



Des systèmes complexes pour l'observation de l'activité humaine

Christophe Courtin

► To cite this version:

Christophe Courtin. Des systèmes complexes pour l'observation de l'activité humaine. Informatique [cs]. Univ Savoie Mont Blanc, 2022. <tel-04961309>

HAL Id: tel-04961309

<https://hal.science/tel-04961309v1>

Submitted on 21 Feb 2025

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.



HAL Authorization



HABILITATION À DIRIGER DES RECHERCHES

Spécialité : **Sciences informatiques**

Présentée par

Christophe COURTIN

préparée au sein du **Laboratoire SYMME**
dans l'**École Doctorale SIE (Sciences, Ingénierie, Environnement)**

"Des systèmes complexes pour l'observation de l'activité humaine"

H.D.R. soutenue publiquement le **25 novembre 2022**,
devant le jury composé de :

Monsieur Éric BONJOUR

Université de Lorraine, Nancy (PR 61), Rapporteur

Madame Nathalie GUIN

Université Claude Bernard, Lyon 1 (MCF HDR 27), Examineur

Monsieur Pierre MARET

Université Jean Monnet, Saint Étienne (PR 27), Rapporteur

Monsieur Christophe ROCHE

Université Savoie Mont Blanc, Chambéry (PR 27), Examineur

Monsieur Kavé SALAMATIAN

Université Savoie Mont Blanc, Annecy (PR 27), Rapporteur



Remerciements

Mes remerciements vont tout d'abord à mon épouse Emmanuelle pour sa grande patience, ainsi qu'à ma famille et notamment mes enfants Juliette, Caroline et Lucas, pour leur soutien.

Je remercie les membres du jury qui me font l'honneur de rapporter et d'examiner ce document. Je remercie Nathalie Guin pour ses interactions scientifiques dans les réseaux de recherche (i.e. Clusters I.S.L.E. ¹, A.R.C.6 ²) depuis plusieurs années et ses conseils avisés pour la présentation de mes travaux.

Je remercie Pierre Maret pour ses qualités scientifiques et humaines avec qui j'ai eu plaisir à co-encadrer une thèse. Je remercie Christophe Roche et Kavé Salamatian pour nos interactions et leur vision humaine du métier de chercheur en informatique.

Je tiens également à remercier Éric Bonjour d'avoir accepté d'apporter, entre autres, une dimension industrielle essentielle aux travaux que nous menons sur les compétences et l'innovation.

Je tiens à exprimer une sincère reconnaissance pour tous les collègues qui ont facilité mon activité de chercheur, comme Laurent Tabourot pour mon intégration au laboratoire SYMME à un moment critique de ma carrière, et Georges Habchi pour m'avoir incité à passer cette H.D.R. Je dois également beaucoup à mes collègues de toujours, Christine Ferraris et Laurence Vignollet, qui m'ont fait confiance il y a vingt-deux ans en participant à mon recrutement et avec lesquelles j'entretiens de très bonnes relations depuis tout ce temps. Je conçois mon travail de chercheur comme une collaboration continue avec des collègues qui sont essentiels à ma capacité d'innovation. C'est en collaborant avec, entre autres, Stéphane Talbot, Miguel Tomasena, Jean-Luc Maire, Lilia Gzara, Jean-Jacques Girardot, Alain Mille, ..., que j'ai réalisé bon nombre de travaux.

Je n'oublie pas d'exprimer ma gratitude aux collègues doctorantes et doctorants, Hajer Chebil, Soumaya Yahiaoui, Olivier Pouydebat et Kamel-Eddin Bemmami, qui forment le ciment de nos travaux de recherche. Le bal a été ouvert par Hajer, qui est restée une amie chère. En présentant ce travail, j'honore mon engagement auprès de l'École Doctorale qui m'a octroyé la direction de la thèse de Kamel, dont le travail est exceptionnel.

Je tiens à exprimer toute mon estime pour les collègues lyonnais et grenoblois qui nous ont soutenus quand notre laboratoire SysCom était au creux de la vague il y a une dizaine d'années, et sans lesquels nos recherches se seraient interrompues. Je remercie enfin bon nombre d'autres personnes avec qui j'ai le plaisir de travailler (et pas seulement ...) au quotidien, Pascal Francescato, Yann Meyer, Philippe Saffré, ..., mais que notre trop vaste environnement professionnel ne me permet pas de citer de manière exhaustive.

1. projet « Personnalisation des Environnements Informatiques pour l'Apprentissage Humain »

2. « T.I.C. et usages informatiques innovants »

Table des matières

Remerciements	i
Table des matières	iii
Introduction générale	1
1 Activité professionnelle	5
1.1 Activités de recherche	6
1.1.1 Contexte des activités de recherche	6
1.1.2 Thématiques, problématiques et questions de recherche	7
1.1.3 Stratégie de recherche	18
1.1.4 Démarche scientifique	22
1.2 Curriculum vitæ	26
2 Traces numériques	33
2.1 Introduction	34
2.2 Traces d'interaction numériques	36
2.2.1 Système à base de traces	36
2.3 Analyse des traces d'interaction	41
2.3.1 Formalisme de construction des séquences	42
2.3.2 Règles de recherche des séquences basées sur les gabarits	42
2.4 Activité d'observation	46
2.4.1 Station d'observation	47
2.4.2 Instrumentation	47
2.4.3 Rétroaction	49
2.5 Expérimentation d'une station d'observation	51
2.5.1 Les objectifs de l'expérimentation	51
2.5.2 Le contexte de l'expérimentation	52
2.5.3 Les outils logiciels de l'expérimentation	53
2.5.4 Le déroulement des opérations	60
2.5.5 La collecte des traces	61
2.5.6 L'analyse des traces	62
2.5.7 Les résultats de l'expérimentation	63
2.6 Conclusion	66
3 Corpus de traces	67
3.1 Introduction	68
3.2 Interopérabilité avec CARTE	69
3.2.1 A.P.I. <i>Collect</i>	70
3.2.2 A.P.I. <i>Query/Notify</i>	71
3.3 Corpus de traces d'activité	72
3.3.1 Contexte de recherche	74
3.3.2 Partage et exploitation des corpus de traces	75

3.3.3	Approche PROXYMA	77
3.3.4	Plate-forme BEATCORP	87
3.4	Expérimentation d'une plate-forme de <i>benchmarking</i>	90
3.4.1	Les objectifs de l'expérimentation	90
3.4.2	Le contexte de l'expérimentation	90
3.4.3	Le déroulement des opérations	90
3.4.4	Les résultats de l'expérimentation	91
3.5	Conclusion	97
4	Des traces aux compétences	99
4.1	Introduction	100
4.2	Travaux de recherche	102
4.2.1	Les compétences	102
4.2.2	Le positionnement de notre instrumentation	107
4.2.3	Les outils de traitement de données pour les compétences	112
4.3	Expérimentation de recherche de compétences	119
4.3.1	Les objectifs de l'expérimentation	119
4.3.2	Le contexte de l'expérimentation	119
4.3.3	Les résultats de l'expérimentation	120
4.4	Expérimentation d'identification de situations d'innovation	130
4.4.1	Les objectifs de l'expérimentation	130
4.4.2	Le contexte de l'expérimentation	131
4.4.3	Les résultats de l'expérimentation	132
4.5	Conclusion	136
	Conclusion générale et perspectives	137
	Bibliographie	I
	Table des figures	XI
	Liste des tableaux	XII
A	Annexe 1 : activité scientifique	XV
B	Annexe 2 : trace d'interaction	XXI
C	Annexe 3 : gestion des données personnelles	XXV

Introduction

Ce document présente l'évolution du travail de recherche en science informatique que nous mènerons ces prochaines années à travers la description de nos travaux passés et actuels à l'université Savoie Mont Blanc dans les laboratoires « Systèmes Communicants » (SysCom), puis « SYstèmes et Matériaux pour la Mécatronique » (SYMME). Ces travaux s'inscrivent dans une collaboration pluridisciplinaire que nous avons menée initialement dans la thématique de l'observation de l'activité humaine à travers les « Environnements Informatiques pour l'Apprentissage Humain » (E.I.A.H.³) et, depuis quelques années, dans la thématique de la reconnaissance des compétences individuelles et collectives en y intégrant le domaine industriel.

Nos travaux étant centrés sur l'humain, nos problématiques sont très liées à de nombreux autres domaines scientifiques proches : sciences de l'éducation, sciences de l'information et de la communication, sciences de gestion, sociologie, etc. Nous considérons les systèmes informatiques comme des systèmes interactifs intelligents qui enrichissent l'activité au service de l'humain, en répondant à ses attentes et en respectant ses valeurs. Dans une approche systémique, nos objets de recherche impliquent des concepts relatifs à la dynamique des activités humaines réalisées dans les environnements numériques : concept de trace numérique d'activité contextualisée et concept de compétence mobilisée.

Nos problématiques de recherche sur l'observation de l'activité humaine portent sur l'explicitation de celle-ci à partir des traces numériques ou numérisées et sur le partage de corpus de traces d'activité et d'outils d'analyse dans des contextes définis. Les activités que nous étudions s'inscrivent dans les environnements de Travail Collaboratif Assisté par Ordinateur (T.C.A.O.)⁴ et plus particulièrement dans le domaine de l'apprentissage⁵. Nos travaux de recherche ont contribué à la modélisation et à la spécification de Systèmes à Base de Traces (S.B.T.⁶), avec notre système BEATCORP⁷ et le format de trace CARTE⁸, puis des systèmes de corpus de traces contextualisées, avec notre approche PROXYMA¹⁰.

Lors de notre intégration au laboratoire SYMME en 2014, nos travaux se sont inscrits dans le domaine de la mécatronique et en particulier dans l'axe « optimisation des processus » associé à l'industrie du futur. Nous menons des études sur le rôle fondamental du capital humain

3. « *Technology Enhanced Learning Environments* » (T.E.L.E.) ou « *Learning Management Systems* » (L.M.S.)

4. « *Computer-Supported Collaborative Work* » (C.S.C.W.)

5. « *Computer-Supported Collaborative Learning* » (C.S.C.L.)

6. « *Trace-Based Systems* » (T.B.S.)

7. « *BEenchmarking platform for Analysis of Traces CORPora* »

8. « *Collection, activities Analysis and Regulation based on Traces Enriched* » publié dans la conférence « EC-TEL » (Courtin and Talbot 2009) disponible sur https://rd.springer.com/chapter/10.1007/978-3-642-04636-0_52 classée A+ selon l'A.T.I.E.F.⁹ <http://www.atief.fr/ressources/classement/conferences-internationales>

10. « *PROXY for Multiple Analyses* » publié dans les revues « *International Journal of Information and Education Technology* » (Chebil, Courtin, and Girardot 2012b) et « *International Journal of Learning Technology* » (Chebil, Courtin, and Girardot 2015)

dans l'inventaire des valeurs d'une organisation, et plus particulièrement dans son positionnement par rapport à ses concurrents. La gestion des compétences est devenue un facteur essentiel de la performance des processus industriels. Avec la transformation numérique des organisations, nous orientons donc nos travaux vers la construction d'outils logiciels qui apportent une connaissance approfondie des compétences mobilisées. Nous travaillons donc dans la thématique de la reconnaissance des compétences individuelles et collectives dans les domaines de l'apprentissage et de l'industrie du futur. Nos problématiques de recherche ont ainsi été élargies à l'explicitation des compétences individuelles et collectives à partir de l'activité humaine effective et contextualisée avec, entre autres, une application aux activités d'innovation d'exploitation et d'exploration.

De par le lien qui unit les traces d'interaction, les activités et les compétences, cette orientation de nos travaux touche de nouvelles questions de recherche relatives aux compétences en tant qu'objet de recherche, mais aussi aux liens avec les traces d'activité. Nous avons étudié la contribution des S.B.T., et notamment des systèmes CARTE et PROXYMA, à la construction de Systèmes à Base de Compétence (S.B.C.). Nous faisons le constat du potentiel, mais aussi de certaines limites des traces d'interaction, que nous présentons au travers d'expérimentations. Nos investigations nous ont conduit également à considérer les possibilités offertes par les diverses techniques d'instrumentation des outils logiciels et celles de l'intelligence artificielle pour les analyses de données. Nous montrons notamment la capacité de l'approche PROXYMA à intégrer un large spectre de sources d'informations pour identifier des compétences. En réponse à nos questions de recherche sur la nature des traces d'interaction numériques permettant de formaliser des compétences, nous avons proposé un modèle structurel et un modèle dynamique de compétence, ainsi que l'architecture SBS-SEARCH¹¹ pour opérationnaliser un système à base de compétence. Dans le cadre d'un projet A.N.R., des développements logiciels sont programmés (20 mois/H) à partir d'octobre 2022 pour intégrer les modules du S.B.C. dans une plate-forme Web accessible librement.

Les travaux à venir s'inscrivent, entre autres, dans le projet P.I.A.¹² « Territoires d'innovation pédagogique » / « Campus des métiers et des qualifications » dans la thématique de la reconnaissance des compétences individuelles et collectives dans les domaines de l'apprentissage et de l'industrie du futur. La problématique de recherche de ce projet concerne l'articulation entre les processus managériaux pour la gestion des compétences dans l'entreprise et les processus de reconfiguration continue des formations professionnelles. La réponse aux questions de recherche touchera à la fois les verrous scientifiques relatifs à l'observation de l'activité humaine et à l'explicitation des compétences individuelles et collectives. Des investigations dans les travaux scientifiques des domaines des E.I.A.H. et de l'industrie du futur sous-tendront ces travaux de recherche dans les années à venir.

11. « *Skill-Based System - Skills Extract, Analyze, Regroup, Consult and Historize* »

12. « Programme d'Investissement d'Avenir »

Présentation du plan

Ce document est structuré en quatre chapitres.

Le premier chapitre intitulé « Activité professionnelle » présente les thématiques, les problématiques et les questions de recherche que nous traitons depuis plus de deux décennies au sein de deux laboratoires. Nous associons également nos principaux résultats (modèles, architectures, P.O.C.¹³, prototypes) qui ont été proposés en réponse aux questions de recherche à travers les projets financés.

Nous présentons ensuite la stratégie qui structure notre recherche et la logique qui a permis d'articuler les travaux réalisés au cours du temps autour de collaborations avec d'autres équipes de recherche. Les domaines de recherche explorés sont en concordance avec mon parcours scientifique personnel initié par le « travail collaboratif assisté par ordinateur » (T.C.A.O.) dans le cadre de ma thèse intitulée « Interfaces coopératives pour les dossiers médicaux multimedias ». Cette dimension collaborative s'est ensuite inscrite dans l'« Éducation » avec les E.I.A.H. au sein du laboratoire « Systèmes Communicants » (SysCom), puis dans l'« industrie du futur » au sein du laboratoire « SYstèmes et Matériaux pour la MécatroniquE » (SYMME). Les domaines d'application successifs ont en commun la capacité à fournir des situations écologiquement fondées, c'est-à-dire présentant les caractéristiques des situations authentiques. Nous avons donc adopté une démarche scientifique analogue dans les différents projets de recherche, qui privilégie les opérations sur des terrains d'expérimentation avant d'élaborer des modèles théoriques.

Le deuxième chapitre intitulé « Traces numériques » aborde la caractérisation des systèmes à base de traces numériques (S.B.T. ou T.B.S. pour *Trace-Based Systems*) issues notamment des activités collaboratives d'apprentissage. Les activités du laboratoire SysCom étant intrinsèquement pluridisciplinaires (science informatique et science de l'information et de la communication), nos travaux se sont orientés initialement vers les problématiques liées à la conscience du travail de groupe (*group awareness*) et la régulation de l'activité à partir des traces d'interaction. La diversité des applications s'appuyant sur l'analyse des traces d'activité a révélé le problème de l'hétérogénéité des formats que nous avons tenté de solutionner avec le format générique CARTE (*Collection, activity Analysis and Regulation based on Traces Enriched*).

Le troisième chapitre intitulé « Corpus de traces » est consacré à la modélisation des corpus de traces d'interaction associés à des contextes de recherches expérimentales. Ce travail répond au besoin de disposer de données contextualisées pour analyser des situations réelles et en particulier dans le domaine de l'apprentissage. Les traces ayant fait l'objet d'une abondance de travaux dans de nombreux domaines d'application, aucun format n'est passé à la postérité, ce qui nous a conduit à adopter une nouvelle approche générique intitulée PROXYMA (*PROXY for Multiple Analyses*) pour modéliser les corpus de traces d'interaction sans format pivot et spécifier une architecture de la plate-forme BEATCORP (*BEenchmarking platform for Analysis of Traces CORPora*) de gestion et application à la construction de corpus de référence.

13. « *Proof Of Concept* »

Le quatrième chapitre intitulé « Des traces aux compétences » aborde les notions relatives à la représentation des compétences identifiées à travers l'activité au moyen des traces numériques. Les travaux sur les compétences s'inscrivent dans une nouvelle thématique spécifique qui est intrinsèquement liée à celle de l'observation de l'activité humaine. Les travaux présentés tendent à positionner les systèmes à base de traces et les systèmes de corpus de traces, à travers des expérimentations, en mesurant leur intérêt et leurs limites scientifiques et techniques. Nous présentons des propositions de modèles de compétence et d'architectures de systèmes à base de compétences, intégrant la notion d'activités tracées. Ce dernier chapitre présente les applications potentielles, dans le domaine de l'industrie du futur, mais aussi dans celui de la formation professionnelle avec une approche par compétences.

Nous concluons en récapitulant le cheminement scientifique des travaux par strates successives pour mener jusqu'aux usages des systèmes numériques utilisés dans les activités professionnelles. Les travaux ont montré l'intérêt de la convergence entre les observations dans le domaine de l'apprentissage et les activités d'innovation dans le domaine de l'industrie du futur. Les perspectives des travaux sont déjà engagées dans un projet P.I.A. qui permettra de concrétiser la convergence précitée.

1

Activité professionnelle

Les systèmes d'observation de l'activité humaine sont des systèmes complexes pour lesquels il paraît difficile d'appliquer une logique déductive et réductionniste pour conceptualiser les phénomènes observés. En d'autres termes, les effets observés ne peuvent être associés à des causes supposées sans une connaissance explicite des situations réelles. Nous privilégions donc une logique inductive qui s'appuie sur des faits observés à travers des éléments temporellement situés que nous appelons des traces numériques d'activité humaine. On peut également dire que notre approche consiste à rechercher les formes opératoires de l'activité, qui correspondent à un registre pragmatique de conceptualisation et non épistémique. C'est par une meilleure connaissance d'une partie du réel que nos modèles pourront être efficaces au-delà des connaissances théoriques qui auront contribué à la conceptualisation. Les modèles conceptuels permettront alors de transmettre sous forme schématisée les principes fondamentaux et les fonctionnalités de base des systèmes complexes qu'ils représentent.

Sommaire

1.1	Activités de recherche	6
1.1.1	Contexte des activités de recherche	6
1.1.2	Thématiques, problématiques et questions de recherche	7
1.1.3	Stratégie de recherche	18
1.1.4	Démarche scientifique	22
1.2	Curriculum vitae	26

1.1 Activités de recherche

1.1.1 Contexte des activités de recherche

Laboratoire : « Systèmes Communicants » (SysCom) / équipe « Observation »

Axe de recherche : Observation de l'activité

Thématiques de recherche : La modélisation des traces d'interaction, l'analyse des corpus de données d'interaction, la conscience de l'activité (*awareness*)

Champ d'activité : les « Environnements Informatiques pour l'Apprentissage Humain » (E.I.A.H.)

En 2000, nos travaux de chercheurs s'inscrivent dans les problématiques du **laboratoire « Systèmes Communicants » (Équipe d'Accueil E.A. et Équipe de Recherche Technologique Éducation E.R.T.E. SysCom)** sur les Environnements Informatiques pour l'Apprentissage Humain (E.I.A.H.), dans la continuité de mes travaux de doctorant en *Computer-Supported Collaborative Work* (C.S.C.W.) appliqués au domaine médical. L'activité du laboratoire SysCom étant intrinsèquement pluridisciplinaire (science informatique et science de l'information et de la communication), nous nous orientons initialement vers les problématiques liées à la conscience de l'activité (*awareness*). En 2005, j'intègre la nouvelle équipe « observation » du laboratoire, et nous étudions les problématiques relatives aux traces d'interaction, issues des systèmes numériques et de la transcription des activités collaboratives d'apprentissage non instrumentées (à partir de vidéos, de questionnaires, etc.). En 2006, l'A.E.R.E.S. souhaite le regroupement des petites structures, et préconise l'intégration du laboratoire SysCom dans le Laboratoire d'« Informatique, Systèmes, Traitement de l'Information et de la Connaissance » (LISTIC) de l'université de Savoie. Dans le même temps, l'A.E.R.E.S. demande au laboratoire LISTIC de diminuer le nombre de ses axes de recherche, ce qui implique pour mes collègues et moi-même une intégration individuelle sans être en mesure de poursuivre nos thématiques, que l'agence reconnaît pourtant pour leur qualité et leur rayonnement. Étant personnellement impliqué dans des projets (A.C.I. ACTEURS, A.N.R. PROCOGEC, Cluster ISLE, P.P.F. APPRENTICE), je poursuis mes travaux de recherche dans l'équipe SysCom reconnue par l'université de Savoie. Pendant huit ans, je maintiens le cap grâce à mes partenariats grenoblois (LIG, ICAR), lyonnais (LIRIS) et stéphanois (RIM), en co-encadrant notamment une thèse avec un financement obtenu auprès du Cluster ISLE sur la création de corpus de traces d'activité dans les E.I.A.H. Il convient de noter que mener une activité de recherche dans ce type de structure locale nécessite beaucoup d'énergie personnelle, et que malgré cela, mon activité tendait à s'essouffler faute de moyens suffisants.

Laboratoire : « SYstèmes et Matériaux pour la MécatroniquE » (SYMME)

Axe de recherche : Qualité Industrielle

Thématique de recherche : l'analyse des compétences à partir de l'activité effective

Champs d'activité : l'Industrie 4.0 et l'Éducation

En 2014, cinq membres de l'équipe SysCom intègrent le **laboratoire « SYstèmes et Matériaux pour la Mécatronique » (SYMME)** pour renforcer les problématiques informatiques liées à la mécatronique. Pour ma part, je m'appuie sur mes compétences acquises dans le domaine de l'apprentissage pour élargir mes problématiques de recherche dans les domaines de l'industrie et de l'énergie (recherche de compétences pour l'évaluation du capital immatériel d'une entreprise étendue, traçage de l'utilisation des systèmes énergétiques pour l'étude des comportements d'utilisateurs). J'encadre actuellement deux thèses sur les données d'interaction dans les systèmes collaboratifs d'ingénierie, avec un focus particulier sur d'une part l'évaluation des compétences associées à ces données et d'autre part les contextes d'innovation d'exploitation et d'exploration. Cette intégration au sein du laboratoire SYMME redonne un véritable élan à ma recherche, en apportant un cadre de travail « normal », dans un laboratoire disposant d'une dynamique positive. L'approche par les traces représente une nouvelle compétence dans le laboratoire qui me vaut d'autres implications dans des appels à projets obtenus (FRESBE pour l'axe « énergie » en collaboration avec le laboratoire LISTIC de l'USMB, ARC6, AAP générique USMB et A.N.R. pour l'axe « industrie » et P.I.A. pour l'axe « Formation et Industrie »).

1.1.2 Thématiques, problématiques et questions de recherche

Nous présentons dans cette partie les deux **thématiques** qui ont sous-tendu nos travaux de recherche depuis environ deux décennies : l'observation des activités humaines et la reconnaissance des compétences individuelles et collectives.

La première thématique s'est inscrite dans le domaine des environnements informatiques pour l'apprentissage humain (E.I.A.H.) et la deuxième dans l'industrie du futur et plus récemment couplée à la formation professionnelle. Les deux thématiques ont en commun l'activité humaine à partir de laquelle nous collectons des traces que nous exploitons en vue d'explicitier nos observations.

Thématique « Observation des activités humaines »

Notre laboratoire de recherche « Systèmes Communicants » (SysCom) était structuré en deux équipes : « Observation » et « Scénarios pédagogiques ». Nos travaux s'inscrivant dans l'équipe « Observation », nous avons opté pour observer l'activité humaine à l'aide des traces laissées par les systèmes informatiques. Cette idée est apparue dans les années 80 par le constat de la production de traces (événements liés aux interactions de l'utilisateur) dans les logiciels. Notre première utilisation des traces numériques d'interaction a été motivée par le besoin d'indicateurs (e.g. *feedback*) dans un éditeur partagé pour faciliter la collaboration (cf. Annexe A). La trace est devenue un objet d'étude à part entière qui a fait l'objet de nombreux travaux depuis quelques décennies. L'orientation de la majorité des travaux a été la création de nouveaux formats adaptés aux domaines d'application considérés pour intégrer les données afférentes. Il existe des spécifications standard pour le partage des données, telles que le modèle *SCORM*¹ (*Sharable Content Object Reference Model*), les métadonnées *DCMI*² (*Dublin Core Metadata Initiative*), les métadonnées *LOM*³ (*Learning Object Metadata*), et plus récemment xAPI (spécification succédant à *SCORM* (Bakhouyi et al. 2020)) pour décrire les ressources respectivement issues du réseau et du domaine éducatif.

Nous traitons dans cette thématique la **problématique** de l'explicitation de l'activité humaine à partir des traces numériques ou numérisées. Cette problématique présuppose la capacité

1. <https://scorm.com/>

2. <http://dublincore.org/>

3. <https://www.ieeeltsc.org/working-groups/wg12LOM/lomDescription/>

à collecter des traces informatiques et la capacité de celles-ci à représenter l'activité humaine. Les traces fournies par les collecticiels se limitant généralement aux fichiers de *logs* fournis par les serveurs Web, nous étudions l'idée d'une instrumentation des systèmes collaboratifs en vue d'établir un historique des actions réalisées. Les travaux que nous avons menés entre 2004 et 2007 dans le volet 3 « Instrumentation pour l'observation des activités collectives » du projet « A.C.T.E.U.R.S. »⁴ ont permis de poser les bases d'une instrumentation des systèmes collaboratifs d'apprentissage. Au-delà des verrous techniques qui peuvent être associés à cette instrumentation (e.g. système propriétaire avec le code source non accessible), nous avons étudié la **question de recherche** sur la représentation des traces d'activité que nous pouvons exprimer de la manière suivante :

Comment représenter les traces d'activité pour permettre à un système de les analyser en tenant compte de l'hétérogénéité de leur format ?

Ce volet du projet « A.C.T.E.U.R.S. » s'intéresse à l'instrumentation des dispositifs d'apprentissage collectif, tels que les environnements numériques de travail ou tout autre environnement offrant des modalités d'apprentissage collectif à distance ou hybrides, synchrones ou asynchrones. Le constat est que les systèmes pour l'apprentissage humain ne disposaient pas de fonctionnalités satisfaisantes pour permettre un suivi de l'activité, et pour rétroagir automatiquement à partir de la reconnaissance de certaines situations particulières (e.g. échec, retard, etc.). Les nombreuses sources de traçage possibles, issues des dispositifs d'apprentissage collaboratif existants, étaient destinées à la construction d'observations adaptées à différents types d'observateurs (enseignants, apprenants, experts, etc.). La problématique générale de ce volet consistait à définir un modèle intégrant un format générique des traces à partir de ces sources, et à l'aide de transformations de ces traces, d'en produire de nouvelles destinées à des interprétations à différents niveaux d'abstraction, ou à des rétroactions sur les dispositifs concernés.

Dans nos premiers travaux de recherche, nous avons étudié cette question en proposant un format pivot nommé CARTE (*Collection, activities Analysis and Regulation based on Traces Enriched*) composé de méta-données et de paramètres qui représentent respectivement les caractéristiques communes à tous les formats (e.g. la date, la description) et les éléments spécifiques liés aux outils utilisés (e.g. le destinataire pour un outil de communication). La structuration par imbrication du format CARTE (i.e. encapsulation des traces nommées « signaux » ou « séquences » selon les niveaux d'abstraction) permet de produire des traces à l'aide de transformations à différents niveaux d'abstraction et d'accéder à différentes interprétations. Ce format pivot est détaillé dans la partie 2 du document.

Sur la question de l'instrumentation, nous basons notre recherche sur les hypothèses suivantes :

- les techniques d'observation traditionnelles (annotation de transcriptions visuelles, sonores ou textuelles, interprétation de traces numériques d'activités) sont insuffisantes pour disposer de résultats d'analyse pertinents en nombre suffisant : elles doivent être enrichies de fonctionnalités automatiques et adaptables permettant aux différents types d'observatrices et d'observateurs de disposer des données appropriées ;
- les enseignantes et les enseignants ne sont pas suffisamment impliqués dans l'élaboration des dispositifs d'observation des situations d'apprentissage collectif, pour prendre en compte les pratiques existantes de suivi et d'accompagnement, adaptées aux situations rencontrées sur le terrain : les enseignantes et les enseignants doivent pouvoir concevoir facilement des modèles d'utilisation qui permettent la transformation des traces en vue d'une interprétation par l'observateur lui-même, c'est-à-dire avec un niveau d'abstraction adapté ;

4. n° PS-2004-008 « Action Concertée Incitative » (A.C.I.) « Activités Collectives et Tutorat dans l'Enseignement Universitaire : Réalités, Scénarios »

- les apprenants eux-mêmes sont impliqués dans l'élaboration des situations d'apprentissage collectif, et doivent donc être impliqués dans la modélisation et l'exploitation d'observations de leurs propres pratiques en situation d'activité collective et également d'observation du collectif.

Le projet permet la spécification, la conception, et le développement d'une méthodologie d'instrumentation et l'élaboration de solutions techniques afférentes, permettant l'amélioration significative de l'observation et de l'analyse des situations collectives. Dans notre approche, nous plaçons les traces au cœur de notre modélisation, considérant que tout ce qui est fourni par les sources de traçage est transformable en vue d'être exploité à différents niveaux d'abstraction. Nous intégrons donc notre format de traces dans un système à base de traces (S.B.T.) du même nom CARTE (*Collection, activity Analysis and Regulation based on Traces Enriched*) (Courtin 2010), qui représente le dénominateur commun à toutes les phases de l'observation. Les transformations précitées sont réalisées grâce à des modèles d'utilisation des outils logicielles élaborés en collaboration avec les enseignantes et les enseignants. Le système à base de traces est détaillé dans la partie 2 du document.

Pour mettre à l'épreuve la généricité de notre format de traces CARTE, nous l'avons utilisé pour analyser des traces issues d'apprentissage collaboratifs instrumentés, avec des outils d'analyse ayant des formats d'entrée spécifiques. Pour analyser ces traces en vue d'automatiser certaines tâches, nous avons couplé notre instrumentation avec un outil d'analyse existant en 2009 dans le cadre de l'action de recherche « 3A »⁵ du projet pluri-formation (P.P.F.) APPRENTICE. De notre collaboration dans ce projet avec le laboratoire « ICAR »⁶ à Lyon et le département « RIM »⁷ de l'École des Mines de Saint-Étienne, nous avons utilisé l'outil d'analyse TATIANA⁸ (Dyke, Lund, and Girardot 2009). L'outil est utilisé par des chercheuses et des chercheurs pour caractériser, annoter, structurer et synchroniser des traces d'utilisation (e.g. fichiers de *logs*, traces d'interaction de *chat* ou de constructions graphiques, transcriptions de sources audio et vidéo, etc.) au cours d'activité d'apprentissage multimodal.

La première partie du projet consiste à développer deux passerelles entre les formats de trace des systèmes CARTE et TATIANA, la première faiblement couplée via des mécanismes de conversion des formats pour préserver l'indépendance des outils (cf. Figure 1.1), la deuxième fortement couplée pour exploiter la capacité d'interprétation par les outils des traces dans leur format d'origine en important leur modèle d'utilisation CARTE (cf. Figure 1.2).

Suite aux résultats probants de cette première action 3A, nous entamons une nouvelle collaboration avec le « LIESP »⁹ de l'INSA à Lyon dans le cadre de l'action de recherche « FATMA »¹⁰ du projet pluri-formation (P.P.F.) APPRENTICE. Cette nouvelle action vise à concevoir un système d'aide à l'analyse automatique de traces issues de communications médiatisées (i.e. forums de discussion, messageries instantanées, ...), dédié à l'évaluation des activités d'apprenants par des tuteurs.

Nous proposons alors de coupler l'outil « TrAVis »¹¹ d'analyse de traces de communications médiatisées au format CMC, développé au laboratoire LIESP, avec notre S.B.T. expérimental basé sur le modèle CARTE, qui analyse et structure automatiquement les traces en fonction de règles d'analyses prédéfinies pour construire des indicateurs d'*awareness* graphiques. Dans ce contexte, le système ainsi constitué d'analyse de traces issues d'une situation d'apprentissage collaboratif

5. « Automatic Analysis Assistant »

6. « Interactions, Corpus, Apprentissages, Représentations » (UMR 5191 ICAR)

7. « Réseaux, Internet, Multimédia »

8. « Trace Analysis Tool for Interaction ANALysts »

9. « Laboratoire d'Informatique pour l'Entreprise et les Systèmes de Production » (EA 4125 LIESP)

10. « Forum Analysis for Tutoring Mediatized and Assisted »

11. « Tracking data Analysis and Visualization tools »

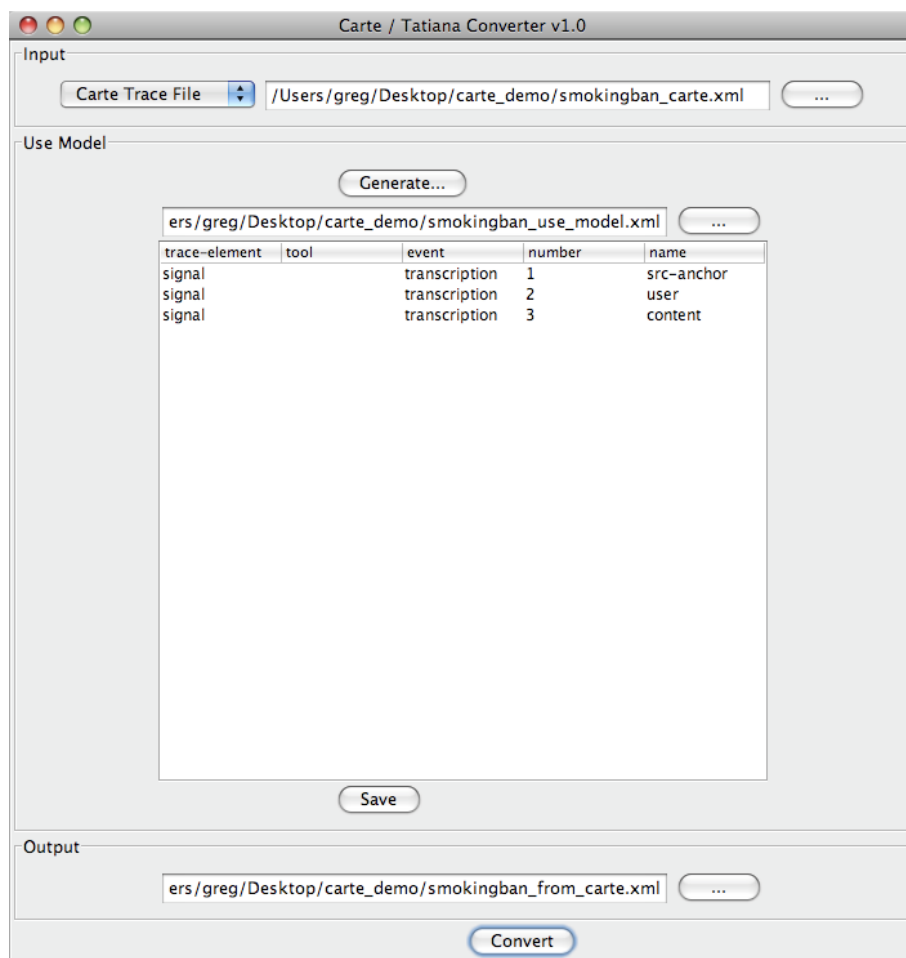


FIGURE 1.1 – *couplage faible entre CARTE et TATIANA*

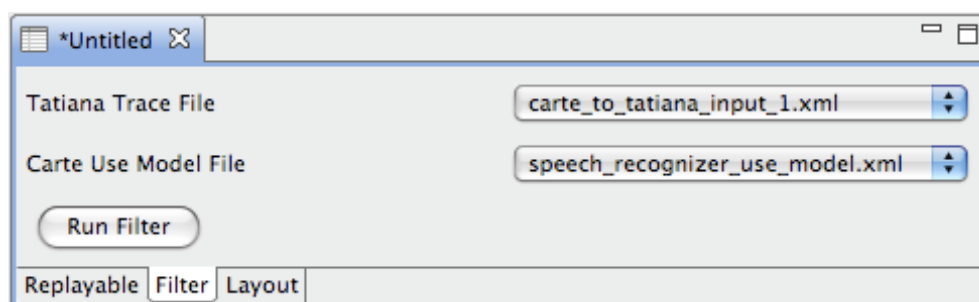


FIGURE 1.2 – *couplage fort entre CARTE et TATIANA*

a pour vocation d'aider les tuteurs à mieux suivre ou évaluer les activités des apprenants. Les apprenants, quant à eux, peuvent utiliser le système d'analyse de traces pour prendre conscience des activités des autres participants, ainsi que leurs propres activités afin, par exemple, d'effectuer une analyse réflexive.

