



HAL
open science

REVIVOS: REalité VIRTuelle pour la construction de VOcabulaires Spécialisés

Nicolas Férey, Nioche Claire Toffano, Taillez Oriane Matte, Rachid Gherbi,
W. Turner

► **To cite this version:**

Nicolas Férey, Nioche Claire Toffano, Taillez Oriane Matte, Rachid Gherbi, W. Turner. REVIVOS: REalité VIRTuelle pour la construction de VOcabulaires Spécialisés. 18es Journées Francophones d'Ingénierie des Connaissances, Jul 2007, Grenoble, France. not specified. hal-00509922

HAL Id: hal-00509922

<https://hal.science/hal-00509922>

Submitted on 17 Aug 2010

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.

REVIVOS: REalité Virtuelle pour la construction de Vocabulaires Spécialisés

Nicolas FÉREY¹, Claire TOFFANO-NIOCHE¹, Oriane MATTE-TAILLEZ,
Rachid GHERBI¹ et William TURNER¹

¹LIMSI-CNRS UPR3251

Université Paris XI – Batiment 508 et 502 bis – 91403 Orsay Cedex
{Nicolas.Ferey, William.Turner, Rachid.Gherbi}@limsi.fr
claire.toffano-nioche@u-psud.fr

Résumé : Cet article vise à établir une méthodologie permettant d'aider à la comparaison de classifications réalisées par différents experts sur un ensemble de termes identiques. Les termes spécifiques ont été extraits de façon automatique à partir d'un corpus spécialisé à l'aide d'un outil d'extraction de terminologie et ont été regroupés en classes, définies librement par chaque expert. Notre méthode s'appuie sur un environnement de visualisation en Réalité Virtuelle (RV) au sein duquel sont représentées les différentes classifications. L'objectif de ce travail est d'étudier comment une méthode de visualisation RV permet de faciliter l'émergence d'une classification consensus, en révélant visuellement les différences de points de vue et les points communs de plusieurs classifications.

Mots clés : terminologie spécialisée, visualisation de graphe, classification de termes, Réalité Virtuelle, socio-informatique, Pratiques Collectives Distribuées.

1 Introduction

Le LIMSI-CNRS se dote d'un dispositif de Réalité Virtuelle (RV) qui pourrait être utile à nos recherches en socio-informatique (Turner, 2007). Cet article a pour objectif de présenter un programme visant à tester cette hypothèse d'utilité. L'objectif, à terme, est de construire un environnement RV d'assistance à la construction collective de classifications lexicales afin d'aider à la fabrication d'ontologies pour la gestion des pratiques collectives distribuées (Turner *et al.*, 2006). Nous définissons ici l'ontologie comme étant « la spécification explicite et formelle d'une conceptualisation commune » (Gruder, 1993 ; Guarino et Giaretta, 1995). L'article est divisé en trois parties. La première permettra d'expliquer pourquoi, avec l'objectif d'organiser un processus de clarification des points de vues, nous pensons que la RV sera utile pour fabriquer des vocabulaires partagés entre plusieurs acteurs. La deuxième servira à la présentation de notre démarche expérimentale pour la visualisation des données réunies. Enfin, la troisième présentera visuellement les différences de points de vue et les points communs entre plusieurs classifications lexicales. Nous cherchons à définir une sémantique d'interprétation des graphes de concepts représentés en milieu RV et à en évaluer son

intérêt par rapport à l'aide qu'elle apporte à l'organisation du travail de construction collective des vocabulaires spécialisés. Ce vocabulaire spécialisé, ou terminologie, doit être le reflet de la compréhension partagée par les communautés de pratique et est caractérisé par l'usage de mots spécifiques à un métier. Ces mots constituent les données observables des concepts du domaine, il semble donc naturel de baser la construction de la terminologie sur la base des matériaux qui les contiennent – les corpus – (Roche, 2005). Nous présentons un exemple concernant une spécialité de recherche en génétique, mais les évaluations envisagées n'ont pas encore été réalisées. Sans cette étude d'usage, nous ne pouvons pas conclure définitivement ici à l'utilité de notre démarche pour la construction collective d'une classification lexicale. Nous présentons l'état d'une recherche en cours, à savoir, la sémantique d'interprétation que nous pensons adopter pour organiser un travail collectif de terminologie en milieu RV.

2 Des objets numériques

Selon la définition donnée dans l'ouvrage collectif le *Traité de réalité virtuelle*, "La finalité de la RV est de permettre à une personne (ou à plusieurs) une activité sensori-motrice et cognitive dans un monde artificiel, créé numériquement, qui peut être imaginaire, symbolique ou une simulation de certains aspects du monde réel". (Fuchs & Moreau, 2006). Pourquoi choisir d'entreprendre un travail collectif de construction terminologique dans un monde artificiel, créé numériquement ? Notre réponse à cette question se trouve dans la précision des intentions communicatives qui en résulte.

