

---

## Fouille de données issues d'études psychologiques liées au vieillissement : extraction de règles graduelles

**Lisa Di Jorio\*** — **Sandra Bringay\*,\*\*** — **Denis Brouillet\*\*** —  
**Anne Laurent\*** — **Sophie Martin\*\*** — **Maguelonne Teisseire\*,\*\*\***

\* LIRMM – UM2 – CNRS UMR 5506, 161 rue Ada 34392 Montpellier Cedex 5  
{nom}@lirmm.fr

\*\* Université Paul Valéry – Montpellier 3, 17, rue abée de l'Epée 34090 Montpellier  
denis.brouillet@univ-montp3.fr, sophiesmr@aol.com

\*\* UMR TETIS – CEMAGREF, 500 rue J.F.Breton 34093 MONTPELLIER Cedex 5  
maguelonne.teisseire@cemagref.fr

---

*RÉSUMÉ. La mémoire humaine n'est pas un simple réceptacle du passé. Elle permet de comprendre le présent et de se projeter dans l'avenir. Elle est l'identité même de l'être. Les patients atteints de la maladie d'Alzheimer souffrent de pertes de mémoire qui réduisent leurs fonctions cognitives et donc leurs relations sociales, leurs interactions, leurs capacités à décider par eux-mêmes, et finalement leur autonomie. Cet article est issu d'une collaboration entre une équipe de psychologues, spécialistes de la mémoire et de la maladie d'Alzheimer et une équipe d'informaticiens, spécialistes de la fouille de données. Cette collaboration vise à développer des méthodes et des outils de fouille adaptés aux données issues d'études psychologiques en vue d'en extraire des connaissances nouvelles.*

*ABSTRACT. Alzheimer disease impacts memory and reduces cognitive functions and social relationships, interactions, self-decision making and autonomy. In this paper, we describe how data mining tools can help psychologists to analyze the data they collect. This work is the result of a collaboration between psychologists and computer scientists. It aims at developing methods and tools to handle the data collected by psychologists to extract relevant knowledge.*

*MOTS-CLÉS : Fouille de Données, règles d'associations graduelles, troubles cognitifs, maladie d'Alzheimer*

*KEYWORDS: Data mining, gradual association rules, Cognitive Diseases, Alzheimer Disease*

---

## 1. Introduction

La mémoire humaine n'est pas un simple réceptacle du passé. Elle permet de comprendre le présent et de se projeter dans l'avenir. Elle est l'identité même de l'être. Les patients atteints de la maladie d'Alzheimer souffrent de pertes de mémoire qui réduisent leurs fonctions cognitives et donc leurs relations sociales, leurs interactions, leurs capacités à décider par eux-mêmes et finalement leur autonomie. Cet article s'inscrit dans le cadre du projet pluri-disciplinaire PEPS (CNRS) LAMAL : *Langage, Mémoire et Alzheimer : une approche des maladies neurodégénératives fondée sur la densité des idées. Fouille de données pour l'extraction de corrélations entre différents indices neuropsychologiques*. Ce projet vise à montrer comment les méthodes et outils de gestion et extraction de connaissances peuvent être utiles dans le cadre du traitement de données liées au vieillissement. Dans le cadre du projet LAMAL, nous avons travaillé sur des données liées au vieillissement acquises à partir de tests destinés à l'évaluation des capacités cognitives. Ces données sont de deux types : des productions langagières (récits) issues d'interviews structurées de patients et des données tabulaires numériques évaluant les performances de la mémoire des patients (scores).

Dans cet article, nous nous focalisons sur le second point (fouille des données tabulaires). Une des limites des travaux menés actuellement par les chercheurs psychologues cogniticiens est que les corrélations mises en évidence ont été établies de manière manuelle et sont donc non exhaustives. Notre objectif est de co-construire des modèles, méthodes et instruments, reposant sur les sciences et technologies de l'information et de l'ingénierie et permettant aux psychologues de formuler puis de vérifier de nouvelles hypothèses sur les liens entre la dégradation de la mémoire, l'âge et la maladie d'Alzheimer. Une difficulté, courante pour les médecins qui diagnostiquent la maladie, est de distinguer les dégradations (symptômes) liées à l'âge et qui sont en quelques sortes normales et les dégradations liées à la maladie d'Alzheimer ou à d'autres maladies comme la dépression. Pour chaque malade, nous connaissons son âge ainsi que la présence/absence de pathologie de type Alzheimer. Nous distinguons parmi la population les individus dits MCI (i.e. Mild Cognitive Impairment, stade préclinique de maladie d'Alzheimer) des dépressifs. Un bilan neuropsychologique standard a été effectué, nous permettant d'évaluer l'efficacité des registres pragmatique et mécanique (e.g. Echelle de Mattis, RI48, Empans, TMT etc.). À partir de différents scores, nous extrayons des règles graduelles de la forme "Plus/Moins A1, ..., Plus/Moins An". Par exemple, la règle suivante a été extraite : "Plus le nombre d'identifications au test RI 48 augmente et plus le souvenir du moment de la journée et de la disposition spatiale d'un moment neutre est bon, alors plus le test du MMS (Mini Mental Score) est réussi".

Nous rapportons dans cet article les principales caractéristiques liées à la fouille des données issues d'études psychologiques liées au vieillissement. La méthodologie employée pour extraire des règles graduelles est décrite<sup>1</sup> et les premiers résultats obtenus.

---

1. Nous invitons le lecteur intéressé par les détails de ces méthodes à se reporter aux articles cités.

nus sont présentés. Nous décrivons également la collaboration entre les chercheurs en informatique et en sciences humaines cherchant à mettre au point une telle méthodologie. L'une des questions récurrentes dans le dialogue Informatique et Psychologie est celle de la dépendance/complémentarité entre ces deux disciplines. Il s'agit en effet de considérer soit (1) les modèles issus des théories informatiques sont (peuvent être) utilisés comme modèles psychologiques (ou qu'à l'inverse sans les modèles psychologiques, l'informatique ne peut rien dire des données qu'elle permet de découvrir), soit (2) dès lors qu'il n'est pas possible d'observer directement les processus qui sont à l'origine des comportements humains, le recours à la simulation (dans notre cas à l'extraction de règles/motifs à partir d'un flux de données disparates) est indispensable à condition qu'il s'agisse d'un dialogue, plus exactement d'une co-construction dans l'analyse intelligente des données ou dans la construction de tel ou tel algorithme ou heuristique.

## 2. Description des données issues d'études psychologiques liées au vieillissement

La maladie d'Alzheimer est une maladie neurodégénérative du tissu cérébral qui entraîne la perte progressive et irréversible des fonctions mentales. Un diagnostic précoce permet d'assurer une prise en charge optimale du patient dès les premiers signes de la maladie comme des troubles de la mémoire et du comportement. Or, ce diagnostic est complexe car ces signes se confondent avec les signes liés au vieillissement ou à d'autres maladies comme la dépression. La psychologie cognitive permet un diagnostic de l'état d'avancement de la maladie, en s'appuyant sur l'étude des grandes fonctions psychologiques de l'être humain : la mémoire, le langage, l'intelligence, le raisonnement.