Techniquement, nous disposons :

- de l'outil TrAVis, fondé sur des technologies Web, dédié aux tuteurs et aux apprenants afin de leur permettre d'analyser et de visualiser leur activité de communication médiatisée sous la forme d'indicateurs graphiques construits à partir des traces d'utilisation ;
- du S.B.T. CARTE qui permet d'analyser et de structurer automatiquement les traces d'interaction lors d'une activité dans un environnement informatique.

L'outil résultant permet de :

- convertir les formats de traces CARTE et CMC (cf. Figure 1.3) ;
- fusionner (coupler) les modèles de deux outils dans le but d'enrichir les traces ;
- construire de nouveaux indicateurs graphiques dans l'outil TrAVis à partir des traces enrichies et injectées par le S.B.T. CARTE.

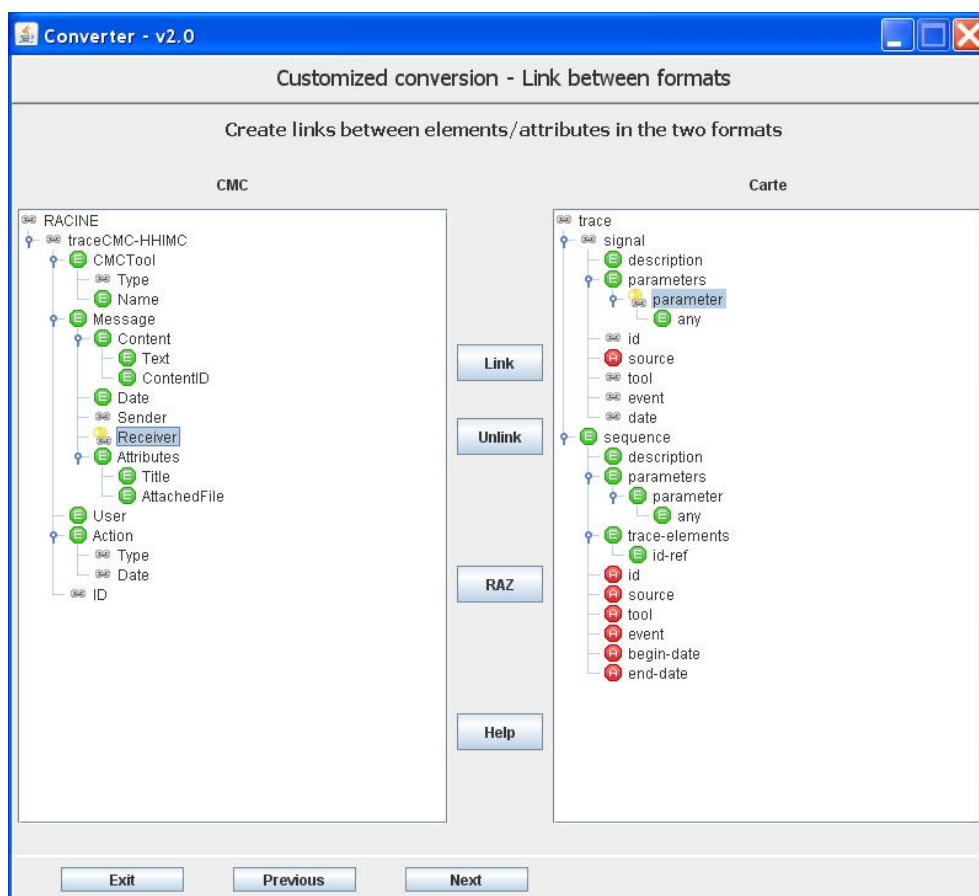


FIGURE 1.3 – conversion entre CARTE et CMC

À partir de ces premiers résultats, nous avons poursuivi nos travaux dans le cadre du projet « Personnalisation des E.I.A.H. »¹² du Cluster de recherche « I.S.L.E. »¹³ financé par la Région Rhône-Alpes. Ce projet traite de l'importance de la personnalisation dans le développement des E.I.A.H., en se basant sur l'utilisation des traces d'interaction pour suivre le déroulement de l'activité d'apprentissage. Notre travail s'est inscrit dans une tâche transversale à quatre autres tâches fortement liées par un objectif commun : « la personnalisation des E.I.A.H. ». Nous avons pour objectif de proposer un modèle générique de partage de corpus de traces d'activité prenant en compte l'hétérogénéité de ces traces due aux différents environnements d'apprentissage et la variété des domaines d'application. Le modèle proposé devait servir de base pour la proposition d'une architecture de plate-forme, prenant en compte les problèmes d'interopérabilité, qui permette le partage entre chercheuses et chercheurs de corpus de traces d'activité d'apprentissage et d'outils d'analyse de ces corpus, ainsi que les analyses réalisées sur ces corpus.

Dans ce projet, nous traitons de la **problématique** du partage de corpus de traces d'activité et d'outils d'analyse entre chercheuses et chercheurs utilisant dans leurs travaux des environnements informatiques pour l'apprentissage humain. Les quatre tâches du projet « Personnalisation des E.I.A.H. » (modélisation des traces, couplage entre les traces et les scénarios pédagogiques, diagnostic des connaissances à partir des traces et utilisation des traces pour améliorer le déroulement des interactions tuteur/apprenant médiatisées par ordinateur) ont engendré l'utilisation d'outils d'analyse spécifiques pour mener des expérimentations. Les traces d'interaction au sein des E.I.A.H., qui sont utilisées pour observer l'activité d'apprentissage, sont généralement collectées de manière ad-hoc pour répondre à des besoins de recherche parfois très liés au domaine d'application. Par ailleurs, les chercheuses et les chercheurs développent généralement des outils d'analyse fortement couplés aux environnements d'apprentissage pour répondre à des besoins d'analyse spécifiques (Martínez-Monés et al. 2005). Le partage, la réutilisation et la comparaison des résultats entre différentes analyses sont donc difficiles.

Si les deux expériences avec les outils d'analyse de traces « TATIANA » et « TrAVis » se sont avérées concluantes, nous faisons le constat des efforts importants engagés dans chacune des actions et de ceux qui seraient nécessaires pour étendre la conversion entre notre format CARTE et la multitude de formats de traces issues des nombreux outils d'analyse existants. Nous avons donc pris le parti de nous abstraire des formats spécifiques des traces dont la diversité continue d'augmenter avec la multiplicité des outils, pour nous placer à un niveau intermédiaire pour interagir avec ces différents formats, ce qui constitue l'originalité de nos travaux. Afin de concevoir un tel système intergiciel (*middleware*) capable de capitaliser, d'analyser et de visualiser des corpus de traces maintenues dans leur format d'origine, nous avons posé trois **questions de recherche** que nous présentons ci-dessous.

Comment modéliser un corpus de traces d'interaction d'apprentissage en vue de son partage et de son analyse, en prenant en considération l'importance de la contextualisation des données partagées ?

Cette première question de recherche adresse la modélisation d'un corpus de traces d'activité et pose naturellement la question de la représentation et de la structure des données à partager. Pour être partageable, nous considérons qu'il est nécessaire d'introduire des métadonnées de description du corpus et de ses composants. Pour que les traces soient interprétables, nous associons au corpus des éléments de contexte, tels que des ressources relatives à une expérimentation d'apprentissage ou à des analyses antérieures.

12. axe « *Benchmarks* et consensus de validation »

13. « Informatique, Signal, Logiciel Embarqué »

Comment garantir une interopérabilité entre les corpus partagés et les outils d'analyse sans imposer une nouvelle représentation des traces d'interaction comme condition à la réalisation du partage ?

Cette deuxième question de recherche traite du problème de l'interopérabilité entre les corpus partagés et entre les outils d'analyse spécifiques développés au sein des laboratoires de recherche. Dans le domaine de la recherche et en particulier en sciences humaines, il est coutume d'associer des données aux publications scientifiques en vue d'offrir la possibilité de vérifier les résultats, de les réutiliser ou de les comparer à l'aide d'autres outils d'analyse. L'absence du contexte de production des traces et la difficulté pour aligner celles-ci avec les formats d'entrée des outils d'analyse, sont un frein à la réutilisation de données d'expérimentation ou d'autres outils d'analyse. Les travaux d'analyse réalisés sur des corpus partagés et l'intérêt de partager les données correspondantes, pour faire des analyses cumulatives ou comparatives et de les intégrer dans de nouveaux corpus, font l'objet de la troisième question de recherche.

Comment représenter et intégrer les travaux d'analyse réalisés sur un ou plusieurs corpus de manière qu'ils soient réutilisables par d'autres chercheuses et chercheurs pour vérification, comparaison, ou analyse complémentaire ?

Les trois questions de recherche précitées sont liées à trois motivations principales justifiant le besoin de partage de corpus de traces d'activité et d'outils d'analyse :

1. l'accès à des corpus de traces d'activité d'apprentissage contextualisées partagés entre équipes de recherche de différentes disciplines ;
2. la réutilisation et la comparaison de résultats avec différents outils d'analyse ;
3. l'accès aux analyses réalisées sur des corpus pour vérification, comparaison et capitalisation.

Pour répondre à la problématique d'interopérabilité entre les corpus de traces et les outils d'analyse, nous avons orienté nos travaux de recherche vers un nouvel objet d'interopérabilité qui préserve les traces dans leur format d'origine et les outils dans leurs fonctions initialement prévues. Nous avons engagé ce travail entre 2009 et 2013 avec la thèse d'Hajer Chebil financée par le projet « Personnalisation des E.I.A.H. », de l'axe « *Benchmarks* et consensus de validation »¹⁴. Dans le cadre de cette thèse, nous avons proposé une nouvelle approche baptisée « approche par proxy » que nous avons appelée PROXYMA (*PROXY for Multiple Analyses*) (Chebil, Courtin, and Girardot 2012a) (Chebil, Courtin, and Girardot 2012b) permettant le partage de corpus de traces d'apprentissage contextualisées et leurs analyses avec des outils partagés. Cette thèse permet, à travers l'approche PROXYMA, la création de trois modèles (modèle de corpus, modèle sémantique, modèle opérationnel) (Chebil, Courtin, and Girardot 2015). Ce travail consiste à proposer un modèle de corpus de traces de référence pour des usages liés à la recherche (validation de modèles, *proof of concept*, expérimentations, etc.). En nous basant sur cette approche, nous avons proposé l'architecture de la plate-forme BEATCORP (*BEenchmarking platform for Analysis of Traces CORPora*) de partage de corpus de traces d'activités d'apprentissage et d'outils d'analyse de ces corpus (Courtin and Tomasena 2016). Les questions d'hétérogénéité des formats de traces produites par les environnements d'apprentissage et attendues par les outils d'analyse sont prises en compte dans la modélisation du système de partage des corpus de traces d'interaction d'apprentissage contextualisées. En d'autres termes, aucune nouvelle représentation des traces n'est imposée dans les corpus de traces issues des E.I.A.H. et pour interagir avec les outils d'analyse. L'approche PROXYMA et l'architecture BEATCORP sont détaillées dans la partie 3 du document.

14. Cluster de recherche « Informatique, Signal, Logiciel Embarqué » (I.S.L.E.), de la Région Rhône-Alpes.

Thématique « Reconnaissance des compétences individuelles et collectives »

La possibilité d'exploiter et de croiser des traces d'interaction générées par des systèmes d'information (e.g. environnements numériques de travail, plates-formes d'apprentissage, réseaux sociaux d'entreprise, etc.) ouvre des perspectives dans divers champs applicatifs liés aux activités. Dans l'axe « Qualité industrielle » du laboratoire « SYstèmes et Matériaux pour la MécatroniquE » que nous avons intégré en 2014, les travaux de recherche adressent, entre autres, des méthodes pour l'amélioration des processus industriels (e.g. LEAN¹⁵) et avec l'introduction de l'industrie du futur, l'étude des organisations industrielles, la gestion des ressources humaines et des informations dans un contexte collaboratif. Faisant le constat que la révolution numérique a un impact sur l'ensemble des dimensions de l'entreprise, les individus doivent s'adapter aux évolutions qu'elle impose, notamment celles des outils et processus de production, de collaboration et de communication. Nous étudions alors en 2016 la notion de compétence exprimée à travers l'activité dans le cadre du projet de recherche CompERe « Dispositif innovant pour l'accroissement des compétences au sein d'entreprises en réseau ». L'enjeu poursuivi à travers ce projet est d'améliorer la performance et la compétitivité des entreprises en donnant les moyens à leurs acteurs d'acquérir les compétences et les connaissances nécessaires à leur adaptation à ces évolutions. Il s'agit pour cela de favoriser d'une part la coopération entre les individus et d'autre part l'exploitation des ressources disponibles. Ce travail a consisté à étudier un dispositif technique intégré qui, à partir des profils de compétences construits automatiquement, fournit des éléments qui pourront être exploités par des systèmes de recommandation de personnes ou de ressources en fonction des besoins exprimés. Ce dispositif contribue à l'amélioration des performances des organisations étendues (i.e. réseaux d'entreprises, mais aussi toute forme de réseaux professionnels) et s'inscrit donc dans le thème « Optimisation des processus » de l'axe « Qualité industrielle » du laboratoire SYMME.

La **problématique** de l'explicitation des compétences individuelles et collectives à partir de l'activité humaine effective et contextualisée est traitée dans cette thématique.

Les travaux sur la modélisation des compétences tracées pendant les activités d'innovation, que nous avons menés dans le cadre de la thèse d'Olivier Pouydebat entre 2018 et 2022, s'inscrivent dans la perspective d'ingénierie d'un système automatisé de reconnaissance des compétences tracées pendant les activités d'innovation d'exploitation et d'exploration. Le lien entre les traces d'activité, les compétences et la performance se situe à l'intersection des domaines constitués par l'analyse informatique des traces modélisées menant aux compétences, et l'analyse en sciences de gestion des spécificités des compétences menant à la performance. Nous avons alors traité la **question de recherche** suivante :

Quelles sont les traces d'interaction numériques qui permettront de formaliser, à l'aide de transformations des traces, des compétences relatives aux activités d'innovation d'exploitation et d'exploration ?

L'enjeu de cette thèse est d'examiner et de comprendre l'imbrication possible des concepts liés aux compétences, à l'innovation et aux techniques d'extraction d'informations à partir des traces informatiques. Ce travail a abouti à la proposition de l'architecture SBS-SEARCH¹⁶ d'un système à base de compétences dans le but d'être implémenté. Dans un souci de simplification pour répondre à la question de recherche associée aux activités d'innovation d'exploitation et d'exploration, le modèle de compétence proposé par Olivier Pouydebat est une représentation structurelle des compétences centrée sur le triptyque savoirs (i.e. connaissances), savoir-faire (i.e. aptitudes) et savoir-être (i.e. attitudes), évaluées et reconnues à partir des tâches et des activités.

15. méthode de gestion « au plus juste » de la production

16. *Skill-Based System - Skills Extract, Analyze, Regroup, Consult and Historize*

Les expériences réalisées avec l'outil de suivi des tâches « KBskill¹⁷ » (i.e. outil KANBOARD¹⁸ avec les compétences intégrées) ont permis de valider l'efficacité d'un système basé sur l'architecture SBS-SEARCH pour tracer des compétences dans divers contextes d'innovation. En revanche, l'identification des compétences collectives n'est pas supportée par ce dispositif qui repère des compétences individuelles et dont la somme ne représente pas la dimension collective, laquelle repose sur les relations interpersonnelles et sur des compétences partagées ou complémentaires. L'architecture SBS-SEARCH est présentée dans la partie 4 du document.

Nous avons conscience du caractère abstrait de l'objet « compétence » et que son identification tient à sa manifestation et à sa reconnaissance à travers l'activité contextualisée. Des chercheurs en sociologie, psychologie et informatique ont tenté de définir et de caractériser la « compétence » pour mieux l'identifier et l'évaluer. La plupart de ces chercheurs s'accordent à dire que les compétences sont difficiles à identifier, mais qu'elles peuvent être déterminées en les déduisant de l'observation et de la traçabilité des actions humaines, plus ou moins représentatives de la réalité. Tout en considérant les recherches actuelles davantage descriptives, nous maintenons une approche analytique pour intégrer dans nos modèles la dynamique de construction des compétences et recueillir les éléments nécessaires à la dimension collective. Cependant, le processus d'identification nécessite en amont une modélisation pertinente des compétences afin de mieux interpréter les traces des activités auxquelles elles sont associées. Dans les travaux suivants, nous tentons donc d'aller au-delà de la structuration des compétences proposée dans les travaux d'Olivier Pouydebat, pour nous orienter vers une vue dynamique de la construction des compétences à travers l'activité et sa dynamique.

Nous poursuivons nos travaux dans le projet « CatCap » (A.N.R. 2019-2023) intitulé « Capture des Traces opérationnelles des acteurs d'entreprise pour construire le Capital humain et définir les Processus gagnants ». Suite à des collaborations entre les laboratoires SYMME (U.S.M.B.) et G-SCOP (Grenoble I.N.P.) au sein du club P.L.M. (*Product Life-cycle Management*), nous avons été sollicités pour participer au montage d'un projet en relation avec nos travaux sur les compétences tracées. Le laboratoire PACTE (U.G.A.) a été associé au projet pour couvrir les aspects de la sociologie du travail et apporter une expertise dans l'observation et l'analyse de trace d'activités, l'identification et la qualification de compétences et l'évaluation des pratiques. L'entreprise C.M.D.L./MANASLU¹⁹ est partenaire du projet pour opérationnaliser les solutions réalisées. La Figure 1.4 illustre la structure du projet.

Après avoir étudié les nombreux modèles de compétence existants, nous avons identifié un certain nombre de verrous scientifiques concernant, entre autres, la mesure de l'implication des individus dans des domaines de compétences à partir de leurs activités, ou l'identification des ressources associées à un niveau de compétence, ou encore la dimension collective des compétences associées à un groupe d'individus. La question des compétences collectives a très peu été traitée dans la littérature scientifique. Il nous est apparu nécessaire de définir le lien entre les compétences individuelles et collectives, ainsi que la dynamique de construction et d'évolution de ces compétences. Pour tenter une représentation conforme aux orientations que nous avons présentées précédemment, nous avons posé la **question de recherche** suivante :

Quelles sont les caractéristiques des compétences individuelles et collectives et comment les représenter pour mettre en exergue les liens qui les unissent ?

17. développé en interne

18. <https://kanboard.org/> est un outil support à la gestion de projet basé sur la méthode Kanban

19. <http://cmdl.fr/> société d'ingénierie et de conseil ayant une expertise technique pointue dans les domaines de l'énergie du bâtiment, et s'appuyant sur une méthodologie originale mise au point par le CEA INES.

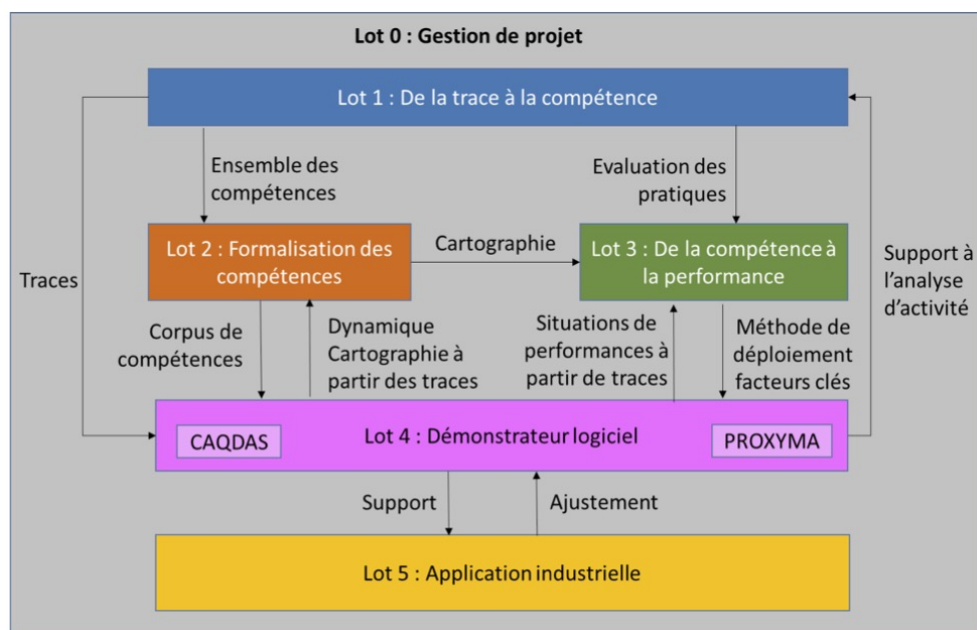


FIGURE 1.4 – structure du projet CaTCaP

Dans les travaux menés dans le cadre de la thèse de Kamel-Eddin Bemmami (2019-2022), la compétence se construit à travers l'activité à l'aide d'un processus de type « schème »²⁰ en mobilisant et en combinant les ressources environnementales et personnelles. Le modèle est basé sur sept concepts : les ressources personnelles, les ressources environnementales, l'activité, le contexte, la performance, l'objectif et le schème. Une fois les compétences formalisées, il s'agit de fournir une représentation de ce capital humain qui soit à la fois partageable par l'ensemble des acteurs (donc communicable, lisible et intuitive), et exploitable (s'appropriant à divers traitements de recherche et de mise à jour). Parmi les formalismes de représentation de connaissances proposés dans la littérature (représentation logique, règles de production, réseaux et graphes sémantiques, ontologies), nous entendons utiliser une représentation graphique pour ses capacités de parcours de graphe, de *clustering*, etc.

Pour résumer, il s'agit de proposer dans ce projet des méthodes permettant :

- d'observer et identifier les activités et situations porteuses d'apprentissage ;
- d'extraire les compétences à partir de cette observation ;
- de formaliser et représenter les compétences ;
- d'identifier et évaluer les pratiques par une validation sociale (e.g. regard des pairs, ...).

Le modèle de compétence de Kamel-Eddin Bemmami est présenté dans la partie 4 du document.

Nous démarrons un projet P.I.A. « Territoires d'innovation pédagogique » / « Campus des métiers et des qualifications », avec une thèse intitulée « Adaptation continue des formations professionnelles à l'évolution des besoins en compétences dans les entreprises », que je co-encadre à partir d'octobre 2022. L'industrie du futur se caractérise par des changements profonds dans le management des entreprises, dans l'informatisation et l'analyse des données, et notamment par une plus grande flexibilité dans les activités professionnelles en vue d'augmenter la capacité d'adaptation des entreprises (Hecklau et al. 2016) (Graczyk-Kucharska et al. 2018). Pour s'adapter à la

20. enchaînement d'actions pour réaliser une activité

demande des marchés, l'anatomie de l'entreprise devra être à géométrie variable. Une telle adaptabilité implique une régulation suivant les lois de la systémique. Pour que cette reconfiguration dynamique soit possible, la formation doit être associée au système de l'entreprise par un couplage entre les activités professionnelles en son sein et les processus de construction des formations en dehors de celle-ci. Si cette vision systémique de l'entreprise couplée avec la formation semble répondre aux besoins actuels d'agilité et de flexibilité, elle doit aussi s'accorder avec la volonté croissante de mettre l'individu au cœur du système, et de tirer le meilleur de ses compétences pour réguler l'activité de l'entreprise. En effet, depuis plusieurs années, les recruteurs se focalisent sur les talents plutôt que sur la volonté d'apparier des candidats et des profils de postes prédéfinis. L'individu peut alors développer son potentiel et s'épanouir dans son activité professionnelle, où il se sent reconnu et confiant.

Les deux visions précitées (vision systémique et vision centrée utilisateur) présentent un paradoxe qui constitue un verrou scientifique :

- adapter la formation professionnelle à l'évolution de l'activité de l'entreprise en vue de s'adapter à la demande des marchés ;
- adapter l'activité de l'entreprise à son potentiel de compétences en vue de les mobiliser avantageusement et d'augmenter la performance.

La vision centrée sur les individus, qui s'oppose à la logique prescriptive classique centrée sur les fiches de postes, favorise l'émergence de talents et minimise le *turn-over* coûteux pour les entreprises.

Les verrous scientifiques sous-jacents se situent à l'intersection des domaines constitués par l'analyse informatique des compétences (Belkadi, Bonjour, and Dulmet 2007) (Bemmami et al. 2021) (Rey 2015) et l'analyse en sciences de gestion des spécificités des compétences menant à la performance (Boucher and Burlat 2003). Les enjeux pour les entreprises et les organisations en charge de la formation sont colossaux en tant qu'ils répondent à des stratégies de développement qui sont primordiales dans un contexte concurrentiel à l'échelle mondiale. L'évolution vers l'industrie 5.0 augure l'émergence de nouvelles compétences qu'il conviendra d'intégrer dans les programmes de formation à une fréquence adaptée à cette évolution. De nouveaux mécanismes devront donc être mis en place pour coupler les formations à l'activité industrielle pour répondre de manière continue aux besoins du marché de l'emploi. La **problématique** de l'articulation entre les processus managériaux pour la gestion des compétences dans l'entreprise et les processus de reconfiguration continue des formations professionnelles sera abordée dans ce projet. Généralement, le processus de reconfiguration de la formation professionnelle est une activité collaborative que nous pouvons analyser avec le modèle classique du trèfle (*clover model*), qui définit trois classes de fonctionnalités intitulées « production », « communication » et « coordination », auquel nous ajoutons la classe « régulation ». Nous traiterons alors la **question de recherche** suivante :

Quels sont les indicateurs caractérisant l'adaptabilité de l'entreprise dans un contexte donné qui permettront de formaliser, à l'aide d'un mécanisme de régulation, les transformations nécessaires dans l'offre de formation ?

Ce travail s'inscrit dans la perspective d'ingénierie d'un système de régulation dynamique du processus de reconfiguration de la formation professionnelle. Ce projet fera le trait d'union entre des travaux des E.I.A.H. et de l'industrie du futur et, notamment, ceux liés aux cartographies de compétences du projet A.N.R. « CaTCaP » en se nourrissant des avancées scientifiques actuelles.

1.1.3 Stratégie de recherche

À ce jour, on constate que l’approche par les traces d’interaction, que nous avons développée depuis plus de dix ans, s’inscrit avantageusement dans d’autres domaines d’application que l’Éducation. Les travaux de recherche sur les traces s’inscrivent dans une sorte de chaîne cinématique représentée Figure 1.5 où s’articulent les modèles de traces²¹, les modèles de corpus de traces²², les modèles de compétences basés sur les traces et les systèmes de recommandations de compétences construits à partir des analyses dans les modèles précédents. Notre contribution initiale dans les systèmes à base de traces (S.B.T.) élaborés conjointement avec nos partenaires lyonnais avant la création des Clusters (2005-2010), a été poursuivie à un niveau d’abstraction supérieur par la création d’un système de *benchmarking* pour les analyses de corpus de traces (BEATCORP). Nos résultats obtenus dans la communauté E.I.A.H. intéressent particulièrement le *Learning Analytics* et le *Data Analytics*. Aujourd’hui, nous menons des recherches sur les systèmes à base de compétences (S.B.C.) issues des traces d’interaction, et notamment en puisant dans les données des réseaux sociaux d’entreprises et du *Big Data*.

Notre stratégie actuelle en termes de recherche consiste à formaliser la démarche globale d’exploitation des compétences basées sur les traces d’activité des utilisateurs (collecte, transformation, analyse, partage, présentation, requête), en vue, entre autres, d’alimenter des systèmes de recommandation. Je dirige actuellement une thèse sur la modélisation d’un système visant à évaluer le capital immatériel des entreprises défini à partir des compétences des professionnels impliqués, et nous co-encadrons une autre thèse sur la caractérisation des compétences individuelles et collectives tracées pendant les activités d’innovation d’exploitation et d’exploration. Ces deux thèses ont comme dénominateur commun l’approche par les traces numériques d’activités effectivement réalisées. Je co-encadre une troisième thèse intitulée « Adaptation continue des formations professionnelles à l’évolution des besoins en compétences dans les entreprises » à partir d’octobre 2022. Ce travail s’inscrit, entre autres, dans la continuité des travaux précités, mais devra également s’appuyer sur les avancées scientifiques actuelles dans les domaines de l’éducation et de l’industrie du futur.

Notre objectif actuel consiste à impulser une dynamique autour de l’équipe de recherche composée de plusieurs doctorants, dont les objectifs scientifiques s’inscrivent en partie dans ceux des projets A.N.R. « Capture des Traces opérationnelles des acteurs d’entreprise pour construire le Capital humain et définir les Processus gagnants » (2019-2023) et P.I.A. « Territoires d’innovation pédagogique » (2021-2024). Nos activités de recherche adressent plusieurs problématiques qui s’articulent avec celles des thèses passées et en cours que nous co-encadrons. Mes connaissances au cœur desdites problématiques m’amènent à co-encadrer généralement les doctorants in situ (au Bourget-du-Lac) avec des quotités majoritaires.

Nos problématiques de recherche ne sont pas cantonnées à un seul domaine de recherche car elles adressent des systèmes complexes, dont la spécification intrinsèquement composite se nourrit des avancées de plusieurs domaines. Nos activités de recherche s’inscrivent ainsi au croisement des domaines de recherche et périmètres d’application suivants : « environnements informatiques pour l’apprentissage humain », « ingénierie des connaissances » dans le cadre des systèmes interactifs, « analyse de données », « interactions humain-machine », etc.

Nos collaborations ont été tissées au sein d’un réseau de recherche structuré initialement en Cluster « Informatique, Signal, Logiciels Embarqués » I.S.L.E. / projet « Personnalisation

21. e.g. C.A.R.T.E. (*Collection, activity Analysis and Regulation based on Traces Enriched*); spécification *eLearning* de format xAPI

22. PROXYMA (*PROXY for Multiple Analyses*); PSLC DataShop (*Pittsburgh Science of Learning Center DataShop*); MULCE (*MUltimodal contextualized Learner Corpus Exchange*)

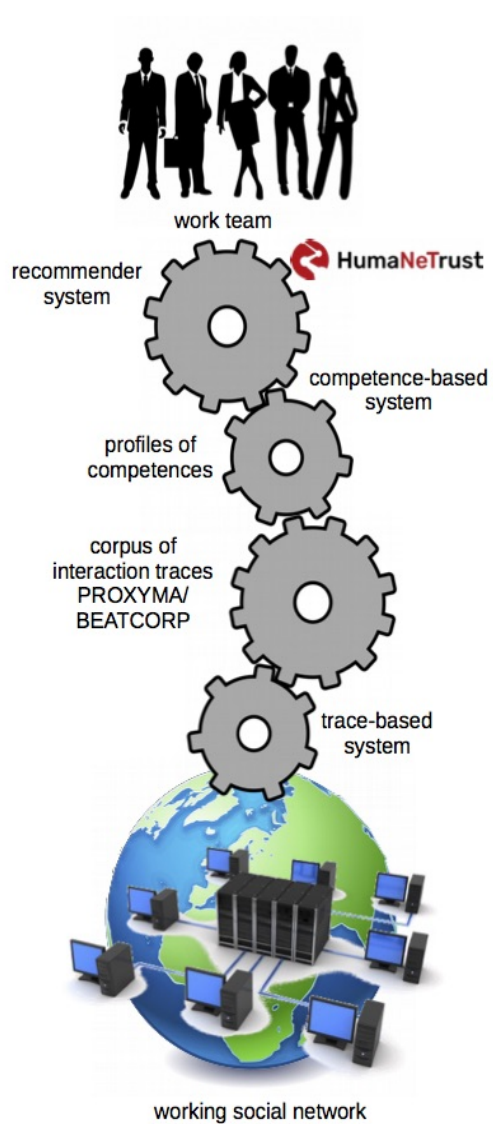


FIGURE 1.5 – *fil conducteur de recherche du bas vers le haut*

des Environnements Informatiques pour l'Apprentissage Humain » financé par la Région Rhône-Alpes depuis 2005. En 2010, le réseau a été restructuré en communautés de recherche académique A.R.C.6 – « T.I.C. et usages informatiques innovants ». À travers ces réseaux de recherche et les projets contractuels (cf. Figure 1.6), nous collaborons avec les chercheuses et chercheurs des laboratoires LISTIC, IREGE, LLSETI (USMB), LIRIS, DISP (Lyon), LIG, ICAR, G-SCOP, PACTE (Grenoble), RIM, Hubert Curien (Saint-Étienne) (liste non exhaustive).

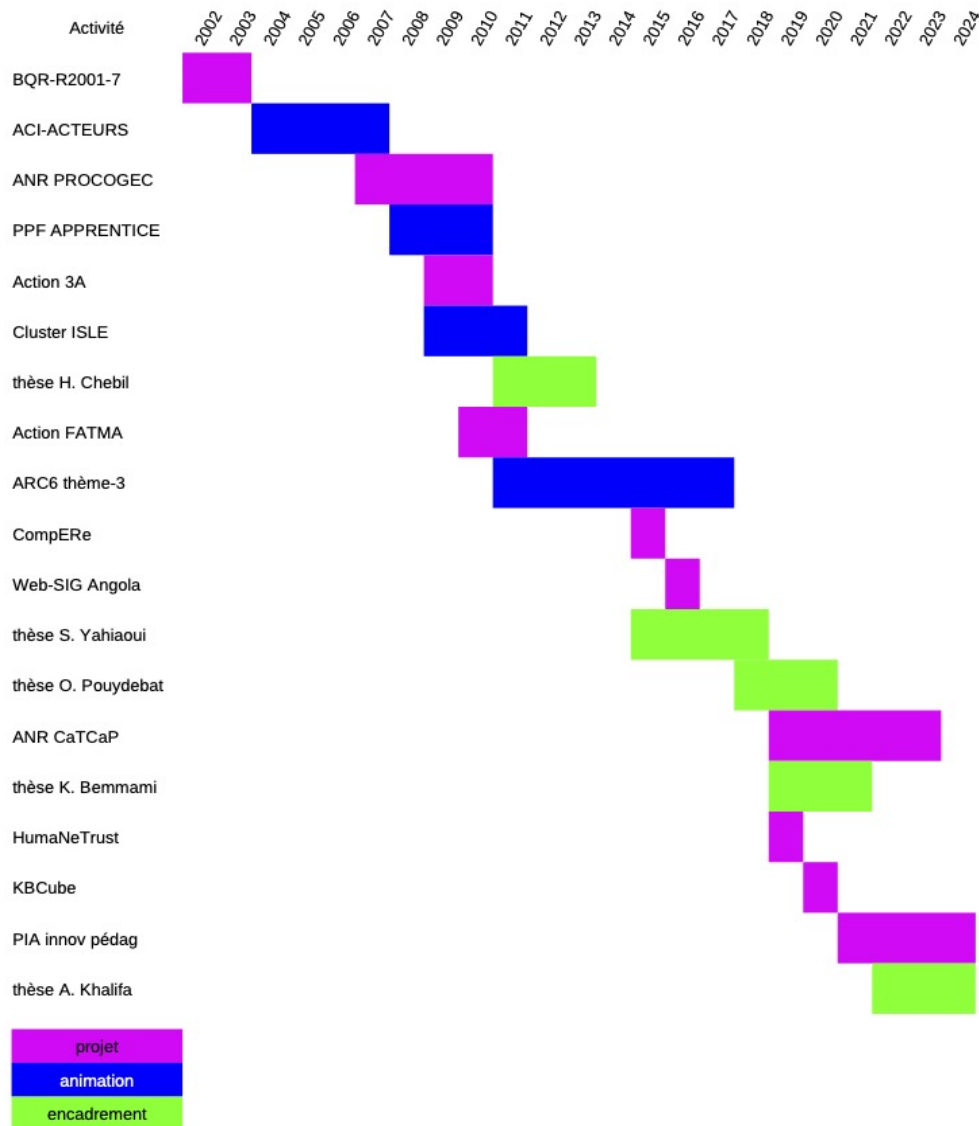


FIGURE 1.6 – frise chronologique des activités de recherche

Nos résultats significatifs ont été publiés dans des conférences, revues et ouvrage.

La construction des corpus de traces d'interaction nécessite en amont l'étude de la **structure intrinsèque des traces** elles-mêmes. Partant du constat de l'hétérogénéité des formats de traces issues d'applications diverses, et de la variété des informations qu'elles contiennent, nous proposons une structuration générique des traces, appelée CARTE (*Collection, activity Analysis and Regulation based on Traces Enriched*), pouvant s'adapter à des contenus variés. Ce travail qui a préfiguré le système à base de traces (S.B.T.), a été sélectionné parmi les publications de la conférence IADIS CELDA (*Cognition and Exploratory Learning in Digital Age*) en 2008, pour être complété et intégré en tant que chapitre de l'ouvrage « Learning and Instruction in the Digital Age » (Springer) en 2010.

Nous partons du postulat que le partage des corpus de traces et des outils d'analyse associés sont les conditions nécessaires à **l'activité d'observation des activités collaboratives d'apprentissage**. Nous spécifions donc l'architecture d'une station d'observation, et développons un *framework* composé d'A.P.I. (*Application Programming Interface*) pour collecter, exploiter, interroger et notifier les traces, liant les analyses d'activités collaboratives aux traces d'activités réalisées au sein d'une plate-forme collaborative. Ce travail a été publié dans la conférence internationale EC-TEL (*European Conference on Technology Enhanced Learning*) en 2009.

Les objectifs de la modélisation PROXYMA (*PROXY for Multiple Analyses*) sont principalement le **partage des corpus de traces hétérogènes et celui des outils externes d'analyse**. Notre approche reconnue comme étant novatrice dans la communauté des E.I.A.H., puisqu'elle ne propose pas un nième format pivot, a été présentée à la conférence internationale IEEE ICALT (*Conference on Advanced Learning Technologies*) à Athènes en 2014.

Le travail précité concernant la modélisation PROXYMA d'un système à base de **corpus de traces d'interaction**, que nous avons finalisé dans le cadre d'un encadrement de thèse, a été présenté dans un article de la revue de rang A « *International Journal of Learning Technology* » en 2015.