Cette réponse fait référence au travail de Sperber et Wilson (1986) sur ce qui rend la communication possible. A la suite de Grice (1957), ces auteurs estiment que la communication, pour être effective, nécessite la production d'un stimulus par le locuteur qui (a) informe l'auditeur de quelque chose et (b) informe l'auditeur de son intention de l'informer de quelque chose. Ainsi que le souligne Dupuy (1992, p. 75), le propre de la communication réside dans l'intention (b). Or, c'est précisément ce qui explique la difficulté d'organiser des interactions dans des situations de communication ouverte qui sont celles (par exemple) engagées sur Internet entre des personnes qui sont éloignées géographiquement, socio-culturellement et économiquement les unes des autres. Les messages échangés sont le plus souvent interprétés en leur attribuant l'intention (a), et plus rarement l'intention (b) (Hinds et Kiesler, 2002). Ces recherches confirment l'utilité d'Internet pour échanger des informations mais montrent ses limites en tant qu'espace d'interaction et de communication. Il nous semble que la raison en est la difficulté de construire collectivement, et à distance avec d'autres, un vocabulaire qui code convenablement l'intention des émetteurs de documents d'informer leurs lecteurs de quelque chose. La véritable intention communicative est l'intention (b) mais, pour que cette intention soit comprise, les mots employés doivent être chargés d'une histoire. Le lien social est fabriqué dans la durée par des actions qui visent à percer l'opacité des intentions communicatives.

Afin d'organiser ces actions de clarification intentionnelle, les sociologues insistent sur l'intérêt de placer des objets au sein de l'environnement cognitif utilisé en commun par les personnes impliquées dans une activité collective (Bowker *et al.*, 1997). Le rôle joué par ces objets est celui d'un référentiel externe qui permet à chacun de se positionner. En effet, sans l'objet externe, les partenaires d'une relation s'engagent dans une « spéculativité » du type « je pense qu'il pense... » (Dupuy, 1992), nuisible à la coopération parce que rien dans les relations directes entre individus ne permet d'y mettre fin. Cette opacité inhérente aux relations sociales est diminuée lorsqu'elles s'organisent autour d'objets externes parce que le comportement est rapporté à ce qui existe dans le monde. L'objet médiatise nos relations : nous savons ce que nous voulons faire avec l'objet et le comportement de l'autre nous dira s'il est d'accord avec notre analyse. C'est dans ce dialogue autour de l'objet que se construit une histoire commune susceptible de faire naître une confiance dans l'intention communicative de l'autre.

Cela dit, si l'objet externe est utile pour clarifier l'opacité des relations sociales, cette clarification n'est jamais totale car le jeu de la spéculation se déplace pour se situer au niveau des propriétés attribuées à l'objet. Les philosophes de la science ont montré à quel point ces propriétés peuvent être perçues comme infinies. Elles varient non seulement intrinsèquement comme le montre Bachelard (1973) en proposant sa théorie des approximations successives pour s'approcher de la réalité objective, mais également en vertu des relations qui leur sont attribuées dans des cadres théoriques divers comme le montrent les travaux de Kuhn (1970). Le jeu de la spéculation prend alors la forme « je pense qu'il pense qu'il peut faire ceci avec l'objet parce que... ».

Au LIMSI, nous avons beaucoup travaillé sur cette malléabilité intrinsèque des objets informationnels et le rôle que cette malléabilité peut jouer dans la fabrication (ou non) des liens sociaux. Un objet informationnel est pour nous un objet médiateur du type numérique. Le programme REVIVOS, *RE*alité *VI*rtuelle pour la construction de *VO*cabulaires *Sp*écialisés, s'inscrit dans le prolongement de cette recherche. Nous pensons que le fait d'être plongé dans un objet informationnel, et d'être capable de l'observer depuis l'intérieur, devrait en principe nous aider à réduire sa malléabilité intrinsèque et donc augmenter son utilité pour tisser des liens sociaux. La raison en est la capacité augmentée qu'auraient des partenaires à un travail coopératif sur l'objet de fabriquer une sémantique qui non seulement informe les uns et les autres de quelque chose (intention (a) ci-dessus), mais les convainc réciproquement de leur intention de coopérer (intention (b)). Ce texte présente nos premières conclusions quant à l'usage des objets numériques RV pour animer la construction d'une sémantique collective.

