été choisi.

### 6.3 Résultats des tentatives de classification

Dans les sections qui suivent, nous utilisons comme annoncé le classifieur neuronal incrémental sur les données collectées. Le but que l'on se fixe est de faire apprendre au réseau à associer l'odeur qui convient à une liste de descripteurs donnée.

Les résultats sont présentés dans la table 6.6 pour chacun des trois groupes de sujets définis par le protocole expérimental: le groupe superficiel GS, le groupe profond GP et le groupe superficiel-profond GSP. L'objectif étant de savoir si le classifieur parvient à discriminer les odeurs à l'aide des descripteurs, les prototypes créés sont testés avec la base de généralisation.

*Présentation  
des résultats*

Plusieurs expériences sont conduites: en utilisant tous les exemples disponibles; uniquement les exemples pour lesquels la réponse du sujet a été jugée correcte (catégorie  $c$ ); et en utilisant tous les exemples pour lesquels la réponse du sujet n'était pas fautive (catégories  $c$  et  $n$ ). L'utilisation de ce dernier jeu de réponses permet d'utiliser plus d'exemples qu'en se limitant aux bonnes réponses (catégorie  $c$ ) sans pour autant dégrader significativement la base d'exemples. En effet, le fait que le sujet ne soit pas parvenu à nommer l'odeur ne signifie pas forcément que les descripteurs qu'il a choisis soient mauvais. La table 6.5 reprend et explique l'entête de chaque colonne.

*Expériences  
réalisées*

Entête	Descriptif
Catégorie	Catégorie de la réponse ( $c$ , $w$ , ou $n$ ) des exemples appris
Nb ex.	Nombre d'exemples appris de la catégorie
<b>Nb proto.</b>	Nombre de prototypes créés par le classifieur (en caractères gras)
Succès	Nombre de bonnes réponses du classifieur
Erreurs	Nombre de mauvaises réponses du classifieur
Rejets	Nombre d'exemples testés et rejetés par le classifieur

TAB. 6.5 – Légende de la table 6.6 des résultats

Groupe	Catégorie	Nb ex.	Nb proto.	Succès	Erreurs	Rejets
GS	cwn	850	<b>764</b>	14 (7%)	178	8
	c	435	<b>392</b>	2 (2%)	95	9
	cn	642	<b>582</b>	6 (4%)	138	15
GP	cwn	825	<b>655</b>	27 (14%)	160	13
	c	440	<b>342</b>	22 (20%)	83	4
	cn	521	<b>402</b>	25 (19%)	98	8
GSP	cwn	825	<b>632</b>	34 (17%)	157	9
	c	459	<b>343</b>	28 (25%)	81	2
	cn	585	<b>441</b>	29 (20%)	112	6

TAB. 6.6 – Résultats de la tentative de classification des odeurs sur la base de test, pour chaque groupe (GS, GP, et GSP), et en fonction de la catégorie des réponses

#### Analyse des résultats

Ces résultats montrent que le classifieur ne parvient pas à nommer correctement les odeurs : les taux de succès sur les bases de test sont au mieux autour de 20 %.

#### Résultats relatifs

On constate que les performances du classifieur sont beaucoup plus faibles pour le groupe superficiel GS que pour les deux groupes de sujets ayant eu la possibilité d'effectuer un traitement plus profond. Cette observation est confirmée par les résultats de Sebban et Zighed [SZ96b] du laboratoire ÉRIC. Ce phénomène trouve une réponse à la fois aux niveaux des sujets et du classifieur :

#### Traitement cognitif des sujets

- Au niveau des sujets, le phénomène est certainement lié à une meilleure possibilité de discrimination, à l'aide d'un plus large choix de descripteurs, de la part des sujets auxquels les odeurs ont été présentées.

#### Traitement algorithmique du classifieur

- Au niveau du classifieur, le phénomène s'explique par une nette différence de dimension des espaces d'entrée. Cette dimension, directement liée au nombre de descripteurs proposés lors de l'expérience, varie de 3 pour GS à 19 pour GSP (voir table 6.2).