L'orientation de nos travaux de recherche vers **l'analyse des données d'interaction en masse** à l'aide de la plate-forme BEATCORP, rapproche nos problématiques de recherche de la communauté des Environnements Informatiques pour l'Apprentissage Humain (E.I.A.H.) de celles du *Big Data* et des communautés *Learning analytics* et *Data Analytics*. Nous avons donc publié un article long à ce sujet dans la conférence internationale IEEE/ACM BDCAT (*Big Data Computing, Applications and Technologies*) à Shanghai en 2016.

L'équipe de recherche formée par les doctorants et les collègues du laboratoire SYMME et des laboratoires partenaires du projet A.N.R. CaTCaP, contribue à la dynamique engagée autour des **problématiques liées aux compétences** basées sur les traces numériques liées aux activités des individus. Les résultats adressant les domaines des E.I.A.H. et de l'industrie du futur, plusieurs articles de revue et conférence ont été publiés en informatique (« *Data-Mining and Knowledge Discovery* » D.M.K.D. / « International Symposium on Electrical, Electronics and Information Engineering » I.S.E.E.I.E.) en février 2021 et en génie industriel (I.F.A.C. revue en ligne en « Sciences de l'ingénieur », « Conférence Internationale de Génie Industriel » C.I.G.I. QUALITA'21 en mai 2021 et I.F.A.C. « Manufacturing Modelling, Management and Control » en juin 2022).

La prochaine étape consiste à spécifier l'architecture d'un système de **recommandations de compétences** à partir des activités réalisées au sein des organisations. Des travaux sont déjà engagés et une publication est en cours de soumission.

1.1.4 Démarche scientifique

Méthode scientifique

Nos travaux de recherche dans les environnements informatiques pour l'apprentissage humain ont été conduits avec une démarche scientifique intrinsèquement imprégnée d'objectivité, c'est-à-dire à partir de données d'interaction avec les outils logiciels. Ces travaux menés dans le réseau de recherche Cluster « Informatique, Signal, Logiciels Embarqués » I.S.L.E.²³ avaient pour principal objectif la personnalisation des E.I.A.H. Les enjeux applicatifs de ce travail consistaient à apporter aux divers utilisateurs des outils utiles, utilisables et efficaces pour le suivi réflexif des activités et la gestion des apprentissages. Les enjeux théoriques de notre approche par les traces étaient la modélisation, la formalisation et le diagnostic, et les enjeux méthodologiques la validation via un *benchmark* (tâche 5 du projet « Personnalisation des E.I.A.H. » répartie sur les laboratoires SysCom²⁴, LIRIS²⁵ et LIG²⁶). Nous qualifions de « **niveau micro** » le niveau de nos analyses de traces d'interaction laissées par des individus dans les systèmes numériques.

Les recherches que nous effectuons actuellement sur les compétences mobilisées, que nous associons aux activités humaines réalisées, accentuent la dimension humaine dans nos travaux, mais nos analyses se différencient de celles de la sociologie en ce qu'elles n'intègrent pas d'éléments subjectifs, auxquels les sociologues peuvent difficilement se soustraire en tant qu'ils sont eux-mêmes des sujets potentiels de l'étude. Notre ambition est d'acquérir une connaissance scientifique à partir d'observations empiriques pour construire des modèles génériques dont l'objectivité garantirait la fiabilité.

Selon le principe d'abduction, que nous appliquons depuis de nombreuses années, nous tentons de répondre à des questions de recherche pour construire notre connaissance scientifique, en nous appuyant sur l'observation empirique et la construction d'hypothèses en vue d'élaborer des lois générales à travers la modélisation. Pour répondre à nos questions de recherche, nous devons donc identifier des sources de données pertinentes et mesurables objectivement. Nous exploitons plusieurs sources, telles que la littérature scientifique présentant les travaux existants, le recueil d'informations auprès des participants au travers d'enquêtes et les études de terrain à l'aide d'une instrumentation numérique.

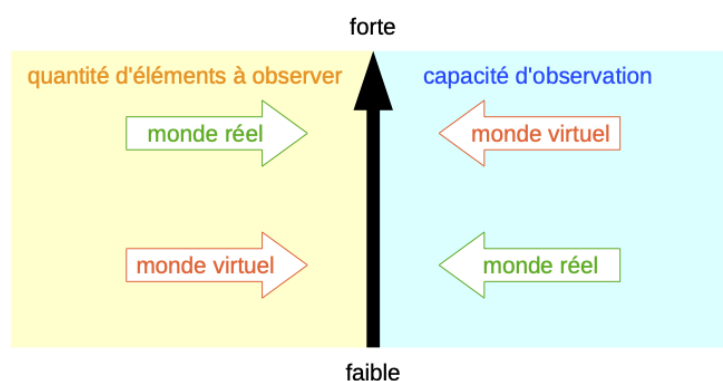
Les observations sur le terrain que nous effectuions initialement pour la personnalisation des E.I.A.H. étaient cantonnées aux traces d'interactions avec des outils d'apprentissage spécifiques, pour répondre à des objectifs principalement d'adaptation des parcours d'apprenants. Nous avons souligné dans une étude sur l'instrumentation d'un environnement d'apprentissage collaboratif (Courtin and Talbot 2005) la difficulté pour observer les activités lorsqu'elles étaient effectuées en partie dans le monde réel, ce qui génère des « zones d'ombre » dans l'observation. Depuis plusieurs années, l'accroissement de la virtualisation des activités de communication, de production et de coordination augmente la capacité d'observation avec un panel de données plus large et plus riche, comme le souligne la Figure 1.7 où la capacité d'observation est proportionnelle à la virtualisation des activités. Pour choisir des sources de données représentatives, il convient de vérifier qu'elles sont utilisées par le panel d'utilisatrices et d'utilisateurs étudié. Par exemple, en vue d'identifier les sources de communication à privilégier, nous avons mené une enquête de terrain auprès d'un panel d'entreprises qui a révélé que 83% d'entre-elles utilisent un réseau social pour communiquer et que LinkedIn est le plus utilisé (Yahiaoui et al. 2017).

23. projet « Personnalisation des Environnements Informatiques pour l'Apprentissage Humain » financé par la Région Rhône-Alpes

24. Systèmes Communicants, université de Savoie

25. <https://liris.cnrs.fr/>

26. <https://www.liglab.fr/>

FIGURE 1.7 – *observation et virtualisation*

Aujourd'hui, nous assistons à une augmentation exponentielle de l'utilisation du numérique (e.g. dans l'industrie du futur, dans l'enseignement) qui conduit au développement d'applications davantage ciblées sur des usages sociaux. Par contre, l'augmentation du volume des données issues des systèmes informatiques, que l'on nomme *big data*, ne suffit pas à enrichir la connaissance sur les situations observées sans une augmentation de la valeur exploitable de ces données. En termes économiques, cette transformation en vue d'une exploitation pour des services prend le nom d'intermédiation. De façon concomitante, nous assistons à un stockage massif de données souvent personnelles du fait de la nature des applications précitées. L'influence inéluctable du numérique sur notre société modifie les structures sociales que nous pouvons observer au « **niveau macro** » ou sociétal, comme étant le résultat d'interactions sociales de « **niveau micro** » ou individuel. Nous identifions ce « **niveau macro** » d'observation comme étant une voie d'investigation pour le développement d'outils de pilotage stratégique, dans le prolongement des travaux de la thèse d'Olivier Pouydebat concernant les activités d'exploitation et d'exploration qu'une entreprise souhaiterait développer dans une perspective d'activité d'innovation au sein d'un marché. Avec l'arrivée du *big data*, les traces brutes, c'est-à-dire non encore traitées, viennent alimenter les systèmes de fouille ou de forage de données de l'intelligence artificielle. Olivier Pouydebat souligne la forte augmentation ces dix dernières années des publications scientifiques alliant l'intelligence artificielle et les traces. Les techniques de l'intelligence artificielle pour extraire de l'information à partir des traces rendent possible la réalisation de systèmes automatisés de décisions basés sur les performances et la productivité des employés (Harrower 2019) (Nicolaescu et al. 2020).

Dans nos travaux actuels sur les compétences, nous nous intéressons à un niveau d'analyse intermédiaire, dit « **niveau méso** », c'est-à-dire relatif aux organisations (e.g. entreprises, équipes, etc.). Cette dimension inclut des données individuelles et organisationnelles, ainsi que des relations inter-personnelles. Les expérimentations que nous avons menées, et qui sont présentées dans ce document, permettent de valider l'exploitation des différents niveaux d'abstraction des traces pour en déduire des compétences. Nous avons souligné dans nos travaux antérieurs les verrous scientifiques associés à la difficulté pour recueillir des informations significatives sur les activités réalisées à travers leurs traces, laquelle est accrue par la multiplicité et l'indépendance des outils utilisés pour communiquer, coordonner ou produire, voire réguler l'activité. Nous avons également présenté les verrous techniques rencontrés pour la mise en œuvre d'une instrumentation, tels que l'impossibilité d'accéder au code ou la configuration de modèles d'utilisation pour expliciter les traces.

L'analyse de l'activité humaine est elle-même une activité humaine généralement construite sur le raisonnement. L'intelligence artificielle en tant qu'elle simule l'intelligence humaine propose de nombreux algorithmes (systèmes experts, systèmes neuronaux, algorithmes stochastiques,

apprentissage profond, ou apprentissage automatique). Contrairement aux algorithmes procéduraux, les algorithmes d'intelligence artificielle sont adaptés à la résolution de problèmes complexes.

Lorsque l'on étudie l'activité humaine, on est tenté par une approche déductive, basée sur le procédé syllogistique utilisé dans les systèmes experts de la programmation logique. La connaissance préalable des vérités associées aux prémisses conduit à la réalisation de systèmes prédictifs, ce qui tend à compromettre l'émergence de nouvelles connaissances. En revanche, l'inférence basée sur des traces témoins du fait observé que les événements afférents ont pu laisser, ouvre le champ des connaissances. L'activité humaine constitue un système complexe composé d'entités reliées entre elles de manière complexe. Étant donnée cette complexité, il n'existe pas à ce jour de modèle général satisfaisant qui permette par raisonnement logique de tirer des conclusions spécifiques à une situation observée sur la base de prémisses. De plus, une méthode déductive conduirait à une théorisation prescriptive de l'activité, ce qui irait à l'encontre de l'approche que nous défendons pour l'observation de l'activité. Nous privilégions donc la méthode inductive, plus adaptée à l'observation de situations spécifiques, pour en dégager des conclusions générales.

D'un point de vue méthodologique, la méthode inductive comprend quatre étapes :

1. observation de situations authentiques ou écologiquement situées (expérimentations);
2. analyse des données observées;
3. dérivation inductive pour construire une généralisation (modèles);
4. vérification sur des situations de même type.

Nos travaux ont donc été majoritairement menés avec une démarche d'acquisition de connaissances au cours d'observations, et ont mené à la construction de modèles selon un principe inductif. De ces nouvelles connaissances, nous tirons par déduction, selon le principe abductif, les prémisses pour arriver à une conclusion qui correspond aux observations. La logique se rapportant aux inférences, c'est-à-dire à l'admission d'une proposition en raison de son lien avec une proposition préalablement tenue pour vraie, nous inférons des vérités sur la base des traces d'activités effectives, dont on suppose qu'elles sont les causes probables du fait observé. De par la quantité importante de traces générées par les systèmes interactifs, le principe abductif que nous suivons contribue à la restriction du nombre d'hypothèses susceptibles d'expliquer les éléments observés. Nous adoptons ce choix de sobriété numérique dans notre approche scientifique, car il contribue au respect des valeurs d'éthique (limitation aux informations utiles et non intrusives, limitation des volumes de stockage des données) tout en préservant des objectifs de performance.

Terminologie et définitions

Les notions abordées dans nos travaux de recherche reposent généralement sur des objets concrets ou abstraits, parfois existants sous une autre forme dans divers contextes. La diversité des définitions associées aux termes représentant ces notions, peut conduire à exprimer des idées différentes avec les mêmes termes intrinsèquement polysémiques. Les chercheuses et les chercheurs ont pour habitude d'arranger leurs définitions de manière à justifier leurs propres solutions, et parfois à supposer en leur faveur ce qui fait foi en matière de recherche. Cette diversité n'est pas tant un mal à déplorer qu'un résultat inévitable et, jusqu'à un certain point, naturel de l'état d'imperfection des sciences²⁷. Considérant qu'une définition ne peut être que provisoire au début d'une recherche, selon le philosophe et logicien britannique John Stuart Mill (Mill 1856), nous nous positionnons explicitement sur la définition des termes dans nos travaux de recherche avec des possibilités d'évolutions relatives à nos découvertes.

27. traduit de « *A System of Logic, Ratiocinative and Inductive* » de John Stuart Mill

Conformément à notre stratégie de recherche présentée dans la partie 1.1.3, nous construisons notre projet de recherche autour d'un sujet central qui est la réalisation d'un système complexe pour observer l'activité humaine médiée au travers des environnements numériques de travail collaboratif (« Travail Collaboratif Assisté par Ordinateur » ou T.C.A.O.). Notre travail de recherche pour traiter ce sujet consiste à définir les nombreux concepts afférents en vue de les intégrer dans nos modèles. Ces modèles théoriques sont déclinés en modèles structurels, sémantiques et opérationnels de type « *Platform Independent Models* » (P.I.M.) pour illustrer le fonctionnement du système complexe. Nous spécifions ensuite les modèles de type « *Platform Specific Models* » (P.S.M.) à l'aide d'architectures logicielles pour implanter techniquement nos solutions en vue de les expérimenter dans les contextes étudiés.

Pour être pragmatique et pour illustrer notre démarche, notre conceptualisation et notre conception peuvent s'exprimer au travers d'exemples concrets : i) des étudiants (les acteurs) communiquent, se coordonnent et écrivent des définitions en anglais de mots de vocabulaire (les actions) au cours d'une séance de travaux dirigés (le contexte) par un enseignant d'anglais (le bénéficiaire) qui analyse leur portfolio, afin d'évaluer leurs compétences tant du point de vue de leurs connaissances que de leur capacité à s'organiser pour réaliser les tâches (le résultat); ii) des individus (les acteurs) réalisent des activités « métiers » (les actions) dans un environnement de travail (le contexte), générant ainsi une cartographie de leurs compétences qu'une chercheuse (la bénéficiaire) exploite à l'aide d'un système de recommandations de compétences (le résultat) pour le montage d'un projet de recherche, afin de constituer une équipe de collaborateurs qui travailleront sur ledit projet.

Dans un cadre de travail médié par ordinateur, notre système complexe doit comporter tous les éléments (concepts, objets, etc.) constitutifs des modèles élaborés de type P.I.M., qui doivent ensuite être représentés dans les modèles de type P.S.M. Nous présentons donc Figure 1.8 les principales notions liées aux éléments précités (acteurs/actions/contextes/bénéficiaires/résultats) qui jalonnent les situations d'expérimentation de nos travaux de recherche.

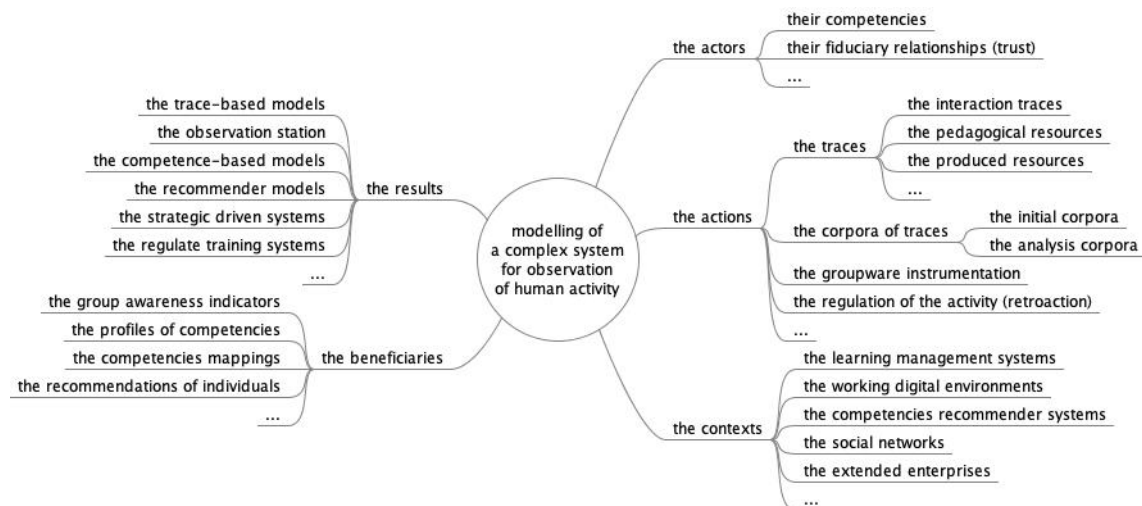


FIGURE 1.8 – notions d'un système complexe d'observation de l'activité humaine

1.2 Curriculum vitæ

Animation scientifique

En 2014, je suis élu au conseil du laboratoire SYMME (pendant 4 ans), et nommé « délégué à la communication » (jusqu'en 2022), où j'ai la charge de glaner et d'organiser l'information scientifique pour permettre sa diffusion au bénéfice de l'activité de recherche interne au laboratoire. Cette fonction consiste à accompagner la direction dans les actions d'animation et de communication scientifiques au sein du laboratoire. La principale tâche est la rédaction d'une Lettre électronique mensuelle (via le progiciel MailchimpTM) rappelant principalement les séminaires internes, les soutenance de thèses et d'H.D.R., les appels à projets, la production scientifique (articles et produits), et tous les autres événements internes du laboratoire.

La fonction est étendue aux trois laboratoires de Polytech Annecy-Chambéry (SYMME, LISTIC et LOCIE) pour la refonte de leur site Web par une entreprise extérieure. En endossant le rôle de chef de projet, je démarcher des agences de développement et intégration Web, je suis le projet avant que les sites soient à nouveau pris en charge au sein de l'université deux ans plus tard, suite à la nouvelle offre de service de la Direction du Numérique.

De 2011 à 2017, en tant que membre du comité scientifique du thème 3 « Mondes numériques pour l'humain et la société : conception, comportements et usages. », du projet « *Academic Research Community* » (A.R.C. 6), financé par la Région Auvergne Rhône-Alpes, je participe aux expertises des dossiers de propositions d'action d'animation scientifique, de manifestations scientifiques, de colloques, de congrès et de séminaires.

J'ai organisé une journée scientifique sur le Technolac (Bourget-du-Lac) le 10 octobre 2014, comprenant la programmation scientifique (cf. Annexe A) et la logistique (sessions de présentations, Pechakucha de posters, convivialités, ...), réunissant une quarantaine de chercheuses et de chercheurs. Le thème de la journée est les « Systèmes de recommandation et réseaux sociaux », avec des interventions sur, entre autres, la densification des réseaux sociaux, les méthodes de hashtag pour la recommandation, les algorithmes de bandits, les smart-cities, les réseaux complexes, etc.

Responsabilités scientifiques

2022-25 - **membre du projet** « Territoires d'innovation pédagogique », Programme d'Investissement d'Avenir (P.I.A.), université Savoie Mont Blanc.

2020 - **porteur** du projet de recherche KBCube « Cube communicant pour interagir avec la méthode Kanban » (A.A.P. U.S.M.B.), université Savoie Mont Blanc.

2019-23 - **co-responsable scientifique** du projet « Capture des Traces opérationnelles des acteurs d'entreprise pour construire le Capital humain et définir les Processus gagnants (CaTCaP) », Agence Nationale de la Recherche (A.N.R.), laboratoire SYMME, Polytech Annecy-Chambéry, université Savoie Mont Blanc.

2019 - **porteur** du projet de recherche HumaNeTrust « un système de recommandations de compétences au sein de réseaux socio-professionnels » (A.A.P. U.S.M.B.), université Savoie Mont Blanc.

Depuis 2018 - **membre du comité** d'éthique pour la recherche, université Savoie Mont Blanc.

2017 - **porteur** du projet de recherche Web-SIG Angola « Développement d'un outil Web-SIG pour estimer la performance des énergies renouvelables : le cas des villes de l'Angola » (A.A.P. « Fédération de REcherche Sur l'efficacité énergétique des BâtimEnts » FRESBE), université Savoie Mont Blanc.

2016 - **porteur** du projet de recherche CompERe « Dispositif innovant pour l'accroissement des compétences au sein d'entreprises en réseau » (A.A.P. U.S.M.B.), université Savoie Mont Blanc.

2015-19 - **membre élu** au conseil de laboratoire, laboratoire SYMME, Polytech Annecy-Chambéry, université Savoie Mont Blanc.

2014-22 - **délégué** à la communication, laboratoire SYMME, **membre nommé** du comité de direction.

2014 - **organisateur** journée scientifique « Mondes numériques pour l'humain et la société : conception, comportements et usages »/ « Systèmes de recommandation et réseaux sociaux », Le Bourget-du-Lac, université Savoie Mont Blanc, action financée par l'« *Academic Research Community* » (A.R.C. 6 – T.I.C. et usages informatiques innovants), Région Auvergne Rhône-Alpes.

2012-13 - **expert scientifique et technique** de projets de société dans le cadre du crédit d'impôt recherche ou du statut de Jeune Entreprise Innovante (Délégation Régionale à la Recherche et à la Technologie).

2011-17 - **co-responsable scientifique** du thème 3 « Mondes numériques pour l'humain et la société : conception, comportements et usages. », projet « *Academic Research Community* » (A.R.C. 6), Région Auvergne-Rhône-Alpes.

2010-14 - **membre élu** du conseil d'administration de l'« Association des Technologies de l'Information pour l'Éducation et la Formation » (A.T.I.E.F.).

2010-11 - **responsable scientifique** de l'action de recherche « *Forum Analysis for Tutoring Mediatized and Assisted* » (F.A.T.M.A.) du projet pluri-formation (P.P.F.) APPRENTICE.

2009-11 - **co-responsable scientifique** du projet « Personnalisation des E.I.A.H. », de l'axe « *Benchmarks* et consensus de validation », Cluster de recherche « Informatique, Signal, Logiciel Embarqué » (I.S.L.E.), Région Rhône-Alpes.

2009-10 - **responsable scientifique** de l'action de recherche « *Automatic Analysis Assistant* » (3A) du projet pluri-formation (P.P.F.) APPRENTICE.

2008-10 - **membre du comité** de pilotage du projet pluri-formation (P.P.F.) APPRENTICE « APPREndre avec les TICE ».

2007-10 - **membre** (quotité 10%) du projet « PROgiciel COLlaboratif de GEstion des Connaissances (PROCOGEC) », Agence Nationale de la Recherche (A.N.R.), laboratoire SysCom, U.F.R. « Sciences Fondamentales Appliquées », université Savoie Mont Blanc.

2004-07 - **membre du comité** de direction du projet de recherche « Action Concertée Incitative » (A.C.I.) « Activités Collectives et Tutorat dans l'Enseignement Universitaire : Réalités, Scénarios » (A.C.T.E.U.R.S.) n° PS-2004-008 - axe « Instrumentation pour l'observation des activités collectives »- subvention A.N.R. « Fonds National de la Science » (F.N.S.).

2002-2003 - **responsable scientifique** du projet « Bonus Qualité Recherche » (B.Q.R.) n°R2001-

Encadrements scientifiques

2022-25 - co-direction (quotité 50%) de la **thèse d'université** « Adaptation continue des formations professionnelles à l'évolution des besoins en compétences dans les entreprises », Amina KHALIFA, université Savoie Mont Blanc, Chambéry.

2019-22 - direction dérogatoire (quotité 60%) de la **thèse d'université** « Modélisation d'un système à base de compétences individuelles et collectives », Kamel Eddin BEMMAMI, université Savoie Mont Blanc, Chambéry.

2018-21 - co-direction (quotité 70%) de la **thèse d'université** « Des compétences individuelles et collectives tracées pendant les activités d'exploitation et d'exploration à la performance », Olivier POUYDEBAT, université Savoie Mont Blanc, Chambéry.

2017-20 - co-direction (quotité 33%) de la **thèse d'université** « Les données d'interaction dans les systèmes collaboratifs d'ingénierie comme base d'évaluation et du dynamisme », Soumaya YAHIAOUI, université Savoie Mont Blanc, Chambéry.

2009-13 - co-direction (quotité 50%) de la **thèse d'université** « Corpus de traces d'activité dans les environnements informatiques pour l'apprentissage humain : modélisation, étude d'une plate-forme de gestion et application à la construction de corpus de référence », Hajer CHEBIL, université Savoie Mont Blanc, Chambéry.

2021 - supervision **stage de Master** « Création de cartographie dynamique des compétences », Estelle PLEINET, laboratoire SYMME, Polytech Annecy-Chambéry, université Savoie Mont Blanc, Le Bourget-du-Lac.

2016 - encadrement **stage de Master** « Création d'un outil Web-SIG pour estimer la performance des énergies renouvelables : le cas des villes en Angola », Fabien TONNERRE, laboratoire SYMME, Polytech Annecy-Chambéry, université Savoie Mont Blanc, Le Bourget-du-Lac.

2011 - encadrement **stage d'ingénieur** INSA de Lyon « Convertisseur de traces informatiques », Adrien GENDARME, laboratoire SysCom, U.F.R. Sciences Fondamentales Appliquées, université Savoie Mont Blanc, Le Bourget-du-Lac.

Responsabilités pédagogiques et administratives

Mon emploi de Maître de Conférences étant en partie associé à l'ouverture du département « Services et Réseaux de Communication » (S.R.C. devenu M.M.I.), **mon implication au sein de l'I.U.T. de Chambéry** a commencé dès mon arrivée par la mise en place de la formation et de l'organisation des études. Par la suite, j'ai eu la charge d'organiser les ressources informatiques pour l'ensemble de l'I.U.T. en créant la Commission Informatique et Réseaux (C.I.R.) faisant le lien entre les instances décisionnaires, les ressources humaines techniques et les usagers. J'ai poursuivi **mon implication dans le fonctionnement de l'université** en participant activement pendant quatre ans au Conseil des Études et de la Vie Universitaire (C.E.V.U.), puis **mon implication au sein du Ministère de l'Enseignement Supérieur et de la Recherche (M.E.S.R.)** en participant pendant dix ans aux activités d'expertise de licences professionnelles pour la Direction Générale de l'Enseignement Supérieur et de l'Insertion Professionnelle (D.G.E.S.I.P.).

Ci-dessous les responsabilités occupées pour la pédagogie et l'administration :

Depuis 2021 - **chargé de mission** des Systèmes d'Information (S.I.) de l'I.U.T. de Chambéry à l'université Savoie Mont Blanc.

Depuis 2020 - **membre du réseau national** Approche Par Compétences (A.P.C.) pour la mise en place du Bachelor Universitaire de Technologie (B.U.T.) / Assemblée des Directeurs d'I.U.T. (A.D.I.U.T.)

2013-19 - **directeur des études** D.U.T.2 du département « Métiers du Multimédia et de l'Internet » (M.M.I.) de l'I.U.T. de Chambéry.

2012-13 - **correspondant C2i** niveau 1 de l'I.U.T. de Chambéry à l'université de Savoie.

2010-14 - **membre nommé** de la Commission Nationale d'Expertise des Licences Professionnelles (C.N.E.-L.P.) / Direction Générale de l'Enseignement Supérieur et de l'Insertion Professionnelle (D.G.E.S.I.P.) / Ministère de l'Enseignement Supérieur et de la Recherche (M.E.S.R.).

2009-10 - **responsable** de l'insertion professionnelle et des poursuites d'études du département « Services et réseaux de communication » de l'I.U.T. de Chambéry.

2008-12 : **membre élu** du Conseil des Etudes et de la Vie Universitaire (C.E.V.U.) de l'université de Savoie.

2007-08 - **responsable de l'alternance** du département « Services et réseaux de communication » de l'I.U.T. de Chambéry.

2006-13 - **responsable de la Commission** Informatique et Réseaux (C.I.R.) de l'I.U.T. de Chambéry

2005-07 - **président de la commission** de la V.A.E. pour la L.P. Activités et Techniques de Communication (A.T.C.) de l'I.U.T. de Chambéry.

2004-10 - **membre associé** de la Commission Nationale d'Expertise des Licences Professionnelles (C.N.E.-L.P.) / Direction Générale de l'Enseignement Supérieur et de l'Insertion Professionnelle (D.G.E.S.I.P.) / Ministère de l'Enseignement Supérieur et de la Recherche (M.E.S.R.).

2002-05 - **vice-président** MCF de la commission de spécialistes C.N.U. 27 de l'université de Savoie.

2001-07 - **chef du département** « Services et réseaux de communication » de l'I.U.T. de Chambéry

2000-01 - **directeur des études** à la création du département « Services et réseaux de communication » de l'I.U.T. de Chambéry.

Activités pédagogiques

Depuis 2000, je fais mon service d'enseignement dans le département « Services et Réseaux de Communication » (S.R.C.), devenu « Métiers du Multimédia et de l'Internet » (M.M.I.). Étant arrivé à la création du département S.R.C. à l'**I.U.T. de Chambéry**, il s'agissait de créer ex nihilo en 6 mois tous les enseignements qui m'incombaient pour la première année de D.U.T. et pour le D.E.S.S. d'Informatique géré par les membres du laboratoire SysCom. L'équipe initiale du D.U.T. étant composée de 3 MCF, 1 PRAG et 1 PRCE, dont 4 collègues étaient nouvellement recrutés, j'ai eu la charge de la direction des études dès mon arrivée. Cette fonction recouvrait la création des emplois du temps, accompagnée du démarchage des vacataires pour compléter les enseignements, le suivi des absences des étudiants, la construction d'outils pour le recrutement des nouvelles promotions (encore utilisés aujourd'hui). Le départ de la cheffe de département m'a conduit à ladite fonction dès la deuxième année (2001-2007), et donc à la direction de l'équipe pédagogique et au suivi financier. En 2004, j'ai participé à la création de la licence professionnelle « Activités et Techniques de Communication » du département, laquelle a été réhabilitée jusqu'à l'arrivée des Bachelor Universitaire de Technologie (B.U.T.).

Les principales matières enseignées furent initialement « Algorithmique et programmation » et « Réseaux » en D.U.T., et « Interface Humain-machine » en D.E.S.S. « informatique » (modélisation du processeur humain vs. modélisation des interfaces), et enfin « Réseau » en D.E.S.S. « Télécom réseau » (modélisation de réseau avec le logiciel OPNET IT Guru™). D'autres enseignements sont venus compléter cette palette à l'I.U.T., comme « création de contenus en relation avec les nouveaux usages » (avec les problématiques du W3C et du WAI), « IHM et Web » (modèle MVC en Java et Python), et « Systèmes interactifs et communication » (avec l'Internet des objets). L'enseignement « I.H.M. » en MASTER 2 de l'**U.F.R. Sciences et Montagne** est désormais intégré à un module intitulé « Intelligence ambiante », qui traite entre autres de l'Internet des objets. J'ai également dispensé ce cours en anglais à l'**Universitatea De Vest Din Timisoara**, en Roumanie en 2018. J'enseigne enfin la conception orientée objet (C.O.O.) et la programmation orientée objet (P.O.O.) en MASTER 1 pour des étudiants suivant le parcours « Télécom Réseau ». Enfin, j'interviens sur des travaux pratiques dans le module « Bases de données » (INFO502) dans les filières EIT et BATiment FISE 3 de **Polytech Annecy-Chambéry** depuis 2021.

Pour consacrer davantage de temps à mon activité de recherche et notamment les encadrements de thèses, j'ai interrompu mes fonctions administratives pour la formation M.M.I. à partir de la rentrée 2019. Par contre, j'ai intégré le réseau national « Approche Par Compétences » pour la mise en place du « Bachelor Universitaire de Technologie » (B.U.T.) en tant qu'il allie mes activités de recherche et d'enseignement pour la mise en pratique du suivi des compétences. Pour optimiser mon temps de travail tout en conservant les bénéfices liés à la mise en place de la classe inversée dans mes enseignements, je m'appuie fortement sur les fonctionnalités de la plate-forme d'apprentissage Moodle™ (ateliers, devoirs, forums, tests et leçons).

Le directeur de l'I.U.T. a également souhaité me confier, à partir de novembre 2021, une mission sur les systèmes numériques afin d'opérationnaliser mon implication depuis plus de dix ans dans la composante sur ce domaine. Enfin, mon bureau dans le laboratoire SYMME du bâtiment 2 (Polytech Annecy-Chambéry), qui est mitoyen à la salle des doctorants, facilite le suivi de leur travail de recherche depuis juillet 2018.

Emplois

Depuis 2000 - **Maître de Conférences** en Informatique (HC depuis 2017 sur contingent national), université Savoie Mont Blanc, Le Bourget-du-Lac.

1998-2000 - **Attaché Temporaire d'Enseignement et de Recherche** (A.T.E.R.) - École Nationale Supérieure des Sciences Appliquées et de Technologie (E.N.S.S.A.T.), université de Rennes, Lannion.

1995-98 - **moniteur** en informatique, université de Rennes, Rennes.

Formation

1998 - **Thèse d'université** « Interfaces coopératives pour des dossiers médicaux multimédia », université de Rennes 1, Rennes.

1994 - **Diplôme d'Études Approfondies (D.E.A.)** « Informatique médicale », université René Descartes, Paris.

1993 - **Master of Science** « *Software System Technology* », université de Sheffield, Sheffield, Angleterre.

1992 - **Maîtrise de Sciences et Technologies** « Micro-Informatique, Microprocesseurs Électroniques (M.I.M.E.) », université du Maine, Le Mans.

2

Traces numériques

Dans Ichnos-Anthropos ou le carnet de l'Homme-trace qui fédère les recherches sur la trace produites dans les différentes disciplines, le mot « trace » est utilisé comme si son sens allait de soi. Pourtant, il paraît difficile de convenir d'une définition partagée et satisfaisante pour un mot utilisé par des géographes, des sociologues, des sémiologues, des psychologues, des linguistes, des anthropologues, des informaticiens et des chercheurs en sciences de l'information et de la communication. Les sachants entreprennent alors de produire des textes qui travaillent le sens de la notion de trace et qui explicitent sa déclinaison dans leurs recherches (Galinon-Mélénec 2011).

Sommaire

2.1	Introduction	34
2.2	Traces d'interaction numériques	36
2.2.1	Système à base de traces	36
2.3	Analyse des traces d'interaction	41
2.3.1	Formalisme de construction des séquences	42
2.3.2	Règles de recherche des séquences basées sur les gabarits	42
2.4	Activité d'observation	46
2.4.1	Station d'observation	47
2.4.2	Instrumentation	47
2.4.3	Rétroaction	49
2.5	Expérimentation d'une station d'observation	51
2.5.1	Les objectifs de l'expérimentation	51
2.5.2	Le contexte de l'expérimentation	52
2.5.3	Les outils logiciels de l'expérimentation	53
2.5.4	Le déroulement des opérations	60
2.5.5	La collecte des traces	61
2.5.6	L'analyse des traces	62
2.5.7	Les résultats de l'expérimentation	63
2.6	Conclusion	66

2.1 Introduction

L'introduction de la technologie dans les activités d'apprentissage collaboratif a suscité l'intérêt des chercheurs en sciences humaines et sociales, qui étudient les différentes formes d'activités cognitives et sociales concernées (Corbel, Girardot, and Lund 2006). Les chercheurs en sciences humaines et sociales sont confrontés à une complexité de l'analyse de l'interaction humaine telle qu'il est nécessaire de les assister avec la technologie, notamment pour les tâches impliquant de très nombreux paramètres ou pour les tâches chronophages. En fait, l'analyse de l'interaction humaine avec un système informatique reste une activité humaine avec un haut niveau d'abstraction. Parallèlement à ces travaux de recherche en sciences humaines et sociales, il existe d'autres travaux en science informatique relatifs à la conception de systèmes d'apprentissage collaboratif assisté par ordinateur, et à l'observation de l'activité humaine basée sur les traces d'interaction.

Nos travaux adressent donc deux champs d'activité :

1. les activités de recherche en sciences humaines et sociales et en science informatique ;
2. les activités d'apprentissage.

Par extrapolation, les outils que nous présentons dans la partie 2.5 page 51 dans le domaine de l'apprentissage peuvent être configurés pour répondre à des besoins du premier champ d'activité. En d'autres termes, les résultats de l'étude dans le domaine de l'observation des activités d'apprentissage collaboratif, que nous présentons dans ce chapitre, exposent la capacité de la technologie pour l'assistance automatique des activités des chercheurs en sciences humaines et sociales, et notamment pour la compréhension des processus d'analyse des interactions socio-cognitives.

Selon les concepts et les principes de la théorie de l'activité (Kaptelinin and Nardi 2018), une activité est structurée sur trois niveaux (cf. Figure 2.1) : les activités sont composées d'actions, lesquelles sont à leur tour composées d'opérations. Ces trois niveaux exposent les liens chez les humains entre les motivations (i.e. qu'est-ce qui motive l'activité ?) et les objectifs de l'activité (i.e. qu'est-ce qui dirige l'activité ?), ainsi que les opérations qui répondent aux conditions de réalisation.