3 Démarche expérimentale

Dans le cadre du projet *Diaspora Knowledge Network* que coordonne le LIMSI pour le Programme UNESCO sur les Migrations Internationales (www.unesco.org/shs/migration/diaspora), un des objectifs est de proposer des outils permettant de construire des cadres conceptuels de travail qui soient partagés par des scientifiques

collaborant à distance. Une de ces collaborations est le projet SEQXAM qui réunit 2 équipes de chercheurs, l'une en Colombie et l'autre en France, et concerne l'étude de la bactériose vasculaire causée par *Xanthomonas axonopodis* pv. *manihotis* (Xam), un des principaux parasites du manioc. Dans le but de construire un vocabulaire commun aux chercheurs, une première application a consisté en l'exploitation des 11 références bibliographiques sélectionnées par le responsable scientifique français du projet comme étant les documents principaux de l'étude.

3.1. Extraction terminologique

A partir du texte de ces publications, une extraction automatique des termes du domaine a été réalisée à l'aide du logiciel ETIQ (Amrani *et al.*, 2004), spécifiquement conçu pour un travail d'extraction de termes sur des corpus spécialisés. Voici un exemple des 355 termes spécialisés extraits automatiquement autour de l'objet de l'étude, la bactérie Xam (**Fig. 1**) :

X.axonopodis pv. citri,Xac
Xanthomonas
Xanthomonas avirulence (avr) genes
Xanthomonas axonopodis pv. manihotis-cassava
Xanthomonas oryzae pathovar oryzae KACC10331
Xcc 8004

Fig. 1 – Un exemple de termes extraits automatiquement du corpus spécialisé utilisé.

3.2. Classification

Deux experts ont proposé des regroupements de ces termes en plusieurs classes avec la consigne qu'un terme n'appartienne qu'à une unique classe, sans hiérarchie de classes ou de sous-classes. En premier lieu, on peut observer que certains termes sont distribués par les deux experts dans un contexte d'usage sensiblement identique mais qu'ils étiquettent différemment (termes entourés, **Fig 2**). Deuxièmement, le premier expert a classé les termes en 19 classes alors que le second s'est servi de 29. Troisièmement, la distribution des termes dans les classes varie considérablement : si le nombre moyen de termes réunis par le premier expert est de 13 et celui du deuxième est de 14, l'écart type est élevé dans les deux cas : 10 pour le premier et 15 pour le second. Enfin, lors du classement des termes par le premier expert, celui-ci a estimé que 112 termes extraits automatiquement du corpus spécialisé étaient soit trop généraux pour définir la spécificité du domaine (hors sujet), le deuxième expert n'en a écarté que 43. Ces différentes observations suggèrent une différence de compréhension du sujet entre les deux experts.

De nombreuses questions peuvent être soulevées dont, par exemple, la signification des étiquettes choisies pour nommer les classes : « catégories » retenues en vue de structurer l'activité ou, au contraire, simples « mots » arbitrairement choisis pour désigner des éléments semblables ? Que signifient les variations du nombre de classes, du nombre de mots retenus, et de leur distribution au sein des

classes ? Il est difficile de savoir si ces variations s'expliquent par des logiques d'organisation ou de recherche différentes, ou par une plus ou moins grande expertise du domaine.

Exemple de la classification des termes opérée par l'expert 1

| PATHOGENICITY (23) | PATHOGENICITY_EFFECTORS (27) |
|--|--|
| pathogenesis | hrp-genes |
| pathogenic bacteria | identified pathogenicity-related genes |
| pathogenic processes | known virulence factors |
| pathogenicity | molecular and biochemical action of TTSS effectors |
| pathogenicity of phytopathogen <i>Xanthomonas campestris</i> pv. <i>campestris</i> | newly discovered effectors |
| pathogens | novel type III effector proteins |
| Phytopathogenic bacteria | pathogenicity (hrp) genes |
| Plant pathogenic bacteria | phytopathogenic bacterial type III effectors |
| secretion apparatus | TTSS effector functions |
| state-of-the-art capillary electrophoresis instruments | TTSS-secreted proteins |
| Symptoms | Type III secretion system effector proteins |
| Xcc virulence | virulence effector protein |

Exemple de la classification des termes opérée par l'expert 2

| ORGANISM (69) | PATHO-SYSTEM (19) |
|---|------------------------------------|
| pepper-pathogenic <i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> | advances in TTSS biology |
| phytopathogenic bacteria | bacterial infection strategies |
| Phytopathogenic bacteria | components of TTSS |
| phytopathogenic <i>Pseudomonas</i> | distribution of TTSS |
| Plant pathogenic bacteria | Hrp TTSS |
| Plant pathogenic bacteria | Hrp type III secretion system |
| plant pathogenic bacterium <i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> | pathogenic processes |
| plant pathogens | pathosystem |
| Proteobacteria | putative type IV secretion systems |
| <i>Pseudomonas aeruginosa</i> | TTSS biology |
| <i>Ralstonia</i> spp. | TTSS injectosome |
| | TTSS of animal pathogens |

Fig. 2 –Extrait du contenu de deux classes de chaque expert. Les étiquettes des classes sont en majuscules sur fond grisé et suivies du nombre de termes entre parenthèses. Cette visualisation est issue de la plate-forme AGORAE (<http://www.dkn.techcico.fr/index.html/>).