Les psychologues cogniticiens utilisent divers tests d'évaluations. Dans notre cadre, les données analysées relèvent de mesures de la mémoire et du souvenir. Ainsi, chaque patient a passé une dizaine de tests, dont voici les 6 principaux :

- **Mini Mental State Evaluation (MMSE)** (ou test de Folstein). Ce test explore l'orientation spatio-temporelle, l'apprentissage, la mémoire, l'attention, le raisonnement, et le langage. Le patient répond à une série de cinq épreuves notées sur un total de 30 points. Ce test est utilisé dans le cadre d'une suspicion de démence. Il permet d'identifier les premiers signes de troubles de la mémorisation et de suivre l'évolution vers des secteurs épargnés en début de maladie.

- **Memory Characteristics Questionnaire (MCQ)** (Johnson *et al.*, 1988). Ce questionnaire porte sur les caractéristiques phénoménologiques (aspects qualitatifs) de la mémoire autobiographique telles que les détails contextuels, sensoriels ou affectifs. Les sujets doivent produire une description détaillée de six souvenirs précis et spécifiques personnellement vécus. Les souvenirs rapportés doivent s'être déroulés il y a plus d'un an et sur une période de moins d'un jour. Chaque sujet doit rapporter oralement deux souvenirs spécifiques à connotation émotionnelle agréable, désagréable et neutre. Après chaque récit, le sujet quantifie la qualité de son souvenir au travers des seize items de l'échelle de Lickert. Nous avons contrebalancé l'ordre de passa-

tion des souvenirs, en ont découlé six protocoles, reflétant les six ordres possibles de succession des souvenirs.

– **Autobiographical Memory Test (AMT)** (William *et al.*, 1986). Ce test permet de quantifier la spécificité de la mémoire autobiographique selon la valence émotionnelle du souvenir décrit. Les participants se voient proposer dix mots successifs à valence émotionnelle (cinq positifs et cinq négatifs) à la suite desquels ils doivent rappeler un souvenir de ce que ce mot leur évoque. La consigne est de parler d'un souvenir précis et spécifique, un exemple leur est proposé. Le temps de latence entre la présentation et le début du discours est pris en compte.

– **Toronto Alexithymia Scale (TAS)** (Bagby *et al.*, 1994). Ce test est une échelle d'auto-évaluation à 20 items appréciant trois dimensions :

- la difficulté à identifier ses états émotionnels,
- la difficulté à décrire ses états émotionnels à autrui,
- la pensée opératoire (organisation psychique caractérisée par un discours se tenant aux faits, à la raison, où le fantasme est exclu).

– **State-Trait Anxiety Inventory (STAI)**, Spielberger). La STAI est un questionnaire d'auto-évaluation de l'anxiété de l'individu. L'anxiété-état renvoie au niveau d'anxiété transitoire ressenti à un temps donnée de la vie du sujet. En revanche, l'anxiété-trait rend compte de la tendance habituelle (personnalité) de l'individu à être anxieux. Les items 1 à 20 mesurent l'anxiété-état tandis que les items 21 à 40 apprécient l'anxiété-trait du participant. De même que pour la passation de la TAS 20, les sujets présentant un diagnostic de MA (maladie d'Alzheimer) évoqué nous ont demandé de lire les items à leur place pour cause de fatigabilité oculaire.

– **Beck Depression Inventory Second Edition (BDI-II)**. L'inventaire de dépression de Beck révisé est un questionnaire d'auto-évaluation destiné à mesurer la sévérité de la dépression chez des sujets à partir de 16 ans. Il est composé de 21 questions auxquelles le sujet répond sur une échelle entre 0 et 3. L'inventaire a été conçu pour évaluer les symptômes dépressifs correspondant aux critères diagnostiques des troubles dépressifs selon le DSM-IV (version française, 1996). De même que pour la passation de la TAS 20 et de la STAI, les sujets présentant un diagnostic de MA évoqué nous ont demandé de lire les items à leur place pour cause de fatigabilité oculaire.

Trente-trois patients ont été listés et leurs résultats à ces différents tests sont stockés sous format tabulaire. Chaque patient se voit également attribuer un identifiant, un groupe et une classe d'âge. Les résultats aux divers tests sont enregistrés sous forme de score par test, plusieurs "sous scores" sont conservés, ainsi qu'un attribut "total".

À partir de ces résultats, les psychologues souhaitent extraire des connaissances nouvelles. Ne pouvant pas tester toutes les hypothèses de corrélation entre ensembles de variables, les méthodes de fouille de données vont être pertinentes pour automatiser le processus sur l'ensemble des variables. Nous présentons ci-après de manière générale le processus de fouille de données, avant de détailler la méthode utilisée dans le cadre de ce travail.

### 3. La fouille de données

Ces dernières années, l'évolution des outils de stockage a permis d'emmagasiner une grande quantité d'informations. Cependant, ces outils sont en quelque sorte victimes de leur propre puissance, puisqu'il devient dès lors impossible d'analyser de manière manuelle les données. Il est donc nécessaire de proposer des outils de traitements automatiques permettant d'assister les experts dans le traitement de ces données. Cependant, de tels outils se révèlent difficiles à créer, puisqu'ils demandent d'une part la compréhension des attentes d'utilisateurs non informaticiens, et d'autre part la conception de modèles que les machines peuvent traiter dans un temps de réponse adapté. Ainsi des chercheurs de divers horizons tels que l'intelligence artificielle, de l'interface homme-machine, des statistiques et des bases de données, ont collaboré ces dernières années à la création d'une méthodologie appelée "*extraction de connaissances dans les grandes bases de données*" (ECD).

La fouille de données est une étape de ce processus d'ECD. Elle consiste à "*extraire, à partir d'une grande masse de données, des schémas valides, et récurrents*" (Fayyad *et al.*, 1996). Intuitivement, de telles recherches ont un fort impact sur les prises de décisions des experts. Les domaines d'applications sont nombreux : diagnostic médical, profils de clients, mailing, accord de prêts bancaires, reconnaissance de caractères manuscrits, finance, prévision d'évolutions de marchés... L'une des applications les plus connues est l'extraction de comportements fréquents à partir de bases de données commerciales. Par exemple, à partir d'expérimentations menées sur un jeu de données réelles lié aux tickets de carte bleue, il nous est possible d'extraire la règle "Lorsque l'on achète des couches culottes, on achète des bières". La connaissance d'une telle règle a poussé certains décideurs à réorganiser l'emplacement des produits dans le magasin.