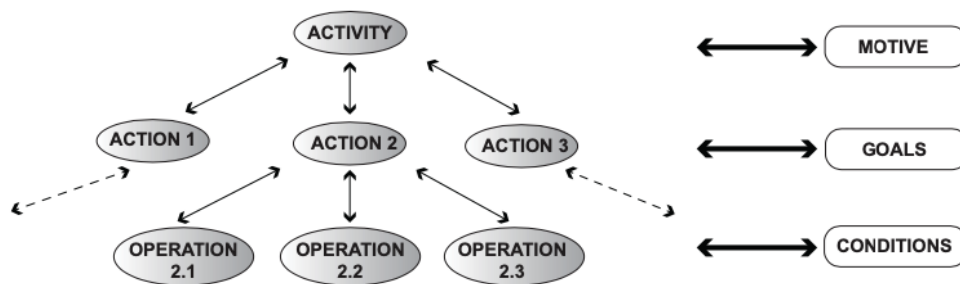


FIGURE 2.1 – Structure hiérarchique de l'activité

Les activités contiennent des actions qui sont la manifestation concrète des tâches fonctionnelles et des tâches « métier » qui s'exécutent dans des systèmes numériques (E.N.T.¹, E.I.A.H.², etc.) (voir Figure 2.2). L'indexation des tâches dans un référentiel de tâches permet de caractériser les traces d'interaction associées aux actions et de construire des motifs (*patterns*) de séquences pour les reconnaître à partir des signaux et des séquences antérieurs (voir modèle de trace).

1. Environnements Numériques de Travail ou *Digital Working Environments* (D.W.E.)
 2. Environnements Informatiques pour l'Apprentissage Humain ou *Technology Enhanced Learning Environment* (T.E.L.E.) ou *Learning Management Systems* (L.M.S.)

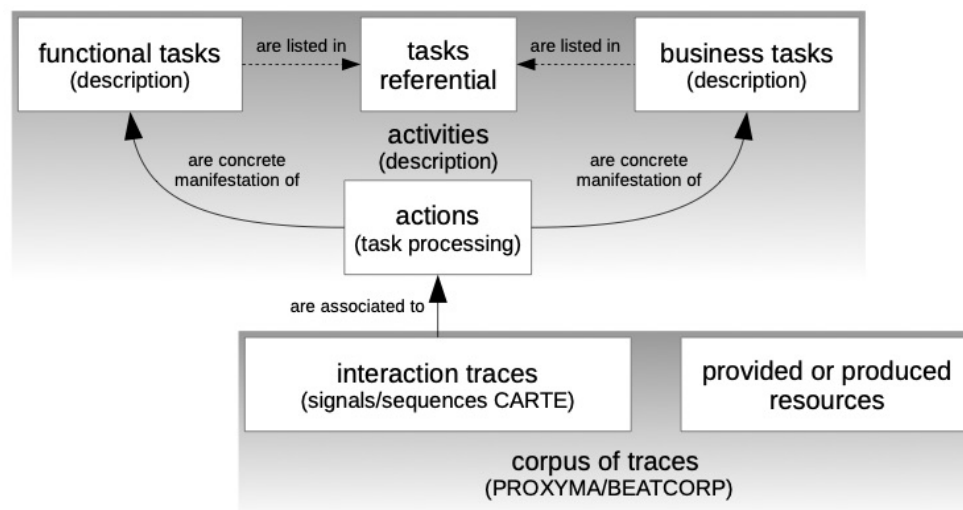


FIGURE 2.2 – Les tâches au sein de l'activité

L'étude socio-cognitive des interactions humaines assistées par ordinateur se base sur les traces logicielles, qui constituent notre objet d'étude (e.g. les données d'un fichier de *logs*, les données d'une instrumentation des outils logiciels) et sur les traces numérisées ou transcrites (e.g. la transcription d'une vidéo). Les pratiques de recherche en sciences humaines et sociales impliquent, entre autres, la création et le partage par les communautés de recherche de corpus de traces utiles à la compréhension des activités réalisées. Celles de l'informatique sont la conceptualisation des traces, leur collecte à partir de diverses sources (e.g. fichier de *logs*, enregistrements vidéo), la représentation à différents niveaux d'abstraction et l'aide à l'analyse des traces d'interaction multimodales et multi-utilisateurs.

Dans le cadre de l'étude des interactions humaines assistées par ordinateur, les chercheurs utilisent généralement trois types de traces (Courtin and Talbot 2009) :

1. les traces d'interaction provenant des outils logiciels utilisés pour les activités de collaboration :
2. les traces calculées (ou créées) qui correspondent à une traduction informatique de traces provenant de sources non informatiques (e.g. des transcriptions d'enquêtes) ;
3. les traces enrichies, c'est-à-dire issues de transformations de traces d'interaction ou calculées ou déjà enrichies au moyen de modèles d'utilisation des outils logiciels dont elles sont issues.

Dans ce chapitre, nous étudions les deux types de traces d'interaction et enrichies, et nous élargissons ce champ d'investigation dans la partie 3 page 67, en y intégrant d'autres types de ressources pour former des corpus de données d'interaction contextualisées.

2.2 Traces d'interaction numériques

L'intérêt que nous portons aux traces d'interaction numériques tient au fait qu'elles sont la mémoire de l'activité humaine menée à travers les systèmes numériques. Cette activité humaine peut être individuelle ou collective, c'est-à-dire intégrant des relations interpersonnelles directes ou médiées par un système numérique. En réalité, la trace numérique est contruite à partir d'une succession d'événements déclenchés pendant l'exécution d'un programme informatique, soit par les actions des utilisateurs, soit automatiquement par le système lui-même à des intervalles de temps prédéfinis. Pour que les traces puissent être utilisées pour représenter l'activité humaine, il convient d'associer ces événements à des modèles d'utilisation des outils logiciels, auxquels s'ajoutent des éléments de contexte et des ressources fournies ou produites.

Selon une vision systémique et donc holistique, le tout est plus que la somme des parties. Une trace composite ne se réduit donc pas à la composition des traces atomiques afférentes et à la somme de leurs propriétés. Si ce constat se traduit généralement pour le travail collaboratif par la somme $1+1=3$, peu conforme aux mathématiques, illustrant ainsi la valeur ajoutée de la collaboration, le tout peut être inférieur à la somme pour les traces d'interaction. Quant aux propriétés d'une trace composite, elles sont propres à celle-ci et dans une moindre mesure aux traces atomiques qui la composent.

Au regard de leur complexité, des systèmes organisés émergent des caractéristiques qui n'existaient pas dans les éléments qui les constituent ([Abdelmalek 2004](#)). Ces caractéristiques issues du principe d'ordre d'une organisation portant le nom d'émergence organisationnelle, nous conduit alors à ne pas limiter la recherche de traces d'activité dans ces systèmes à la seule agrégation de traces d'actions atomiques réalisées en leur sein. La complexité implique la compréhension des relations entre le tout et les parties, dont la connaissance ne suffit pas à la connaissance du tout. En d'autres termes, le paradigme de disjonction dans une logique réductionniste incarnée par la décomposition des activités pour le traçage d'actions atomiques (les parties), doit être complété par le paradigme de conjonction dans une logique holistique, mettant en exergue les relations entre ces actions et les caractéristiques propres au système global (le tout).

2.2.1 Système à base de traces

Ayant placé les traces au cœur de l'observation, nous présentons sur la Figure [2.3](#) le modèle d'un système à base de traces, auquel nous avons contribué à la définition avec d'autres acteurs du domaine des E.I.A.H. ([Settouti et al. 2009b](#)) et notamment par l'ajout d'un modèle de rétroaction que nous décrivons ci-après.

Selon notre approche, nous effectuons des observations à partir d'éléments temporellement situés, appelés « obsel » (*observed elements*), associés à un modèle d'utilisation de l'application ([Settouti et al. 2009b](#)). Ce modèle d'utilisation fournit le registre sémantique nécessaire à l'interprétation des éléments de trace, lesquels deviennent des traces modélisées ou MTrace.

Modèle de collecte

Dans le système à base de traces, le modèle de collecte (*collect model*) de la Figure [2.3](#) décrit la technique de moissonnage de toutes les traces issues des diverses sources disponibles, comme les *logs* ou les outils logiciels des E.I.A.H. Les MTraces sont stockées en général dans une base de données relationnelle, dans un format adapté pour ne pas perdre d'information. À défaut d'existence d'un format standard pour structurer les traces, nous avons défini un modèle simple et générique que nous décrivons ci-dessous. Ce format générique de trace, présenté sur la Figure [2.4](#), prend en compte les sources de trace propres à chaque application. Les éléments communs aux diverses sources (e.g. description) définissent les méta-données, et les éléments spécifiques (e.g. une « page » dans un éditeur de texte, une « table » dans un *chat room*) sont les paramètres optionnels.

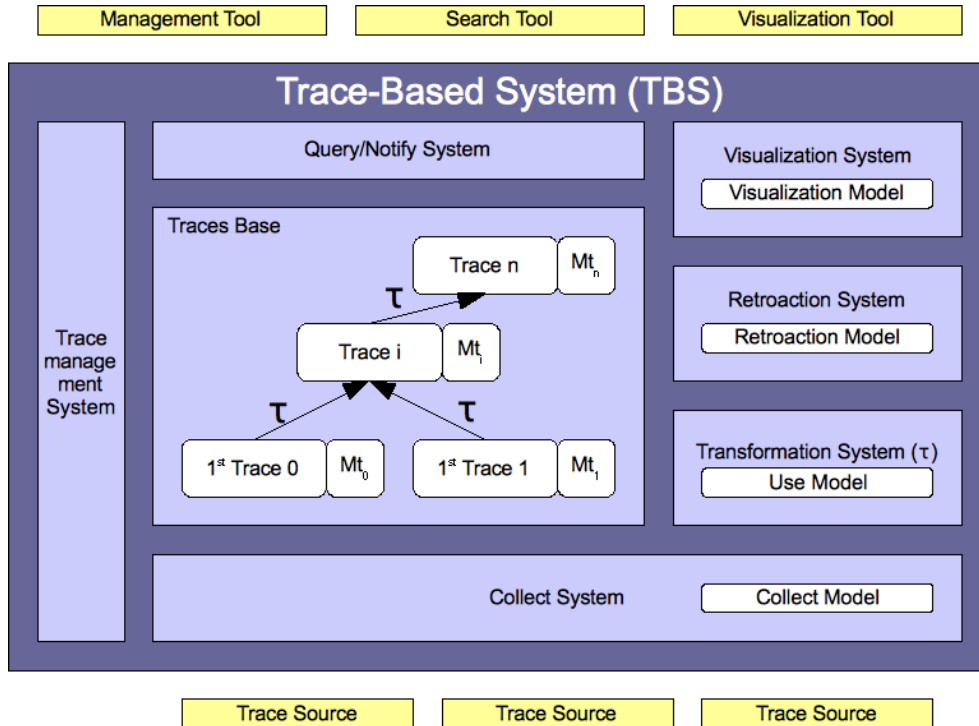


FIGURE 2.3 – Système à base de traces

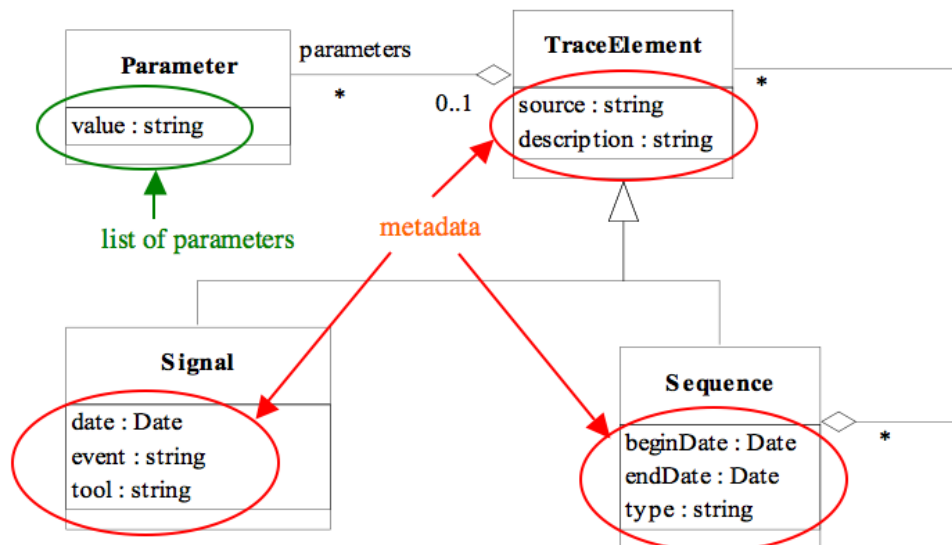


FIGURE 2.4 – Format générique de traces

Nous introduisons deux types de gabarits (*templates*) : le **signal** et la **séquence** (cf. Figure 2.4). Un signal représente généralement des informations basiques correspondant à des actions élémentaires ou atomiques effectuées dans une application (e.g. enregistrer une page). La spécificité des différentes traces recueillies est prise en compte dans une liste variable de paramètres. Nous décrivons ci-dessous la structure des signaux :

- une source, c'est-à-dire l'E.I.A.H. ou la station d'observation à l'origine du signal ;
- une description textuelle du signal (e.g. « sauvegarde de la page ») ;
- une date, celle de l'évènement à l'origine du signal ;
- un évènement logiciel généré par l'outil ;
- un outil, celui où a eu lieu l'action ;
- une liste de paramètres qui constitue la partie variable du signal, pour prendre en compte les concepts spécifiques de l'outil (e.g. « envoyeur » ou « message » pour un chat, « rédacteur » ou « page » pour un éditeur de texte).

Une séquence représente une information enrichie en tant qu'elle est le résultat d'une transformation, mais peut parfois être obtenue directement à un niveau d'abstraction élevé, c'est-à-dire qui fait sens pour une observatrice (e.g. dans un logiciel de gestion de tâches). S'il y a eu transformation, une séquence est une trace composée des signaux et/ou de séquences de niveaux d'abstraction plus faibles (e.g. `edit_page` + `save_page` = `edition`) (cf. Figure 2.5). Le niveau d'abstraction d'une trace d'interaction correspond à son niveau d'interprétation, l'objectif étant d'atteindre celui de l'observateur (e.g. le niveau pédagogique pour l'enseignante, le niveau communicationnel pour le chercheur en sciences de l'information et de la communication, etc.). Il est donc nécessaire de connaître les objectifs de l'observation pour définir les gabarits du modèle d'utilisation qui permettent de tracer les « obsels³ » (Courtin and Talbot 2006). Nous privilégions cette méthode d'analyse des traces de type *top-down* pour des questions à la fois d'atteinte des objectifs d'observation et de performances techniques liées à l'analyse de quantités importantes de données générées par les systèmes informatiques avec la méthode *bottom-up*, comme le soulignent (Smith and Korel 2001).

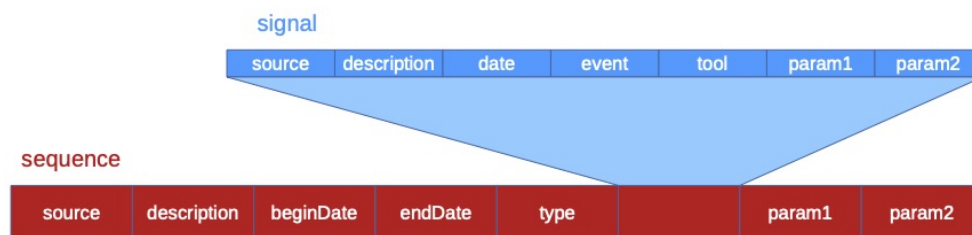


FIGURE 2.5 – Encapsulation d'un signal dans une séquence

Modèle d'utilisation

Les techniques d'*UX Design*⁴ tendent à intégrer les utilisateurs finaux dans la phase de conception des outils logiciels afin de garantir, entre autres, leur utilité et leur utilisabilité. En théorie, nous pourrions en déduire des modèles d'utilisation et ainsi reconnaître les actions effectuées à travers ces outils. Pour autant, les outils n'imposent généralement pas de procédure d'utilisation qui pourrait être associée à un scénario ou à un *workflow*⁵, mais autorisent des usages personnalisés, voire

3. *observed elements*

4. stratégie pour la conception de services logiciels basée sur l'expérience des utilisateurs

5. suite de tâches effectuées par un individu ou un groupe de personnes, accompagnée du passage des informations afférentes d'une étape à l'autre

détournés. Dans notre approche scientifique, nous impliquons l'observatrice dans la définition des modèles d'utilisation, afin qu'elle puisse les configurer conformément à l'utilisation effective des outils, aux objectifs de l'observation et à son niveau d'interprétation des activités réalisées.

La configuration du modèle d'utilisation (*use model*) de la Figure 2.3 consiste à spécifier les gabarits des activités (*activities patterns*) réalisables avec l'application et à les exploiter avec un langage basé sur le principe des expressions régulières. La construction des gabarits de niveau n est réalisée par des opérations de transformations τ d'« obsels » atomiques (des signaux) et/ou d'« obsels » composites (des séquences) de niveau d'abstraction inférieur à n . L'objectif de cette construction est d'atteindre des niveaux d'abstraction interprétables par les observatrices humaines selon leur domaine d'expertise (e.g. sociologie, pédagogie, ergonomie, etc.). À noter qu'avec les mêmes « obsels », il est possible de définir plusieurs jeux de gabarits et donc d'envisager plusieurs types d'analyse, de même qu'il est possible de substituer des outils logiciels de même catégorie (e.g. outil de communication) dans la composition du gabarit sans modifier sa description.

L'extraction des traces d'utilisation des outils logiciels peut être réalisée de diverses manières, selon la nature des applications et leur capacité d'importation et d'exportation des informations (traces d'interaction et données). Par exemple, on peut extraire des séquences relatives à des tâches « métier » (*business task*) dans des logiciels de gestion de tâches (cf. Figure 2.6), c'est-à-dire des séquences explicites en termes de tâches réalisées (e.g. envoi d'un courriel, rédaction d'un rapport). L'extraction de ces traces peut être réalisée soit par une éventuelle option de l'application, soit par l'utilisation d'A.P.I.⁶ parfois fournie avec le logiciel. Les traces ainsi collectées peuvent être récupérées dans des fichiers C.S.V.⁷, J.S.O.N.⁸, ... ou dans des bases de données. À défaut de pouvoir récupérer des traces explicites, les traces brutes (*raw traces*) générées lors de l'utilisation des logiciels (ex. fichier de *logs*) correspondent à des signaux qui peuvent être combinés à l'aide de transformations τ du T.B.S. pour retrouver des séquences décrites dans les gabarits.

Modèle de rétroaction

Nous avons introduit le concept de « régulation » dans (Martel et al. 2004) pour représenter tous les mécanismes relatifs au contrôle social (i.e. le respect des règles sociales, telles que l'interdiction de communiquer pendant une activité) et à l'adaptation des activités aux objectifs pédagogiques en fonction d'un contexte spécifique (e.g. suggérer des nouveaux exercices à des étudiants en difficulté). Dans le modèle classique du travail collaboratif, nommé « *clover model* », divisé en trois espaces (communication, coordination et production) (Ellis and Wainer 1994), la régulation représente un quatrième espace orthogonal aux trois autres pour prendre en compte la dimension sociale des activités (cf. Figure 2.7) (Ferraris and Martel 2000). Nous considérons que tous les mécanismes de configuration manuelle ou automatisée de l'activité dans les environnements informatiques permettant d'assurer la cohérence entre le travail en cours et les objectifs de l'activité, sont inclus dans la régulation (Chabert et al. 2006).

Le mécanisme de rétroaction au sein d'une station d'observation s'appuie sur le concept de régulation de l'activité d'apprentissage, qui sous-tend le modèle de rétroaction (*retroaction model*) de la Figure 2.3.

Nous présentons (sur la figure 2.12 de la partie 2.5) une interface graphique de coordination de l'activité collaborative qui implémente le modèle de rétroaction dans une station d'observation. Nous présentons également (sur la figure 2.13 de la partie 2.5) l'intégration de nouvelles fonctionnalités sous la forme d'*add-on* au sein d'un outil de communication existant, qui est utilisé au cours d'une expérimentation.

6. *Application Programming Interface* qui permet à des applications de communiquer entre elles et de s'échanger mutuellement des services ou des données.

7. *Comma Separated Values*

8. *JavaScript Object Notation*

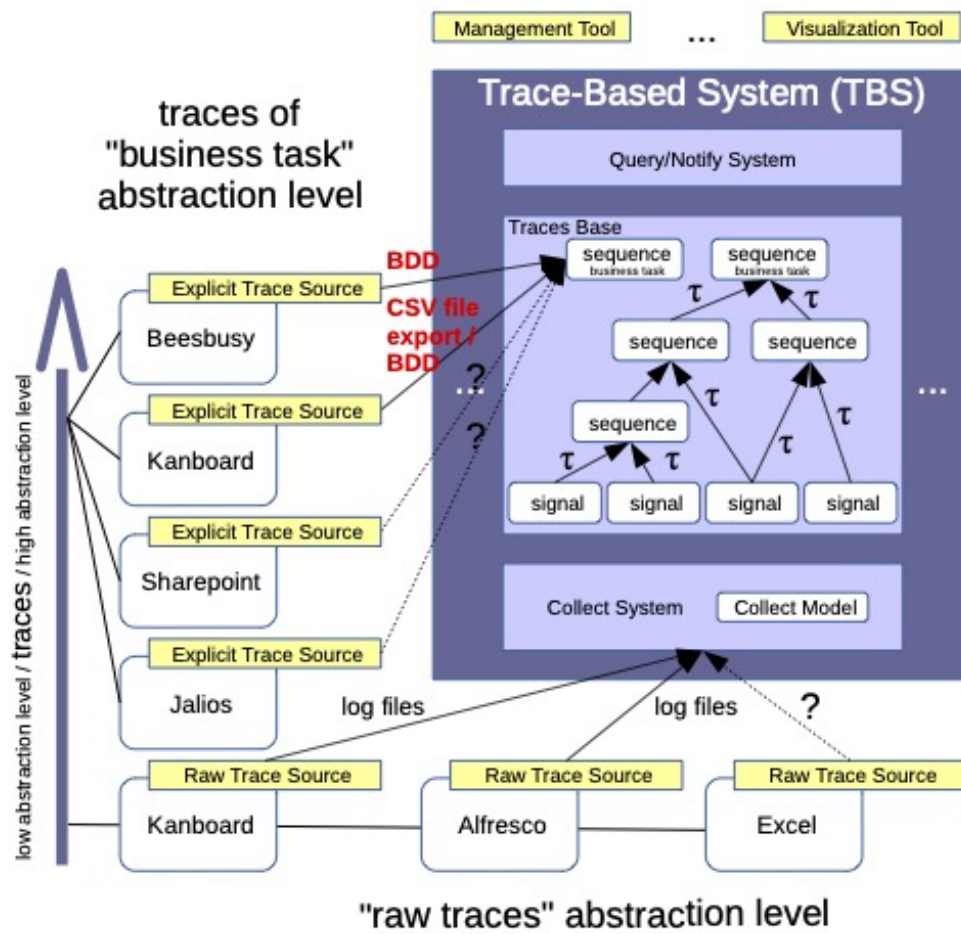


FIGURE 2.6 – Niveaux d'abstraction des traces collectées

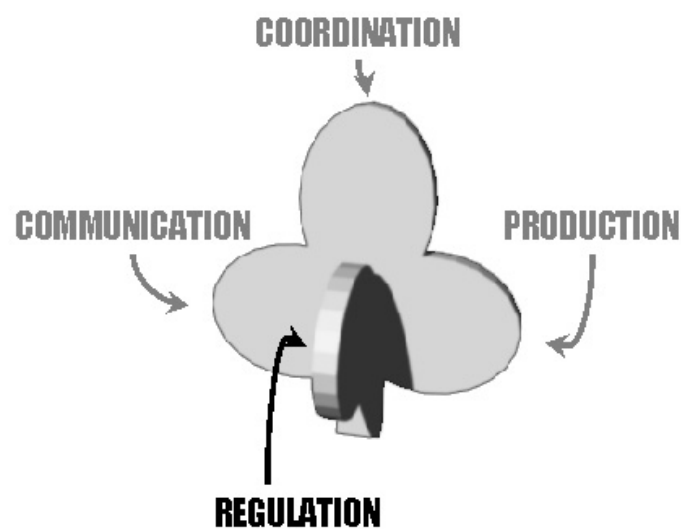


FIGURE 2.7 – Modèle du trèfle augmenté

Système de visualisation

La visualisation des traces numériques dépend du niveau d'observation considéré. Elle est supportée soit par des outils spécifiquement créés pour les observatrices et les observateurs, soit insérée dans les outils logiciels existants qui sont utilisés pour le travail collaboratif.

Nous présentons (sur la figure 2.15 de la partie 2.5) une interface graphique de visualisation des signaux et des séquences, dédiée à des chercheuses et des chercheurs informaticiens et réalisée spécifiquement pour les besoins d'une expérimentation. L'objectif de ce système de visualisation des traces directes est de permettre le suivi technique du fonctionnement de la station d'observation.

Nous présentons enfin (sur la figure 2.14 de la partie 2.5) une interface graphique pour l'un des utilisateurs finaux de l'expérimentation, à savoir l'enseignant d'anglais. Cette interface de visualisation des traces indirectes est créée spécifiquement pour fournir à l'enseignant des indicateurs adaptés à son niveau d'interprétation qui sont de type *group awareness*, servant à connaître la structure des groupes (i.e. la composition des binômes) et leurs activités chronodatées.

Système de requêtes

Pour la réalisation du prototype de station d'observation présentée dans la partie 2.5, nous implémentons des A.P.I.⁹ pour interroger la base de données stockant les traces d'interactions sous la forme de signaux et de séquences.

Pour les besoins d'une expérimentation, nous présentons (sur la figure 2.10 de la partie 2.5) une interface graphique d'écriture des règles pour interroger « à la volée » les traces atomiques (les signaux) et composites (les séquences), qui sont générées et transformées au cours de l'activité collaborative.

2.3 Analyse des traces d'interaction

Pour préparer une analyse, il est nécessaire de définir un modèle d'utilisation pour chaque application utilisée. En pratique, le déroulement d'une analyse s'établit en trois étapes :

1. configuration des gabarits (*templates*) pour définir les paramètres optionnels des signaux (cf. système de requêtes d'écriture des règles) ;
2. exécution des règles pour reconnaître les signaux et les séquences (moteur d'inférence) ;
3. présentation des traces résultantes pour l'interprétation (cf. systèmes de visualisation pour les chercheuses et les chercheurs ou les utilisatrices et les utilisateurs finaux).

Des opérateurs logiques (ET, OU et NON), une relation temporelle (PUIS) et des relations de priorité (des parenthèses) sont intégrés dans les règles pour la composition des gabarits à partir des « obsels ». Les règles permettent soit d'enrichir l'information, soit de filtrer le bruit dans les traces pour répondre aux objectifs de l'observation. Des indicateurs propres à certaines activités peuvent être créés, sous la forme de signaux de haut niveau d'abstraction non générés par le système lui-même (e.g. réussir l'écriture d'une définition).

9. *Application Programming Interface*

2.3.1 Formalisme de construction des séquences

Nous décrivons dans cette partie le formalisme à partir d'exemples issus d'une expérimentation que nous décrivons dans la partie 2.5 page 51. Lors de cette expérimentation, l'utilisatrice « u1 » édite une page « page1 » d'un outil d'édition « wiki » dans l'intervalle de temps [?t0, ?t1]¹⁰ pour rédiger la définition d'un mot en anglais. Cette **séquence de niveau 1** est composée de 2 signaux « *edit_page* » et « *save_page* » et peut être décrite comme suit :

```
["analyser", ?description2, ?t0, ?t1, ?type2,  
  ["wiki", ?description0, ?t0, "edit_page", "wiki", "user1", "page1"],  
  ["wiki", ?description0, ?t1, "save_page", "wiki", "user1", "page1"],  
  ?su2, ?sp2]  
WHERE ?t0<?t1 AND ?su2="user1" AND ?sp2="page1" AND ?type2="editing"
```

Cette séquence est structurée selon le modèle présenté sur la Figure 2.4 dont nous rappelons la spécification ci-dessous :

```
[source, description, beginDate, endDate, type,  
  [source, description, date, event, tool, param1, param2],  
  [source, description, date, event, tool, param1, param2],  
  param1, param2]  
WHERE beginDate<endDate AND param1=param1 AND param2=param2 AND type="editing"
```

2.3.2 Règles de recherche des séquences basées sur les gabarits

Comme dans la plupart des moteurs d'inférence, les règles comportent deux phases : une phase d'évaluation et une phase d'exécution. La phase d'évaluation consiste à choisir les règles qui vont être utilisées en fonction des signaux et des séquences reconnus. Conformément au fonctionnement de la programmation logique, les règles sont déclenchées en fonction de l'adéquation entre une situation donnée et les conditions desdites règles.

Nous décrivons ci-dessous le fonctionnement des règles à travers plusieurs exemples de construction de séquences, les deux premières **séquences de niveau 1**, c'est-à-dire composées uniquement de signaux, servent à construire une **séquence de niveau 2**, c'est-à-dire à partir de signaux et de **séquences de niveau 1**. Enfin, la construction d'une **séquence de niveau 3** qui s'appuie sur les séquences sous-jacentes est également présentée.

Exemple 1 : édition d'une page d'éditeur de texte

Le gabarit des signaux à l'origine de la séquence décrite dans la partie 2.3.1 est présenté ci-dessous :

```
[?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?p1] THEN  
[?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?p2]  
WHERE  
?tool1=?tool2 AND ?t1<?t2 AND ?event1="edit_page" AND ?event2="save_page"  
AND ?u1=?u2 AND ?p1=?p2
```

Ce gabarit est utilisé dans la partie « prédicat » d'une règle qui construit, dans la partie « action », la séquence correspondante présentée ci-dessous :

10. par convention, nous préfixons les variables avec un point d'interrogation

```
["analyser", "editing of a page", ?t1, ?t2, "editing",
  [?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?p1],
  [?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?p2],
  ?u1, ?p1]
```

Nous présentons dans les exemples suivants plusieurs gabarits que nous combinons pour construire des gabarits de plus haut niveau d'abstraction.

Exemple 2 : co-édition d'une page d'éditeur de texte

Partie « prédicat » de la règle

À noter que la co-édition pourrait s'appuyer soit directement sur le repérage des 4 signaux qui la composent, soit sur l'utilisation des 2 séquences d'édition si les gabarits ont été définis. Nous présentons ci-dessous la solution des signaux et dans un deuxième temps celle des séquences de niveau 1 :

```
a(?event, ?u, ?p) AND b(?event, ?u, ?p) AND c(?event, ?u, ?p) OR d(?event, ?u, ?p)
WHERE
a.?event="edit_page" AND b.?event="edit_page" AND a.?u <> b.?u AND a.?p = b.?p
c.?event="save_page" AND d.?event="save_page" AND c.?u <> d.?u AND c.?p = d.?p
AND c.?p = a.?p
```

Partie « Action » de la règle

```
e(?type, ?gu, ?p)
WHERE
e.?type="co-editing" AND e.?gu=U(a.?u, b.?u) AND e.?p=a.?p
```

La partie **Prédicat** de la règle contient 2 gabarits de séquences ayant 2 paramètres chacun ?u et ?p représentant respectivement 2 utilisateurs différents et une même page (e.g. de Wiki), et correspondant à 2 séquences d'édition de la page par les 2 utilisateurs. Les séquences correspondantes sont présentées ci-dessous :

```
[?source3, ?description3, ?t1, ?t2, ?type1,
  [?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?p1],
  [?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?p2],
  ?u1, ?p1] AND
[?source6, ?description6, ?t3, ?t4, ?type2,
  [?source4, ?description4, ?t3, ?event4, ?tool4, ?u4, ?p4],
  [?source5, ?description5, ?t4, ?event5, ?tool5, ?u5, ?p5],
  ?u4, ?p4]
WHERE
?type1="editing" AND ?type2="editing" AND ?u1<>?u4 AND ?p1=?p4
```

La partie **Action** de la règle permet de construire la séquence e de « co-editing » de la page a.?p par 2 utilisateurs a.?u et b.?u, que nous présentons ci-dessous :

```
["analyser", "co-editing of a page", ?t5, ?t6, "co-editing",
  [?source3, ?description3, ?t1, ?t2, ?type1,
    [?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?p1],
    [?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?p2],
    ?u1, ?p1],
```

```
[?source6, ?description6, ?t3, ?t4, ?type2,
  [?source4, ?description4, ?t3, ?event4, ?tool4, ?u4, ?p4],
  [?source5, ?description5, ?t4, ?event5, ?tool5, ?u5, ?p5],
  ?u4, ?p4],
?gu1, ?p1]
```

Exemple 3 : communication dans un chat

Partie « prédicat » de la règle

```
a(?event, ?u, ?r) AND b(?event, ?u, ?r) AND c(?event, ?u, ?r) OR d(?event, ?u, ?r)
WHERE
a.?event="join_room" AND b.?event="join_room" AND a.?u <> b.?u AND a.?r = b.?r
c.?event="leave_room" AND d.?event="leave_room" AND c.?u <> d.?u AND c.?r = d.?r
```

Partie « Action » de la règle

```
e(?type, ?gu, ?r)
WHERE
e.?type="communicating" AND e.?gu=U(a.?u, b.?u) AND e.?r=a.?r
```

La partie **Prédicat** de la règle contient 4 gabarits de signaux a, b, c et d ayant 2 paramètres chacun ?u et ?r représentant respectivement un utilisateur et une salle (e.g. de CoffeeRoom ([Courtin and Talbot 2005](#))), où les signaux a et b correspondent à l'arrivée et à la sortie de la salle du premier utilisateur et les signaux c et d à l'arrivée et à la sortie de cette même salle par le deuxième utilisateur. Les signaux correspondants sont présentés ci-dessous :

```
([?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?r1] THEN
[?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?r2])
AND
([?source3, ?description3, ?t3, ?event3, ?tool3, ?u3, ?r3] THEN
[?source4, ?description4, ?t4, ?event4, ?tool4, ?u4, ?r4])
WHERE
?t1<?t2 AND ?t3<?t4 AND ?t3<?t2
AND ?event1="join_room" AND ?event2="leave_room"
AND ?event3="join_room" AND ?event4="leave_room"
AND ?tool1=?tool2=?tool3=?tool4 AND ?u1=?u2 AND ?u3=?u4 AND ?u1<>?u3
AND ?r1=?r2=?r3=?r4
```

La partie **Action** de la règle permet de construire la séquence e de « *communicating* » dans la salle a.?r par 2 utilisateurs a.?u et b.?u, que nous présentons ci-dessous :

```
["analyser", "communicating in a chat", ?t3, ?t2, "communicating",
  [?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?r1],
  [?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?r2],
  [?source3, ?description3, ?t3, ?event3, ?tool3, ?u3, ?r3],
  [?source4, ?description4, ?t4, ?event4, ?tool4, ?u4, ?r4],
  ?gu1, ?r1]
WHERE
?gu1=U(?u1, ?u3)
```

Exemple 4 : co-rédaction d'une page d'éditeur de texte avec communication dans un chat

Nous définissons l'activité de « co-rédaction » comme étant composée des activités de « co-édition » et de « communication ».

Partie « prédicat » de la règle

```
a(?type, ?gu, ?p) AND b(?type, ?gu, ?r)
WHERE
a.?type="co-editing" AND b.?type="communicating" AND a.?gu = b.?gu
```

Partie « Action » de la règle

```
c(?type, ?gu, ?p, ?r)
WHERE
c.?type="co-writing" AND c.?gu=a.?gu AND c.?p=b.?p AND c.?r=b.?r
```

La partie **Prédicat** de la règle contient 2 gabarits de séquences a et b de niveaux respectivement 2 et 1, ayant 2 paramètres chacune ?gu1 et ?p1 pour la première et ?gu2 et ?r8 pour la deuxième, représentant un groupe d'utilisateurs (« gu » pour *group of users*), une page (« p » pour page) et une salle (ici « r » pour *room*). Les 2 séquences correspondent respectivement à la co-édition de la même page par le groupe d'utilisateurs et à la communication dans la même salle par ce même groupe. Les séquences précitées sont présentées ci-dessous :

```
[?source7, ?description7, ?t5, ?t6, ?type3,
  [?source3, ?description3, ?t1, ?t2, ?type1,
    [?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?p1],
    [?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?p2],
    ?u1, ?p1] AND
    [?source6, ?description6, ?t3, ?t4, ?type2,
      [?source4, ?description4, ?t3, ?event4, ?tool4, ?u4, ?p4],
      [?source5, ?description5, ?t4, ?event5, ?tool5, ?u5, ?p5],
      ?u4, ?p4]
  ?gu1, ?p1]
AND
[?source2, ?description12, ?t10, ?t9, ?type4,
  [?source8, ?description8, ?t8, ?event8, ?tool8, ?u8, ?r8],
  [?source9, ?description9, ?t9, ?event9, ?tool9, ?u9, ?r9],
  [?source10, ?description10, ?t10, ?event10, ?tool10, ?u10, ?r10],
  [?source11, ?description11, ?t11, ?event11, ?tool11, ?u11, ?r11],
  ?gu2, ?r8]
WHERE
?gu1=?gu2 AND ?type3="co-editing" AND ?type4="communicating"
```

La partie **Action** de la règle permet de construire la séquence c de « co-writing » de la page ?p1 avec communication dans la salle ?r8 par le groupe d'utilisateurs ?gu1, que nous présentons ci-dessous :

```
["analyser", "co-writing in a pair", ?t12, ?t13, "co-writing",
  [?source7, ?description7, ?t5, ?t6, ?type3,
    [?source3, ?description3, ?t1, ?t2, ?type1,
      [?source1, ?description1, ?t1, ?event1, ?tool1, ?u1, ?p1],
      [?source2, ?description2, ?t2, ?event2, ?tool2, ?u2, ?p2],
```

```
?u1, ?p1] AND
[?source6, ?description6, ?t3, ?t4, ?type2,
  [?source4, ?description4, ?t3, ?event4, ?tool4, ?u4, ?p4],
  [?source5, ?description5, ?t4, ?event5, ?tool5, ?u5, ?p5],
  ?u4, ?p4]
?gu1, ?p1],
[?source2, ?description12, ?t10, ?t9, ?type4,
  [?source8, ?description8, ?t8, ?event8, ?tool8, ?u8, ?r8],
  [?source9, ?description9, ?t9, ?event9, ?tool9, ?u9, ?r9],
  [?source10, ?description10, ?t10, ?event10, ?tool10, ?u10, ?r10],
  [?source11, ?description11, ?t11, ?event11, ?tool11, ?u11, ?r11],
  ?gu2, ?r8],
?gu1, ?p1, ?r8]
WHERE
?t12=min(?t5,?t10) AND ?t13=max(?t6,?t9)
```

Avec le mécanisme d’encapsulation des séquences, tous les niveaux d’abstraction sous-jacents sont accessibles pour des analyses a posteriori plus approfondies, telles que la recherche d’enchaînements d’actions qui ont été effectués en fonction de tel ou tel résultat et ainsi repérer des séquences « gagnantes », ou détecter des difficultés avec certaines actions. Par exemple, une enseignante peut vouloir évaluer le degré de collaboration au sein d’un binôme d’étudiants. Cet objectif représente un haut niveau d’abstraction qui peut être exprimé par une séquence avec une description spécifique. La séquence étant un élément de trace composite, l’enseignante peut naviguer à travers les informations encapsulées dans celle-ci, si elle souhaite obtenir plus de détails sur la collaboration intra-binôme par exemple.