3.3. Le tiers extérieur

Notre démarche méthodologique a été directement inspirée par les observations de Goody (1979) quant à la spéculation réflexive qui prend son origine dans des listes et des classifications. Selon lui, les classifications projettent une image de la réalité qui est mue par les intérêts et les motivations, plus que par des relations de cause à effet. Elles donnent à voir une lecture de la réalité, un récit, une manière de rendre intelligible une situation. Mais réciproquement, elles enrichissent la construction en cours d'un sens collectif partagé. En effet, chaque nouvelle inscription d'un élément sur une liste est une invitation à en parler, à examiner de quelle manière il renforce la cohérence des éléments déjà réunis ou, au contraire, la remet en cause. Le projet REVIVOS vise à fabriquer un environnement numérique permettant de travailler avec ces objets de manière à canaliser la « spéculativité » qu'ils engendrent mais aussi à ne pas empêcher l'émergence d'un contexte de coopération confiante entre les acteurs. Car le danger est là, comme nous l'avons vu en parlant de la difficulté d'instaurer des espaces de communication et d'interaction sur Internet.

Le même type de démarche que nous envisageons est à l'œuvre dans le domaine de l'apprentissage assisté par ordinateur. Par exemple, l'*Open University* en Angleterre, après avoir adopté le système MOODLE (*Modular Object Oriented Dynamic Learning Environment*) enrichi son environnement d'apprentissage virtuel avec des applications de travail coopératif sur les contenus scolaires. Il s'agit de créer des espaces de communication entre élèves et professeurs qui permettent aux élèves

d'avoir constamment à disposition un système de représentation des obligations du *workflow* imposé par le cours, des listes de connaissances pertinentes dynamiquement mises à jour, des outils pour contacter d'autres élèves (chats, blogs, forums, ...) et enfin, des mécanismes d'évaluation des connaissances. Dans ces configurations, le professeur joue par rapport à la montée en puissance de la spécularité le rôle du tiers extérieur, c'est-à-dire, celui qui, dans la théorie de Sperber et Wilson, garantit ce qu'ils appellent le *mutual manifestness*. Par rapport à l'ensemble de ses élèves, le professeur est présent en tant que point de focalisation externe de leurs interactions. Il se manifeste par ses interventions (corrections, suggestions) et l'existence de ce tiers externe crée les conditions d'une canalisation positive de la spécularité. Les élèves sont en mesure d'anticiper sur ce que veut le professeur, non seulement parce que ce dernier se manifeste de temps en temps, mais surtout parce que le sens de ses interventions devient un objet de discussion entre eux, leur permettant de construire ensemble la sémantique de leur action collective.

Ce qui caractérise les pratiques collectives distribuées qui se nouent sur Internet est l'absence d'un tiers externe qui peut servir de point de focalisation pour l'alignement de comportements communicatifs. Nous pensons au LIMSI que les professionnels de l'information pourraient éventuellement jouer ce rôle (Turner 1997), mais la faiblesse d'une telle hypothèse réside dans une expertise moindre que celle normalement attribuée à un professeur. Ce qui explique pourquoi, au LIMSI, nous cherchons à pallier cette expertise présumée faible en matière de contenus scientifiques et techniques par une expertise procédurale forte en matière de construction du sens collectif qu'il convient à accorder à l'organisation des pratiques sociales sur Internet. Selon nous, cette expertise procédurale suppose une double compétence (Turner, 2007) : technique pour ce qui concerne la configuration d'un espace numérique d'interaction et de communication ; et sociale pour ce qui concerne les interventions qui sont destinées à faire dialoguer les acteurs entre eux. Nos travaux actuels sur REVIVOS reprennent cette double contrainte : d'une part, nous travaillons en vue de créer un espace de visualisation des classifications spécialisées facilement configurable par des non-informaticiens ; et d'autre part, nous cherchons à mettre au point des clés d'interprétation de la visualisation utiles pour la canalisation positive de l'activité collective de construction du sens.