#### Nombre de prototypes créés

Il est important de remarquer que le nombre de prototypes créés par le classifieur est toujours de 100 à 200 fois plus grand que le nombre de classes à distinguer (25). Ce nombre n'est que légèrement inférieur au nombre d'exemples présentés en apprentissage, ce qui signifie que presque tous les exemples conduisent l'algorithme à créer un nouveau prototype. Ce comportement diffère radicalement de ce que nous avons constaté pour le problème d'OCR ou pour la classification des formes d'ondes. Pourtant, les résultats ci-dessus sont les meilleurs qui ont pu être obtenus après réglage des paramètres du réseau séquentiel (c'est-à-dire les seuils  $\Theta_{influence}$  et  $\Theta_{confusion}$ ). C'est donc au niveau des exemples qui constituent la base qu'il faut chercher les causes de ce comportement.

## 6.4 Analyse des données olfactives

Nous allons donc étudier les prototypes créés par le classifieur afin d'analyser la base d'exemples. Seul les principaux résultats obtenus sont présentés, pour plus de détails un rapport de recherche a été rédigé [EPMPR95]. Ce travail a été réalisé en étroite collaboration avec Eberhardt alors en stage au LIP.

Pour commencer, considérons les exemples à apprendre. Quelle que soit l'odeur, les vecteurs exemples sont creux. En effet, un sujet ne désigne en moyenne qu'environ 3 descripteurs. De ce fait, même pour des sujets ayant correctement identifié l'odeur, les profils de descripteurs enregistrés sont parfois très différents d'un sujet à un autre, et parfois très semblables. On comprend mieux pourquoi l'apprentissage des odeurs à partir des profils de descripteurs n'est pas une tâche aisée.

La figure 6.1 montre des exemples de telles variations pour une même odeur. L'odeur considérée est facilement reconnue, il s'agit de l'odeur d'orange. En revanche, on constate sur la figure 6.2 que les sujets n'ayant pas reconnu l'orange, soit en donnant une réponse fausse ( $w$ ), soit en ne donnant aucune réponse ( $n$ ), ont défini des profils de descripteurs parfois assez proches de ceux des sujets ayant bien répondu, et d'autre part assez différents entre eux.

*Seulement 3 descripteurs choisis*

*Données correspondant à une même odeur*

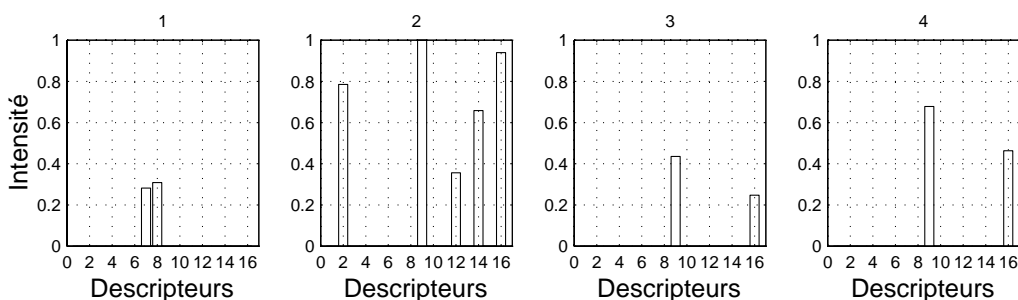


FIG. 6.1 – Quatre exemples de choix de descripteurs faits par quatre sujets distincts, ayant nommé l'odeur correctement, pour l'odeur d'orange