La fouille de données se décompose en diverses disciplines, dont trois ont particulièrement émergé : la catégorisation consiste à classer les éléments en diverses classes prédéfinies. Par exemple, la catégorisation permet de classer automatiquement le produit 'Coca-Cola' dans la classe 'Boissons'. La segmentation permet de partitionner les données en groupes, de telle manière que la similarité entre les données d'un même groupe et la dissimilarité entre les différents groupes soient les plus grandes possible. Par exemple, nous pouvons ainsi constituer des catégories de consommateurs selon leurs comportements. Enfin, la description des données est réalisée à l'aide de règles d'association ou de motif séquentiels qui rendent compte des corrélations fréquentes tenant ou non compte d'une notion d'ordre. Nous détaillons ci-dessous ces méthodes qui ont été utilisées dans ce travail.

La description des données est particulièrement utile dans le cas d'analyses de grandes bases, car elle permet de mettre en évidence des corrélations concernant un grand nombre d'éléments de la base. De plus, des algorithmes de plus en plus performants ont été mis au point pour s'exécuter sur des bases contenant un grand nombre d'attributs. Les règles d'association mettent en jeu les différents attributs de la base, et sont de la forme "Pain, beurre => Nutella" (70%, 95%). Deux mesures sont asso-

ciées à une règle. La première, la fréquence (ici 70%) la représentativité de la règle, c'est-à-dire le nombre d'éléments de la base qui supportent la règle. La seconde est la confiance (ici 95%). Elle mesure la probabilité conditionnelle : sachant qu'on a acheté X, quelle est la proba d'acheter Y ? Dans notre exemple, combien de clients ayant acheté du pain et du beurre ont également acheté du Nutella ? Les motifs séquentiels sont de la forme «<(Lecteur dvd)(dvd1, dvd2)> (80%)», ce qui signifie que 80% des clients de la base ont acheté un lecteur dvd, puis sont revenus plus tard acheter le dvd1 et le dvd2 en même temps. S'ils ne mesurent pas un degré d'implication, ces derniers incluent une notion d'ordre, ici associée au temps.

Les règles d'association et les motifs séquentiels ont connu un très grand succès de par leur adaptation aux bases de données réelles. Aux débuts de la fouille de données, les algorithmes étaient destinés à être appliqués sur des bases commerciales. Mais très vite, des utilisateurs de divers milieux sont devenus demandeurs. Ainsi, la fouille a été appliquée au web (web mining) ou encore à des données industrielles telles que les relevés de capteurs. Ainsi, les algorithmes sont conçus dans un souci de performance d'une part, mais également de généricité (l'algorithme doit pouvoir s'adapter à tout type de jeux de données et tout type de contexte).

Plus récemment, les chercheurs du milieu médical ont eux aussi exprimé le besoin d'analyse automatique de données. Ces données sont de natures variées : questionnaire, expériences biologiques (comme par exemple les puces à ADN), ou encore tests psychiatriques et psychologiques. Cependant, ces nouvelles bases se présentent sous un format différent de celles citées précédemment : elles ne contiennent pas un grand nombre de lignes comparé au nombre d'attributs, mais au contraire un grand nombre d'attributs comparé au nombre de lignes (on parle alors de bases de données inversées). Cela s'explique par la nature même de la discipline. En effet, les tests sont coûteux et ne peuvent être réalisés qu'une seule fois par patient (l'exemple le plus frappant étant l'analyse tumorale, possible uniquement après extraction de la tumeur). Ce format rend les algorithmes d'extraction cités plus haut inefficaces. Enfin, la quantité de résultats obtenus est souvent supérieure à la base de départ, redondante, et difficile à exploiter par les experts.

Dans cet article nous nous intéressons à la description de données, et plus particulièrement à l'extraction de règles graduelles. Les règles graduelles représentent des corrélations fréquentes de variations. Leur intérêt est qu'elles sont faciles à interpréter par des non informaticiens, de par leur sémantique. Par exemple, une règle pourrait être "Plus le mur est proche, plus le pilote automatique doit actionner fortement le frein". Dans la section suivante, nous décrivons de manière plus détaillée les règles graduelles, et donnons l'intuition de l'algorithme d'extraction.

#### 4. Les règles graduelles

Une règle graduelle se présente sous la forme "plus / moins  $A_1$  et ... plus / moins  $A_n$ , alors plus / moins  $B_1$  ... plus / moins  $B_n$ ". Contrairement aux approches de fouille

classiques, qui se basent sur la présence ou absence d'un élément dans la transaction, l'extraction de gradualité nécessite la comparaison des valeurs entre chaque élément. Afin d'illustrer nos propos, nous utiliserons la base de données du tableau 1. Celle-ci représente, pour une voiture lancée sur un mur, la distance à laquelle le conducteur a actionné le frein, ainsi que le risque d'accident.

Voiture	Distance au mur	Force de freinage	Risque d'accident
$V_1$	3	8	9
$V_2$	5	9	8
$V_3$	4	8	7
$V_4$	6	6	6
$V_5$	8	5	5
$V_6$	9	6	4
$V_7$	9	4	4
$V_8$	7	3	4

**Tableau 1.** *Un exemple de base de données*

La gradualité telle que décrite dans cet article peut être associée à la notion de corrélation statistique. Cependant les méthodes statistiques classiques sont alors limitées à la considération de deux attributs dont on recherche s'ils sont ou non corrélés, ce qui pose deux problèmes principaux : (i) il faut choisir les deux attributs dont on évalue le degré de co-variation, et (ii) seulement deux attributs sont considérés.

Récemment, la notion de gradualité a donc été étudiée en utilisant les méthodes liées à l'extraction automatique (Hüllermeier, 2002; Berzal *et al.*, 2007; Laurent *et al.*, 2009; Di Jorio *et al.*, 2009) afin d'extraire automatiquement les ensembles d'attributs liés par une relation de co-variation.

Dans (Hüllermeier, 2002), les auteurs utilisent le calcul de régression linéaire pour décrire le degré de co-variation de variables. Cependant, il reste difficile de manipuler un grand nombre d'attributs et il est impossible de décrire des co-variations mêlant des variations croissantes et décroissantes (plus A augmente, plus B diminue) à partir de bases de données assez conséquentes.

Dans (Berzal *et al.*, 2007), les auteurs décrivent une approche s'appuyant sur la comparaison de lignes deux à deux. De même, l'approche décrite dans (Laurent *et al.*, 2009), fondée sur le *Kendall tau ranking correlation coefficient*, calcule le nombre de paires de lignes ordonnables dans la base de données pour être en accord avec le motif graduel considéré. On parle alors de paires concordantes et discordantes.

L'approche (Di Jorio *et al.*, 2009) est quant à elle fondée non pas sur le décompte du nombre de paires ordonnables, mais sur la recherche de la plus longue liste de lignes que l'on peut ordonner pour qu'elles respectent la gradualité du motif graduel considéré.