2.4 Activité d’observation

Nous accepterions l’idée que le passage d’une situation traditionnelle à une situation médiée par ordinateur serait bénéfique pour tous les participants, si cette dernière maintenait la dimension humaine des activités. Il est à noter que la conception techno-centrée des environnements informatiques limite l’observation des usages (Rabardel 1995). L’apparition des Environnements Numériques de Travail (E.N.T.¹¹) et des Environnements Informatiques pour l’Apprentissage Humain (E.I.A.H.¹²) a accru considérablement les possibilités d’observation de l’activité humaine. L’activité d’observation, qui sous-tend le cœur de nos travaux de recherche, compte plusieurs objectifs. Le premier objectif consiste à fournir aux utilisateurs (élèves, enseignantes ou experts) d’une plate-forme d’E.I.A.H. une vue sur l’activité du groupe, qui incarne la conscience de l’activité du groupe (*group awareness*). Nous avons mené une étude sur les opérations de régulation de l’activité d’apprentissage, incluant la modélisation et l’architecture d’une station d’observation, ainsi qu’une expérimentation du suivi en temps réel (mode synchrone) de l’activité à partir des traces d’interaction des participants. Le deuxième objectif, qui se place au niveau des métiers de l’enseignement, adresse l’utilisabilité du système et l’adéquation avec les activités d’apprentissage. L’objectif à long terme porte sur la mesure de l’impact de l’instrumentation des E.I.A.H. sur les activités réalisées au sein des situations d’apprentissage collectif.

11. *Digital Working Environments* (D.W.E.)

12. *Learning Management Systems* (L.M.S.)

2.4.1 Station d'observation

La station d'observation que nous avons conçue est un système qui interagit avec un environnement informatique pour l'apprentissage humain. Son externalisation est motivée par la volonté de généraliser son fonctionnement avec d'autres systèmes d'apprentissage. Le prototype de station d'observation nommée CARTE (*Collection, activities Analysis and Regulation based on Traces Enriched*), par extension du modèle de système à base de traces (T.B.S.¹³) que nous avons conçu (Courtin 2010), s'appuie sur un format de trace générique afin de s'adapter aux diverses sources de traces.

Nous rappelons l'importance de l'externalisation des opérations d'analyse pour assurer l'indépendance de la station d'observation vis-à-vis des E.I.A.H. La communication entre les outils et la station d'observation CARTE est donc assurée par des A.P.I.¹⁴. Nous avons développé une version de type portail Web de la station d'observation, avec un serveur d'application Web libre, orienté objet Zope¹⁵, écrit en langage Python. Nous avons choisi d'implémenter nos A.P.I. avec la spécification XML-RPC¹⁶ pour interagir de manière transparente avec divers langages de programmation, tels que Java, Zope et Python utilisés pour l'implémentation des outils des E.I.A.H. Si ce choix technique contribue inévitablement à la genericité de notre solution, l'hétérogénéité des technologies utilisées pour les développements d'outils implique certaines adaptations pour récupérer les événements produits pendant leur utilisation.

2.4.2 Instrumentation

Capter des événements afin de les réutiliser ultérieurement n'est pas une idée nouvelle, comme le montrent les travaux de recherche de (Jaczynski and Trousse 1998) sur la navigation Web. Cependant, la plupart des systèmes supportant les situations d'apprentissage collaboratif instrumenté se heurtent à des difficultés d'interprétation pour les utilisateurs finaux des traces collectées. Les travaux de recherche de (Desmarais et al. 1997) sur la production de statistiques sur les traces collectées, à travers des systèmes logiciels dédiés à l'apprentissage humain, établissent que de telles informations sont insuffisantes pour décrire précisément le processus d'apprentissage. D'autres travaux de recherche sur les techniques de fouille de données (*data mining*), associées aux *Web logs*, ont cependant montré qu'il est possible d'évaluer in fine le processus d'apprentissage (Zaiane and Luo 2001).

L'instrumentation consiste à poser des capteurs d'événements logiciels pour alimenter des indicateurs de haut niveau d'abstraction, c'est-à-dire qui expriment des actions relatives au domaine d'application (e.g. rédiger une définition). Le passage des événements bruts aux indicateurs explicites nécessite des transformations effectuées au sein d'un système à base de traces, à l'aide d'un modèle d'utilisation associé à l'application d'où ont été extraites les traces.

L'adaptation d'interface et la personnalisation des parcours basées sur l'instrumentation de l'activité réalisée ont contribué au développement de plates-formes numériques adaptées aux besoins de l'apprentissage. Lors de la mise en place de situations collectives d'apprentissage instrumentées (S.C.A.I.¹⁷), l'enseignant peut planifier des séquences d'activités dans un scénario pédagogique contraint (*workflow*), mais il peut aussi vouloir observer une situation non contrainte (quelles séquences d'actions ont été réalisées ? quelles interactions ont été effectuées ? quels échanges ont eu lieu ? quelles ressources ont été utilisées ? etc.).

13. *Trace-Based System*

14. création d'un *framework* que nous avons développé avec le langage de programmation Java

15. <https://www.zope.org/>

16. « *eXtensible Markup Language - Remote Procedure Calls* » est une spécification d'appel à distance de procédures via H.T.T.P. pour le transport et X.M.L. pour l'encodage

17. *Equipped Collective Learning Situations* (E.C.L.S.) ou *Instrumented Collective Learning Situations* (I.C.L.S.)

L'activité peut être régulée de façon totalement automatisée (scénario conditionné par les activités effectuées, à l'instar des activités « leçon » de Moodle¹⁸) ou manuellement par l'enseignant qui intervient pendant l'activité pour aider, corriger ou stimuler l'apprenant si nécessaire. À la fin de la séance, l'enseignant peut analyser l'activité a posteriori, l'utilité des outils disponibles (statistiques), et la nature des tâches effectuées (production, communication, etc.). À partir de là, il procède à l'évaluation du processus de l'activité, pour ajuster son organisation et ainsi l'améliorer. Les activités d'**observation** et de **régulation** sont essentielles au suivi pédagogique pour aider les apprenants et adapter l'environnement à l'individu. (Pernin and Lejeune 2004) soulignent la nécessité d'automatiser ces deux activités traditionnellement réalisées par l'être humain.

Si certains logiciels fournissent, peu ou prou, d'éléments de traces sous forme de fichiers de *log* ou de bases de données, voire de fonctionnalités intégrées (historique des actions, rapport statistique, etc.), ces éléments ne sont généralement pas exploités pour adapter leur fonctionnement aux usages effectifs comme, par exemple, la création de parcours différenciés d'apprenants dans les E.I.A.H. Notre objectif est donc d'appareiller les systèmes sur la base des actions effectuées en leur sein, pour fournir des extensions de fonctionnalités (*add-on*) à partir de traitements externes. L'externalisation des opérations de collecte et de traitement contribue à la généralisation de notre solution, mais se heurte à la difficulté pour rétroagir vers le logiciel. Dans le meilleur des cas, des A.P.I.¹⁹ fournissent des scripts pour échanger avec l'application, ce qui est indispensable pour les applications propriétaires fermées.

Dans une expérimentation décrite dans la partie 2.5 page 51, nous instrumentons l'outil de co-édition « Jibiki » (Mangeot and Nguyen 2009) et l'outil de communication « CoffeeRoom » (Courtin and Talbot 2005). L'outil « Jibiki » développé en Python est intégré dans un portail Zope, alors que l'outil « CoffeeRoom » est développé en Java. Nous avons d'abord voulu démontrer que nous pouvions instrumenter des outils, qu'ils soient ou non intégrés à notre portail Zope. Ensuite, nous avons prévu d'utiliser ces outils dans différentes expériences, mais jusqu'à l'instrumentation, nous manquions d'informations utilisables sur ce qui se passait à l'intérieur pendant ces expériences.

Nous avons utilisé deux approches différentes pour l'instrumentation. Avec le wiki, nous avons légèrement modifié l'I.H.M.²⁰. Ainsi, lorsque l'utilisateur clique sur des boutons ou des liens, le code supplémentaire que nous avons inséré dans l'I.H.M. envoie des signaux (e.g. « *create_page* », « *edit_page* », « *save_page* ») à notre collecteur de traces. L'I.H.M. du « CoffeeRoom » est plus sophistiquée que celle du wiki, car l'outil fonctionne en mode client/Serveur où le serveur reçoit des informations de chaque client avant de les transmettre vers les autres. L'option retenue est donc d'instrumenter le serveur central du « CoffeeRoom » pour l'envoi des signaux (e.g. « *connect* », « *create_table* », « *send_message* »). À l'instar du « Jibiki », le « CoffeeRoom » dispose d'une A.P.I. pour effectuer les rétroactions. De la même manière, une méthode générique reçoit des paramètres tels que la référence du serveur qui centralise les messages et sur lequel se trouvent déjà les méthodes correspondant aux différentes (rétro)actions. Cette technique permet à la fois d'être indépendant de l'outil de communication (dès lors qu'il contient des espaces de discussion séparés) et de prendre en compte les spécificités de l'outil choisi (ici, le « CoffeeRoom »). L'A.P.I. de rétroaction liée au serveur du « CoffeeRoom » permet un accès à distance via le protocole R.M.I.²¹.

18. <https://moodle.org/>

19. *Application Programming Interface* qui permet à des applications de communiquer entre elles et de s'échanger mutuellement des services ou des données.

20. Interface Humain-Machine

21. *Remote Method Invocation* est une A.P.I. Java permettant de manipuler des objets distants, c'est-à-dire instanciés sur une autre machine du réseau, de manière transparente pour l'utilisateur

Nous avons présenté ci-dessus une instrumentation des différents outils, qui permet à l'analyste de mettre en œuvre les deux types de régulation, à savoir : des interventions humaines à partir d'informations issues des « obsels » (*awareness*) et des rétroactions déclenchées par une ou plusieurs règles prédéfinies par l'observateur lui-même. Dans les deux cas, le système est basé sur l'exploitation des traces d'activité qui sont collectées, puis analysées dans la station d'observation au moyen de modèles d'utilisation prédéfinis.

2.4.3 Rétroaction

La station d'observation CARTE²² que nous avons produite opérationnalise le système à base de traces présenté Figure 2.3, en s'appuyant sur le format de trace CARTE présenté Figure 2.4, associé au modèle d'utilisation présenté dans une sous-partie de la partie 2.2.1 page 38. Les traces modélisées, ou MTraces, sont soit issues du module de collecte de la station d'observation, soit produites par le module d'analyse à l'aide de transformations (τ) à partir de MTraces de plus bas niveaux d'abstraction. L'objectif de ces transformations est d'atteindre un niveau d'abstraction adapté aux observatrices humaines, qui peuvent ainsi interpréter les activités effectuées dans un domaine d'application spécifique. À défaut de disposer d'indicateurs pertinents sur l'utilisation des applications logicielles, les observatrices sont impliquées dans la spécification des modèles d'utilisation de ces applications. Au-delà de la simple observation, les modèles d'utilisation peuvent modifier le schème²³ d'une activité réalisée avec l'application, pour lequel il n'existe pas de fonctionnalité prévue au sein de celle-ci (e.g. création d'une notification quand une opération est terminée). En d'autres termes, la station d'observation permet, entre autres, de modifier le fonctionnement d'une application pour des besoins d'analyse, sans ajouter les procédures afférentes dans celle-ci et ainsi comparer des situations avec et sans ces fonctionnalités. Nous partageons l'idée de (Mille, Caplat, and Philippon 2006) précisant que les actions des participants dépendent des réactions de l'environnement dans un contexte d'activité donné. L'**originalité de notre approche** tient essentiellement à l'indépendance entre les modèles d'utilisation et les applications logicielles observées, et donc à l'adaptation des analyses à divers objectifs d'observation. En plus d'explicitier le fonctionnement des applications logicielles, les modèles d'utilisation servent à intégrer des règles sociales dans le déroulement des activités (e.g. bloquer les intrusions interdites dans des groupes de discussion du « CoffeeRoom » (Courtin and Talbot 2005)).

L'outil d'analyse de la station d'observation CARTE, dont le fonctionnement est décrit Figure 2.8, fournit des traces de types « *awareness* » et « **rétroaction** » (incluse dans la régulation). Dans le premier cas, l'observatrice reçoit à travers une interface de type « tableau de bord » des informations relatives aux « obsels » décrits par le modèle d'utilisation. En interprétant ces informations, l'observatrice (une enseignante, une étudiante, etc.) a conscience de l'activité en cours et agit en conséquence pour atteindre les objectifs fixés. Dans le second cas, les traces déclenchent l'exécution d'actions destinées à faciliter l'utilisation (e.g. proposition de la tâche suivante dans un schème reconnu), ou à enrichir le schème d'une activité (e.g. avertissement automatique suite à la modification d'un document).

Au démarrage de la station d'observation, une base de règles (au format XML) dédiée à la régulation est chargée. L'analyste est donc prêt à recevoir des signaux soit collectés au moyen de l'A.P.I.²⁴ « *collect* » (e.g. l'action « sauvegarder une page »), soit produits par l'analyste lui-même (e.g. l'action « début d'une co-écriture ») (Courtin and Talbot 2007).

22. *Collection, activities Analysis and Regulation based on Traces Enriched*

23. désigne une séquence d'actions pour réaliser une activité

24. *Application Programming Interface*

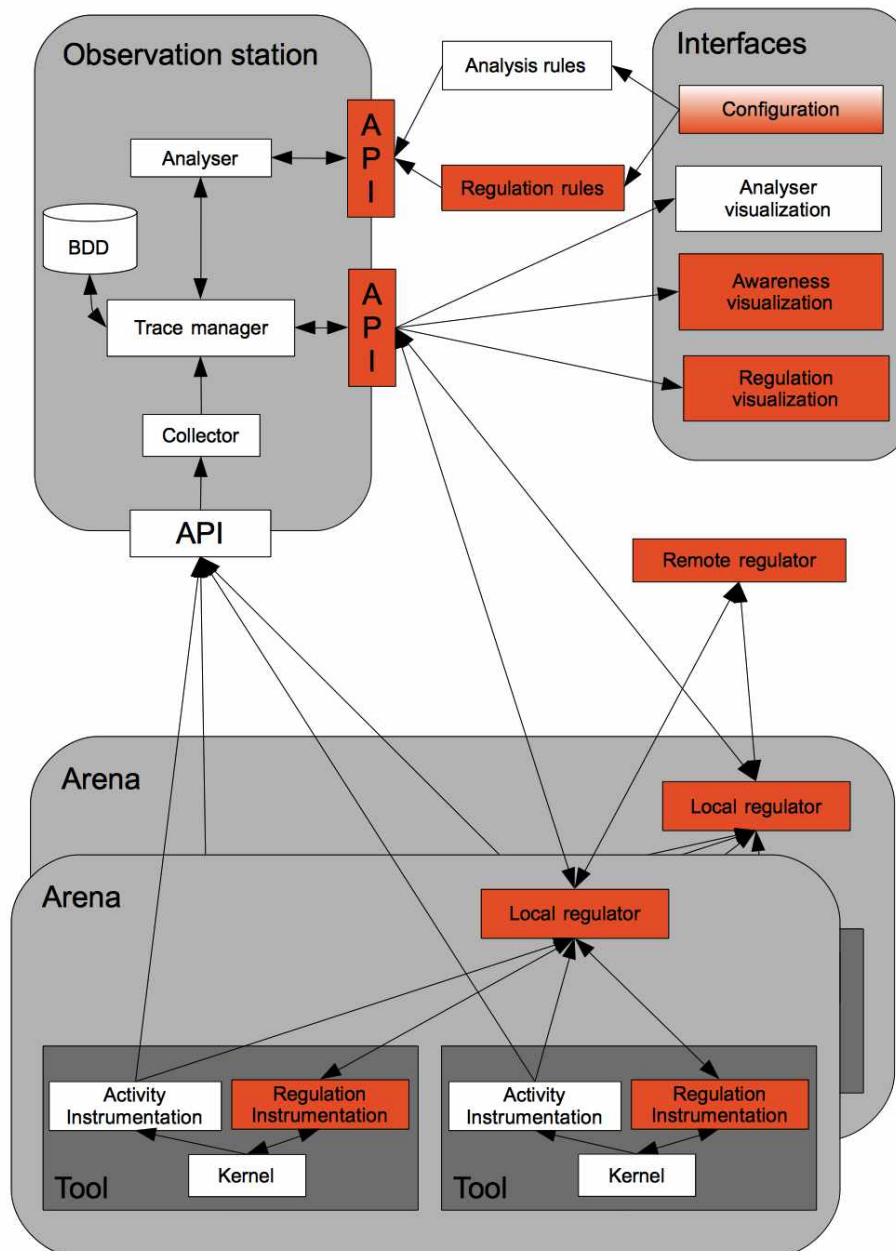


FIGURE 2.8 – Architecture de la station d'observation

Les rétroactions (ou actions entrantes automatiques) que nous implémentons pour l'expérimentation présentée ci-dessous sont :

- créer ou détruire une table dans le « CoffeeRoomn » à partir des traces issues de l'outil de coordination ;
- connecter ou déconnecter un utilisateur ;
- ajouter un utilisateur à une table selon la définition des binômes ;
- renvoyer à sa table un utilisateur qui tente d'accéder à une table où il n'est pas inscrit ;
- envoyer un message à un table pour, par exemple, informer un partenaire de la validation d'une action.

2.5 Expérimentation d'une station d'observation

2.5.1 Les objectifs de l'expérimentation

Nous considérons qu'une activité collaborative d'apprentissage requiert l'investissement mutuel des participants, et implique des intentions éducatives visant à aider ou à faciliter le processus d'apprentissage. Partant de ce postulat, cette expérimentation tente de répondre à plusieurs questions de recherche sous des angles de vue variés (informatique, pédagogique, etc.).

D'un point de vue informatique, nous testons le mécanisme de régulation de l'activité, que nous avons conçu à partir de la spécification d'un système à base de traces, et donc l'expressivité des modèles sous-jacents, des objectifs que ce sont également fixés ([Laflaquière et al. 2006](#)) et ([Settouti et al. 2009a](#)) pour leur propre ébauche de « *framework* » de S.B.T.

D'un point de vue pédagogique, nous cherchons à mesurer les effets de la régulation ([Talbot and Pernelle 2003](#)) sur la perception des utilisateurs de la S.C.A.I.²⁵ et sur leur investissement dans l'activité d'apprentissage. Les travaux sur la dimension « *group awareness* » de ([Gutwin and Greenberg 2002](#)) mettent en exergue l'importance des indicateurs d'« *awareness* » dans les environnements numériques de travail collaboratif. Nos travaux tentent d'aller plus loin en proposant d'automatiser à la fois la production de ces indicateurs et le déclenchement des actions afférentes qu'auraient pu effectuer les utilisateurs sur la base de ces indicateurs.

Nous supposons que s'il existe un mécanisme de régulation qui offre la possibilité d'adapter le profil des utilisateurs à leur propre activité, nous nous attendons à ce que ces utilisateurs aient le sentiment que leur activité sera mieux perçue par la communauté des participants. Nous considérons donc que si la régulation améliore les interactions avec l'environnement informatique, elle peut avoir une influence sur la participation des utilisateurs.

Dans le cadre de cette expérimentation, nous traduisons ces objectifs sous la forme d'une spécification pour laquelle nous listons les grandes lignes ci-dessous :

- fournir un moyen d'informer sur les actions des autres participants (e.g. les actions des membres d'un groupe) ;
- permettre l'enrichissement des fonctionnalités d'un outil logiciel sans y intégrer de nouvelles fonctions (e.g. l'envoi automatique d'un courriel lors de la reconnaissance d'une action spécifique, comme demander de l'aide à l'enseignant) ;

25. Situation Collective d'Apprentissage Instrumenté

- contrôler le respect des règles sociales nécessaires au bon déroulement de la séance d'apprentissage (e.g. interdire à un étudiant de changer d'espace de discussion dans l'outil de communication);
- simplifier la réalisation des opérations en automatisant certaines tâches (e.g. l'automatisation de la configuration des outils logiciels à partir de la définition des binômes par l'enseignant).

2.5.2 Le contexte de l'expérimentation

Le cadre de notre expérimentation est celui d'une situation collective d'apprentissage instrumentée (S.C.A.I.), en temps réel, menée dans le cadre d'un cours d'anglais à l'université. La séance a lieu dans une salle de classe équipée d'ordinateurs utilisés par les élèves pour communiquer entre eux, pour coordonner leurs activités et pour produire des contenus. Les trois activités relatives au travail des élèves (co-édition, communication et coordination) doivent être réalisées à l'aide d'une interface Web instrumentée, pour suivre le déroulement des opérations avec le consentement éclairé des participants. Ce consentement est indispensable puisque la station d'observation trace également le contenu du canal de communication utilisé au sein des groupes, ainsi que la répartition du travail entre les individus.

L'expérimentation de la station d'observation CARTE mesure la faisabilité du mécanisme de régulation basé sur l'instrumentation des outils numériques d'apprentissage. Nous présentons un prototype couplé avec un E.I.A.H., qui permet aux participants de communiquer entre eux, d'organiser des groupes de travail et de produire des contenus en collaborant. La station d'observation trace les activités d'apprentissage en temps réel et fournit aux observateurs (élèves, enseignants ou experts) des informations sous la forme de *feedbacks* en vue de suggérer des opérations à travers l'E.I.A.H. Par exemple, l'enseignante peut utiliser ces *feedbacks* pour intervenir pendant la séance en vue d'aider, de corriger ou de stimuler les étudiants si nécessaire. Une fois la séance terminée, l'enseignante peut analyser les traces d'activité pour évaluer l'utilisabilité des outils logiciels utilisés, la nature des tâches réalisées (e.g. les productions, les échanges entre les pairs, etc.) et ainsi envisager des ajustements en vue d'améliorer l'activité et d'atteindre les objectifs de réussite.

Au cours de l'expérimentation, nous utilisons deux types de trace dans notre S.B.T.²⁶ :

- les traces d'interaction atomiques (les signaux) en provenance du serveur Web « Jibiki » utilisés pour les activités collaboratives et du *chat-room* « CoffeeRoom »;
- les traces composites ou enrichies (les séquences) à l'aide de transformations dans la station d'observation « CARTE ».

Certains « obsels » proviennent des outils logiciels eux-même et d'autres sont générés par la station d'observation, ce qui implique que leur spécification soit définie par l'observateur avant de démarrer la séance de travail. La Figure 2.9 illustre l'architecture du système global et souligne les protocoles utilisés pour le transport des traces.

26. Système à Base de Traces ou *Trace-Based System*

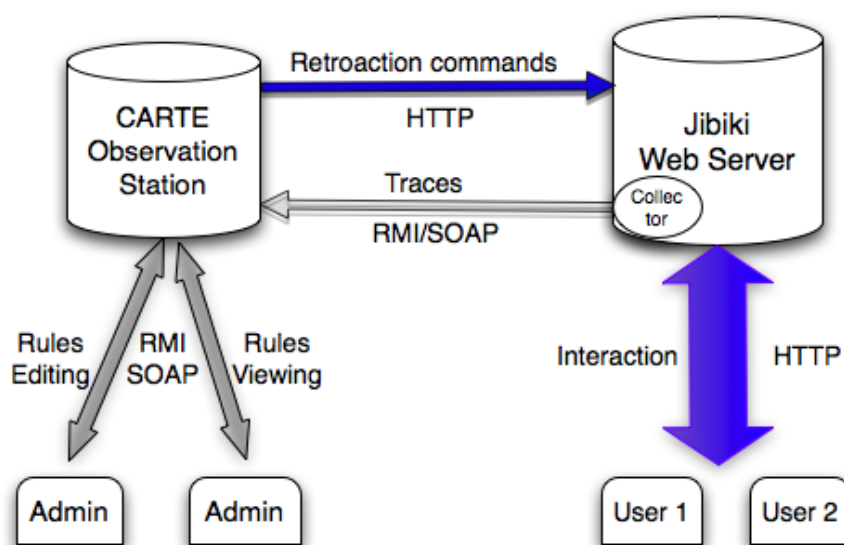


FIGURE 2.9 – Architecture logicielle pour l'expérimentation

2.5.3 Les outils logiciels de l'expérimentation

L'expérimentation repose sur cinq outils logiciels pour configurer l'observation, rédiger de façon collaborative, communiquer, coordonner et observer :

- **L'outil d'édition des règles** pour configurer les gabarits répondant aux modèles d'utilisation des outils de la S.C.A.I.
- **L'outil de production**, développé avec Zope/Python, de type « Wiki », est l'éditeur de texte Web collaboratif asynchrone nommé « Jibiki », intégré dans un portail Zope. Cet outil a été développé dans le cadre du projet de recherche en linguistique « MotÀMot » pour créer des dictionnaires multilingues (Mangeot and Nguyen 2009) (Courtin, Talbot, and Mangeot 2010).
- **L'outil de communication** « CoffeeRoom », développé en Java, de type « chat » est structuré sous la forme de tables de discussion (Courtin 2010).
- **L'outil de coordination** pour structurer les groupes (formation des binômes).
- **L'outil d'observation** CARTE des traces d'activité.

Nous partons du postulat que notre S.C.A.I.²⁷ répond à la spécification des modèles de collecticiels²⁸. Nous nous basons sur le modèle classique du trèfle (*clover model*) présenté dans la sous-partie de la partie 2.2.1 page 39, augmenté par l'introduction d'un quatrième espace nommé « régulation ». Les outils logiciels ont été choisis en tant qu'ils couvrent les différents espaces du modèle du trèfle augmenté. L'adaptation de ces outils logiciels, qui sont capables d'interopérer les uns avec les autres et de fournir des fonctionnalités adaptées aux situations collaboratives d'apprentissage, a été rendue possible grâce à l'utilisation des traces qu'ils produisent. En d'autres termes, la spécification des outils logiciels d'apprentissage n'implique aucun ajout de procédures pour prendre en compte les nouvelles actions, sinon du code pour détourner certains événements consécutifs à des actions tracées, sans pour autant perturber le fonctionnement initial des outils.

27. Situation Collective d'Apprentissage Instrumenté ou *Equipped Collective Learning Situation*

28. *groupware*

L'**outil d'édition des règles** permet de spécifier les gabarits des séquences à reconnaître. Nous utilisons une DTD²⁹ pour décrire les règles de l'analyseur (e.g. DTD et exemple de règle en Annexe B) et l'interface graphique de la Figure 2.10 pour les configurer. Dans l'exemple, à la règle intitulée « EditionSansParler » correspond une séquence de niveau 1, c'est-à-dire uniquement composée de signaux. La partie gauche des règles qui représente la spécification du modèle d'utilisation, est basée sur l'utilisation potentielle d'opérateurs logiques (ET, OU, NON), de relation temporelle (PUIS) et de priorité (parenthèses). Chaque fois qu'un signal est reçu, le module d'analyse vérifie s'il contribue à la condition d'une ou plusieurs règles. Quand il s'agit d'une règle de régulation, cette dernière produit une séquence d'actions qui seront exécutées sur les outils au moyen de l'A.P.I. de rétroaction (contenant toutes les méthodes de rétroaction prédéfinies), lorsque la condition de la règle est remplie. La règle de l'exemple sur la Figure 2.10 génère une rétroaction intitulée « activeMotProf » qui consiste à informer l'enseignant qu'une définition est terminée, sans qu'il y ait eu d'échange entre les étudiants. Il lui appartient ensuite d'apprécier l'activité collaborative au sein du binôme pour éventuellement intervenir en temps réel auprès des étudiants. À noter que cette notification n'existe pas dans la version non instrumentée de l'outil de co-édition.

L'interface graphique actuelle de l'outil d'édition des règles n'étant pas *friendly*, nous assistons l'enseignant dans cette opération, en attendant qu'une réflexion en *UX Design* et les développements informatiques afférents soient engagés.

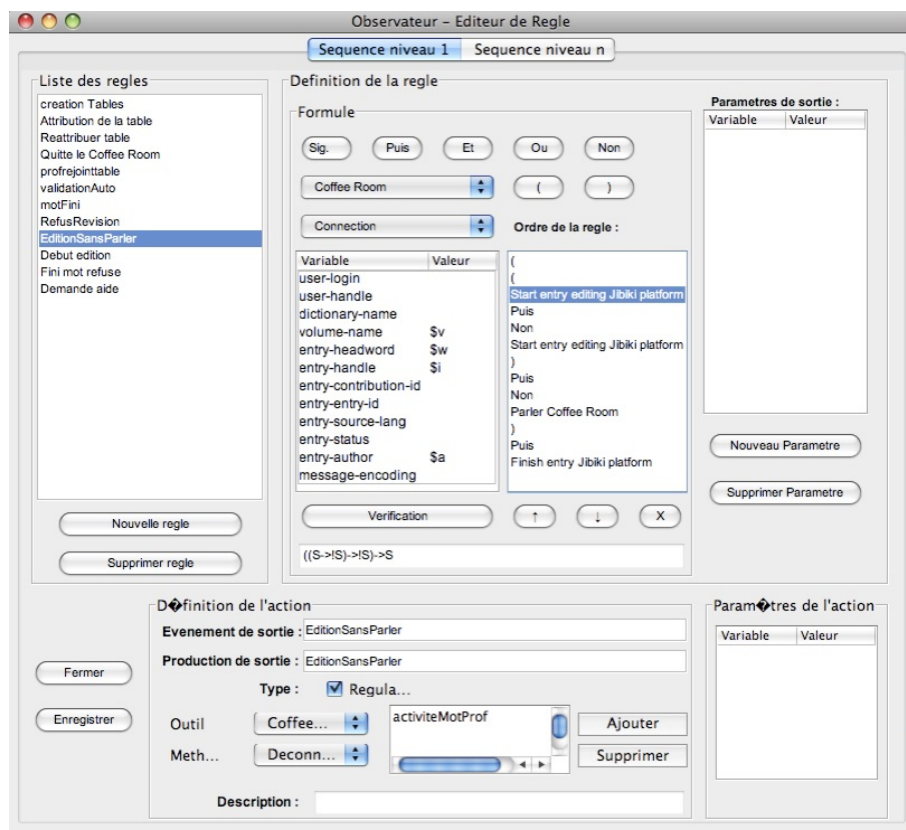


FIGURE 2.10 – Outil d'édition des règles

29. Document Type Definition

L'**outil de production** est l'éditeur « Jibiki » (cf. Figure 2.11) conçu pour la création de corpus multilingues de définitions au moyen d'un client Web (Mangeot and Chalvin 2006). Dans cet outil, chaque participant est enregistré avec un rôle auquel sont associées des actions (e.g. création d'une nouvelle définition, validation d'une définition, administration du corpus). Comme « Jibiki » est un client Web, nous avons choisi d'utiliser des requêtes H.T.T.P. pour rétroagir depuis le module de rétroaction de la station d'observation. Une A.P.I.³⁰ a été créée pour interpréter des requêtes génériques de rétroaction recevant les paramètres suivants :

- ?actionType qui représente l'action à déclencher ;
- ?parametersList qui contient tous les paramètres nécessaires pour que la requête soit envoyée.

Par exemple, la déconnexion correspond à la requête H.T.T.P. composée des paramètres suivants :

- ?actionType = JIBIKI_DECONNECTION ;
- ?parametersList :
 - Jibiki address = http ://yeager.univ-savoie.fr ;
 - Port used by the Jibiki = 8999 ;
 - Corresponding Java method for the deconnection = LoginUser.po ;
 - Session identifier ?jsession = KBK4LJQ9zTQa13uAasQTcf63 ;
 - Parameter ?logout = yes ;

Pour l'expérimentation, nous avons instrumenté l'éditeur pour détecter l'ensemble d'actions suivant :

- la déconnexion d'un utilisateur (e.g. l'étudiant « Albert ») ;
- la création d'une entrée (e.g. le mot « gender ») ;
- l'édition d'une définition (e.g. « le sexe d'une personne ou d'un organisme ») ;
- la soumission d'une définition pour être vérifiée par l'enseignant ;
- la mise en veille d'une définition à corriger ;
- l'examen d'une définition pour prendre en compte les remarques de l'enseignant ;
- la validation d'une définition quand elle est correcte ;
- l'intégration de la définition dans le corpus quand elle est validée.

Un **outil de coordination** a été spécialement développé pour les besoins de l'expérimentation en termes d'organisation des groupes (cf. Figure 2.12). L'enseignant coordonne les groupes à l'aide de cet outil de structuration des groupes, qui forme les binômes qui évolueront dans les outils de co-édition et de communication. La spécification des fonctionnalités est basée sur les traces d'utilisation issues des outils de co-édition et de communication, pour respecter le principe d'indépendance des outils, tout en permettant une interdépendance fonctionnelle. En d'autres termes, les interactions avec les autres outils sont réalisées uniquement par les règles de rétroaction définies dans le module d'analyse de la station d'observation. En pratique, après que l'enseignant a défini les binômes, les traces collectées par la station d'observation déclenchent des règles dans lesquelles les rétroactions permettent de créer les tables correspondantes dans le « CoffeeRoomn ». Il convient de noter qu'avec ce mécanisme intermédiaire, le remplacement du « CoffeeRoom » par un autre outil de communication structuré n'affecterait pas la mise en œuvre de l'outil de coordination.

30. *Application Programming Interface*

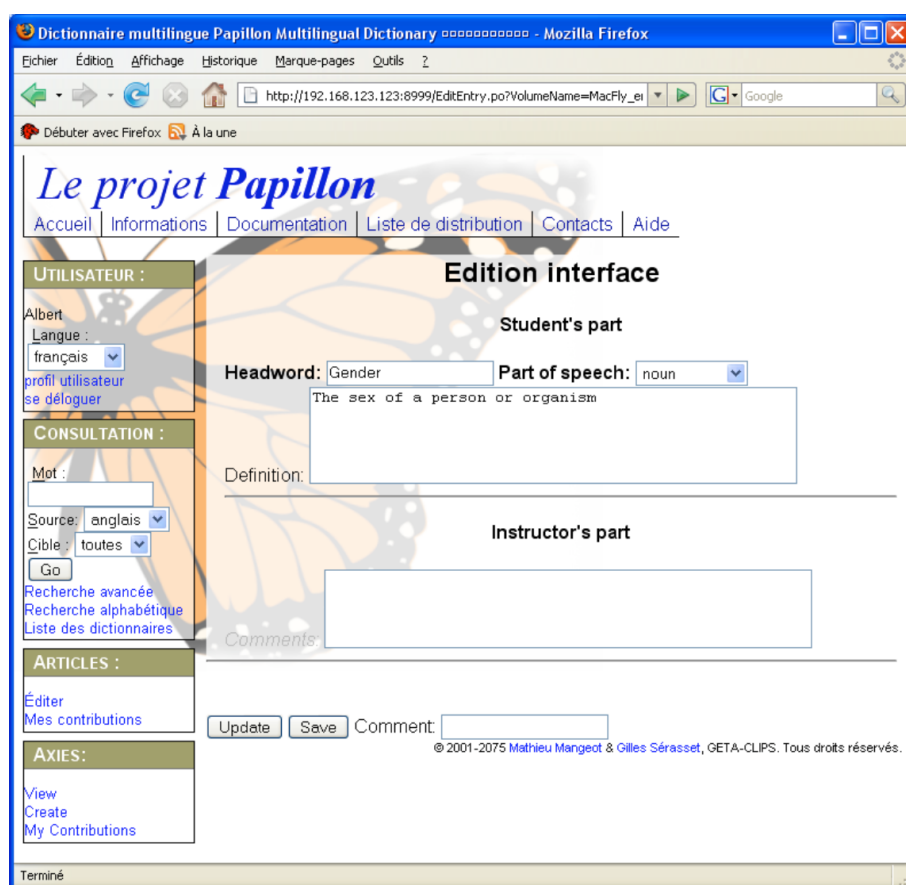


FIGURE 2.11 – Outil de co-édition « Jibiki »

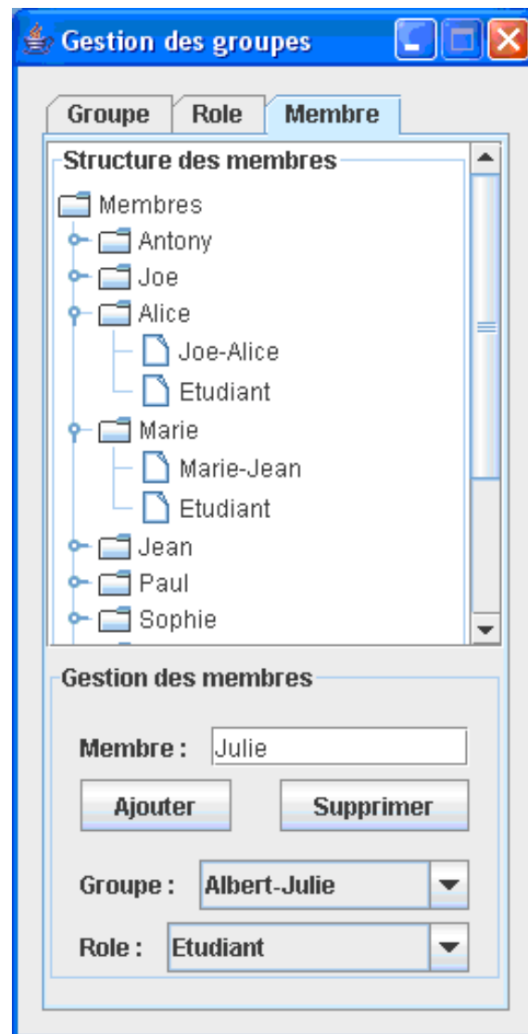


FIGURE 2.12 – *Outil de coordination*

L'**outil de communication** « CoffeeRoomn », de type *chat-room*, est structuré sous forme d'espaces de discussion appelés « tables » (cf. Figure 2.13). Les participants sont automatiquement inscrits dans l'outil avec des rôles prédéfinis par l'enseignant au début de la séance, de même que pour la constitution des binômes. Par exemple, lors de la préparation de la classe, l'enseignant définit les binômes à l'aide de l'outil de coordination, puis édicte des règles de contrôle social à l'aide de l'éditeur de règles de la station d'observation. Dès lors, toute action qui enfreint ces règles sociales (e.g. Albert tente un déplacement non autorisé vers une autre table du « CoffeeRoom », cf. Figure 2.14) entraîne des rétroactions, comme l'envoi d'avertissements dans l'outil d'« *awareness* » (e.g. l'enseignant Antony est averti de la tentative d'intrusion du participant Albert dans un autre binôme), ou l'action sur l'outil dans lequel la transgression a été commise (e.g. le participant est renvoyé à sa table).