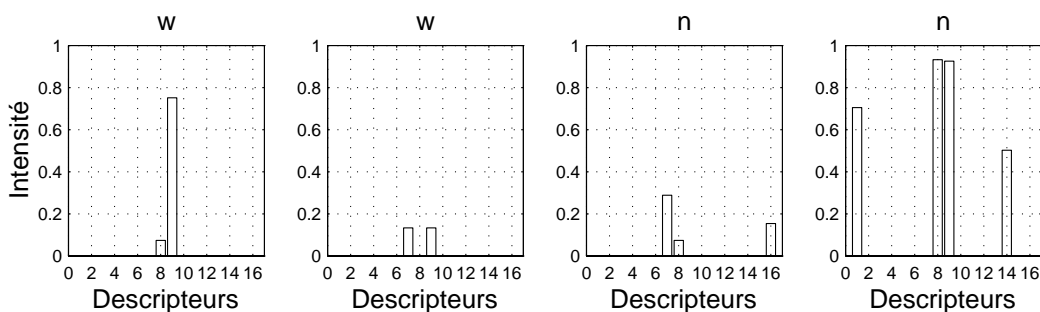


FIG. 6.2 – Quatre exemples de choix de descripteurs faits par quatre sujets distincts, ayant mal répondu ( $w$  = faux,  $n$  = pas de réponse), pour l'odeur d'orange

*Données correspondant à des odeurs différentes*

La difficulté de l'apprentissage apparaît encore plus nettement lorsque l'on compare certains exemples correspondant à des odeurs distinctes. En effet, la figure 6.3 met en valeur une forte similarité entre des données correspondant à différentes odeurs de fruits. La figure 6.4 montre que l'on peut même rencontrer des profils semblables pour des odeurs de nature plus distincte, telles que la menthe et le caramel.

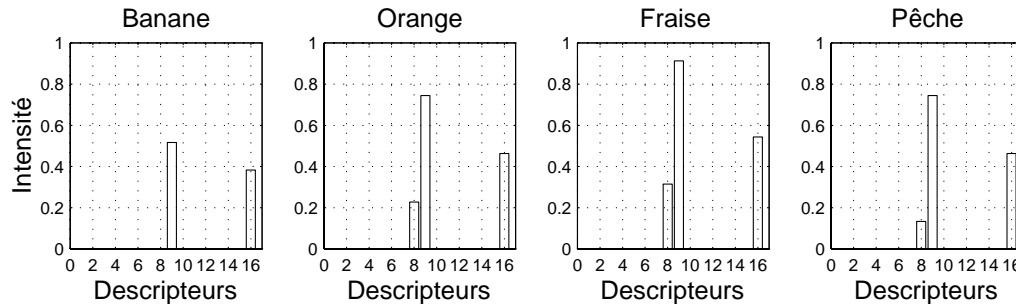


FIG. 6.3 – Quatre exemples très similaires pour quatre odeurs de fruit distinctes

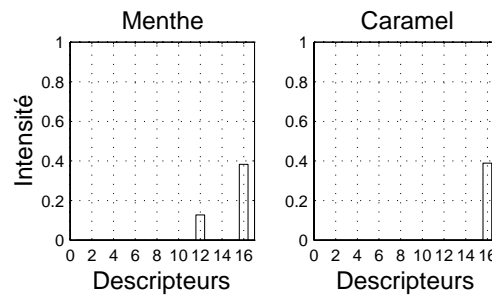


FIG. 6.4 – Deux exemples assez similaires pour deux odeurs très différentes

*Remarque*

Notons que certains exemples doivent quasiment se trouver sur une même droite de l'espace d'entrée (bien que celui-ci soit de dimension 16) alors qu'ils correspondent pourtant à des odeurs distinctes. Il suffit de reprendre les exemples d'orange numérotés 3 et 4 sur la figure 6.1 et de voir que l'odeur de banane de la figure 6.3 vient juste s'insérer entre leurs deux profils. On comprend alors mieux pourquoi le classifieur incrémental crée autant de prototypes.