En reprenant l'exemple décrit ci-dessus issu du tableau 1, si nous voulons extraire "Plus la distance au mur diminue" (notée  $Mur^-$ ), nous allons comparer et

ordonner les valeurs sur l'attribut Mur, et obtenir l'ensemble ordonné suivant :  $\{V_1, V_3, V_2, V_4, V_8, V_5, V_6, V_7\}$ . Intuitivement, la gradualité sur un seul attribut est facile à extraire. Cependant, cela devient plus compliqué dans le cas de plusieurs attributs, car il peut exister différentes solutions. Par exemple, si l'on cherche les ensembles de voitures respectant la règle "Plus le mur est proche, plus le freinage est fort.", nous avons  $S_1 = \{V_1, V_3, V_4, V_5, V_7\}$  ou  $S_2 = \{V_2, V_4, V_5, V_7\}$ . Si l'on considère qu'une règle est d'autant plus vraie que le nombre d'objets de la base de données qui peuvent être ordonnés pour la valider l'est, alors l'existence de ces différents ordonnancements possibles devient primordiale. On parle alors de support. Dans notre cas, nous considérons que le support est le nombre d'éléments du plus grand ensemble ordonné. Dans notre exemple, nous choisissons le premier ensemble  $S_1$  et déduisons la fréquence de la règle, soit  $5/8 = 62.5\%$ .

D'autre part, il est important de noter que si un ensemble n'est pas maximal pour une règle  $A_1^+ A_2^+$ , il peut le devenir pour une règle  $A_1^+ A_2^+ A_3^+$ , ce qui nous oblige à conserver tous les choix possibles. Nous adoptons donc une stratégie basée sur les treillis qui offre le double avantage de conserver l'ordre et d'être représenté de manière binaire. La représentation binaire permet l'utilisation de l'opération logique 'ET', qui compte parmi les plus performantes en terme d'exécution machine.

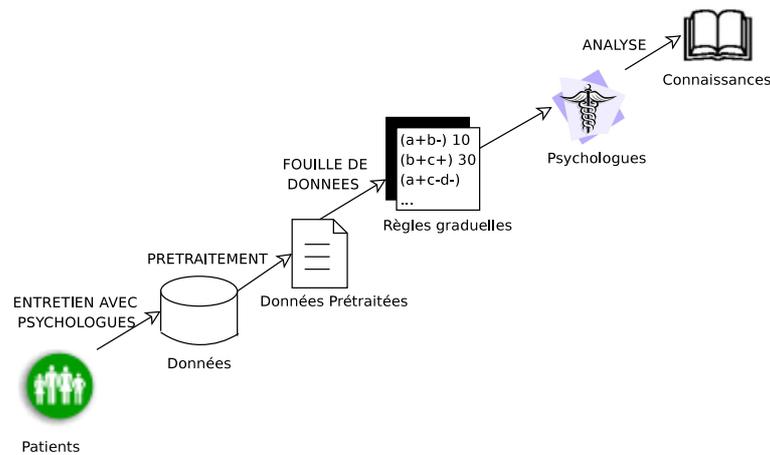
Les règles graduées ont été conçues de manière générique, c'est-à-dire qu'elles peuvent être appliquées à n'importe quelles bases, du moment qu'elles ne contiennent que des valeurs munies d'un ordre (numériques notamment). L'un de leur principaux avantages est d'être adapté aux bases contenant un faible nombre de lignes et un grand nombre d'attributs. Ainsi, en utilisant les structures adaptées (des treillis dans notre cas), les informations à conserver en mémoire sont moindres, puisqu'elles sont représentées sous une forme binaire. Ainsi, il se trouve que ces règles sont parfaitement adaptées aux bases médicales.

La seconde force des règles est leur aspect sémantique. En effet, de telles règles correspondent aux recherches effectuées par la plupart des chercheurs en sciences humaines. Par exemple, les biologistes analysant les interactions entre gènes recherchent des corrélations du genre "plus le gène X et le gène Y se sur-expriment, alors plus le gène Z se sous-exprime", ce qui signifie que l'activation des gènes X et Y a un effet inhibiteur sur le gène Z. Une règle d'association classique aurait simplement extrait que les gènes X Y et Z interagissaient entre eux, sans capturer la nature de leur interaction. De la même manière, on retrouve en analysant les tests effectués par les psychologues des règles du type "plus le patient a un caractère contrôlant, plus son score au test x est élevé".

Comme nous l'avons vu, l'extraction de règles graduées est très pertinente dans le cadre de données issues des études des psychologues cogniticiens. Cependant, cette extraction ne peut se faire qu'à partir de bases de données préparées à cet effet. Dans la section suivante, nous détaillons le processus global d'extraction, depuis la collecte des données jusqu'à la restitution des résultats.

## 5. Mise en œuvre

La figure 1 décrit la méthode adoptée pour rechercher des règles graduelles à partir des données associées aux études psychologiques sur le vieillissement. Celle-ci se décompose en quatre principales étapes : la collecte des données, la préparation des données, l'extraction et enfin l'interprétation. Nous décrivons ici en détail chacune de ces étapes.



**Figure 1.** Méthodologie employée

Dans un premier temps, des entretiens ont été réalisés entre les psychologues et des personnes en stade MCI, stade préclinique de maladie d'Alzheimer. Ces données sont recueillies sous un format audio puis transcrites sous un format textuel.

La fouille de données ne peut s'appliquer directement sur les données brutes. En effet, il est souvent nécessaire de prétraiter les données afin de les préparer à la fouille. Dans notre cas, les données fournies par les psychologues sont dans un format tabulaire. Dans un premier temps, un prétraitement est nécessaire afin de nettoyer les données. Cette étape est souvent réalisée avec l'aide de l'expert. Ensuite, un jeu de données est généré dans le format d'entrée de l'algorithme. Les valeurs manquantes sont représentées par un chiffre haut (par exemple 2000 si les tests sont notés de 0 à 100).

Par exemple, soit le tableau 5, représentant un échantillon des données issues de tests psychologiques. Ici, les patients possèdent un identifiant (première colonne du tableau), un groupe, une classe d'âge, et un score au test psychologique du MMS. L'identifiant constitue l'élément de comparaison. Le groupe n'est pas exploitable tel quel. En effet, chaque patient a un groupe différent, et qui ne correspond pas à un élément graduel. Cet attribut ne sera donc pas inclus dans la fouille. En revanche, la classe d'âge est un élément graduel, puisque les différentes valeurs sont 'jeune', 'moyennement âgé' et 'âgé'. Nous affectons la valeur 0 à jeune, 1 à moyennement âgé

Sujet	Groupe	Classe d'âge	Score MMS
1	1	âgé	5
2	2	âgé	11
3	Sujet1A	âgé	9
4	MA2	J	15
5	MA3	MA	10

**Tableau 2.** Echantillon exemple de résultats de tests psychologiques

Sujet	Classe d'âge	Score MMS
1	2	5
2	2	11
3	2	9
4	0	15
5	1	10

**Tableau 3.** Echantillon obtenu après sélection

2 5
2 11
2 9
0 15
1 10

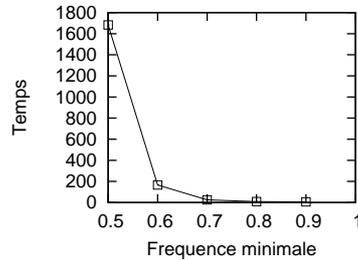
**Tableau 4.** Jeu de données généré

et 2 à âgé, ce qui permet de refléter “l'âge augmente” ou “l'âge diminue”. Après sélection, nous obtenons le tableau 3 qui sera transformé en tableau 4 pour lui appliquer l'algorithme de fouille.