4 Méthode de visualisation RV

Nous avons donc à comparer de deux classifications différentes. Nous proposons une méthodologie qui permet à la fois de comparer ces deux classifications mais aussi de proposer une série d'actions pour fusionner ces deux classifications en une seule. Nous n'avons pas réalisé la fusion des classifications, puisque le début de cette histoire appartient aux experts et qu'il est apparu des choix que nous ne pouvons faire à leur place. Cependant, nous avons mis au point puis exploré cette possibilité de fusion de classifications grâce à leur visualisation dans l'environnement de RV. Et ce sont ces étapes que nous décrivons ensuite : i) le calcul d'une mesure exprimant le nombre de termes en commun entre les classes de chaque expert prises deux à deux,

ii) la représentation de cette mesure de comparaison et de la taille des classes à l'aide d'un graphe multivalué et iii) la visualisation du graphe organisé dans l'espace 3D d'un environnement virtuel et interactif qui permettra la modification dynamique des classifications.

4.1 Mesure de comparaison

Dans un premier temps, nous avons utilisé les notions d'inclusion, d'intersection, et d'exclusion, pour comparer deux à deux les classes de termes fournies par les deux experts. La notion d'inclusion de la classe A dans la classe B, signifie que tous les termes constituant la classe A appartiennent à la classe B. La notion d'intersection entre deux classes, signifie que deux classes partagent plusieurs termes, sans pour autant être incluses l'une dans l'autre. Enfin la notion d'exclusion décrit deux classes n'ayant aucun terme en commun. Cependant, ces notions sont insuffisantes pour effectuer des comparaisons fines entre deux classes et le critère de la taille, en nombre de termes, de chaque classe paraît un critère important à prendre en compte (deux classes sont d'autant plus similaires que leur taille est proche). Nous avons donc choisi une mesure pour comparer deux classes qui prend en compte à la fois la taille des deux classes, et la taille de leur intersection (Equation 1).

$$comp(A, B) = \frac{|A| \times |B|}{|A \cap B|^2} \quad (1)$$

4.2 Structure de données des classes et de leurs relations

Nous disposons donc comme données d'entrée, de la liste des termes spécifiques du corpus, et de deux listes de classes décrites par une étiquette, regroupant les termes selon l'expertise de chaque expert. Nous connaissons d'autre part, la taille en nombre de termes de chaque classe, l'identité de l'expert à l'origine de chaque classe, et nous disposons de relations d'inclusion, d'intersection, ou d'exclusion, pour chaque paire de classes ainsi que d'une métrique de comparaison.

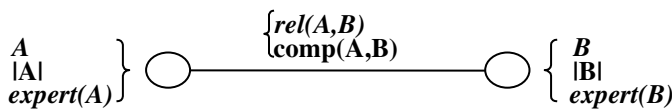


Fig. 3 – Graphe multivarié à deux sommets et appliqué à la comparaison de deux classes A et B, de taille |A| et |B|, en relation d'inclusion ou d'intersection $rel(A,B)$, et dont la mesure de comparaison vaut $comp(A,B)$. Les valeurs qualitatives sont en italiques contrairement aux valeurs quantitatives.

Nous avons choisi une structure de données spécifique de graphe multivarié pour représenter et visualiser ces données : chaque sommet et chaque arête est porteur d'un ensemble d'attributs qualitatifs, et/ou quantitatifs. Dans notre étude (**Fig. 3**), chaque sommet, représentant une classe, est porteur de l'identité de l'expert qui a défini cette classe, de la taille de cette classe, et de l'étiquette de cette classe. Les

arêtes reliant ces sommets, sont porteuses du type de relation, inclusion ou intersection, et de la mesure de comparaison entre ces deux classes. La relation d'exclusion entre deux classes n'est pas considérée comme une arête, car par définition, ces classes ne possèdent aucun terme en commun.

4.3. Dessin du graphe

Une fois ces données organisées sous forme de graphe multivarié, il s'agit de le dessiner, en traduisant les informations quantitatives et qualitatives, par des caractéristiques graphiques correspondantes du dessin du graphe. Un dessin de graphe est donc composé de sommets, caractérisés par une couleur, une forme et/ou une transparence, et d'arêtes, caractérisées-elles aussi par une couleur, une forme, une transparence, une épaisseur, et une longueur. Nous choisissons ici de représenter une classe par un sommet dont la forme sphérique ou carrée représente le propriétaire de la classe, et dont le volume symbolise le nombre de termes qui la constitue. Deux classes sont reliées par une arête si elles possèdent des termes en communs (relation d'intersection), ou lorsqu'elles sont incluses l'une dans l'autre (relation d'inclusion). La couleur noire des arêtes permettra de différencier les relations d'inclusion de celles d'intersection en gris. Enfin, la métrique de comparaison quantitative entre deux classes est représentée par la longueur de l'arête reliant ces deux classes. Ainsi deux classes sémantiquement proches, c'est-à-dire qui possèdent un grand nombre de termes en commun, seront spatialement proches dans la représentation du graphe. La caractéristique quantitative de la longueur de l'arête dans la représentation visuelle implique de résoudre des contraintes de distance entre plusieurs sommets interconnectés. Comme ces contraintes ne sont en général pas satisfaites dans l'espace euclidien, nous avons exploité *Graph3DExplorer* (Ferey, 2006) qui utilise un algorithme de type masse/ressort (Eades, 1984 ; Herman *et al.*, 2000) et dessine un graphe qui approche de manière optimale ces contraintes de distance.