Avec l'instrumentation, l'outil de communication « CoffeeRoom » envoie des signaux paramétrés au gestionnaire de traces en réponse aux actions effectuées par les participants (e.g. envoi d'un message sur une table). L'instrumentation peut aussi révéler certains « actes de communication » comme, par exemple, une « conversation » qui correspondrait à de nombreux échanges de messages entre les mêmes participants.

Ainsi, cet outil de communication, lorsqu'il est instrumenté, voit ses fonctionnalités enrichies par l'utilisation des traces d'activité d'un groupe au sein d'un ensemble d'outils (prise en compte d'une partie du contexte).

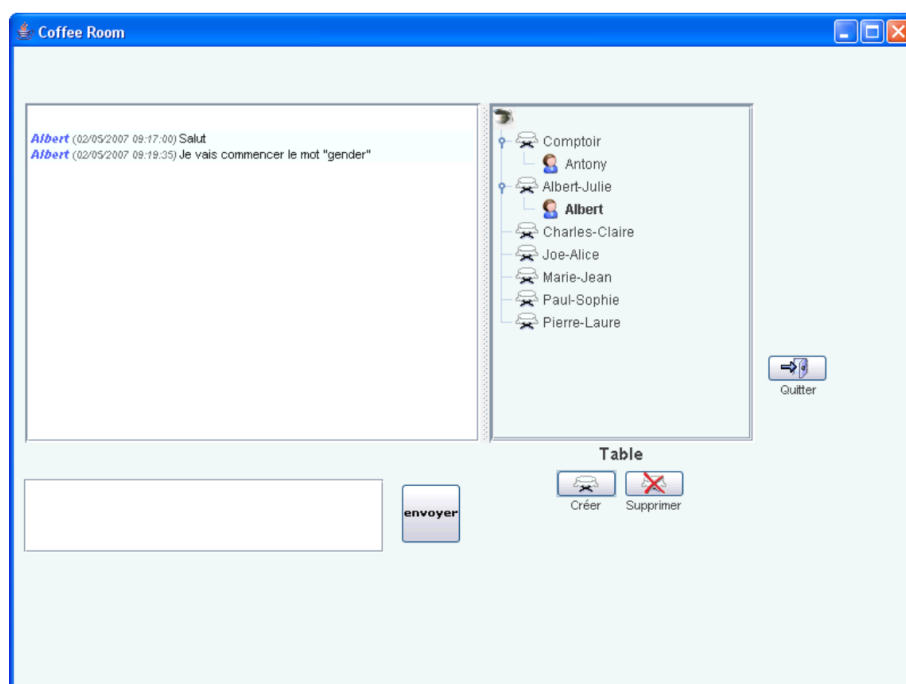


FIGURE 2.13 – Outil de communication « CoffeeRoom »

Un **outil d'observation** de type « *awareness* » a été conçu (cf. Figure 2.14) pour afficher les éléments du contexte de l'activité en cours (e.g. la structure du groupe, le nom des élèves en difficulté, les actions qui viennent d'être réalisées, etc.). Ces informations issues d'« obsels » définis par l'observateur lui-même en accord avec son niveau d'interprétation, sont appelées des indicateurs.

Un autre outil d'observation a été conçu (cf. Figure 2.15) pour permettre aux chercheurs de visualiser les signaux collectés et les séquences produites par l'analyseur à l'aide des règles prédéfinies (Talbot and Courtin 2008). L'outil affiche l'ensemble des séquences et des signaux qu'elles contiennent, avec leur description sur la partie gauche et leurs paramètres sur la partie droite.

2.5.4 Le déroulement des opérations

L'expérimentation se déroule pendant une séance de cours d'anglais à l'université, où les étudiants travaillent en binôme sur des ordinateurs séparés. Le scénario de l'activité étant très simple, il n'est pas nécessaire de le spécifier avec un outil de script. L'activité se déroule comme suit :

- Phase préliminaire :
 - l'enseignant génère les groupes à l'aide de l'outil de coordination ;
 - l'enseignant distribue les sujets avec les mots à définir.
- Phase de travail :
 - les étudiants sont libres d'organiser leur travail en binôme avec le « CoffeeRoom » ;
 - les étudiants définissent les mots avec le « Jibiki » et les soumettent à l'enseignant ;
 - l'enseignant vérifie les définitions, les annote ou les valide directement ;
 - parfois, les étudiants demandent de l'aide à l'enseignant.

L'activité d'apprentissage à réaliser, qui consiste à rédiger des définitions en anglais pour expliciter le vocabulaire issu d'un texte, est divisée en plusieurs étapes synthétisées ci-dessous :

1. les binômes (x_1, x_2) et (y_1, y_2) discutent respectivement de deux définitions ;
2. les individus x_1 et y_1 écrivent séparément chaque définition ;
3. au même moment, les individus x_2 et y_2 échangent successivement au sujet de chaque définition ;
4. les binômes initiaux (x_1, x_2) et (y_1, y_2) synthétisent chaque définition ;
5. les individus x_2 et y_2 remplacent leur pair au sein de leur binôme pour finir la rédaction de chaque définition.

Dans ce scénario, les actions des étapes 1, 3 et 4 sont effectuées dans le canal de communication (« CoffeeRoom ³¹ » (Courtin and Talbot 2005)) et les actions 2 et 5 avec l'éditeur de texte (« Wiki ³² »). Nous décrivons ci-dessous plus en détail les opérations de co-édition et de communication :

1. les binômes (x_1, x_2) et (y_1, y_2) discutent de l'une des deux définitions, ce qui implique qu'un membre de chaque binôme x_1 et y_1 crée une table thématique dans le « CoffeeRoom », et les autres membres x_2 et y_2 rejoignent leur table respective pour échanger des messages au sein de leur binôme ;
2. les individus x_1 et y_1 écrivent séparément chaque définition, ce qui signifie que chaque membre x_1 et y_1 rejoint l'une des 2 pages, l'édite et la sauvegarde ensuite ;
3. au même moment, les individus x_2 et y_2 échangent successivement au sujet de chaque définition, autrement dit les membres x_2 et y_2 de chaque binôme discutent de chaque définition ;
4. les binômes initiaux (x_1, x_2) et (y_1, y_2) synthétisent chaque définition, c'est-à-dire que tous les membres x_1, x_2, y_1 et y_2 rejoignent leur table respective ;
5. les individus x_2 et y_2 remplacent leur pair au sein de leur binôme pour finir la rédaction de chaque définition, en rejoignant leur page respective pour l'éditer, puis la sauvegarder.

Durant la séance d'apprentissage, l'enseignant observe à distance la progression du travail des élèves. Le mécanisme de rétroaction de la station d'observation génère des indicateurs de fin de tâche (e.g. une définition terminée). L'enseignant peut ainsi anticiper les évaluations, voire informer les élèves qu'ils devraient revoir leurs définitions avant la fin de la séance. Une fois la séance terminée, l'enseignant peut analyser le déroulement des opérations au-delà du seul résultat pour, par exemple, évaluer les compétences mobilisées par les élèves (cf. travaux de la partie 4 page 99).

31. une messagerie instantanée où les espaces de discussion sont représentés par des tables

32. un éditeur de texte collaboratif de pages Web

2.5.5 La collecte des traces

L'instrumentation s'appuie sur les traces d'interaction liées aux évènements logiciels qui fournissent des informations sur l'activité collaborative de type « *awareness* » selon le modèle « classique » défini par (Gutwin and Greenberg 2002) et dont la version « avancée » est introduite dans la partie 2.2.1 page 39. Le modèle « classique » compte quatre types d'« *awareness* » : « *workspace awareness*³³ », « *informal awareness*³⁴ », « *group-structural awareness*³⁵ » et « *social awareness*³⁶ ». Ces informations sont plus riches que les *log* du système (*access.log* du serveur HTTP Apache) qui contiennent essentiellement des informations sur les accès par adresses I.P. (*Internet Protocol*) aux différentes pages Web et sur les erreurs de fonctionnement (*error.log*). L'analyse des *log*, très utilisée pour les opérations de S.E.O. (*Search engine Optimization*), ne permet pas de caractériser les opérations effectuées dans la mesure où elle ne dispose pas du modèle d'utilisation de l'application. La spécificité de notre instrumentation est l'introduction d'un modèle d'utilisation de l'application afin d'obtenir des traces plus élaborées, c'est-à-dire ayant un niveau d'abstraction proche de celui de l'observateur humain (e.g. enseignant), qu'on appelle des indicateurs (e.g. pour déterminer le succès ou l'échec d'un exercice).

La Figure 2.16 illustre les informations sur les statistiques d'utilisation des outils collaboratifs qui peuvent être obtenues à partir des *logs* du système, mais sans que les actions effectuées dans ces outils ne soient identifiées du point de vue de l'observation par l'enseignant. Par exemple, nous ne pouvons pas déterminer que le premier binôme est formé par Laure et Antony, et le deuxième par Catherine et Pierre. En d'autres termes, si nous pouvons identifier la position de chaque individu dans l'E.I.A.H., nous ne pouvons pas le faire au niveau des outils eux-mêmes.

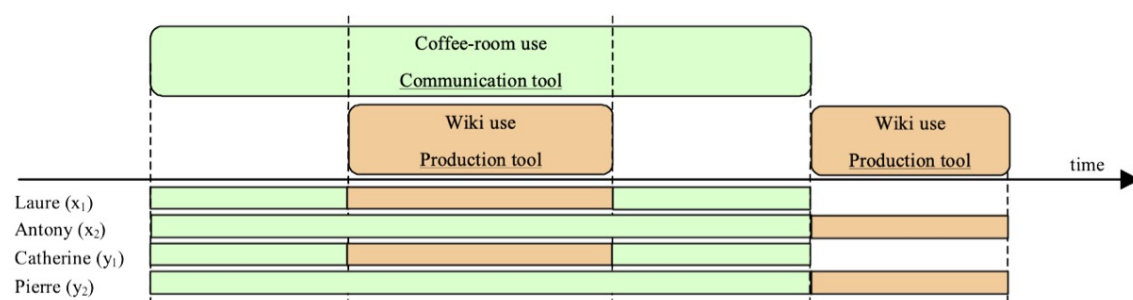


FIGURE 2.16 – Observation avec les logs

La Figure 2.17 présente les éléments de traces fournis par la station d'observation, dont la granularité est celle des actions atomiques (les signaux) effectuées dans les outils de co-édition et de communication, décrites ci-dessous :

* Actions de l'étape (1)

- (a) création d'une table pour la première définition et entrée sur celle-ci (actions indissociables)
- (b) création d'une table pour la deuxième définition et entrée sur celle-ci (actions indissociables)
- (c) entrée sur la table de la première définition
- (d) entrée sur la table de la deuxième définition

33. « the information of what happens in the shared workspace where productions are made », notion complexe simplifiée dans ce document

34. « the general sense of who's around and what they are up to »

35. « knowledge about such things as people's roles and responsibilities, their positions on an issue, their status, and group processes »

36. « the information that a person maintains about others in a social or conversational context »

- (e) envoi d'un message
- * Actions des étapes (2) et (5)
 - (h) entrée sur la page de la première définition
 - (i) entrée sur la page de la deuxième définition
 - (l) édition de la page
 - (n) sauvegarde de la page
- * Actions des étapes (3) et (4)
 - (c) entrée sur la table de la première définition
 - (d) entrée sur la table de la deuxième définition
 - (e) envoi d'un message

L'enchaînement des signaux (h), (l) et (n) forme une séquence d'édition de la première définition, et celui des signaux (i), (l) et (n) de la deuxième définition. L'enchaînement des signaux (c) et (e)* forme une séquence de discussion sur la première table, et celui des signaux (d) et (e)* sur la deuxième table.

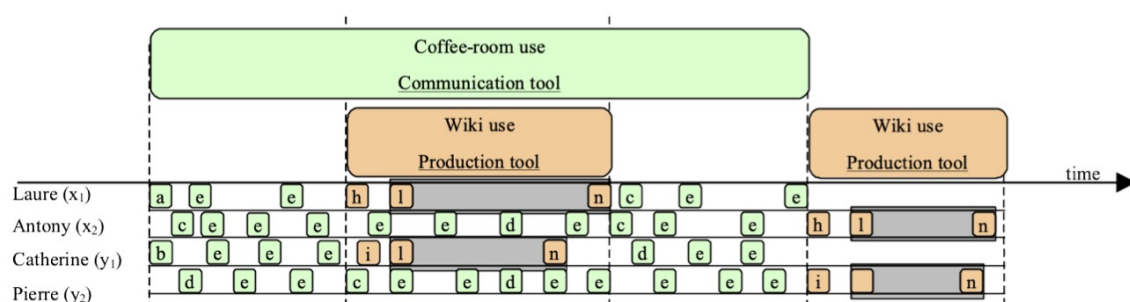


FIGURE 2.17 – Observation avec l'instrumentation

2.5.6 L'analyse des traces

Lorsque des signaux répondent aux conditions des règles, celles-ci s'appliquent et construisent une séquence de niveau supérieur à celui de ses composants. Par extrapolation du mécanisme, les séquences de niveau n peuvent être créées à partir des signaux et des séquences précédemment créées de niveau $n-1$ (Talbot and Courtin 2008).

Quelle que soit la technique utilisée, nous observons que certaines données ne peuvent pas être fournies par le système lui-même au moment où les actions commencent, comme par exemple, la communication entre deux participants sur un sujet donné. Ces informations significatives ne sont pas fournies par un événement de type « ouvrir un outil de communication », car le deuxième participant ne l'a pas encore rejoint. Par conséquent, notre système d'analyse des traces génère explicitement des traces de haut niveau d'abstraction avec une structure de faible granularité, c'est-à-dire des signaux (sur fond coloré sur la Figure 2.18). La génération de ces signaux est possible en mode asynchrone pour une analyse a posteriori. Dans ce cas, nous identifions l'analyseur comme étant la source du signal (comme pour les séquences), et non les outils eux-mêmes. Ces nouveaux signaux s'ajoutent aux signaux collectés par l'instrumentation de la S.C.A.I. (sur fond blanc), qui correspondent au niveau de traces dites brutes ou MTraces de premier niveau dans la colonne de gauche intitulée « Row trace level ». En bref, la technique d'instrumentation permet d'enrichir les traces produites par le système, avec des signaux et des séquences.

Les gabarits d'analyse de cette expérimentation, utilisés pour la construction des séquences dans la colonne du milieu intitulée « Interaction level », sont décrits dans la partie 2.3 page 41.

Nous constatons que le formalisme de construction des gabarits permet de changer de niveau de granularité (e.g. séquences de niveaux 1 et 2) sans changer de niveau d'abstraction (e.g. dans le niveau d'interprétation « Interaction level »).

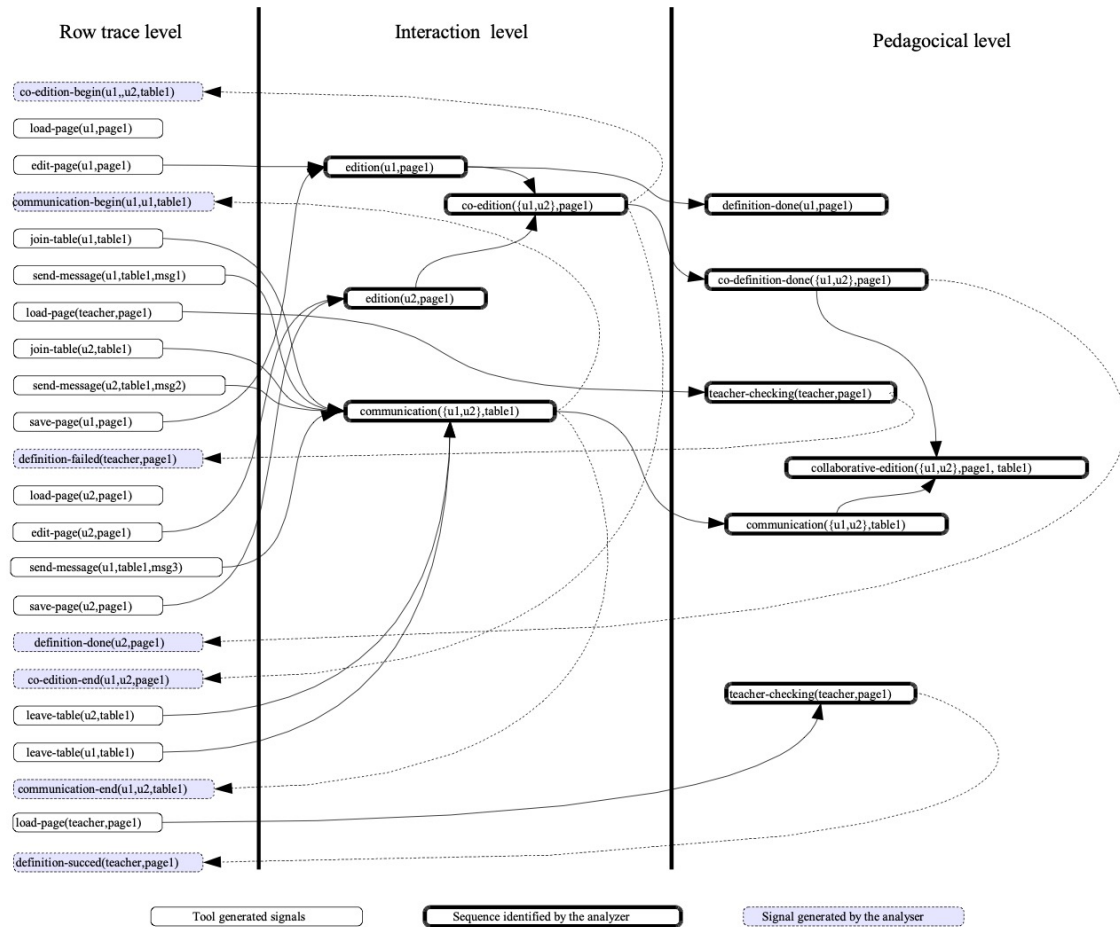


FIGURE 2.18 – Reconnaissance de séquences et génération de signaux

2.5.7 Les résultats de l'expérimentation

Pour cette expérimentation, nous avons fixé un certain nombre d'objectifs que nous rappelons ci-dessous :

- s'assurer de l'expressivité des modèles du système CARTE ;
- tester le module de régulation de l'activité ;
- mesurer les effets de la régulation ;
- automatiser la production d'indicateurs ;
- enrichir les fonctionnalités des applications ;
- simplifier les activités en déclenchant des actions ;
- fournir un contrôle social au sein de l'activité.

Les premiers résultats de l'expérimentation confortent l'**expressivité de notre modèle CARTE** pour exprimer les traces atomiques (les signaux) et composites (les séquences) issues des actions menées à travers la S.C.A.I. L'architecture et les choix d'implémentation de la station d'observation permettent de respecter le principe d'indépendance des outils de la S.C.A.I., qui reprennent

leur fonctionnement de base quand les instrumentations de collecte et de rétroaction sont déconnectées. Cette expérimentation montre que notre formalisme permet non seulement d'exprimer des modèles d'utilisation de haut niveau d'abstraction intra-outils, mais qu'il permet également de définir des modèles d'utilisation de plus haut niveau pour des activités inter-outils (e.g. pour l'activité de « co-rédaction d'une page d'éditeur de texte avec communication dans un *chat* »). La structuration des traces composites (les séquences) à partir des traces atomiques (les signaux) et composites de niveaux inférieurs, permet d'explorer les actions réalisées pour l'activité et donc d'observer son déroulement pour évaluer, par exemple, la stratégie du binôme ou les compétences des membres (cf. travaux de la partie 4 page 99).

L'expérimentation présentée dans ce chapitre succède à une expérimentation d'une S.C.A.I. avec les mêmes outils, mais sans le module de rétroaction (Courtin 2010). Cette dernière expérimentation révèle la capacité du modèle d'utilisation à spécifier les actions de régulation de l'activité en cours lors d'une S.C.A.I. Nous comparons ainsi deux situations similaires, avec et sans le mécanisme de rétroaction, afin d'évaluer les avantages de la régulation dans les activités d'apprentissage collaboratif. Nous énumérons brièvement ci-dessous certaines limites observées lors de la première expérimentation sans la régulation :

- le manque d'information sur l'activité au sein du binôme (e.g. un étudiant n'est pas averti de l'acceptation d'une définition envoyée par son partenaire) ;
- la difficulté à identifier les étudiants qui ont besoin d'aide ;
- la difficulté à faire respecter les règles (e.g. pas de communication entre les binômes) et à prendre en compte les caractéristiques du contexte pédagogique ;
- la répétition de certaines tâches propres aux outils (e.g. dans le « Jibiki », la validation de la définition par l'enseignant est toujours suivie d'une introduction dans le corpus ; deux actions sont nécessaires) ;
- le manque de synchronisation des actions effectuées dans différents outils (e.g. gestion des groupes par l'enseignant dans le « Jibiki » sans répercution de la configuration dans le « CoffeeRoom »).

Avec la régulation, les objectifs pédagogiques fixés par l'enseignant sont évaluables à partir des outils d'observation des traces, modulo des améliorations sur la présentation des informations devant être moins technique et plus *friendly*. Pour cette expérimentation, la collaboration entre les participants au sein d'un binôme fait partie des objectifs pédagogiques définis par l'enseignant. Pour autant, une étude sur ordinateurs ne permet pas à elle seule d'évaluer les effets de la régulation sur la qualité de l'apprentissage. Afin de présenter les résultats de la régulation, nous proposons de synthétiser le mécanisme de rétroaction et de présenter **les effets de la régulation** sur l'activité d'apprentissage. Par exemple, nous avons observé des nouveaux comportements chez les étudiants, comme la discussion après la validation d'un mot afin d'en comprendre la bonne définition. En d'autres termes, ces nouvelles fonctionnalités virtuelles peuvent accroître la collaboration au sein des binômes d'étudiants. Cette observation renforce l'idée que l'enrichissement de l'interactivité, sur la base d'informations émanant du contexte (activité en cours, difficultés rencontrées, etc.), donne lieu à de nouveaux comportements.

Pour effectuer une rétroaction sur les outils, l'outil d'édition des règles présente une liste d'actions à effectuer dans la partie inférieure (cf. Figure 2.10). Ces actions sont déclenchées après que la séquence a été reconnue et ajoutée au S.B.T. Le fait que les rétroactions produisent des traces utiles pour reconnaître d'autres activités, renforce l'idée que les séquences correspondantes doivent être également stockées dans le S.B.T. Dans notre système, les actions sur le « Jibiki » lui sont envoyées sous forme de requête H.T.TP, alors que celles dédiées au « CoffeeRoom » ou à l'outil d'« *awareness* » sont réalisées avec l'appel de méthodes à distance (R.M.I. en langage Java).

Du point de vue de l'architecture du système, il n'y a pas de lien direct entre les outils, toutes les actions sont déclenchées à partir des MTraces qui sont produites par la station d'observation. Un avantage majeur de ce niveau d'abstraction qu'est le modèle d'utilisation, est qu'il préserve l'indépendance des outils, qui sont donc interchangeables. Par exemple, l'action qui consiste à renvoyer un participant sur sa table dans le « CoffeeRoom », est associée à l'événement « *join* » avec deux paramètres : le nom de l'espace de discussion (ici, la table) et l'identifiant de l'utilisateur. Avec un autre outil de communication dont la structure comporte des espaces de communication portant des noms différents (e.g. des salles), la règle serait alors la même, ce qui nous dispenserait de modifier le modèle d'utilisation configuré par l'observateur. On notera également la pertinence de passer par ce niveau d'abstraction, qui place la trace au cœur du système et qui garantit ensuite la présence dans la trace de toutes les étapes ayant conduit à une rétroaction, cette dernière générant elle-même une nouvelle trace. Par exemple, lorsqu'un participant tente de changer de table (événement « *join* » avec un autre nom de table), la trace produite par cette tentative déclenche une règle qui exécute une action sur le « CoffeeRoom » (« *join* » sa propre table) et une autre sur l'outil d'« *awareness* » (notification à l'enseignant et à lui-même), ce qui génère deux nouvelles traces.

Du point de vue fonctionnel, le modèle d'utilisation intégrant le mécanisme de rétroaction implique une instrumentation « entrante » pour les outils concernés, qui complète l'instrumentation « sortante » qui assure la connexion entre les « sources de traces » et la première trace du module de collecte du système CARTE. L'introduction de la rétroaction augmente le couplage entre les différents outils logiciels, ce qui a pour conséquence d'**enrichir leurs fonctionnalités**. Par exemple, le fait d'associer l'outil d'« *awareness* » au « Jibiki », qui a été initialement conçu pour faire de l'édition collective avec un couplage très faible, nous permet d'avertir le binôme de la validation de la définition et de visualiser la bonne version. Nous avons **testé le module de régulation** et constaté l'efficacité de l'outil d'analyse pour une utilisation en temps réel pendant les séances d'apprentissage. Les règles utilisées dans cette expérience permettent :

- d'avoir un retour d'information (*feedback*) sur les activités des autres participants (e.g. un étudiant est averti dans l'outil d'« *awareness* » lorsque son partenaire a terminé une définition) ;
- de changer la façon dont les outils sont perçus en raison de l'interactivité augmentée ;
- **d'automatiser la production d'indicateurs** et de **simplifier les activités** en déclenchant certaines tâches, notamment pour l'enseignant (e.g. dans le « Jibiki », en une seule action, la validation de la définition par l'enseignant est automatiquement suivie d'une introduction dans le corpus) ;
- **renforcer le contrôle social** au niveau du groupe (e.g. en limitant la possibilité de communication entre les étudiants appartenant à des groupes différents).

L'expérimentation confirme que le système CARTE peut soutenir la régulation de base dans une S.C.A.I. Cependant, nous avons remarqué certaines limites dans la version actuelle du prototype pour exprimer des actions spécifiques de régulation telles que : avertir l'enseignant lorsqu'un binôme d'étudiants a complété *n* définitions dans une période de temps donnée. En effet, l'itération et la dimension temporelle (autre que la succession) ne sont pas encore supportées par l'éditeur de règles. En fait, nous contournons cette limitation de l'itération en répétant *n* fois les MTrace (signal ou séquence) correspondantes. Des travaux sont nécessaires pour introduire ces deux dimensions temporelle et itérative directement dans la spécification des règles, à l'aide de nouveaux opérateurs à définir.

2.6 Conclusion

Le travail présenté dans cette partie s'inscrit dans la thématique de l'observation des activités humaines pour traiter de la problématique de l'explicitation de cette activité à partir des traces numériques ou numérisées. Nous avons tenté d'adresser la question de recherche sur la représentation des traces d'activité permettant à un système de les analyser en tenant compte de l'hétérogénéité de leur format. Nous avons donc proposé une station d'observation qui opérationnalise notre modèle de système à base de traces, sur la base du format générique CARTE (*Collection, activity Analysis and Regulation based on Traces Enriched*). L'intérêt de ce travail tient à la fois à la capacité de notre système à apporter des solutions au problème d'hétérogénéité des formats, mais aussi à l'interopérabilité résultante des systèmes instrumentés à l'aide de ce format et à l'enrichissement de leurs fonctionnalités grâce au mécanisme de rétroaction de la station d'observation. Des expérimentations ont été menées dans le cadre de la mise en place d'un banc d'essai dédié aux chercheurs en informatique, en sciences sociales, en didactique, etc. pour le projet « **Personnalisation des E.I.A.H.** » du pôle de recherche Cluster I.S.L.E. (Informatique, Signal, Logiciel embarqué) de la Région Rhône-Alpes, 2008-2011 (cf. travaux de la partie 3 page 67).

Bien que les résultats obtenus aient répondu au-delà de la question de recherche, avec notamment la régulation et les *add-on* dans les outils instrumentés, nous faisons l'hypothèse que la proposition d'un format générique ne nous prémunit pas d'une éventuelle perte d'information, liée à la structure sémantique non reprise du format initial, lors de la transformation de celui-ci vers notre format générique. Par ailleurs, ce format est perçu comme étant un n^{ème} format pivot, sachant que des formats spécifiques sont proposés dans certains domaines d'application précisément pour la richesse de leurs spécificités. Dans la partie 3, nous présentons une nouvelle orientation de nos travaux, qui reste compatible avec l'utilisation du format générique pour les situations particulières (e.g. en l'absence de format d'export ou pour des traces « fabriquées »).

3

Corpus de traces

L'hétérogénéité des formats de traces d'interaction fait écho à la multitude des domaines qui les utilisent et donc des contextes afférents. Pour relever le défi du partage des traces contextualisées et des outils d'analyse spécifiques, en respectant la richesse des formats existants, nous prenons la voie de la conciliation des egos. Notre approche se singularise en tant qu'elle considère les données et les outils dans leur format natif dans des corpus, allant jusqu'à les rendre interopérables.

Sommaire

3.1	Introduction	68
3.2	Interopérabilité avec CARTE	69
3.2.1	A.P.I. <i>Collect</i>	70
3.2.2	A.P.I. <i>Query/Notify</i>	71
3.3	Corpus de traces d'activité	72
3.3.1	Contexte de recherche	74
3.3.2	Partage et exploitation des corpus de traces	75
3.3.3	Approche PROXYMA	77
3.3.4	Plate-forme BEATCORP	87
3.4	Expérimentation d'une plate-forme de <i>benchmarking</i>	90
3.4.1	Les objectifs de l'expérimentation	90
3.4.2	Le contexte de l'expérimentation	90
3.4.3	Le déroulement des opérations	90
3.4.4	Les résultats de l'expérimentation	91
3.5	Conclusion	97

3.1 Introduction

Dans la partie 2 page 33, nous présentons le traitement des traces dans un système indépendant, appelé « Station d’observation », conçu à partir d’un modèle théorique d’observation. Nos choix de conception sont motivés par la volonté de s’adapter à tout type d’environnement informatique pour l’apprentissage humain (E.I.A.H.). Pour autant, les communautés de recherche utilisent déjà un certain nombre d’outils logiciels d’analyse des traces d’activité dédiés à leur domaine, voire spécifiques à leurs données. Par ailleurs, ces communautés entretiennent la culture du partage des données motivée, entre autres, par la nécessité d’évaluer des modèles théoriques avec d’autres jeux de données, et pour limiter les coûts humains engendrés par de la mise en place de nouvelles expérimentations. Nous mesurons donc le paradoxe entre les pratiques actuelles et les besoins des chercheuses et des chercheurs, et donc la nécessité de concevoir des solutions technologiques qui ouvrent des perspectives favorables pour leur pratique professionnelle.

Nous avons identifié trois familles d’outils logiciels pour réaliser des analyses de traces d’activité (Courtin and Talbot 2009).

La première famille correspond aux **plates-formes collaboratives** qui contiennent les outils de production, de communication, de coopération et de régulation selon les trois dimensions du modèle du trèfle (Ellis and Wainer 1994) augmenté de la régulation (Chabert et al. 2006) (cf. partie 2.2.1). Ces outils logiciels produisent intrinsèquement des traces qui, selon la situation, peuvent être directement exploitées pour des observations généralement en mode asynchrone. C’est le cas de la plate-forme d’apprentissage collaboratif assisté DREW © (*Dialogical Reasoning Educational Web*) (Corbel, Girardot, and Lund 2006) qui exploite ses propres traces. Elle est conçue pour analyser des interactions humaines avec un système informatique en rejouant les activités qui ont été réalisées à l’aide d’outils logiciels ou décrites à partir d’autres sources (e.g. la transcription de vidéo). Elle permet à la chercheuse et au chercheur de gérer les traces d’interactions médiatisées par ordinateur découlant de l’utilisation des outils internes (e.g. le *chat* de DREW © ou l’éditeur de texte partagé de DREW ©). En d’autres termes, DREW © peut lire ses propres traces (séquences d’événements au format X.M.L.¹) et reproduire l’activité, au moyen d’un rejoueur interne (mode *replay*), dans les outils logiciels (e.g. l’éditeur de texte partagé de DREW ©) qui étaient initialement utilisés (Jaillon, Corbel, and Girardot 2002).

La deuxième famille correspond aux **environnements d’analyse spécifiques** qui comprennent des outils logiciels destinés à assister les observateurs dans la gestion, la synchronisation, la visualisation et l’analyse des traces d’activité. Dans ces environnements, la plupart des actions susmentionnées sont effectuées manuellement sur les traces. Le système TATIANA © (*Trace Analysis Tool for Interaction ANALysts*) (Dyke, Lund, and Girardot 2009) d’analyse des interactions humaines avec un système informatique correspond à la description ci-dessus. Ce système ayant été conçu par la même équipe de recherche, le lecteur de DREW © peut être synchronisé et contrôlé par un lecteur de l’outil d’analyse TATIANA ©. Ainsi, les chercheuses et les chercheurs peuvent ajouter des indicateurs dans le rejoueur de DREW ©, par exemple une couleur pour chaque participant afin de différencier les contributions. Ce couplage entre un environnement d’analyse et une plate-forme collaborative permet la création manuelle d’indicateurs épistémiques (traces de haut niveau) à partir de l’observation d’actions (traces de bas niveau). Si le système DREW © répond de manière satisfaisante aux attentes des chercheuses et des chercheurs en sciences humaines et sociales, les analyses sont en revanche restreintes aux outils spécifiques de la plate-forme.

La troisième famille, qui est au cœur de notre étude, concerne les **stations d’observation génériques**, c’est-à-dire les environnements qui répondent à la spécification du modèle théorique

1. *Extensible Markup Language*

d'observation que nous appelons un S.B.T. (Système à Base de Traces) (Settouti et al. 2009b) (Courtin 2010). L'architecture que nous proposons pour la station d'observation CARTE tend à se soustraire des contraintes d'interopérabilité des outils logiciels, mais se heurte à la contrainte d'alignement des formats hétérogènes de traces produites par ces outils. Malgré la généralité de notre format de trace CARTE, nous n'avons pas la garantie que la transformation d'un format de traces issues d'un outil propriétaire vers ce format générique ne perde de la richesse sémantique des traces d'origine.

Nous faisons donc le constat que le potentiel d'une station d'observation des traces d'activité est insuffisant pour assurer des analyses avec tout type de traces, et pour répondre aux besoins des communautés de recherche de partage des outils d'analyses et des résultats d'expérimentation. L'idée émergeant de notre analyse de la situation est la construction de corpus de traces d'activité, permettant le partage des traces et des outils d'analyse. Après avoir étudiées les solutions existantes au niveau international, nous présentons notre travail qui a été réalisé dans le cadre de la thèse d'Hajer Chebil (2013) au laboratoire « Systèmes Communicants » (SysCom) de l'université de Savoie², en co-tutelle avec l'École Nationale Supérieure des Mines de Saint-Étienne. Ce travail est effectué dans le cadre du projet « Personnalisation des E.I.A.H. » du pôle de recherche Cluster I.S.L.E. (Informatique, Signal, Logiciel embarqué) de la Région Rhône-Alpes³.