L'étape suivante est l'étape de fouille, qui consiste à lancer l'algorithme GRITE sur le jeu de données obtenu. GRITE prend en compte deux paramètres : la fréquence minimale et la confiance minimale. Le choix de ces paramètres est une étape délicate du processus : pour pouvoir obtenir des résultats, il faut parfois baisser la fréquence. Cependant, un seuil trop bas entraîne la génération de trop nombreuses règles, difficiles à interpréter puisqu'elles ne concernent qu'un petit ensemble de la base. Ainsi, la seule manière de déterminer le seuil adéquat est empirique : baisser le support, présenter les résultats aux experts, puis rebaisser ou remonter en fonction de leur analyse. Le choix de la confiance minimale est moins difficile. En effet, pour être valide, une règle, même ayant une faible fréquence, doit être vraie pour un maximum de patients la supportant.

Enfin, les résultats obtenus sont présentés sous forme de règles aux psychologues. Ceux-ci donnent alors deux types de retour :

– Les résultats sont satisfaisants, c'est-à-dire que les psychologues retrouvent des connaissances qu'ils avaient déjà, et dans des proportions non surprenantes. Les autres règles peuvent alors être analysées pour découvrir des nouveautés.



**Figure 2.** Temps d'exécution en fonction de la fréquence minimale (sec)

– Les résultats s'avèrent difficiles à analyser, pour diverses raisons : trop de règles, pas assez de règles connues, trop d'informations redondantes ou de connaissances évidentes. Dans ce cas, il est nécessaire d'identifier les causes et de les corriger. Cela passe par la reprise des deux étapes de prétraitement.

## 6. Résultats

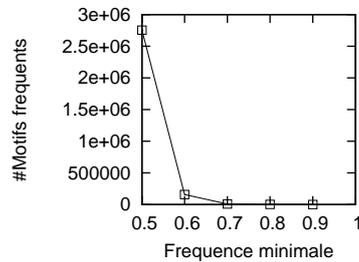
Nous décrivons les résultats selon deux points de vue : du point de vue de la performance (combien de règles extraites, avec quel support ?) et du point de vue pertinence (les règles ont-elles un sens ? Sont-elles utiles pour les experts ?).

### 6.1. Analyse quantitative des règles graduelles

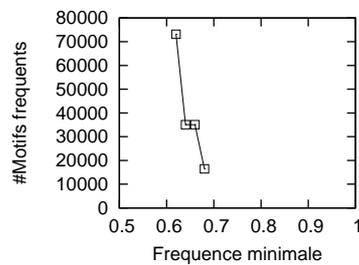
Le jeu de données généré une fois les étapes de prétraitement effectuées contient 122 attributs pour 33 lignes. Dans un premier temps, une extraction de règles d'associations classiques a été pratiquée avec l'algorithme Apriori (Agrawal *et al.*, 1994). Cependant, la nature numérique des résultats ne permet pas une extraction efficace : soit le support est élevé et l'on obtient un nombre très réduit et évident de règles (par exemple, tout le monde est âgé), soit le support est relativement faible, et dans ce cas, le nombre de règles est trop volumineux pour que celles-ci soient exploitées. Ces règles n'apportent guère plus d'informations que les outils statistiques précédemment employés par les experts.

L'algorithme est hautement satisfaisant en terme de temps d'exécution et d'utilisation mémoire. La figure 2 montre le temps d'exécution pour différents seuils de fréquence minimale, allant de 90% à 50%. Le temps d'exécution augmente de manière exponentielle, et est en accord avec le nombre de motifs extraits. Par exemple, il faut moins de 10 secondes pour extraire environ 300 règles contre 30 minutes pour en extraire plus de deux millions.

Le jeu de données est très riche en corrélations et est donc sensible à la baisse de support. La figure 3 représente le nombre de motifs extraits en fonction de la fréquence



**Figure 3.** Nombre de motifs extraits en fonction de la fréquence minimale



**Figure 4.** Nombre de motifs extraits en fonction de la fréquence minimale affinée

minimale. À 90%, environ 270 règles sont extraites. Ce nombre passe à plus de deux millions pour un support à 50%. La figure 3 montre une cassure entre un support à 60% et un support à 70%, où le nombre de règles extraites passe de 1500 à 15000. Nous avons donc analysé de manière plus fine en ajustant les supports entre 60% et 70%. Les résultats du nombre de motifs extraits sont présentés à la figure 4. Clairement, le nombre de motifs extraits augmente de manière exponentielle.

Afin de présenter des résultats pertinents aux experts, nous avons utilisé les critères suivants :

- éliminer les jeux contenant trop de redondance. La redondance des informations est due au fait que lorsque qu'une règle telle que  $(A^+B^+C^+)$  est fréquente, alors  $(A^+B^+)$  le sera,  $(A^+C^+)$  également etc... Cela pourrait être résolu en ne proposant que les règles maximales, cependant la perte d'information peut être conséquente, puisque la fréquence de  $(A^+B^+C^+)$  peut être largement inférieure à celle de  $(A^+B^+)$ .

- éliminer les jeux contenant trop de règles contradictoires. Une règle contradictoire est de la forme  $(A^+B^+)$  70% et  $(A^+B^-)$  50%. Cela est du au fait qu'une partie de la base a des valeurs similaires pour l'attribut B. Or il se trouve que pour mesurer l'augmentation ou la diminution d'une valeur, nous utilisons les opérateurs  $\geq$  et  $\leq$ . Cela amène à considérer des sous-ensembles de patients communs aux règles

contradictaires.

Ces choix ont mené à la présentation des résultats obtenus pour une fréquence minimale de 70%. Nous avons, en accord avec les experts, décidé de fixer un seuil de confiance minimal élevé (80%), afin de ne fournir que les règles les plus pertinentes.

## 6.2. Analyse sémantique des règles graduelles

Traditionnellement, les analyses effectuées par les psychologues sont de nature statistiques : inférentielles telle que l'ANOVA, ou descriptives comme les corrélations. Nous discutons ci-après l'apport des règles graduelles par rapport à ces méthodes.