4.4. Clés d'interprétation de la visualisation

Une fois une représentation sous forme de graphe obtenue, nous avons utilisé la plate-forme MUSE (Multi-User Stereoscopic Environment) et les outils d'interaction immersive mis au point par l'équipe de RV&A « VENISE¹ » de notre laboratoire, afin de pouvoir naviguer dans le graphe dans un contexte immersif (i.e. de réalité virtuelle). Afin de structurer cette navigation, plusieurs patrons du graphe visualisé semblent prometteurs pour canaliser la spéculation et faire dialoguer des acteurs de manière coopérative et confiante entre eux. D'une part, ces patrons visuels mettent en évidence des relations simples entre les classes des deux experts et, d'autre part,

¹ VENISE (Virtual ENvironment for Immersive Simulation and Experiments) est une action de recherche pluridisciplinaire en Réalité Virtuelle et Augmentée (RV&A) du LIMSI-CNRS dont l'une des applications phares est l'exploitation des techniques immersives de la RV au domaine de la Bio-informatique (site web : <http://www.limsi.fr/venise>).

ils facilitent la prise de décisions en termes de modifications à effectuer pour aboutir à une classification commune. Ainsi, en partant des patrons visuels des plus simples aux plus compliqués (Fig. 4), nous identifions très rapidement les classes incluses les unes dans les autres (patron 1). Ceci suggère une spécialisation plus forte de la classe incluse dans l'autre. La sémantique que nous attribuons à cette relation est celle d'une différence d'expertise entre les experts. La décision à prendre dans ce cas pour aboutir à une classe consensus dépend du degré d'expertise recherché dans la classification finale : choisir entre la conservation des deux classes ou leur fusion. Dans le cas d'une intersection forte (patron 2), quand deux classes partagent un nombre important de termes, la discussion sur une éventuelle fusion entre ces deux classes s'imposera sans doute. Le résultat dépendra des termes indépendants *a priori* peu nombreux, puisque les classes sont proches.

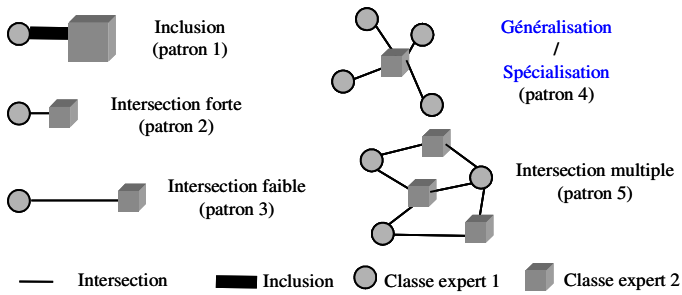


Fig. 4 – Patrons visuels identifiés pour structurer le travail collectif dans l'espace

Il est probable que dans le cas d'une intersection faible (patron 3) entre deux classes, lorsque les deux classes partagent peu de termes, il sera nécessaire de clarifier deux points de vue potentiellement différents sur l'organisation de l'activité à entreprendre. Dans ce cas, une démarche utile pourrait être d'identifier et d'extraire les termes communs, et de choisir si ces termes peuvent constituer une classe consensus. La sémantique que nous attribuons aux patrons en étoiles (patron 4) est également celle d'une différence de spécialisation des experts, moins forte que dans le cas de l'inclusion (patron 1), mais qui conduit au même processus décisionnel. Pour terminer, les regroupements de classes très proches (patron 5) sont plus difficiles à traiter de façon systématique, mais pourraient n'être qu'une expression d'interrelation semblable à celle d'une interaction forte (patron 2). Une proposition de scénario décisionnel est de discuter dans un premier temps de la fusion des classes en commençant par les plus proches, en extrayant les termes non partagés, et en les distribuant de façon consensuelle dans d'autres classes.

5 Scénario vers une classification commune

Nous proposons dans cette section un scénario vers une classification commune qui s'appuie sur les clés d'interprétation précédemment évoquées. Notre méthode de visualisation met très rapidement en évidence des différences de granularité entre les deux classifications visualisées et à propos desquelles les acteurs de la tâche de

classification devront prendre une décision (fusion, maintien ou création d'une nouvelle classe) en fonction des objectifs de la classification finale. Nous avons jusqu'ici interprété ces différences de granularité comme des conflits de spécialisation entre les experts. Il pourrait exister d'autres types de conflits, en particulier issus de différences de points de vue sur l'activité à réaliser (par exemple : classer en fonction des types de pathologies, de symptômes, ou encore de causes de la pathology). Ces différences de points de vue concernant la classification à réaliser doivent être résolues avant le début de l'activité et l'étape initiale de notre approche consiste donc à définir un consensus de la tâche à réaliser.