## 6.5 Distances et coefficients de corrélation-croisée

La mesure utilisée par le classifieur est la distance euclidienne que nous avons déjà utilisée avec succès sur le problème des formes d'ondes. Cette mesure peut être directement appliquée aux données de la base d'exemples. Ainsi, la figure 6.5 permet de visualiser la distance entre deux odeurs au regard de la base d'exemples : l'orange et l'éther. La figure étant un peu complexe à lire, il est utile de comprendre sa construction : (i) on choisit une odeur, par exemple l'orange ; (ii) on calcule, pour chaque exemple de cette odeur, sa distance minimale à tous les autres exemples ;



(iii) on place une marque dont l'abscisse indique le numéro de l'odeur la plus proche et dont l'ordonnée est la valeur de cette distance minimale.

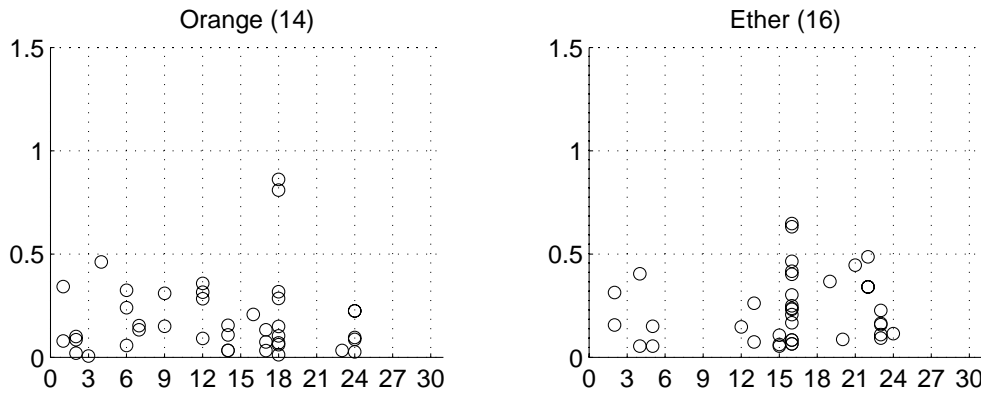


FIG. 6.5 – Mesures de distances pour les différents exemples d'une odeur donnée

Analysons la figure obtenue. Si plusieurs exemples correspondant à une même odeur sont à la fois très proches entre eux et très éloignés des autres odeurs, alors leurs marques sont situées sur une colonne ayant pour abscisse le numéro de cette odeur. Les graphiques montrent que les marques sont beaucoup plus dispersées qu'on ne l'aurait souhaité. C'est par exemple le cas pour l'odeur d'orange (numéro 14) qui se confond avec plusieurs autres fruits. En revanche les exemples correspondant à l'éther (colonne 16), odeur plus caractéristique, sont assez massivement regroupés sur la colonne de l'éther (numéro 16) et sont aussi assez nombreux sur la colonne 23 qui correspond au vernis à ongles, odeur assez proche de l'éther, en effet.

*Interprétation*

Afin de mieux mettre en valeur les confusions qui semblent nombreuses d'après l'étude des distances, dans le cas de l'orange, nous allons calculer les coefficients de corrélation-croisée de l'orange et des autres odeurs, en considérant tous les exemples quelle que soit leur classe ( $c$ ,  $w$  ou  $n$ ). Le résultat obtenu est présenté sur la figure 6.6. On constate que les odeurs les plus corrélées à l'orange sont les autres fruits : le fruit de la passion (numéro 6), l'orange (14) bien sûr, la pêche (17), la fraise (18), la banane (25), et le citron (30) ; un peu au-dessous, viennent l'odeur de rose (7), l'anis (24), la verveine (28) et l'amande amère (29) ; dans une moindre mesure, on rencontre aussi la menthe (2), le caramel (9), la citronnelle (12) et la cannelle (27). En revanche, les odeurs d'éther (16), de camphre (5) ou de goudron (26) sont nettement décorrélées de celle de l'orange.

*Coefficients de corrélation-croisée  
Résultat sur l'orange*

Notre étude des données met en évidence plusieurs raisons pour lesquelles essayer de classer les odeurs à l'aide des descripteurs est un problème « difficile à apprendre ». Elle explique les difficultés rencontrées par le classifieur incrémental. Enfin, nous pouvons à présent imaginer une modification de l'algorithme d'apprentissage du classifieur destinée à le rendre moins sensible aux exemples difficiles.