L'ANOVA permet essentiellement de mesurer les interactions entre plusieurs variables de même nature (échelle d'intervalle) en comparant leurs attributs deux à deux. L'ANOVA ne permet pas de regrouper les variations communes à plusieurs attributs pour un sous ensemble de patients précis sans avoir recours à la moyenne. De plus, l'ANOVA ne permet pas de savoir dans quel sens la différence entre les groupes est significative (par exemple le test donne-t-il des résultats inférieurs chez les malades ?) sans avoir recours à l'observation des moyennes ; ni de donner la puissance de cette différence. De plus, elle n'autorise pas le traitement de données négatives.

Les règles graduelles corrélent quant à elles des variations de sens différents entre plusieurs variables.

La recherche de corrélations permet de mesurer la force de relation entre deux variables. Les valeurs obtenues sont comparées par couple d'attributs. De nouveau, ces analyses n'autorisent pas les analyses de corrélations sur un ensemble de variables contrairement aux règles graduelles.

Voici un extrait des résultats obtenus avec GRITE (fréquence minimale à 80% et confiance minimale à 100%) :

- (81.82%, 100%) “Plus le patient est âgé et moins le souvenir des sons d'un bon moment est bon, alors moins le test du MMS est réussi”
- (87.88%, 100%) “Plus le souvenir de la disposition spatiale des personnes et du moment de la journée est bon, alors plus le souvenir de l'endroit est bon”
- (81.82%, 100%) “Plus le nombre d'identifications au test RI 48 augmente et plus le souvenir du moment de la journée et de la disposition spatiale d'un moment neutre est bon, alors plus le test du MMS est réussi”

Ici, le premier pourcentage représente la fréquence de la règle, et le second la confiance. Nous discernons deux types de règles : les règles dites de validation, qui n'apportent pas de nouvelles connaissances mais qui confirment ce que l'expert connaît déjà, et des nouvelles règles, qui apportent une information inconnue, méconnue ou surprenante. Dans cet exemple, la première règle est une règle de validation déjà connue pour les psychologues : une personne âgée qui a des difficultés à se souvenir aura un faible score au test d'efficacité globale. Les deux autres règles,

en revanche, amènent de nouvelles corrélations liant dimensions spatiales et temporelles aux souvenirs d'un lieu ainsi qu'à la réussite à un test de mémoire épisodique et d'efficience globale.

Finalement, l'avantage des règles graduelles vient des liens inédits qu'elles permettent d'établir entre les souvenirs (objet de l'étude) et les résultats aux divers tests (validation clinique). Grâce à de telles règles, les psychologues peuvent établir des profils type du stade préclinique de la maladie d'Alzheimer. Par exemple, il peut être utile de différencier le type de souvenir (bon, neutre, mauvais), afin de détecter des signes précurseurs : quels types de souvenirs s'effacent en premier ?

Néanmoins, la production de règles graduelles ne se suffit pas en elles-mêmes car elle conduit de fait à la mise en évidence de certains liens triviaux ou non pertinents. Par exemple, dire que l'efficience de la mémoire d'un patient alzheimer dépend du stade d'évolution de sa pathologie (e.g. score au MMSE) revient à démontrer, par analogie, que quelqu'un de grand n'est pas petit ! Ce type de relation, ne reposant sur aucune variabilité dans les données, n'autorise aucune mise à l'épreuve d'hypothèse. L'utilité de la production de règles graduelles repose donc nécessairement sur la collaboration avec des experts qui eux seuls sont à même d'apprécier les relations ayant un sens psychologique. En retour, l'apport majeur pour les psychologues est que l'apparition de règles inattendues permet de repenser le fonctionnement intellectuel dans son ensemble. Classiquement, la cognition est modélisée sous forme de registres indépendants. Son fonctionnement dépend de règles (computations) appliquées sur des représentations stockées en mémoire (symboles), autrement dit, la cognition est *computo-symbolique*. Cette conception, ainsi que les statistiques traditionnelles, nous conduisent à considérer que le tout est égal à la somme des parties, ces dernières étant cloisonnées et insensibles à ce qui les entoure. Or, il existe bien une autre conception du fonctionnement intellectuel selon laquelle le tout est plus que la somme des parties. Cette dernière s'intéresse, non aux registres, mais aux processus cognitifs mêmes. Dans ce contexte, le comportement observable d'un individu n'est pas le fait de la simple efficience d'un mécanisme intellectuel mais de l'expression d'un processus au regard de dimensions jusque là obliées : le contexte, le corps, la sensorimotricité, ou encore l'intentionnalité, les priorités de l'individu ou la temporalité par exemple. C'est pourquoi les théories de la complexité sont mieux à même d'expliquer comment l'individu s'adapte à son environnement et comment l'environnement se modifie en retour. Incrire le vieillissement dans cette perspective impose que l'on puisse analyser des données de différentes natures (i.e., qualitatives et quantitatives) et leur interactions réciproques. La production de règles graduelles permet ici la mise en évidence de liens ignorés, voir niés, par les statistiques traditionnelles. Finalement, ces règles nous permettent d'interpréter cliniquement les relations entre tests plutôt que le résultat à chaque test. Elles permettent de dégager des typologies de fonctionnement au sein d'une population dont l'hétérogénéité rend caduque nombre de tentative réalisées avec les statistiques classiques. En résumé, les implications futures de la collaboration entre informaticiens et psychologues cogniticiens sont multiples notamment dans le champ de l'ergonomie cognitive, de l'adaptation de l'environnement aux besoins des usagers, de la compréhension des phénomènes sous-jacents aux pathologies étudiées.

## 7. Aspect collaboratif

Ce projet s'inscrit dans le cadre du projet multidisciplinaire LAMAL : Langage, Mémoire et Alzheimer : une approche des maladies neurodégénératives fondée sur la densité des idées. L'objectif est de fouiller des données issues d'études psychologiques pour extraire des corrélations entre différents indices neuropsychologiques. Dans cette section, nous présentons l'impact de la collaboration entre des chercheurs informaticiens et psychologues cogniticiens sur la co-construction de nouvelles méthodes dans les deux disciplines ainsi que les difficultés rencontrées.

### 7.1. Informatique et psychologie cognitive

Pendant longtemps, l'informatique a été utilisée soit dans un but de simulation de modèles explicatifs par la reproduction automatique de processus humains, soit en tant qu'outils d'analyses et de traitement de données. La simulation de la pensée humaine a été étudiée par l'intelligence artificielle (e.g. (Wagman, 1993)). Il s'agit alors d'utiliser les théories proposées par la psychologie cognitive afin de raisonner et/ou prendre des décisions. Entre autres, la psychologie cognitive permet de sélectionner quels attributs privilégier pour modéliser un système ou un environnement donné. Dans le cas inverse, la psychologie a recours à l'informatique lorsque l'on ne peut pas observer directement le comportement. Par exemple, pour des questions d'éthique, il n'est pas possible d'observer directement l'implication comportementale de telle ou telle structure cérébrale en les lésant directement (expérimentation animale) dès lors qu'il s'agit d'un être humain. La simulation permet alors de générer de nouvelles hypothèses de recherche pour les experts.