Ensuite, une première navigation permet d'observer les étiquettes des classes. En effet, de rares cas concernent l'identité ou quasi identité de deux classes, qu'il suffira de réunir en une seule classe (ex. **Fig. 5** : proximité spatiale et terminologie forte des deux classes au centre).

Après, les conflits les plus simples à résoudre sont ceux associés à une différence de granularité propre à chaque classification. C'est le cas des patrons visuels de type inclusion (détection rapide grâce au noir vs gris des arêtes, **Fig. 6**). Selon la granularité de la classification finale, la classe « CANCER », pourrait être fusionnée avec la classe « PLANT DISEASE », étant incluse dans celle-ci. La classe « PLANT DISEASE » pourrait aussi être divisée afin de conserver la classe « CANCER », et en attribuant les termes restant de la classe « PLANT DISEASE » à une autre classe

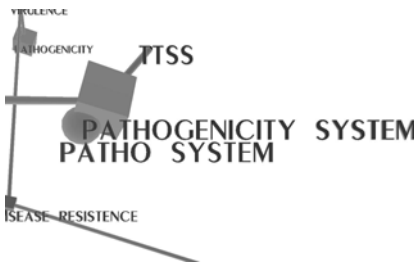


Fig. 5 – Equivalence sémantique PATHO/GENICITY SYSTEM

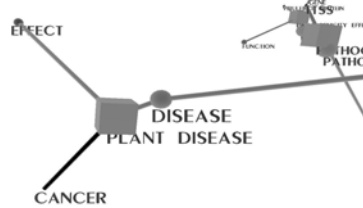


Fig. 6 – Inclusion "CANCER" dans "PLANT DISEASE"

existante ou à une nouvelle classe. D'autre part, l'équivalence forte entre les classes « PLANT DISEASE » et « DISEASE », ainsi que le choix de leur étiquette, indiquerait une relation hiérarchique où la classe « PLANT DISEASE » désignerait un concept plus spécialisé que celui associé à la classe « DISEASE ». Cependant on peut soulever l'incohérence de cette interprétation, puisque la taille de la classe « DISEASE » est plus petite que la classe spécialisée « PLANT DISEASE ». Dans ce cas, l'incohérence pourrait être résolue en fusionnant ces deux classes. Le conflit à résoudre ici porte plutôt sur les termes non partagés par ces deux classes et en nombre marginal étant donnée leur forte proximité.

A la résolution des conflits simples, succède l'analyse des ensembles de classes formant des clusters denses. Ces ensembles présentent parfois une organisation en étoile (**Fig. 7**) témoignant d'une différence d'expertise des deux classifications. La classe « PATHOGENICITY EFFECTOR » du premier expert (carré) est une classe

très proche des classes plus petites du second expert (sphères). Pour organiser cet ensemble de termes proches, le premier expert a utilisé une seule classe alors que le second en a utilisé plusieurs. Les décisions (conservation ou fusion des classes spécialisées ?) dépendent de la pertinence de ces classes spécialisées par rapport au corpus, ainsi que de la spécificité de la classification commune souhaitée.

Enfin, il reste des clusters de classes sans organisation particulière, plus difficiles à interpréter (Fig. 8). Leur résolution passe dans un premier temps par la comparaison deux à deux des classes, des plus proches aux plus éloignées, en étudiant les termes non partagés. La redistribution des termes dans d'autres classes, éventuellement nouvelles, permet d'envisager la fusion des classes comparées. En parallèle, la vérification de la pertinence des classes les plus petites, par rapport aux objectifs en termes de granularité de la classification finale, permet d'en supprimer certaines. Ces décisions sont beaucoup moins systématiques et objectives que celles des cas précédents, et cette étape nécessite discussion entre les acteurs.



Fig. 7 – Cluster dense organisé en étoile

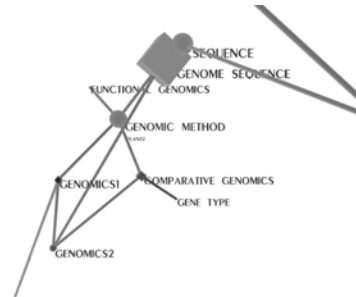


Fig. 8 – Clusters sans organisation

Notre méthode n'a pas pour objectif de résoudre automatiquement des divergences de point de vue mais justement de les exposer, de les concrétiser par la visualisation des deux classifications en même temps dans l'environnement RV. C'est ensuite aux acteurs présents ensemble dans le dispositif, d'échanger leurs points de vue pour aboutir à une action. L'animateur, notre tiers extérieur, doit avoir la capacité de rendre visible les différentes dimensions du problème. Son action n'est pas uniquement *top-down*, c'est-à-dire orientée par la codification d'une procédure d'interaction suffisamment experte pour lancer la discussion sur les différents points suggérés ; elle est également *bottom-up* puisque son rôle est de faire « monter » en discussion des questions d'incompréhension qui nécessitent (souvent) une modification de granularité.