3.2 Interopérabilité avec CARTE

Avant de poser les fondations d'un système de gestion de corpus de traces d'activité, nous explorons la capacité de notre station d'observation, présentée dans la partie 2, à interagir avec des outils d'analyse spécifiques. Notre étude porte sur l'environnement TATIANA © en tant qu'il est utilisé pour assister des chercheuses et des chercheurs dans l'analyse des traces issues de contextes complexes, car multimédias, multimodaux (synchronisation de plusieurs sources de traces) et multi-utilisateurs (Dyke, Lund, and Girardot 2009). Comme expliqué ci-dessus, les chercheuses et les chercheurs peuvent utiliser TATIANA © pour créer des « rejouables » et des analyses à l'aide d'un langage spécifique pour définir des filtres/règles (Dyke et al. 2007). Un « rejouable » peut être affiché sous forme d'un tableau avec plusieurs colonnes telles que Heure / Utilisateur / Message / Outil, ou sous forme graphique en utilisant des feuilles de style. Les filtres, basés sur des scripts XQuery⁴, sont utilisés pour convertir les données dans un format X.M.L.⁵ (avec la description de leur structure), puis dans un format compréhensible par TATIANA © (« format d'affichage » ou format pivot ou séquence d'événements). Le « rejoueur » de TATIANA © lit un fichier de traces et interprète chaque événement comme l'aurait fait l'outil utilisé pendant l'observation. Les traces sont rejouées au sens « analyse » du terme, mais aussi au sens « action » du terme, c'est-à-dire dans les outils de l'environnement collaboratif (e.g. l'outil de *chat* de DREW ©). TATIANA © est conçu pour interagir avec les relecteurs de plates-formes collaboratives. Un script d'import permet de sélectionner une ou plusieurs sources dans un fichier d'entrée (e.g. une discussion dans un *chat*, une transcription vidéo). Il convient de noter que les tâches relatives à l'analyse (e.g. la catégorisation, l'annotation, l'argumentation), réalisées manuellement, peuvent être compliquées et répétitives. L'automatisation de certaines de ces tâches dévolues aux chercheuses et aux chercheurs, nécessiterait une interopérabilité avec une station d'observation pour effectuer des trans-

2. ancienne dénomination de l'université Savoie Mont Blanc

3. ancien nom de la région

4. W3C. XQuery 1.0, An X.M.L., <http://www.w3.org/TR/xquery/>

5. *Extensible Markup Language*

formations de traces grâce à des modèles d'utilisation prédéfinis des outils. Nous présenterons ci-dessous l'A.P.I.⁶ qui permet ce couplage avec la station d'observation CARTE.

La station d'observation CARTE est un environnement dédié à la collecte de traces issues de sources diverses, à leur structuration selon un format générique, à leur transformation selon des objectifs d'analyse prédéfinis et à leur réinjection vers des outils d'analyse ou vers les outils dont elles sont issues (principe de rétroaction de CARTE (Courtin 2010)) pour permettre la régulation de l'activité. Dans notre contexte d'étude, les chercheuses et les chercheurs (i.e. les observatrices et les observateurs) peuvent utiliser la station d'observation CARTE pour suivre les interactions dont le traitement doit viser à faciliter leurs propres tâches dans leur système d'analyse dédié. CARTE permet la collecte en temps réel de traces afin de faire des analyses, soit en parallèle avec différents modèles d'utilisation, soit successives (donc cumulatives) afin d'augmenter les niveaux d'abstraction des traces résultantes. Ainsi, certaines tâches d'analyse élémentaires répétitives qui sont effectuées avec TATIANA © peuvent être automatisées avec CARTE, ainsi que certaines tâches plus complexes issues de l'expérience des chercheuses et des chercheurs. De plus, nous observons que le rejeu de TATIANA © permet l'intégration manuelle de données d'« *awareness* » afin d'améliorer la compréhension des traces. Par le biais du mécanisme de rétroaction, le système CARTE peut alimenter automatiquement des outils d'« *awareness* » externes, qui pourraient être ceux de TATIANA ©. Par conséquent, l'analyste lui-même doit d'abord préciser, dans la station d'observation, à l'aide d'un modèle d'utilisation des outils basé sur des règles, les informations d'« *awareness* » qu'il souhaite recevoir pendant l'analyse avec TATIANA ©. L'analyse des traces peut être effectuée à différents niveaux d'abstraction : fichiers de *logs* ou données d'instrumentation dans les plates-formes collaboratives ; indicateurs épistémiques élaborés par les chercheurs eux-mêmes dans les environnements d'analyse. Pour déterminer les niveaux d'analyse selon (Drulhe 2008), les sociologues considèrent trois critères : le contexte temporel engagé dans l'observation, le volume et la nature des unités d'observation. Il y a ensuite la question de la sélection du matériel d'analyse et des sources dont il provient. Certains types d'analyse, qui consistent à monter dans l'échelle d'abstraction, impliquent des traces d'interaction issues de plate-formes collaboratives, des traces calculées (ou créées) au moyen de modèles d'utilisation dans une station d'observation et des traces enrichies par les techniques de catégorisation, d'argumentation et d'annotation dans les environnements d'analyse.

3.2.1 A.P.I. Collect

D'un point de vue technique, notre A.P.I.⁷ de collecte vise à simplifier le processus de collecte des traces. Cette A.P.I. fournit aux développeurs d'applications des méthodes destinées à les aider lors de l'instrumentation des outils dans la phase d'implémentation ou à les adapter par la suite. Pour notre étude, elle fournit des méthodes spécifiques pour importer et exporter des traces (e.g. à partir du format X.M.L.⁸ de DREW ©) en fonction des objectifs d'interopérabilité (i.e. collecte, exploitation, interrogation, notification) entre la station d'observation et les outils externes. Dans notre prototype CARTE, le service « collecteur » est utilisé pour donner accès à l'A.P.I. de collecte. Deux méthodes *sendSignal* et *sendSequence* sont définies et permettent d'envoyer des signaux ou des séquences à la station d'observation. Comme nous utilisons un serveur J2EE⁹ pour notre système, les A.P.I. sont accessibles soit en tant que services WEB, soit en tant qu'objets java R.M.I.¹⁰

6. *Application Programming Interface*

7. *Application Programming Interface*

8. *Extensible Markup Language*

9. Jakarta EE est une spécification pour la plate-forme Java d'Oracle, destinée aux applications d'entreprise

10. *Remote Method Invocation* est une interface de programmation (A.P.I.) pour le langage Java qui permet d'appeler des méthodes distantes

ou Corba¹¹ (selon la configuration du serveur J2EE). Afin de simplifier l'utilisation de notre système, nous fournissons différentes versions de ces méthodes selon le principe du polymorphisme. Dans les plus simples, un seul paramètre doit être donné : le signal ou la séquence à enregistrer. Dans une deuxième version des méthodes, les signaux ou les séquences peuvent être décrits en utilisant leurs parties internes : horodatage, source, outil, évènement, description, paramètres, etc. Enfin, une version spéciale utilisée par l'analyseur comprend un paramètre supplémentaire, qui permet d'associer un identifiant de session aux traces recueillies et générées au cours de la même analyse. Cet identifiant peut ensuite être réutilisé, par exemple, pour visualiser les résultats d'une analyse donnée ou pour la rejouer.

3.2.2 A.P.I. *Query/Notify*

En utilisant la même approche que celle employée pour la collecte, nous sommes amenés à définir l'A.P.I. d'exploitation des traces afin de normaliser l'utilisation et l'interrogation des traces. Nous identifions deux types de besoins :

1. les besoins en matière d'interrogation de la base de traces ;
2. les besoins de notification - certains outils (e.g. les analyseurs en ligne) doivent être notifiés lorsque de nouvelles traces (signaux ou séquences) sont produites et enregistrées dans la base de traces.

Nous définissons donc des A.P.I. pour chacun de ces deux usages : API d'interrogation et API de notification.

L'A.P.I. *Query*, utilisée pour effectuer des requêtes sur la base de traces, est « *pull-oriented* », i.e. que les outils l'utilisent pour demander à la station d'observation de rechercher des éléments de trace particuliers (signaux ou séquences) et de les renvoyer. Comme pour l'A.P.I. de collecte, celle d'interrogation reste très simple avec un service qui définit seulement les deux méthodes *searchSignal* et *searchSequence*. Sur la base du polymorphisme autorisé par le langage Java, nous définissons deux méthodes d'interrogation :

1. avec un gabarit de signal ou de séquence partiellement instancié ;
2. avec un modèle d'utilisation.

Dans la première méthode, un gabarit de signal ou de séquence partiellement instancié est utilisé pour la recherche : le type et les paramètres des signaux ou des séquences peuvent être définis ou laissés libres. La station d'observation renvoie une liste de signaux ou de séquences qui correspondent au gabarit. Par exemple, nous pouvons rechercher tous les signaux générés par un salon de discussion dans lequel le premier paramètre est égal à « Albert ».

Dans la seconde méthode, les gabarits ne sont pas appliqués directement aux traces, mais sont définis et interprétés à l'aide d'un modèle d'utilisation défini au préalable. L'interprétation des traces est ensuite utilisée pour déterminer celles qui doivent être renvoyées. Par exemple, nous pourrions rechercher toutes les séquences où un paramètre peut être interprété en utilisant un modèle d'utilisation particulier, comme un nom d'utilisateur qui a pour valeur « Albert ». Dans cette dernière forme, le format du gabarit est très similaire à celui utilisé dans l'analyseur.

11. *Common Object Request Broker Architecture* est une architecture logicielle pour le développement de composants

L’A.P.I. *Notify*, utilisée pour effectuer des notifications sur les outils logiciels, est « *push-oriented* ». Les outils (analyseurs, outils de visualisation, ...) agissent comme des consommateurs de traces et sont enregistrés dans la station d’observation, qui agit comme un producteur de traces et notifie les outils lorsque de nouveaux signaux ou séquences arrivent. Le service de notification de la station d’observation définit deux méthodes, *addSignalListener* et *addSequenceListener*, qui sont utilisées respectivement pour enregistrer les consommateurs de signaux et de séquences. Du côté de l’environnement d’analyse, les outils doivent implémenter les interfaces Java¹² *SignalListener* ou *SequenceListener* qui définissent respectivement les méthodes *sendSignal* et *sendSequence* que la station d’observation utilise pour envoyer les signaux et les séquences qu’elle reçoit. En ce qui concerne l’A.P.I. d’interrogation, les *listeners* (i.e. les consommateurs de traces) ont la possibilité de spécifier, à l’aide d’un gabarit, le type de traces dont ils veulent être notifiés. Comme nous l’avons vu précédemment, ce gabarit doit être utilisé pour faire correspondre les traces directement ou par le biais d’un modèle d’utilisation particulier.

Ces A.P.I. simplifient grandement la connexion d’outils tiers externes, tels que les traqueurs d’activité ou les afficheurs de traces, et améliorent l’extensibilité du système. Des filtres plus sophistiqués peuvent être mis en œuvre à l’aide des filtres simples décrits ci-dessus et être utilisés pour améliorer progressivement la complexité des filtres et des demandes que les outils pourraient utiliser pour interagir avec la station d’observation.

3.3 Corpus de traces d’activité

Les traces d’activité sont souvent analysées en ligne et/ou hors ligne pour adapter les supports d’apprentissage aux profils des apprenants (Settoui et al. 2011) (Glahn, Specht, and Koper 2007), pour réorganiser (Choquet and Iksal 2007) ou évaluer (Nogry, Jean-Daubias, and Guin-Duclosson 2012) les environnements d’apprentissage, pour créer et exploiter à distance des situation actives d’apprentissage (Guéraud et al. 2004), pour évaluer le travail des élèves (Guin, Jean-Daubias, and Nogry 2007) ou pour analyser les processus d’apprentissage (Lallé, Luengo, and Guin 2013). Le modèle théorique CARTE, de système à base de traces qui conduit à la spécification de la station d’observation, constitue une première étape pour la modélisation des systèmes à base de corpus de traces de référence, ouverts à la communauté internationale des chercheurs. L’objectif de tels corpus se trouve à l’intersection de nombreuses problématiques centrées sur un objet de recherche, appelé trace, qui provient de diverses sources et peut prendre différentes formes, tout en respectant une spécification permettant l’interopérabilité avec d’autres traces et la prise en compte de toutes les spécificités inhérentes à son utilisation (propriété, éthique, classification, etc.).

Construire une expérimentation dans une situation d’apprentissage authentique, est une tâche complexe et longue qui implique généralement différentes personnes (apprenants, tuteurs, développeurs, etc.). Cette complexité motive le partage de corpus de traces entre chercheurs de différentes disciplines pour accéder aux données relatives aux expérimentations d’apprentissage précédentes (Chanier and Ciekanski 2010) (Stamper et al. 2010). En outre, l’accès aux données de recherche permet de reproduire les résultats publiés (King 2007), voire de comparer les résultats relatifs à différentes analyses sur les mêmes données (Reffay and Betbeder 2009), tout en augmentant la visibilité des chercheuses et des chercheurs.

12. une interface Java est une classe complètement abstraite, sans attribut, qui sert à regrouper des signatures de méthodes apparentées, i.e. ayant un corps vide

Les problématiques associées au partage de corpus de traces d'activité et d'outils d'analyse dépassent celles des systèmes à base de traces, en tant qu'elles adressent, entre autres, des questions relatives à l'hétérogénéité structurelle des traces et aux divers formats d'entrée/sortie des outils d'analyse. Bien que la station d'observation propose un format générique de trace, nous n'excluons pas que la transformation au format CARTE de traces déjà structurées dans un format natif en vue d'être analysées, puisse générer une perte d'information. En effet, la conversion de la trace peut impliquer une perte sémantique si certaines données n'existent pas dans le format cible. Comme le partage implique, entre autres, la prise en compte d'utilisations diverses, et non planifiées à l'avance, nous considérons que le critère de préservation de la richesse sémantique des traces collectées est très important. Les projets IA JEIRP (Martínez, De La Fuente, and Dimitriadis 2003), PSLC Datashop (Stamper et al. 2010), REDiM (Choquet and Iksal 2007), MULCE (Chancier and Ciekanski 2010), Undertracks (Bouhineau et al. 2013), et CAM-CIM (Wolpers et al. 2007), prennent en compte l'hétérogénéité des traces en introduisant plusieurs formats de représentation des données, pour augmenter leur flexibilité. Nous constatons que cette flexibilité ne peut atteindre l'exhaustivité au regard du nombre important d'outils différents produisant leur propre format de trace. Les projets PSLC Datashop et MULCE proposent explicitement de partager les analyses effectuées sur des corpus de traces.

Pour éviter un nième format pivot et la perte potentielle d'information dans les traces, nous posons donc comme contrainte de modélisation d'un système à base de corpus de maintenir les traces des diverses sources dans leur format natif. En d'autres termes, nous ne proposons pas un nouveau format de représentation des données, qui doit être suivi par les chercheuses et les chercheurs, pour partager leurs corpus et par les concepteurs d'outils d'analyse en tant que modèle d'entrée de données. Notre idée principale est de permettre aux chercheuses et aux chercheurs de représenter leurs données collectées et les données d'entrée de leurs outils d'analyse en utilisant les modèles et les formats qui répondent à leurs besoins. Nous faisons de cette contrainte l'originalité de notre modèle au vu des modèles existants dans la littérature scientifique.

L'étude de l'état de l'art sur les corpus de traces et le partage des outils d'analyse nous a montré le faible nombre de projets existants traitant de cette question. Les projets étudiés peuvent être divisés en trois groupes en fonction du problème de partage considéré :

1. ceux qui traitent de la question du partage des corpus ;
2. ceux qui se concentrent sur le partage des outils d'analyse ;
3. ceux qui considèrent les deux types de partage.

Les approches existantes (Giguet et al. 2009) (Reffay and Betbeder 2009) (Koedinger et al. 2008) (Bouhineau et al. 2013) (Butoianu et al. 2010) sont basées sur une idée générale commune qui consiste à proposer une représentation spécifique (modèle et format sémantique) qui doit être obligatoirement suivie par les chercheuses et les chercheurs souhaitant partager leurs données. En ce qui concerne le partage des outils d'analyse, cette représentation imposée des données est utilisée pour définir les entrées attendues par les outils d'analyse. En d'autres termes, pour utiliser les services fournis par les plates-formes de partage existantes, les utilisatrices et les utilisateurs doivent effectuer un travail préalable de conversion pour formater leurs données selon la représentation imposée. Les projets existants ont choisi cette approche en raison de l'absence d'une représentation standard des traces d'apprentissage. En effet, l'hétérogénéité liée aux environnements d'apprentissage, aux scénarios pédagogiques, aux disciplines et aux besoins d'analyse rend très difficile la définition d'une représentation standard.

Sur la base de ce constat, nous avons travaillé sur une approche différente qui évite d'imposer une énième représentation qui ne peut convenir à tous les besoins de représentation de données. À défaut de convertir toutes leurs données dans un format pivot, les chercheuses et les chercheurs

doivent fournir un ensemble de métadonnées permettant la description des ressources et leurs interrogations futures. Dans le modèle que nous proposons, lorsqu'une utilisatrice ou un utilisateur souhaite interroger le contenu d'un corpus afin d'effectuer une analyse, un composant ¹³ est utilisé pour jouer le rôle d'intermédiaire entre la chercheuse ou le chercheur, le corpus et l'outil d'analyse.

3.3.1 Contexte de recherche

Ce travail fait partie d'un projet ¹⁴ impliquant différentes équipes de recherche dont l'objectif commun est la personnalisation des E.I.A.H. et dont les objectifs respectifs sont :

- l'adaptation des scénarios pédagogiques (Ferraris et al. 2009);
- l'étude et l'amélioration de la communication entre les participants (May, George, and Pré-vot 2011);
- l'adaptation du comportement des systèmes de tutorat intelligents (S.T.I.) en fonction des profils des apprenants en diagnostiquant leurs modèles de connaissances (Lallé, Luengo, and Guin 2013) (Settouti et al. 2011).

L'objectif de notre travail est de permettre aux chercheuses et aux chercheurs utilisant des E.I.A.H. dans leurs travaux de recherche de partager des corpus de traces d'apprentissage contextualisées et des outils d'analyse pour traiter ces corpus. Parmi les nombreuses questions de recherche possibles, nous en sélectionnons trois que nous présentons ci-dessous :

- Comment créer et partager un modèle de corpus de traces d'apprentissage, afin de permettre des analyses en tenant compte du contexte ?
- Comment définir l'interopérabilité entre les différents modèles de corpus et les formats de traces d'entrée/sortie spécifiques des outils d'analyse, sans imposer une nouvelle représentation des traces comme condition de partage ?
- Comment représenter et intégrer les travaux d'analyse réalisés sur un ou plusieurs corpus, afin qu'ils puissent être réutilisés par d'autres chercheuses ou chercheurs à des fins de vérification, de comparaison ou de nouvelles analyses ?

Nous avons identifié un certain nombre de verrous à traiter afin de répondre à nos questions de recherche :

- l'**hétérogénéité des traces collectées** (traces numériques, enregistrements audio/vidéo), leur format et leur granularité ;
- le **contexte des traces collectées** (Chanier and Ciekanski 2010) pour permettre à d'autres chercheuses ou chercheurs de les réutiliser et de reproduire les résultats d'analyse.
- l'**anonymat des traces** qui peuvent contenir des données sur l'identité des participants (cette considération n'étant pas traitée à ce niveau-là, nous considérons que les chercheuses et les chercheurs fournissent des données déjà anonymisées et que les participants sont avertis que les traces d'activité sont collectées pour des objectifs de recherche exclusivement) ;
- le besoin de décrire **différents types d'analyses** effectuées avec différents environnements d'analyse.

13. décrit dans la partie 3.3.3

14. intitulé « **Personnalisation des E.I.A.H.** » du pôle de recherche Cluster I.S.L.E. (Informatique, Signal, Logiciel embarqué) de la Région Rhône-Alpes

Afin de guider notre conception, nous définissons un certain nombre de critères qui représentent les caractéristiques attendues d'un système à base de corpus de traces d'activité. Le premier critère est la prise en compte de l'**hétérogénéité des traces d'activité**, en préservant la richesse sémantique de leur format initial (adaptation totale) ou en s'adaptant à plusieurs formats (adaptation partielle). Par extrapolation, le deuxième critère est le **partage des corpus de traces**, en incluant leur contexte de production et/ou d'analyse, pour permettre des analyses croisées. Le troisième critère est le **partage des outils d'analyse**, i.e. leur réutilisation pour d'autres corpus que ceux pour lesquels ils ont été conçus et donc l'adaptation des formats d'entrée/sortie. Les trois critères réunis adressent les questions de recherche relatives au partage et à l'interopérabilité des corpus de traces et des outils d'analyse, et sont essentiels aux conditions de reproductibilité des analyses.

Les différents systèmes existants ont un dénominateur commun : ils proposent tous un format de trace spécifique comme solution pour le partage des données et/ou des outils d'analyse. Quelle que soit la solution retenue, elle induit un temps de traitement incompressible soit pour la transformation des traces effectuée dans les systèmes existants, soit pour l'adaptation des traitements en vue d'exploiter les traces séparément ou simultanément dans notre proposition. En d'autres termes, les chercheuses et les chercheurs sont impliqués dans les phases de configuration des traitements pour les analyses. La valeur ajoutée du système que nous proposons tient davantage à la précision des analyses qu'à ses performances en temps d'exécution. Notre proposition tend à améliorer la flexibilité d'un tel système, par le biais d'une approche incrémentale et participative. En attendant qu'une réflexion en UX Design et les développements informatiques afférents soient engagés pour améliorer l'utilisabilité du système, nous accompagnons les chercheuses et les chercheurs dans les phases de configuration.

3.3.2 Partage et exploitation des corpus de traces

La mise à disposition de données pour procéder à des analyses par d'autres chercheuses ou chercheurs, avec des outils différents, implique l'unification ou la transformation des formats afférents.

Nous pouvons citer l'exemple du projet *PSLC Datashop* (Stamper et al. 2010) qui propose une plate-forme offrant (1) un entrepôt de données d'interaction d'apprenants utilisant des *I.T.S.*¹⁵ et (2) un ensemble d'outils d'analyse, de *reporting* et de visualisation permettant d'analyser ces données. Ce projet a proposé un format pour représenter les données d'interaction, qui doit être respecté pour pouvoir utiliser les outils offerts.

Le projet MULCE (Chanier and Ciekanski 2010), dans un objectif de partage de données d'interaction issues de situations d'apprentissage collaboratives, propose l'utilisation des formats standards existants pour la représentation et l'échange des données. Par ailleurs, ce projet met en évidence l'importance de la contextualisation des données d'interaction partagées. En effet, ces données ne sont pas suffisantes pour une chercheuse ou un chercheur n'ayant pas participé à l'expérimentation. Il est indispensable pour une chercheuse ou un chercheur de comprendre le contexte de production des données pour les analyser. Pour ceci, le projet MULCE propose une structure d'un corpus d'apprentissage et d'enseignement dans lequel les interactions médiatisées sont accompagnées par un composant « contexte ». Ce contexte est divisé en deux parties : le scénario pédagogique et le protocole de recherche, ce dernier étant facultatif, il permet d'exprimer les questions de recherche derrière l'expérimentation, ainsi que les constituants et le déroulement du processus de recherche. Il utilise les standards *Dublin Core*¹⁶ et *OLAC*¹⁷ pour la représentation

15. *Intelligent Tutoring Systems*

16. <http://dublincore.org/specifications/dublin-core/dcmi-terms/>

17. *Open Language Archives Community* <http://www.language-archives.org/>

des métadonnées dans le corpus. Il préconise l'utilisation du standard *IMS Learning Design*¹⁸ pour représenter le contexte de l'expérimentation. Le standard *IMS Content Packaging*¹⁹ est utilisé pour emballer et échanger les corpus.

L'exploitation des corpus de traces d'interaction peut prendre des formes très variées et notamment selon qu'il s'agit d'analyse en temps réel ou différé. Dans le premier cas, l'analyse des traces est généralement intégrée à la plate-forme d'apprentissage (e.g. projet Aplusix (Nguyen-Xuan et al. 2002)), ce qui limite l'exploitation à celle-ci. C'est aussi le cas de l'outil *TrAVis* (*Tracking data Analysis and Visualisation*) qui permet le calcul et la visualisation d'indicateurs à partir des traces de communications médiatisées par ordinateur (*chat*, forums, etc.) représentées dans un format propriétaire (May, George, and Prévot 2011). Cet outil permet au tuteur d'analyser les activités des apprenants, et aux apprenants d'avoir des *feedbacks* sur leur activité durant une session d'apprentissage à distance.

Le projet GNOWSIS (Sauermann et al. 2006) s'intéresse à la notion de bureau sémantique (ou *Semantic Desktop*), lequel est constitué de l'ensemble des outils utilisés sur le bureau d'un ordinateur, et permet d'organiser les données et ressources manipulées et leur description à l'aide d'une ontologie partagée afin de faciliter leur gestion et leur recherche. Ce projet utilise le traçage dans une étude statistique de l'activité des utilisatrices ou des utilisateurs pour étudier les concepts qui les intéressent.

U.I.C.O. (*User Interaction Context Ontology*) (Rath, Devaurs, and Lindstaedt 2009) est un modèle de représentation du contexte d'interaction d'une utilisatrice ou d'un utilisateur basé sur une ontologie pour la détection automatique de leurs tâches sur le bureau de leur ordinateur. Ce modèle permet d'améliorer la détection automatique des tâches et doit être intégré au *framework* orienté services « KnowSE » qui vise à offrir à l'utilisatrice ou l'utilisateur des services de gestion des connaissances contextualisés et personnalisés.

Le projet *A.P.O.S.D.L.E.* (*Advanced Process-Oriented Self-Directed Learning Environment*) (Stern et al. 2010) s'intéresse à l'assistance des employés, qui peuvent avoir trois rôles différents : travailleur, apprenant ou expert. Il vise à améliorer leur productivité en appuyant les activités d'apprentissage informel durant le travail de tous les jours et au sein de leur environnement de travail. Ce projet définit trois objectifs principaux : une méthodologie, un *framework* logiciel et une architecture de référence pour les espaces de travail d'apprentissage permettant d'améliorer les processus d'apprentissage pour l'individu et l'organisation. Des outils de suivi de l'activité d'une actrice ou d'un acteur permettent de lui proposer des outils et ressources adaptés à son rôle et besoin.

Dans le cadre de nos recherches, la plupart des analyses sont réalisées en temps différé, sur des traces d'interaction auxquelles il est nécessaire d'ajouter leurs contextes de production et d'observation. Les fonctionnalités de décontextualisation (i.e. extraction de parties d'interactions d'un corpus) et de recontextualisation (i.e. association des parties précitées pour des analyses) sont des notions clés dans les outils dits *C.A.Q.D.A.S.* (*Computer Aided Qualitative Data Analysis Software*) d'analyse de corpus linguistiques. On retrouve ces notions dans le projet ViCoDiLi (*Visualisation de Contenus de Discussions en Ligne*) (Deutsch, Bangou, and Dejean-Thircuir 2007), avec l'outil de visualisation des interactions synchrones (clavardage) et asynchrones (forum, courriel) dans leur contexte de production (i.e. visualiser le lien entre les messages échangés).

18. <http://www.imsglobal.org/learningdesign/>

19. <http://www.imsglobal.org/content/packaging/>

Dans le projet TATIANA (Dyke, Lund, and Girardot 2009), le contexte de production des traces d'interaction est représentée par (1) les traces issues des E.I.A.H. qui doivent être exprimées au format X.M.L., et pouvant aussi concerner des enregistrements audio/vidéo et leur transcription ; (2) les « rejouables » qui sont une représentation au format propriétaire « display » des traces numériques. Il convient de noter que les traces sont généralement considérées comme des données propriétaires fortement couplées à des environnements d'apprentissage spécifiques.

Pour autant, on peut souligner des efforts d'uniformisation avec les bibliothèques d'outils d'analyse du projet IA JEIRP Kaleidoscope (Dimitracopoulou et al. 2004) et d'outils de visualisation et d'analyse de forums du projet CALICO (Giguet et al. 2009).

La capitalisation des résultats d'analyse de traces d'interaction issues de situations d'apprentissage, en vue de procéder à des analyses comparatives, cumulatives ou contrastives, s'exprime par la définition de corpus de traces de référence.

Le projet *P.S.L.C. DataShop* met à la disposition des chercheuses et des chercheurs : (1) un service de dépôt sécurisé des données d'interaction générées dans un I.T.S.²⁰ favorisant ainsi le partage de données de recherche, il est également possible de joindre d'autres documents tels que des articles publiés sur l'étude ; (2) un ensemble d'outils d'analyse et de *reporting* partagés permettant l'analyse des données d'interaction exprimées dans le format propriétaire « *tutor message format* »²¹ proposé par le projet.

Le projet MULCE (Reffay et al. 2008) s'intéresse particulièrement au partage de corpus d'apprentissage et d'enseignement et à leur considération comme objet d'étude scientifique. Ce projet insiste sur l'importance et les bénéfices du partage des données entre chercheuses ou chercheurs et sur la nécessité de contextualisation de ces données. Une structure de corpus d'apprentissage et d'enseignement LETEC est donc proposée préconisant l'utilisation de formats standards existants. Deux types de corpus sont utilisés dans MULCE : (1) corpus global qui contient toutes les données collectées d'une expérimentation ; (2) corpus distinguable pouvant être de trois types suivant les objectifs de la chercheuse ou du chercheur (associer publications scientifiques et données ; rassembler des données prêtes à l'analyse avec la mise en forme pour des outils/logiciels libres ; partager des analyses avec des outils associés). Un autre objectif du projet est aussi de partager des outils d'analyse au sein de la communauté comme conséquence souhaitée du partage d'un format commun de représentation des données. Le projet MULCE donne lieu à un entrepôt pour le partage en ligne de corpus LETEC donnant la possibilité d'effectuer des recherches selon des critères de classification des données partagées, de parcourir des éléments des corpus et de télécharger les corpus globaux ou distinguables mis à disposition.

3.3.3 Approche PROXYMA

Dans (Chebil, Courtin, and Girardot 2015), nous présentons l'approche PROXYMA (*PROXY for Multiple Analyses*), un cadre que nous proposons pour traiter la question du partage et de l'analyse des corpus de traces d'apprentissage contextualisées. Dans cette approche, diverses traces sont partagées avec leur format natif grâce à la définition d'un niveau intermédiaire capable de décrire et d'interroger le corpus. Pour assurer l'interopérabilité sémantique des formats, un nouveau composant nommé « proxy » est proposé pour jouer le rôle d'intermédiaire entre un chercheur et un couple (corpus, outil d'analyse). Ce composant de type *middleware* permet d'effectuer des conversions partielles de données avec des scripts pour répondre aux besoins de l'analyse. La chercheuse peut utiliser les mêmes scripts pour toutes ses autres expérimentations avec le même format de trace, sans avoir à effectuer d'autres conversions.

20. *Intelligent Tutoring Systems*

21. https://pslcdatashop.web.cmu.edu/dtd/guide/tutor_message_dtd_guide_v4.pdf

Dans la présente section, nous expliquons ce que les chercheuses et les chercheurs doivent faire pour partager leurs propres traces et pour utiliser les outils d'analyse existants. Dans un système classique, ils convertissent leurs propres traces dans le format de trace spécifique à l'entrepôt de données. D'autres conversions peuvent être nécessaires afin de correspondre aux données d'entrée des outils d'analyse spécifiques. Toutes ces phases de conversion doivent être effectuées à chaque fois qu'un nouvel ensemble de données doit être analysé ou qu'un nouvel outil d'analyse est utilisé. Quel que soit le système, il convient de noter que les transformations des données d'interaction²² sont inévitables pour interagir avec différents outils d'analyse. Cependant, l'approche PROXYMA est différente dans la mesure où les données d'interaction sont conservées dans leur format d'origine, afin de préserver leur richesse sémantique. Nous distinguons deux modèles de corpus : corpus initial et corpus d'analyse. Les chercheuses et chercheurs créent un corpus de traces d'interaction avec l'un des deux modèles, en suivant trois étapes :

1. la collecte ou la création des différentes ressources qui composent le corpus ;
2. la fourniture d'un ensemble de métadonnées descriptives pour documenter le corpus et les ressources qui le composent (c'est le modèle sémantique) ;
3. la création des différents scripts correspondant aux opérations du modèle opérationnel et qui permettront l'interrogation, l'extraction et la conversion des données du seul corpus qui sont traitées par les outils d'analyse partagés.

L'approche PROXYMA est basée sur trois modèles complémentaires qui sont conceptualisés comme une ontologie O.W.L.²³ (cf. Figure 3.1). Le premier modèle est le « **modèle de corpus** » qui définit les deux types de corpus précités, les différents types de ressources qui peuvent composer un corpus, et un ensemble de métadonnées de description pour documenter le corpus et ses composants. Le deuxième modèle est le « **modèle sémantique** » qui définit une taxonomie de concepts que l'on peut trouver dans les traces collectées et qui sont liés à des données intéressantes pour les analyses. Ce modèle peut être considéré comme un outil sémantique partagé que les chercheuses et les chercheurs peuvent utiliser pour disposer d'une base sémantique commune pour leurs analyses. Le troisième modèle est le « **modèle opérationnel** » qui définit un ensemble d'opérations permettant d'aligner le contenu des corpus partagés, à la fois avec les concepts définis dans le modèle sémantique et avec la représentation des données d'entrée des outils d'analyse.

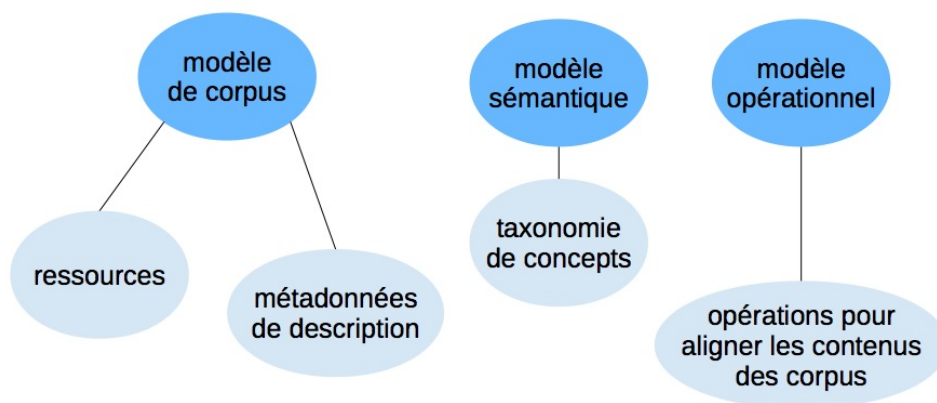


FIGURE 3.1 – Les modèles de l'approche PROXYMA

22. le terme plus général « données d'interaction » permet d'intégrer d'autres ressources dans les corpus que les seules traces d'interaction

23. *Web Ontology Language* est un langage de représentation des connaissances construit sur le modèle de données de R.D.F. (*Resource Description Framework*)

Le modèle de corpus

Ce modèle définit certains types de ressources qui peuvent être partagées au sein d'un corpus (cf. Figure 3.2). La question de la contextualisation étant très importante, nous avons identifié cinq types de ressources différentes qui peuvent enrichir les ressources de traces par des « données contextuelles ». Un corpus est physiquement composé d'un ensemble de différents types de ressources (six types : ressources de traces + ressources contextuelles) et d'une ressource de description. Ce dernier contient les données descriptives modélisées par le modèle de description du corpus qui définit un ensemble de métadonnées pouvant être utilisées pour décrire de manière générale le corpus et les ressources qui le composent.

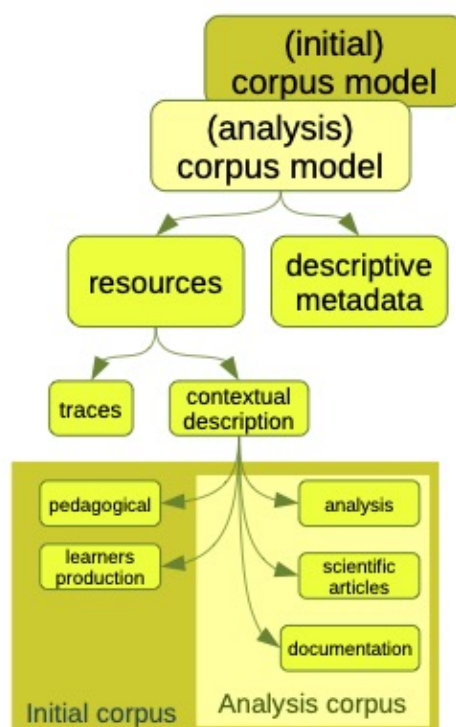


FIGURE 3.2 – Les modèles de corpus de l'approche PROXYMA (1)

Nous définissons deux types de corpus :

1. le **corpus initial** qui contient les ressources relatives à une expérience d'apprentissage et fournies par une chercheuse ou un chercheur souhaitant partager un corpus, et contient toutes les ressources collectées et produites avant le partage du corpus ;
2. le **corpus d'analyse** qui est construit pour répondre à une question de recherche particulière et qui contient des données relatives à des analyses effectuées sur un ou plusieurs corpus à l'aide d'un ou plusieurs outils d'analyse.

Cette différenciation nous permet de séparer :

1. les ressources initiales (traces + ressources contextuelles) collectées et fournies par les chercheuses et les chercheurs, ainsi que les ressources relatives aux analyses effectuées sur le corpus en dehors de la plate-forme de partage ;
2. les analyses effectuées au sein de la plate-forme de partage au moyen des outils d'analyse partagés.