Dans notre étude, la collaboration entre informaticiens et psychologues cogniticiens s'est ouverte avec les questions suivantes : comment les méthodes informatiques peuvent aider les psychologues à conduire une analyse intelligente des données ? Et à l'inverse, de quelle manière les méthodes habituellement utilisées en psychologie cognitive peuvent aider à améliorer les structures informatiques ?

Cette collaboration a abouti à une description originale et la découverte de nouvelles règles portant sur le vieillissement normal et pathologique, source de nouvelles hypothèses actuellement mises en œuvre. Le type de connaissances recherchées par les psychologues a, en retour, obligé les informaticiens à revoir leurs algorithmes d'extraction en les conduisant vers des règles graduelles. Cette remise en cause s'est traduite par des applications dans de nombreux autres domaines.

Par la combinaison constructive de cultures scientifiques, la multidisciplinarité a permis l'identification de thématiques pour lesquelles les outils d'une première discipline permettent de découvrir de nouvelles connaissances ou de nouvelles fonctions dans la deuxième discipline et inversement.

## 7.2. Difficultés rencontrées

Pour étudier les données des psychologues, nouvelles pour les informaticiens, ces derniers doivent les comprendre et adapter leurs techniques. Cela les conduit à se poser de nouvelles questions et développer de nouvelles approches dans leurs propres thématiques. L'enrichissement est alors mutuel.

Toutefois, la mise en place d'une telle collaboration ne se fait pas sans difficultés. L'intégration de technologies pour étudier des données issue de tests cognitifs entraîne chez l'une et l'autre des communautés la nécessité de construire un vocabulaire commun ainsi que le partage de certaines connaissances afin de se comprendre. L'investissement intellectuel est très fort. Ce dialogue indispensable est long à s'installer. Par exemple, durant la phase d'analyse des besoins, il faut savoir transmettre des connaissances aux partenaires sans en donner trop pour ne pas les noyer sous un flot dont ils ne pourront rien traiter. Il est également nécessaire de savoir écouter les partenaires afin d'identifier les éléments importants et nécessaires pour réaliser la collaboration. Enfin, une fois ce dialogue installé, il est nécessaire de bien identifier les problématiques levées de chaque côté.

Dans le cas de cette étude, pour les informaticiens, il s'agit de répondre à la demande croissante d'outils performants sur des bases de données d'un type particulier en fouille de données puisqu'elles contiennent peu de lignes par rapport au nombre d'attributs, alors que la plupart des algorithmes ont été conçus à l'inverse pour traiter des bases de données contenant de très nombreuses lignes (transactions bancaires par exemple) par rapport au nombre d'attributs. On parle alors de bases inversées. De plus, les algorithmes proposés doivent rester objectifs. Les informaticiens doivent donc comprendre en surface les résultats obtenus avec les outils classiques dans le contexte des psychologues, ce qui dépasse de loin la seule dimension d'implémentation technique. Du côté des psychologues, il est nécessaire de comprendre la manière d'utiliser les résultats et de les interpréter au regard de la subjectivité des individus.

La phase de post-traitement des résultats, est également une période forte de la collaboration. En effet, les psychologues doivent trouver parmi la multitude de résultats proposés par les informaticiens des éléments qui font sens pour eux et aider les informaticiens à mettre en forme ses résultats pour qu'ils soient réellement exploitables. Les informaticiens quant à eux doivent comprendre ce que les psychologues recherchent dans les résultats pour éventuellement les retravailler, et relancer des sessions d'expérimentations afin de mieux répondre aux besoins de chacun.

L'enrichissement mutuel de deux disciplines telles que l'informatique et la psychologie cognitive repose donc sur une collaboration de longue haleine mais qui peut être très fructueuse comme le souligne les premiers résultats obtenus dans le projet LAMAL.

En résumé, cette collaboration a permis, pour les psychologues, de repenser à la fois leur recueil de données et certaines théories explicative du vieillissement normal et pathologique, pour les informaticiens d'éprouver les limites d'une analyse pure-

ment automatique des données et la nécessité permanente de remettre en cause les algorithmes pour les adapter aux questions posées.

## 8. Conclusion

L'objectif de ce travail est de montrer l'intérêt d'une collaboration entre psychologues et informaticiens pour analyser comment les méthodes et outils de gestion et extraction de connaissances peuvent être utilisés et être très utiles dans le cadre du traitement de données liées au vieillissement. Les données traitées ont été acquises à partir de tests destinés à l'évaluation des capacités cognitives. Ces données sont de deux types : des productions langagières (récits) issues d'interviews structurées de patients et des données tabulaires numériques évaluant les performances de la mémoire des patients (sous la forme de tableaux excel).

Nous appliquons pour l'analyse des méthodes d'extraction automatique de connaissances non supervisées. Il s'agit alors de décrire les principales tendances présentes dans les données à travers des règles graduelles qui sont l'une des méthodes de fouille de données les plus adaptées aux données numériques présentes dans cette étude. Ces règles permettent d'exprimer des co-variations entre attributs, sous la forme "plus / moins  $A_1$  et ... plus / moins  $A_n$ , alors plus / moins  $B_1$  ... plus / moins  $B_n$ ".

Dans, cet article, nous avons décrit la mise en œuvre de la collecte des données, de l'analyse de ces données par l'extraction de règles originales, les règles graduelles et de l'interprétation de ces règles par les psychologues. Nous avons par ailleurs dressé un premier bilan de ce projet, les apports mutuels de chaque discipline et nous avons noté les besoins de chaque communauté ainsi que les difficultés rencontrées.

Les travaux décrits dans cet article ouvrent la voie à de nombreuses perspectives. En particulier, nous notons que l'application des règles graduelles, si elle permet de découvrir des connaissances pertinentes pour les psychologues, donne de très nombreux résultats. Trois pistes principales sont envisagées pour pallier ce problème :

- La restriction aux seuls motifs graduels clos permettra de réduire les motifs extraits. Cependant, le concept de clos reste difficile à définir dans le cadre de tels motifs.
- L'étude de mesures de qualité associées aux motifs graduels permettra d'ordonner les règles selon les critères de pertinence souhaitées par l'utilisateur final.
- La mise en place de filtres restant à définir permettra aux experts d'avoir différents points de vue sur ces résultats et facilitera leur interprétation.

## 9. Bibliographie

Agrawal R., Srikant R., « Fast Algorithms for Mining Association Rules », *20th International Conference on Very Large Data Bases, (VLDB'94)*, p. 487-499, 1994.