6 Conclusion

REVIVOS vise à expérimenter l'intérêt d'utiliser la réalité virtuelle pour la construction des vocabulaires spécialisés avec l'objectif d'organiser un processus de clarification des points de vues. Cette construction suppose la mise en cohérence de différents points de vue et la capacité de détecter les points communs de plusieurs classifications. Nous avons présenté dans cet article 1) une méthode RV de visualisation lexicale pour aider à la comparaison de classifications réalisées par

plusieurs experts ; 2) une sémantique interprétative des résultats de cette visualisation afin d'organiser un travail coopératif au sein de l'espace numérique et 3) un cadre théorique permettant d'inscrire le projet REVIVOS dans le prolongement des recherches socio-informatiques du LIMSI. Ces recherches concernent plus particulièrement le rôle que jouent des objets informationnels produits par l'informatique (listes, classifications, visualisations,...) dans l'organisation des pratiques collectives distribuées. La réalité virtuelle ouvre des nouvelles perspectives à ce type de recherche. C'est ce que nous avons voulu montrer dans cet article.

Références

- AMRANI A., KODRATOFF Y., MATTE-THAILLIEZ O (2004) A Semi-automatic System for Tagging Specialized Corpora. In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004)*, p. 670-681
- BACHELARD G. (1973) *Essai sur la Connaissance Approchée*, Paris: Vrin
- BOWKER G., STAR L., TURNER W.A., GASSER L.(ED.) (1997) *Social Science, Technical Systems and Cooperative Work : Beyond the Great Divide*, New Jersey, Lawrence Erlbaum Associates
- DUPUY JP. (1992) *Logique des phénomènes collectifs*, Paris, Ecole Polytechnique, Ellipses
- EADES. P. (1984) A heuristic for graph drawing. *Congressus Numerantium*, 42 : p.149-160
- FÉREY N. (2006). Exploration immersive de données génomiques textuelles et factuelles : vers une approche par Visual Mining. *PhD thesis, Université Paris Sud XI*.
- FUCHS PH., (DIR) & MOREAU, G. (COOR) (2006), *Traité de la Réalité Virtuelle*, Paris : Presses de l'Ecole des Mines, www.ensmp.fr/Presses.
- GOODY J (1979), *La raison graphique*, Paris: Editions de Minuit
- GRICE HP (1957) Meaning, *Philosophical Review*, 66, pp 377-388.
- GRUBER T. (1993) A translation approach to portable ontology specifications, *Knowledge Acquisition* 5(2), pages 199-220.
- GUARINO N. & GIARETTA P. (1995) Ontologies and knowledge bases, towards a terminological clarification, in MARS N., eds., *Towards very large knowledge bases: knowledge building and knowledge sharing*, IOS Press, pages 25-32.
- HERMAN M., MARSHALL S., MELANÇON G. (2000) Graph visualisation and navigation in information visualisation: A survey. *IEEE Trans. on Visualization & Computer Graphics*, 6(1)
- HINDS PJ & KIESLER S (ED) (2002) *Distributed Work*, Cambridge, Mass : MIT Press
- KUHN TS. (1970) *The Structure of Scientific Revolutions*, Chicago : Univ. of Chicago Press
- ROCHE C (2005) Terminologie et Ontologie – LAROUSSE – Revue « Langages » -n°157
- SPERBER D., & WILSON D. (1986) *Relevance*, Oxford : Basil Blackwell
- Turner WA. (2007) Elements pour une socio-informatique dans Brossaud C. & Reber, B. (eds) *Les Sciences sociales à l'heure des technologies de l'information et de la communication*, Paris : Hermès, *Traité IC2 Série Cognition*, à paraître
- TURNER WA. (1997) *Pour une nouvelle alliance entre les documentalistes et les scientifiques : Conférence d'Ouverture du Colloque*, Actes du Colloque Information scientifique et technique : nouveaux enjeux documentaires et éditoriaux, Paris : INRA
- TURNER WA., BOWKER G., GASSER L., ZACKLAD M. (ED) (2006) Infrastructures for Distributed Collective Practices, *J. of Comp. Supported Cooperative Work (Special Issue)*, 15(2-3)