*Conclusion sur la base d'exemples*

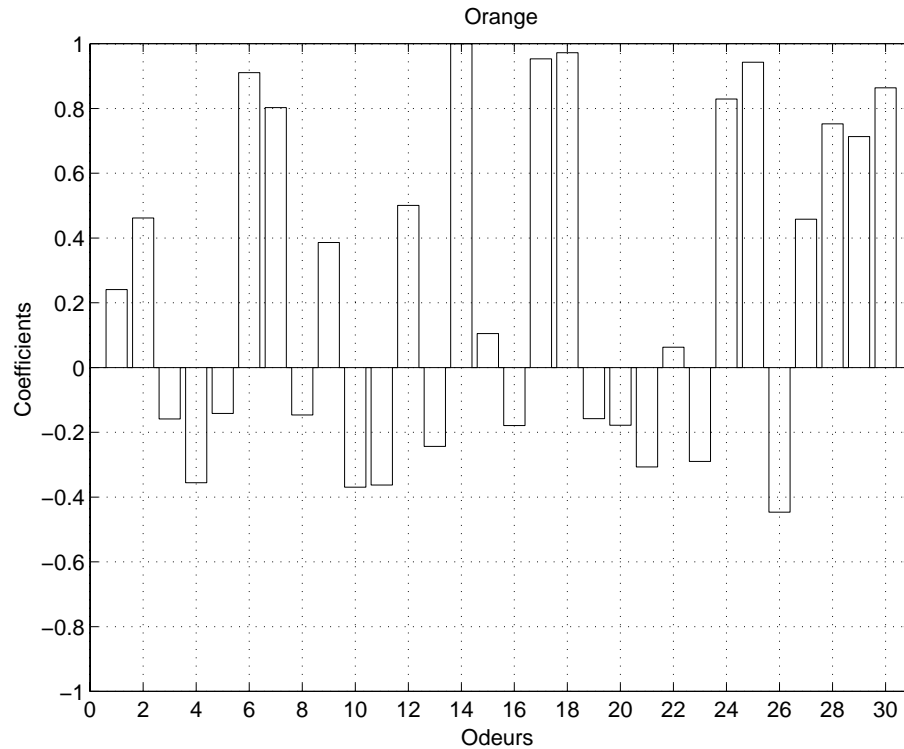


FIG. 6.6 – *Corrélation-croisée de l'orange (14) à toutes les autres odeurs*

## 6.6 Un nouvel algorithme d'apprentissage incrémental

L'analyse des données a mis en évidence le fait que des formes en entrée très proches entre elles correspondaient parfois à des odeurs différentes (par exemple les odeurs de banane et d'orange). Dans de telles situations, il faut éviter de créer des prototypes spécifiques car cela détériore les capacités de généralisation du classifieur. Ce problème est bien connu et porte le nom de « sur-apprentissage ». Toutefois, notons que notre objectif n'est toujours pas de faire de la « reconnaissance d'odeur », nous avons montré que la base comprend suffisamment d'incohérence pour conclure que la tâche est difficile. L'objectif de cette section est le test d'une modification de l'algorithme d'apprentissage du classifieur dans le cas d'une base comprenant des incohérences. Cette nouvelle version du classifieur doit être mise en œuvre dans tous les cas où les ensembles d'exemples correspondant aux différentes classes sont fortement imbriqués.

*Restriction  
dans la  
recherche du  
prototype  
gagnant*

Afin de remédier à l'abondance de prototypes inutiles, nous proposons de modifier l'algorithme d'apprentissage de la façon suivante : désormais, chaque nouvel exemple présenté en entrée est uniquement comparé aux prototypes associés à sa propre classe. Les performances obtenues en appliquant cette nouvelle règle d'apprentissage sont présentées dans la table 6.7 (revoir si nécessaire la légende, table 6.5).