- Bagby R., Taylor G., Parker J., « The Twenty-Item Toronto Alexithymia Scale-II. Convergent discriminant, and concurrent validity. », *Journal of Psychosomatic Research*, vol. 38, p. 33-40, 1994.
- Berzal F., Cubero J.-C., Sanchez D., Vila M.-A., Serrano J. M., « An alternative approach to discover gradual dependencies », *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)*, vol. 15, n° 5, p. 559-570, October, 2007.
- Di Jorio L., Laurent A., Teisseire M., « Extraction efficace de règles graduelles », *Extraction et gestion des connaissances (EGC'2009)*, p. 199-204, January, 2009.
- Fayyad U. M., Piatetsky-Shapiro G., Smyth P., « From Data Mining to Knowledge Discovery : an Overview », *Advances in knowledge discovery and data mining*, vol. 1, p. 1-34, 1996.
- Hüllermeier E., « Association Rules for Expressing Gradual Dependencies », *PKDD '02 : Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery*, Springer-Verlag, London, UK, p. 200-211, 2002.
- Johnson K. M., Foley A. M., Suengas G. A., Raye L. C., « Phenomenal characteristics of memories for perceived and imagined autobiographical events », *Journal of Experimental Psychology : General*, vol. 117, p. 371-376, 1988.
- Laurent A., Lesot M.-J., Rifqi M., « GRAANK : Exploiting Rank Correlations for Extracting Gradual Dependencies », *Proc. of FQAS'09*, p. 382-393, 2009.
- Wagman M., *Cognitive Psychology and Artificial Intelligence*, vol. 1, Praeger Publishers, 1993.
- William J., Broadbent K., « Autobiographical memory in attempted suicide patients », *Journal of Abnormal Psychology*, vol. 95, p. 144-149, 1986.

Article reçu le 6 mars 2009

Accepté après révisions le 18 février 2010

**Lisa Di-Jorio** termine un doctorat en informatique à l'Université de Montpellier 2 sur la fouille de données de santé.

**S. Bringay** a obtenu un doctorat de l'université Jules Vernes de Picardie en 2006. Après deux ans d'ATER au CERIM de Lille (CEntre de Recherche en Informatique Médicale de Lille), elle est maîtresse de conférences à l'Université de Montpellier 3 depuis septembre 2007. Ses centres d'intérêt sont l'ingénierie des connaissances, le web sémantique, les annotations, les ontologies, les activités collaboratives, l'informatique médicale, le dossier patient et la fouille de données.

**Denis Brouillet** est professeur de psychologie cognitive expérimentale à l'Université Paul Valéry de Montpellier. Ses recherches s'intéressent à la mémoire humaine dans une perspective synergétique. La pathologie et en particulier la maladie d'Alzheimer considérée comme modèle, permet de déterminer le ou les facteurs qui participent au fonctionnement normal de la mémoire. L'utilisation de la fouille de données est à ce titre une aide précieuse en faisant émerger des relations inconnues à l'origine d'hypothèses nouvelles.

**Anne Laurent** est maîtresse de conférences au LIRMM. Ses travaux portent principalement sur la fouille de données, la fouille de données complexes, la fouille de

données floue, dans l'objectif de détecter aussi bien des tendances que des exceptions. Anne Laurent est particulièrement intéressée par l'étude de l'utilisation de la théorie des sous-ensembles flous pour obtenir des résultats plus exploitables par l'utilisateur, tout en garantissant le passage à l'échelle.

**Sophie Martin** est Enseignant Chercheur à l'université Paul Valéry et psychologue spécialisé en neuropsychologie en Unité d'Evaluation, de Réentraînement et d'Evaluation Sociale et/ou Professionnelle pour personnes cérébro-lésées. Elle s'intéresse plus particulièrement aux modifications des capacités d'adaptation lors du vieillissement normal et pathologique ainsi qu'à l'issue d'accidents cérébraux ou de traumatismes crâniens. Elle est spécialiste de l'interaction sujet/environnement en tenant compte des ressources cognitives et du support environnemental offert au sujet. D'un point de vue fondamental, le vieillissement et les lésions cérébrales offrent un modèle lui permettant de défendre une vision dynamique de la cognition. D'un point de vue appliqué, les connaissances développées en psychologie cognitive, psychopathologie et psychologie de la santé lui permettent d'adapter l'environnement aux capacités des individus, offrant ainsi des opportunités de reprise de développement.

**Maguelonne Teisseire** est directrice de recherche Cemagref à l'UMR TETIS (Territoires, Environnement, Télédétection et Information Spatiale). Ses travaux de recherche s'inscrivent dans le domaine des grandes bases de données et plus particulièrement dans le domaine de l'extraction de connaissances et la fouille de données complexes (volumineuses et souvent hétérogènes), dynamiques, i.e. sous la forme d'un flot continu et la fouille de données approximatives dans un contexte d'aide à la décision.

**ANNEXE POUR LE SERVICE FABRICATION**  
A FOURNIR PAR LES AUTEURS AVEC UN EXEMPLAIRE PAPIER  
DE LEUR ARTICLE ET LE COPYRIGHT SIGNE PAR COURRIER  
LE FICHER PDF CORRESPONDANT SERA ENVOYE PAR E-MAIL

1. ARTICLE POUR LA REVUE :

*Technique et science informatiques. Volume 29 – n°8/9*

2. AUTEURS :

*Lisa Di Jorio\* — Sandra Bringay\*,\*\* — Denis Brouillet\*\* —  
Anne Laurent\* — Sophie Martin\*\* — Maguelonne Teisseire\*,\*\*\**

3. TITRE DE L'ARTICLE :

*Fouille de données issues d'études psychologiques liées au vieillissement : extraction de règles graduelles*

4. TITRE ABRÉGÉ POUR LE HAUT DE PAGE MOINS DE 40 SIGNES :

*Extraction de règles graduelles*

5. DATE DE CETTE VERSION :

*27 mai 2010*

6. COORDONNÉES DES AUTEURS :

– adresse postale :

\* LIRMM – UM2 – CNRS UMR 5506, 161 rue Ada 34392 Montpellier  
Cedex 5

{nom}@lirmm.fr

\*\* Université Paul Valéry – Montpellier 3, 17, rue abée de l'Épée 34090  
Montpellier

denis.brouillet@univ-montp3.fr, sophiesmr@aol.com

\*\* UMR TETIS – CEMAGREF, 500 rue J.F.Breton 34093 MONTPEL-  
LIER Cedex 5

maguelonne.teisseire@cemagref.fr

– téléphone : 0467418631

– télécopie : 0467418500

– e-mail : laurent@lirmm.fr

7. LOGICIEL UTILISÉ POUR LA PRÉPARATION DE CET ARTICLE :

L<sup>A</sup>T<sub>E</sub>X, avec le fichier de style `article-hermes2.cls`,  
version 1.23 du 17/11/2005.

8. FORMULAIRE DE COPYRIGHT :

Retourner le formulaire de copyright signé par les auteurs, téléchargé sur :  
<http://www.revuesonline.com>