Catégories	Nb ex.	Nb proto.	Succès	Erreurs	Rejets
cwn	825	<b>68</b>	40 (20%)	156	4
c	440	<b>62</b>	31 (28%)	74	4
cn	521	<b>68</b>	31 (24%)	100	0

TAB. 6.7 – Classification pour GP avec le nouvel algorithme d'apprentissage

On constate deux améliorations : (i) le nombre de prototypes créés est fortement réduit : il est maintenant de l'ordre du double du nombre de classes, ce qui est tout à fait satisfaisant ; (ii) les performances en généralisation sont clairement accrues, puisqu'elles dépassent toutes 20% et atteignent même 28% pour la catégorie des réponses correctes.

*Amélioration  
des résultats*

## 6.7 Mise en œuvre d'un réseau d'ondelettes

Les considérations précédentes laissent penser qu'un modèle connexionniste basé sur des séparations par hyperplans aurait beaucoup de difficultés pour discriminer les odeurs. Afin de vérifier que les difficultés sont plus liées à la base d'exemples qu'au modèle connexionniste choisi, Baron [Bar97] a mis en œuvre un autre modèle, le réseau d'ondelettes sur les mêmes données.

La théorie de la décomposition en ondelettes a été reliée au domaine du connexionnisme en 1992 [BZ92]. En effet, la décomposition d'une fonction (équation 6.1) en somme de fonctions obtenues par dilatation et translation d'une ondelette mère peut être écrite sous la forme d'un réseau connexionniste.

*Définition  
du modèle*

L'apprentissage passe par la recherche de coefficients de dilatation et de translation adaptés au problème. Une règle d'apprentissage modifie les coefficients et les poids de sortie, selon un algorithme de gradient minimisant une fonction de coût. La cellule de sortie réalise une somme pondérée de ses entrées, qui sera seuillée dans le cas d'un problème de classification.

*Apprentissage*

$$f(x) \approx \sum_{i=1}^N w_i \cdot h(d_i(x \Leftrightarrow t_i)) \quad (6.1)$$

où  $h$  est appelée *fonction ondelette*,  $d_i$  et  $t_i$  sont respectivement les facteurs de *dilatation* et de *translation*,  $w_i$  sont les poids de la cellule de sortie.

À la lumière des résultats obtenus avec le classifieur incrémental, Baron a mis en œuvre un réseau d'ondelettes sur un problème beaucoup plus simple : distinguer une odeur parmi toutes les autres. En d'autres termes, 1 réseau est dédié à 1 odeur et l'unique cellule de sortie permet de savoir si l'entrée est de la bonne classe (figure 6.7).

*Application  
à la reconnaissance  
des odeurs*

La perspective, si les résultats sont bons, est d'inclure les meilleurs réseaux<sup>2</sup> obtenus pour chacune des odeurs dans une architecture modulaire, basée sur la coopération de sous-réseaux.

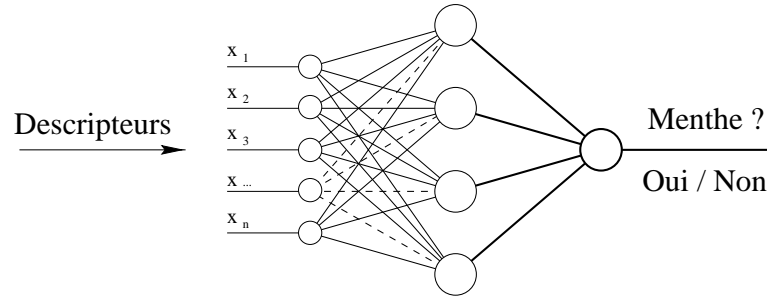


FIG. 6.7 – Exemple d'un réseau de 4 ondelettes dédié à la menthe

*Performances observées*

La figure 6.8 présente les taux de reconnaissance en généralisation. On constate que les résultats sont très bons. Cependant, pour ce problème « une odeur contre les autres », la disproportion des nombres d'exemples pose un problème d'effectif de classes. En fait, le réseau a tendance à apprendre uniquement les « autres » odeurs. Ceci explique en partie des valeurs de taux toujours supérieures à 80 %.

*Intégration des réponses*

Le problème est maintenant de nommer l'odeur, c'est-à-dire de donner une réponse unique malgré l'emploi de plusieurs réseaux, chacun dédié à une odeur (ce problème n'est pas résolu pour le moment). En effet, la décision du « méta réseau » va devenir difficile dès qu'au moins 2 sous réseaux déclarent reconnaître une odeur. Outre le fait que la difficulté augmente rapidement avec le nombre de sous réseaux, notons que l'ajout d'un seul mauvais réseau va perturber tout le système. Il est donc probable que la coopération de 25 réseaux de ce type redonnera des résultats similaires à ceux du classifieur incrémental.

*Performances relatives*

Nous constatons cependant que les taux d'apprentissage les plus bas (inférieurs à 92.5 % et notés par des carrés sur la figure) sont obtenus pour les odeurs 6, 14, 17, 18, 25 et 30 qui correspondent respectivement au fruit de la passion, à l'orange, la pêche, la fraise, la banane et le citron.

La proximité entre ces odeurs a été remarquée dans l'étude des données, notamment dans l'étude de la corrélation entre l'odeur d'orange et les autres odeurs (figure 6.6). Ce phénomène réapparaît ici, bien que l'algorithmique mise en œuvre soit très différente (minimisation d'une fonction de coût contre approche constructive).

2. Le modèle n'étant pas incrémental, un paramètre supplémentaire doit être fixé : la taille du réseau, c'est-à-dire le nombre de cellules ondelettes de la couche interne.

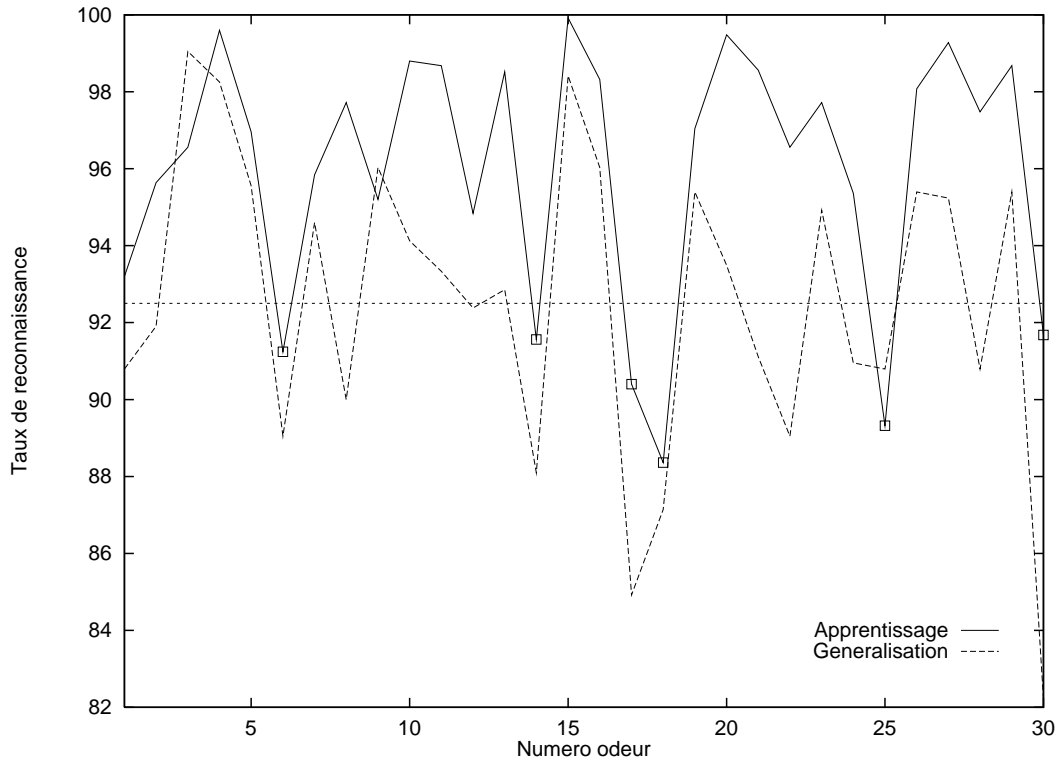


FIG. 6.8 – Performances obtenues pour chaque odeur par réseau d'ondelettes

## Conclusion

Ce chapitre a présenté une série d'expérimentations sur le domaine de l'olfaction. L'étude de la possibilité de discriminer des odeurs à l'aide de simples descripteurs valués a confronté le classifieur à une base d'exemples « difficiles à apprendre ». L'étude du comportement du classifieur, notamment un sur-apprentissage, nous a permis de mettre en évidence des incohérences de la base d'exemple. On a par exemple montré que les difficultés inhérentes à la base de données se retrouvent sur le comportement d'un autre modèle connexionniste, le réseau d'ondelettes.

Face aux difficultés du classifieur sur une base contenant des incohérences, nous avons proposé une amélioration de l'algorithme d'apprentissage du classifieur incrémental. Le nombre de prototypes inutiles est réduit et une odeur sur quatre est bien classifiée. Cette variante de la règle d'apprentissage, moins sensible aux exemples difficiles, peut être appliquée à tous les problèmes de classification pour lesquels les nuages des différentes classes présentent de fortes interférences. Enfin, notons que le nouvel algorithme corrobore les observations faites lors des parallélisations modulaires. En effet, le comportement de notre nouvel algorithme est proche de celui des versions modulaires du classifieur incrémental, avec un important délai entre les diffusions de prototypes ( $\delta$  grand). Pour un module donné, moins de communications implique

*Discrimination des odeurs par les descripteur*

*Un nouvel algorithme séquentiel*

*Lien avec la parallélisation modulaire*

moins de prototypes reçus. Aussi, le module travaille-t-il davantage avec ses propres prototypes. Cette stratégie de « cloisonnement » a été bénéfique aux taux d'erreurs pour le problème d'OCR et pour la classification des formes d'ondes. Une baisse significative des échanges entre modules nuit à la reconnaissance des chiffres manuscrits. En revanche, sur le problème des formes d'ondes, nous avons constaté que les meilleures performances sont atteintes avec très peu de communications ( $\delta = 95$ ). Ces différences de comportement ont été expliquées dans la section 5.8.

*Faisabilité  
d'un  
classifieur  
olfactif*

Bien que ce ne soit pas l'objet de notre travail, il est intéressant de se poser la question de la faisabilité d'une reconnaissance olfactive. Cette question impose quelques remarques. Tout d'abord, si un système arrive à reconnaître des odeurs, ce ne sera pas par l'intermédiaire de quelques descripteurs. En effet, la tâche d'apprentissage proposée à nos modèles connexionnistes est réellement complexe car ceux-ci reçoivent en entrée des informations symboliques, sous forme de listes de descripteurs, et doivent en extraire des représentations internes leur permettant de reconnaître la classe de l'odeur, parmi 25 dénominations possibles. Si l'objectif est réellement la reconnaissance d'une odeur il faut trouver une entrée discriminante. Hélas, contrairement aux matrices de pixels (en image), on ne dispose pas de caractéristiques objectives du *stimulus* olfactif de base [Sch91]. D'une manière plus générale, la modélisation de la perception olfactive est plus difficile que celle de la perception visuelle car notre vocabulaire ne contient pas de termes spécifiques à la qualification et à la dénomination des odeurs [RZ89] (contrairement à « rouge » ou « vert » pour les couleurs, par exemple). Les termes employés, c'est-à-dire la liste des descripteurs, font le plus souvent appel à d'autres modalités perceptives (goût, vision, *etc.*).