Nous identifions cinq types de ressources contextuelles qui peuvent être considérées comme appropriées à partager (cf. Figure 3.3) :

1. des ressources pédagogiques ou du projet²⁴ qui correspondent aux matériaux fournis aux apprenants par l'intermédiaire de l'environnement d'apprentissage (e.g. un cahier des charges, un cahier de spécification, un énoncé d'exercice, un cours, etc.) ;
2. des ressources de production qui correspondent aux productions des apprenants (e.g. un rapport, la résolution d'un problème) ;
3. des ressources d'analyse qui sont liées aux analyses effectuées sur le corpus, et qui peuvent être utilisées pendant l'analyse (e.g. le modèle d'un domaine particulier spécifique à une activité d'apprentissage), qui peuvent aussi être des ressources produites par le processus d'analyse, ou des ressources d'interprétation dans lesquelles les résultats de l'analyse sont interprétés ;
4. des ressources de publication qui correspondent à différentes communications scientifiques qui sont liées au corpus ;
5. les ressources documentaires qui sont utilisées pour documenter un corpus (e.g. la description de l'expérimentation, la description des objectifs d'apprentissage et/ou de recherche, la description de l'analyse, etc.).

Un corpus initial peut contenir tout type de ressources tandis qu'un corpus d'analyse ne peut contenir que des ressources d'analyse, des ressources de publication et des ressources de documentation. En effet, un corpus d'analyse ne contient pas de ressources relatives à la situation d'apprentissage ou la situation professionnelle.

Concernant la partie description du corpus, elle consiste à fournir un ensemble de métadonnées (un sous-ensemble de métadonnées *DublinCore*²⁵ enrichi par des métadonnées orientées apprentissage/enseignement) pour :

1. décrire le corpus de l'expérimentation (e.g. le nom du corpus, la date de création, les mots-clés, le type d'apprentissage (collectif ou/et individuel), l'aspect de l'apprentissage dans l'espace-temps (face-à-face/distance, synchrone/asynchrone), etc.) ;
2. décrire les ressources composant le corpus (e.g. le créateur, le type de ressource, le format, la date de création, etc.) ;
3. décrire les analyses précédentes effectuées sur le corpus en dehors de la plate-forme (cas d'un corpus initial) ou les analyses effectuées au sein de la plate-forme sur un ou plusieurs corpus partagés (cas d'un corpus d'analyse).

Pour décrire les analyses, un ensemble de métadonnées doit être fourni et certaines ressources documentaires peuvent également être référencées. Par exemple, des informations sur les créateurs et les contributeurs du travail d'analyse doivent être fournies, ainsi que des références aux outils d'analyse utilisés, aux objectifs de l'analyse, etc.

Le modèle de corpus permet de partager des corpus tout en conservant la structure initiale des ressources. Il définit un modèle de description qui offre une base minimale pour simplifier l'interrogation de la base de données de corpus. Il convient de noter que les métadonnées descriptives ne sont pas obligatoires, mais qu'elles sont très utiles pour contextualiser le contenu du corpus et

24. le modèle présenté est une adaptation plus récente au domaine de l'Industrie 4.0, en complément du domaine initial de l'éducation

25. format de métadonnées du *Web semantic* développé par la DCMI, *Dublin Core Metadata Initiative*, <http://dublincore.org/specifications/dublin-core/dcmi-terms/>

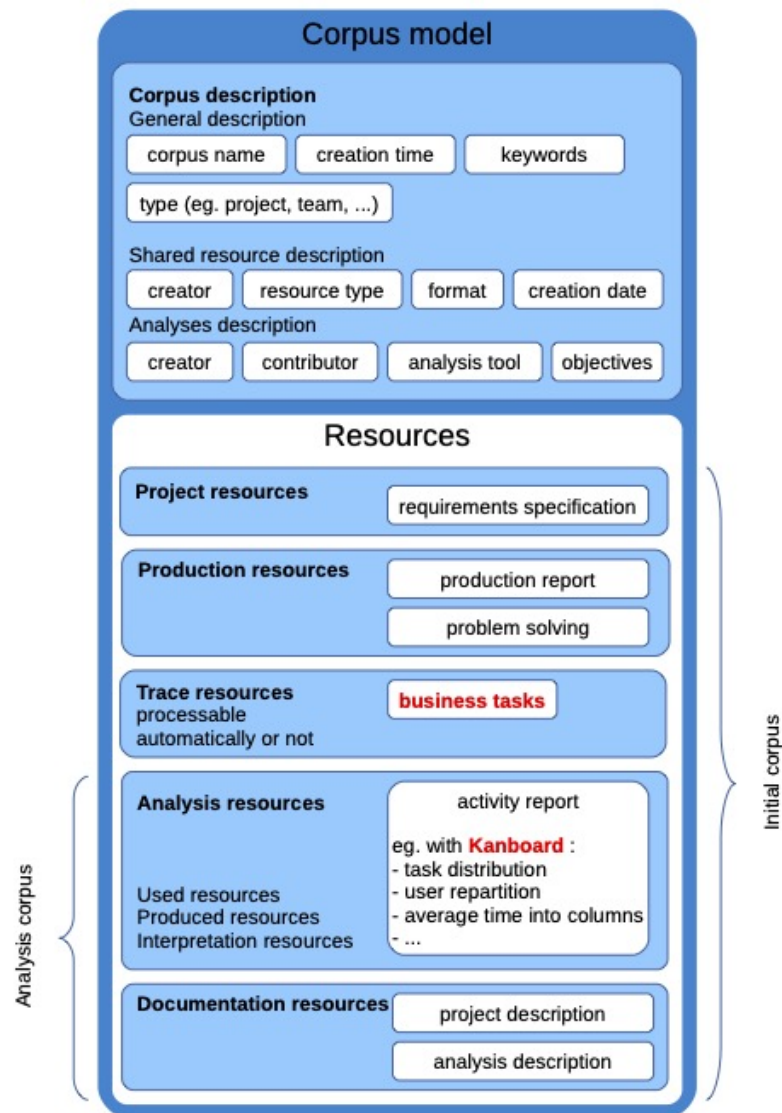


FIGURE 3.3 – Les modèles de corpus de l'approche PROXYMA (2)

aider les chercheuses et les chercheurs qui n'ont pas d'idée sur l'expérience à comprendre les données sans avoir besoin de l'aide de celles et ceux qui ont construit le corpus. En d'autres termes, le corpus devient auto-explicatif. Le modèle de description ainsi défini permet aux chercheuses et aux chercheurs de décrire en détail les analyses précédentes effectuées sur le corpus, ce qui les rend reproductibles.

Nous présentons dans les deux paragraphes suivants le modèle sémantique et le modèle opérationnel qui permettent d'interroger un corpus afin d'analyser ses données.

Le modèle sémantique

Ce deuxième modèle de l'approche PROXYMA définit la sémantique du contenu des corpus partagés. L'objectif de ce modèle est de définir une sémantique qui puisse être commune entre les chercheuses et les chercheurs souhaitant partager et analyser des corpus de traces d'interaction contextualisées. Ce modèle est représenté par le biais d'une taxonomie (cf. Figure 3.4), dans le cadre de l'ontologie descriptive globale qui conceptualise les trois modèles de l'approche.

Une telle taxonomie définit un ensemble de concepts que l'on peut retrouver dans les données collectées lors d'une expérimentation. Les concepts définis dans la taxonomie correspondent à des données trouvées dans des corpus de ressources de traces ou d'autres ressources contextuelles, et qui sont utiles pour un travail d'analyse. Le modèle sémantique peut donc être considéré comme un outil de mise en correspondance sémantique, qui permet aux chercheuses et aux chercheurs d'aligner implicitement les données contenues dans un corpus, ou les données attendues en entrée d'un outil d'analyse d'une part, et les concepts définis dans la taxonomie d'autre part. Il est évident que la définition d'une telle taxonomie est une tâche complexe en raison de la variété des types de concepts qui peuvent être nécessaires. Pour cette raison, nous suggérons de construire la taxonomie de manière incrémentale et participative. Les concepts qui composent la taxonomie devraient ensuite être ajoutés progressivement pour répondre à l'évolution des besoins des chercheuses et des chercheurs. À la différence d'une taxonomie classique, où les concepts sont reliés par un type de relation unique (la relation généralisation/spécialisation), nous avons identifié deux types de concepts qui peuvent avoir deux types de relations (cf. Figure 3.5).

Pour mieux comprendre les différentes catégories de concepts qui composent la taxonomie, nous commençons par présenter la genèse des concepts définis, qui sont d'origines diverses :

- des concepts universels que l'on retrouve dans différents corpus, indépendamment de l'environnement d'apprentissage et du domaine d'application (e.g. des concepts tels que « participant » et « indicateur temporel ») ;
- les concepts liés au domaine d'application des environnements d'apprentissage qui peuvent être spécifiques à une discipline et/ou un niveau ; e.g. Andes (Gertner and VanLehn 2000) est un tuteur de physique destiné à l'enseignement de la physique au collège et au lycée, le tuteur de lecture (Mostow and Beck n.d.) est un *I.T.S.*²⁶ qui aide les enfants à apprendre à lire, Teleos (Luengo et al. 2006) est un *I.T.S.* destiné à la formation des étudiants en chirurgie orthopédique ; la spécificité des environnements d'apprentissage peut impliquer des données tracées spécifiques à un domaine donné ;
- des concepts liés à un certain type d'expérimentation d'apprentissage, un même E.I.A.H. pouvant être utilisé dans différents contextes (e.g. apprentissage individuel ou collaboratif, en face à face ou à distance) ;

26. *Intelligent Tutoring Systems*

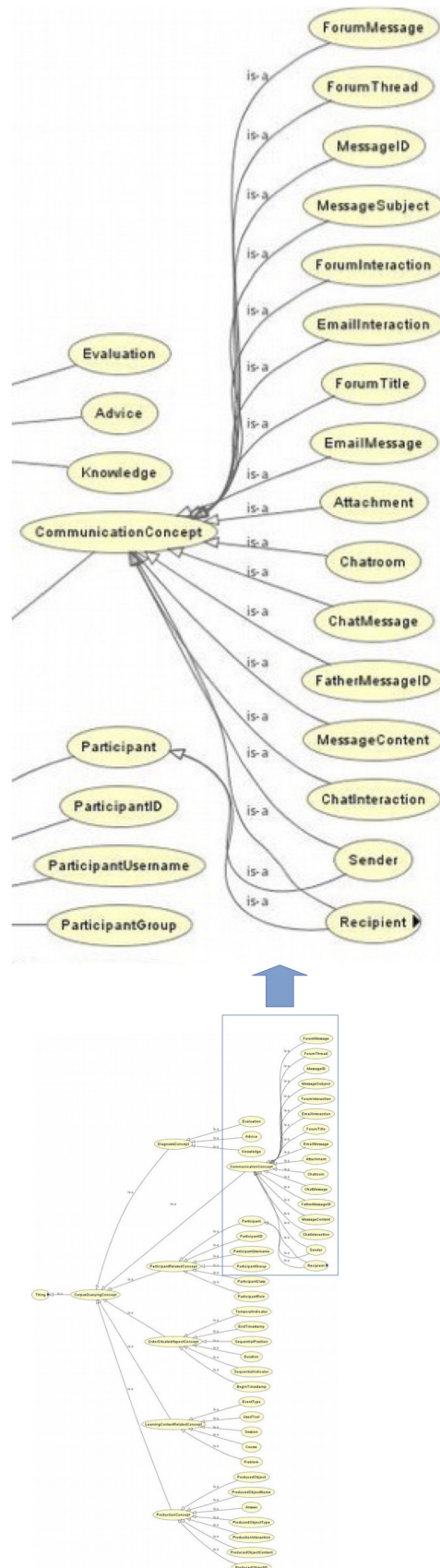


FIGURE 3.4 – Le modèle sémantique de l'approche PROXYMA

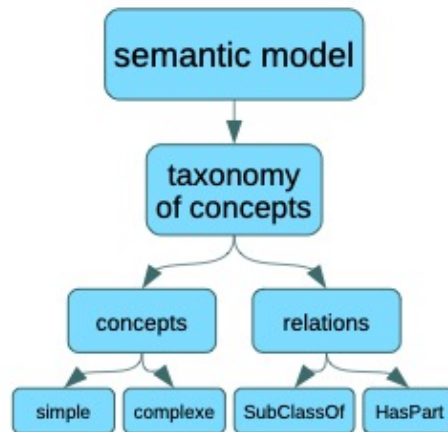


FIGURE 3.5 – Le méta-modèle du modèle sémantique de l'approche PROXYMA

- les concepts liés au travail d'analyse qui est souvent itératif et où les chercheurs peuvent souhaiter capitaliser sur les analyses précédentes (Dyke, Lund, and Girardot 2009); une chercheuse ou un chercheur peut avoir besoin d'interroger des ressources d'analyse produites précédemment (ces ressources peuvent être considérées comme des ressources de traces enrichies par l'analyse); e.g. les concepts liés au diagnostic des activités d'apprentissage sont induits par l'analyse.

Cette diversité des origines des concepts explique que l'ontologie doit rester ouverte pour être enrichie en fonction des besoins de représentation et d'analyse des chercheuses et des chercheurs. Nous proposons de classer les concepts du modèle sémantique de manière thématique sous différentes catégories. Nous avons identifié six catégories pour faciliter la navigation au sein des concepts définis :

1. les concepts liés à l'aspect temporel des traces sont son début, sa fin, sa durée et sa position dans une séquence d'éléments observés situés dans le temps qui découle de l'interaction elle-même, et incluent deux **concepts complexes** *TemporalIndicator* et *SequentialIndicator* qui agrègent respectivement les concepts de début, de fin et de durée, et les concepts de position et de durée;
2. les **concepts simples** liés au contexte d'apprentissage correspondent à un cours, un problème en cours, un type d'événement tel que l'envoi d'un message, une session dans une activité, et un outil logiciel utilisé;
3. les concepts liés à une personne participant aux activités d'apprentissage caractérisées par un identifiant pour être unique, un pseudonyme pour être protégée, un nom de groupe auquel elle appartient lorsque l'activité est collective, et un rôle dans le groupe;
4. les concepts liés aux activités de communication qui dépendent de l'outil logiciel utilisé (e.g. un *chat*) sont des concepts complexes qui regroupent un message de *chat* (qui regroupe à son tour un identifiant de message, un salon de *chat* et un contenu de message), un cours, un problème en cours, un type d'événement tel qu'une connexion dans le *chat*, une session dans le *chat*, l'outil de *chat* spécifique, l'expéditeur du message, et les deux **concepts complexes** *TemporalIndicator* et *SequentialIndicator* décrits ci-dessus;
5. les concepts liés aux activités de production caractérisés par la réponse d'une personne fournie au cours de l'activité d'apprentissage, et l'objet produit qui regroupe à son tour son identifiant, son nom, son type selon le domaine d'application, et son contenu.;
6. les concepts liés au diagnostic sont « connaissance », « évaluation » et « conseil », pour fournir des conseils après l'évaluation des connaissances d'un apprenant, afin de l'aider dans son activité.

La catégorisation proposée n'est pas exhaustive et peut être enrichie en fonction de l'évolution des besoins d'analyse. En raison de l'hétérogénéité des traces et de la diversité des analyses possibles, les chercheuses et les chercheurs peuvent être amenés à ajouter de nouveaux concepts, voire une nouvelle catégorie de concepts. Nous avons identifié deux types d'évolution de modèles sémantiques, chacun pouvant être divisé en deux sous-types :

1. l'ajout d'un nouveau concept générique :
 - l'ajout du concept à une catégorie existante dans la taxonomie consiste à lui donner un nom et une définition textuelle, et éventuellement à utiliser l'une des deux relations pour le relier à d'autres concepts existants ;
 - ou à une nouvelle catégorie générique.
2. l'ajout d'un nouveau concept spécifique (i.e. spécifique à une situation d'apprentissage) :
 - à un module ontologique existant ;
 - ou à un nouveau module spécifique représenté sous la forme d'une mini-ontologie O.W.L.²⁷ qui sera liée par la relation « SubClassOf » à une catégorie de l'ontologie centrale.

Pour mieux tirer parti des recherches antérieures, deux modules similaires ajoutés par des chercheuses ou des chercheurs différents, dans le cadre de travaux de recherche différents, peuvent être alignés sémantiquement pour leur permettre de comparer leurs travaux. Dans (Vargas-Vera and Nagy 2015) et (Kalfoglou and Schorlemmer 2003), différentes techniques d'alignement sémantique sont présentées et peuvent être utilisées séparément ou de manière complémentaire.

Le modèle opérationnel

Le modèle opérationnel définit les mécanismes nécessaires pour interroger la base de données de corpus partagée, afin d'effectuer différentes analyses à l'aide de différents outils et méthodes d'analyse. Ce modèle définit un ensemble d'opérations permettant d'extraire, de convertir, de filtrer et de formater des données de corpus en vue de leur analyse (cf. Figure 3.6). Définir des scripts comme faisant partie du modèle opérationnel, permet à une chercheuse ou un chercheur qui prévoit d'analyser certaines données de corpus, d'effectuer un alignement sémantique des concepts des données de corpus qui l'intéressent avec ceux du modèle sémantique.

Couplé au modèle sémantique, le modèle opérationnel permet d'atteindre l'un des objectifs fondamentaux de l'approche PROXYMA, qui est d'offrir une couche intermédiaire, appelée « proxy », entre des corpus partagés et différents outils d'analyse. Sur la base du modèle sémantique, les différentes opérations que nous définissons permettent de faire correspondre une partie des données partagées au sein d'un corpus avec un couple (sémantique, représentation) ($S_{OutilAnalyse}$, $R_{OutilAnalyse}$) relatif aux données attendues en entrée d'un outil d'analyse.

Les six types d'opérations que nous définissons sont :

1. *Query* pour extraire d'un corpus des données relatives à un concept défini dans le même modèle sémantique ;
2. *Convert* pour convertir les données extraites dans le type de données attendu par un outil d'analyse (e.g. la date) ;
3. *Extract* pour récupérer les données d'une ressource partagée dans un corpus, relatives à un concept complexe nécessaire à l'analyse ;

27. *Web Ontology Language*

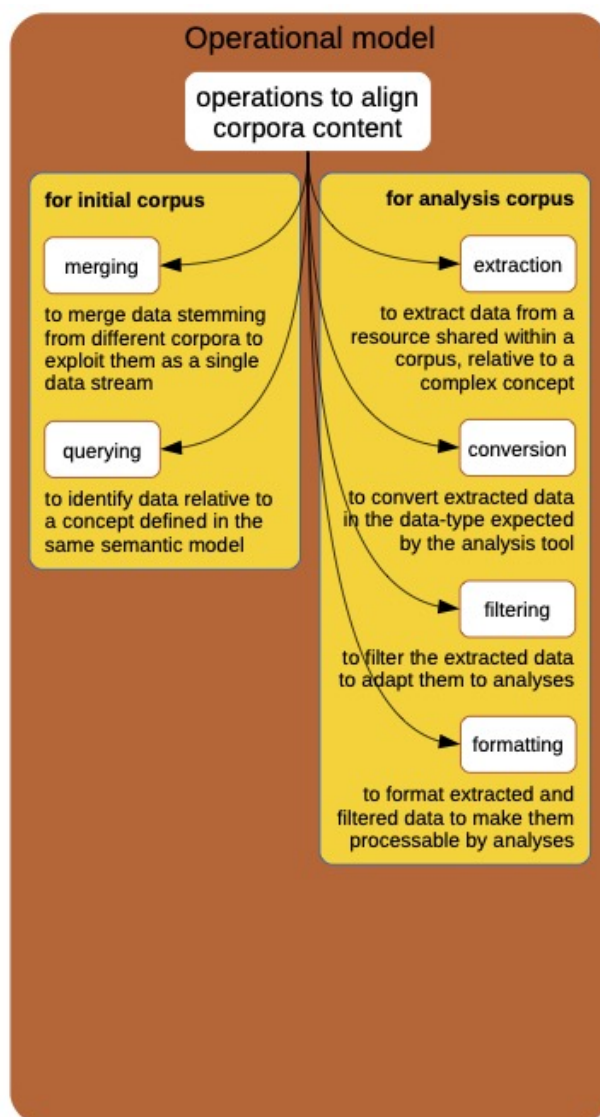


FIGURE 3.6 – *Le modèle opérationnel de l'approche PROXYMA*

4. *Filter* pour filtrer les données extraites afin de les adapter à des besoins d'analyse spécifiques ;
5. *Format* pour mettre en forme les données extraites et filtrées afin de les rendre traitables par une analyse ;
6. *Merge* pour fusionner des données provenant de différents corpus afin de les analyser comme un seul flux de données.

L'architecture que nous avons utilisée comme prototype pour mettre en œuvre l'approche PROXYMA est basée sur une plate-forme de gestion de base de données existante. Ayant choisi de représenter les données interrogeables des corpus en X.M.L.²⁸, nous utilisons eXist-db²⁹, un système de gestion de base de données (S.G.B.D.³⁰) natif X.M.L., comme plateforme d'infrastructure pour le stockage et la manipulation des données. La base de données eXist stocke la base de données de corpus, la base de données de scripts et l'ontologie qui formalise les trois modèles de l'approche. Notre approche et ses trois modèles connaissent une première phase de validation avec la construction d'un corpus initial « OOD-OOP », et d'un corpus d'analyse « ForumCollab » dans lequel les interactions du forum du corpus initial « OOD-OOP » sont analysées (Chebil, Courtin, and Girardot 2014).

3.3.4 Plate-forme BEATCORP

Pour opérationnaliser l'approche PROXYMA, nous proposons une plate-forme de *benchmarking*, nommée BEATCORP (*Benchmarking Platform for Analysis of Traces CORPora*) pour les analyses partagées et/ou croisées de corpus de traces d'activité, dont nous présentons l'architecture Figure 3.7 (Courtin and Tomasena 2016). Le système BEATCORP est construit sur les modèles de l'approche PROXYMA et est dédié à la recherche en informatique, sciences sociales, pédagogie, didactique, etc. Cette plate-forme permet le partage de corpus de traces et l'analyse de ces corpus avec un ensemble d'outils d'analyse et de visualisation partagés.

Beaucoup de questions se posent alors sur : (1) la possibilité d'intégration de traces de natures et de sources différentes ayant des formats syntaxiquement différents, mais surtout ayant des sémantiques différentes ; (2) l'adéquation entre les différents objectifs d'exploitation des corpus et leur contenu, ce qui suppose une catégorisation des corpus ou parties de corpus en fonction de leur contenu ; (3) l'élaboration d'une méthodologie d'exploitation des corpus face à la diversité des pratiques et des disciplines, ceci se traduit par une correspondance entre les besoins et les questions de recherche et les outils d'analyse et de visualisation offerts par la plate-forme de *benchmarking*.

La plate-forme BEATCORP comprend cinq composants :

1. la base de données de corpus pour stocker les corpus partagés ;
2. une ontologie O.W.L.³¹ pour formaliser les trois modèles de l'approche ;
3. une base de données de scripts pour stocker les scripts relatifs aux six opérations du modèle opérationnel permettant d'extraire des données de corpus afin de les analyser ;
4. un moteur de gestion permettant la manipulation des différents composants de la plate-forme ;
5. une application client jouant le rôle d'une interface entre un utilisateur et les autres composants de la plate-forme.

28. *Extensible Markup Language*

29. exist-db.org

30. *Système de Gestion de Base de Données*

31. *Web Ontology Language*

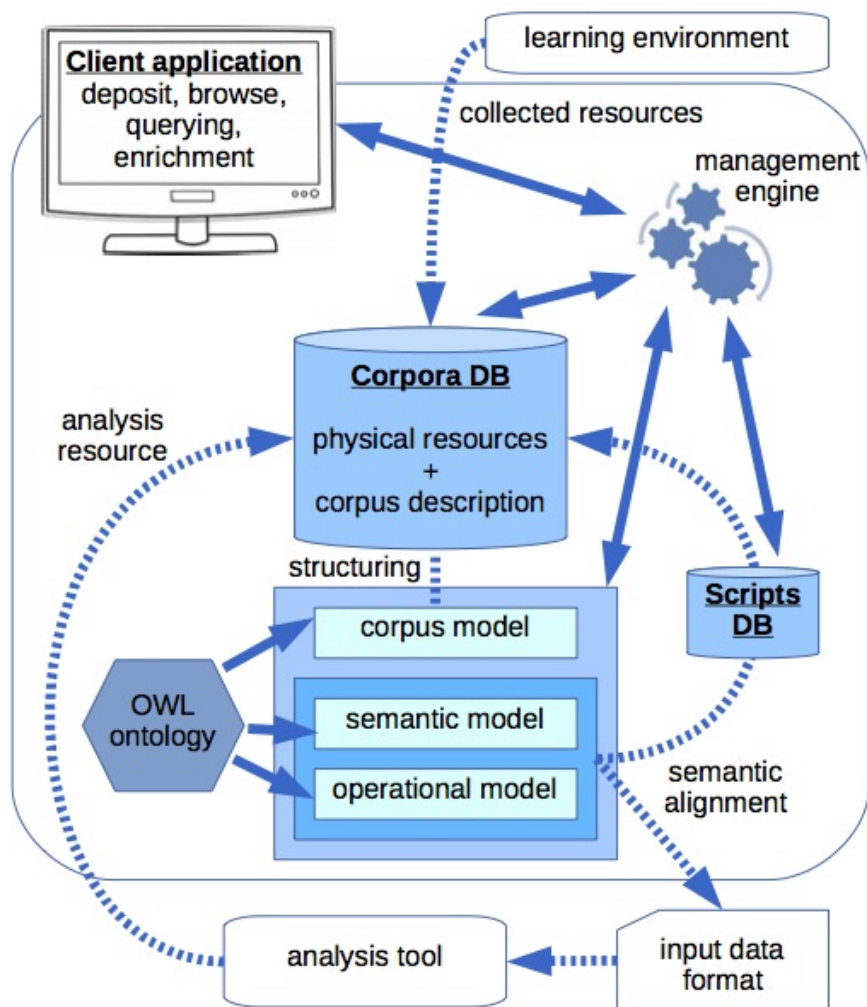


FIGURE 3.7 – L'architecture de la plate-forme de benchmarking BEATCORP

La base de corpus

La base de corpus stocke les corpus partagés entre les observateurs. Comme spécifié par le modèle de corpus, il peut s'agir d'un corpus initial correspondant à une expérimentation d'apprentissage particulière ou d'un corpus d'analyse construit afin de répondre à une question de recherche, en utilisant un ou plusieurs corpus partagés précédemment (initial ou d'analyse). Un corpus est composé de ressources physiques collectées pendant l'expérimentation ou produites par l'analyse et d'une description du corpus et de son contenu. La description d'un corpus suit le modèle de description du modèle de corpus de PROXYMA. Pour analyser les corpus stockés dans la base de corpus à l'aide de différents outils d'analyse, le système doit être capable d'interroger la base de corpus et de la mettre à jour en ajoutant de nouvelles ressources à un corpus et/ou en modifiant la description de celui-ci. La manipulation de cette base de corpus est effectuée par le moteur de gestion qui sera présenté dans la suite.

La base de scripts

La base de scripts contient les scripts relatifs aux différents types d'opérations définis dans le modèle opérationnel, et nécessaires pour interroger les corpus partagés dans la base de corpus, afin d'analyser leur contenu. Cette base de scripts est construite de manière incrémentale pour répondre aux besoins d'analyse des observateurs. Dans nos expérimentations, les données interrogées sont exprimées en X.M.L.³² et les scripts sont écrits en XQuery³³. Pour analyser un corpus, l'observatrice ou l'observateur sélectionne l'outil d'analyse à utiliser. La base de scripts est ensuite sollicitée par le moteur de gestion pour vérifier si les scripts nécessaires sont mis en œuvre, auquel cas l'extraction et la conversion des données sont automatiquement effectuées. À défaut, si certains scripts sont manquants, ils doivent être définis et liés aux concepts concernés. Basé sur les trois modèles de l'approche PROXYMA, le moteur de gestion permet l'exécution des différentes opérations nécessaires à la manipulation de la base de corpus, de la base de scripts et de l'ontologie. Ce moteur joue le rôle de PROXY, une couche intermédiaire entre l'application cliente utilisée par une observatrice ou un observateur et les autres composants de la plate-forme. Il exécute leurs requêtes et y répond. Le moteur de gestion doit mettre en œuvre les fonctionnalités suivantes : insertion, suppression, modification et interrogation des trois composants : la base de corpus, l'ontologie et la base de scripts.

L'application cliente

L'application cliente permet aux observatrices et observateurs d'accéder et de manipuler les autres composants de la plate-forme via le moteur de gestion. Le rôle de l'application cliente est de rendre transparents les aspects techniques de la manipulation et de l'interrogation des données. Cela permet d'améliorer la convivialité pour les non-informaticiens. Par exemple, une observatrice ou un observateur qui crée un nouveau corpus ne verra pas le formalisme du modèle de corpus explicite que nous proposons, mais remplira un formulaire qui est construit sur la base de ce modèle.

Le composant principal de la plateforme BEATCORP est l'**ontologie O.W.L.**³⁴ qui formalise les trois modèles de l'approche PROXYMA : le modèle de corpus, le modèle sémantique et le modèle opérationnel. Nous avons utilisé l'éditeur « Protégé »³⁵ pour construire l'ontologie.

32. *Extensible Markup Language*

33. W3C. XQuery 1.0, An X.M.L., <http://www.w3.org/TR/xquery/>

34. *Web Ontology Language*

35. <https://protege.stanford.edu/>

3.4 Expérimentation d'une plate-forme de *benchmarking*

Nous présentons notre travail à travers une expérimentation réalisée dans une authentique salle de classe à l'université. Pour cette expérimentation, nous utilisons deux outils d'analyse externes avec deux formats de trace différents. Cette étude de cas est considérée comme une première étape pour tenter de valider notre approche. D'autres expériences avec d'autres formats de traces sont nécessaires pour compléter la validation.

3.4.1 Les objectifs de l'expérimentation

L'objectif principal de l'expérimentation est la création d'un corpus pour prouver la faisabilité de notre approche, en utilisant deux environnements d'analyse différents sur le même corpus, et en partageant les données résultantes dans un corpus d'analyse.

L'objectif pédagogique est d'aider les étudiants à réaliser un travail pratique tout en communiquant et en collaborant via un forum de discussion. L'utilisation du forum vise à capitaliser les informations importantes et à éviter de perdre du temps à poser une question à laquelle on a déjà répondu.

Nous attendons également de cette expérimentation un moyen pour :

- mesurer le degré de collaboration élève/élève et élève/enseignant ;
- faciliter l'apprentissage par la recherche et la collaboration.

3.4.2 Le contexte de l'expérimentation

L'expérimentation est une situation d'apprentissage à l'université pendant l'unité d'enseignement « conception orientée objet/programmation orientée objet », à l'issue de laquelle nous construisons un corpus initial, appelé « OOD-OOP »³⁶, afin de valider notre approche PROXYMA.

Nous construisons le corpus d'analyse, appelé « ForumCollab », pour étudier la question de recherche : « Dans quelle mesure l'utilisation du forum de discussion améliore-t-elle la collaboration étudiant/étudiant et étudiant/enseignant et contribue-t-elle à la capitalisation de l'information ? ».

Cette approche répondant aux problèmes (1) d'hétérogénéité (sémantique, formats) des données collectées et des contextes d'apprentissage correspondants, (2) de couplage fort entre les formats de données et les outils d'analyse et (3) de non-reproductibilité des analyses précédentes, nous utilisons deux environnements d'analyse différents : TATIANA³⁷ et CALICO³⁸ (Giguet et al. 2009), pour analyser les données de ce corpus.

3.4.3 Le déroulement des opérations

Les étudiants utilisent deux outils de la plate-forme d'apprentissage Moodle³⁹ : le module de forum pour communiquer entre eux, et le module d'affectation pour envoyer leurs travaux à l'enseignant au fil des séances. L'expérimentation porte sur quatre séances de travaux pratiques et les étudiants doivent envoyer leurs travaux en cours à la fin de chaque séance. Pendant les heures de cours, les participants communiquent en face à face en utilisant le forum, ainsi que par des canaux de communication classiques (illustrations verbales et classiques sur tableau blanc).

36. *Object-Oriented Design/Object-Oriented Programming*

37. <http://code.google.com/p/tatiana/>

38. <https://giguete.users.greyc.fr/epal2011-blondel-giguete.slides.pdf>

39. <https://moodle.org/>

Le reste du temps (i.e. entre les séances), les étudiants travaillent à distance et communiquent entre eux et avec l'enseignant en utilisant le forum. Les étudiants sont encouragés à utiliser le forum pour poser leurs questions. C'est-à-dire que les élèves recherchent des informations auprès d'autres élèves et/ou du professeur, en sachant que le résultat de ces recherches doit être capitalisé dans le forum. À cette fin, le forum a été structuré en six thèmes traitant de différents aspects du travail demandé aux étudiants.

Pour combler les zones d'ombre du numérique, les deux dernières séances ont été filmées, afin de garder une trace des conversations synchrones et de ce qui s'est passé en classe pendant ces séances.

3.4.4 Les résultats de l'expérimentation

Dans le cas du corpus « OOD-OOP », nous ne décrivons pas les analyses du corpus initial, puisque nous utilisons les services de l'approche PROXYMA pour réaliser les analyses que nous décrivons ci-après. Nous créons donc un corpus d'analyse nommé « ForumCollab » pour étudier le rôle des interactions du forum dans l'amélioration de la collaboration étudiant/étudiant et étudiant/enseignant et dans la contribution à la capitalisation d'informations utiles.

Nous distinguons différents types de ressources qui peuvent être partagées au sein d'un corpus. Ces ressources peuvent contenir des traces ou différentes données contextuelles utiles pour le partage et la compréhension du corpus.

Nous distinguons deux types de traces complémentaires :

1. celles qui sont extraites avec une requête SQL dans la base de données relationnelle hébergeant les données de la plate-forme Moodle, et qui correspondent aux interactions entre les étudiants et l'enseignant, qui ont eu lieu dans le forum de discussion ;
2. celles qui correspondent aux sept enregistrements audio/vidéo des deux dernières séances d'apprentissage.

Les corpus contiennent des ressources ayant des formats différents et, surtout, une sémantique différente. Les outils d'analyse attendent des données d'entrée dont la sémantique et les formats sont différents de ceux des données originales du corpus.

Le **corpus initial** contient :

- des ressources de traces, contenant des traces X.M.L. exportées de la base de données de logs de Moodle et des enregistrements audio/vidéo des deux dernières séances ;
- cinq types de ressources pédagogiques fournies aux étudiants par l'enseignant, pour les aider à atteindre les objectifs de l'unité :
 1. trois ressources de cours ;
 2. une ressource expliquant le travail à réaliser par les étudiants ;
 3. une ressource contenant quelques conseils de programmation ;
 4. un tutoriel ;
 5. une ressource contenant une solution de conception pour aider les étudiants dans leur travail de programmation.
- quarante-sept ressources de production correspondant aux productions intermédiaires et finales des élèves ;
- une ressource documentaire décrivant l'expérimentation à l'origine du corpus ;
- deux publications connexes, une thèse de doctorat décrivant l'expérimentation et un article scientifique.

Afin d’analyser l’utilisation du forum, nous avons développé un ensemble de scripts, comme faisant partie du **modèle opérationnel**, pour interroger les ressources de traces afin d’extraire les données relatives aux interactions du forum. Partant du principe que toute ressource de traces numériques peut être convertie en une représentation X.M.L., nous avons fait le choix de représenter les traces numériques à l’aide d’un arbre X.M.L., et d’écrire les scripts en XQuery. Nous listons ci-dessous les scripts, détaillés dans (Chebil, Courtin, and Girardot 2015), permettant l’extraction de données relatives à une partie des concepts du **modèle sémantique** :

- script XQuery d’interrogation du concept simple *BeginTimestamp* ;
- script XQuery d’interrogation du concept complexe *TemporalIndicator* ;
- script XQuery d’interrogation du concept complexe *ForumInteraction* qui consiste à aligner le concept « interaction de forum », défini dans le modèle sémantique, avec les données de ressources du corpus ;
- script XQuery d’extraction retrouvant toutes les interactions du forum à partir des traces numériques que nous avons exportées de la base de données de Moodle, appelant le script XQuery d’interrogation du concept complexe *ForumInteraction* ;
- script XQuery de filtrage récupérant les interactions du forum correspondant à un fil de discussion particulier, et exécutées à la sortie du script d’extraction *ForumInteraction* ;

Si nécessaire, des scripts de conversion des types de données peuvent être définis. Dans notre expérimentation, nous n’avons pas besoin de définir de tels scripts, car les types de données originaux des ressources interrogées sont compatibles avec nos besoins d’analyse.

Nous effectuons quelques analyses sur les traces des interactions des forums du corpus « OOD-OOP ». Pour analyser l’interaction du forum avec TATIANA et CALICO, deux scripts de formatage différents, détaillés dans (Chebil, Courtin, and Girardot 2015), sont utilisés afin de formater les données pour l’entrée de l’environnement d’analyse. De tels scripts effectuent l’alignement entre le concept « interaction forum » du modèle sémantique et la structure de données attendue en entrée des outils d’analyse.

Les ressources physiques composant le **corpus d’analyse** sont :

- une ressource documentaire décrivant la question de recherche étudiée et les analyses effectuées dans le corpus ;
- deux ressources de publication ;
- onze ressources d’analyse (huit sont produites par TATIANA et peuvent être réutilisées par d’autres chercheuses ou chercheurs et ouvertes dans TATIANA, et deux captures d’écran illustrant l’utilisation de la plate-forme CALICO) ;
- une ressource d’interprétation contenant nos interprétations des analyses effectuées.

La Figure 3.8 illustre deux types de visualisation (graphique et tabulaire) des interactions du forum dans TATIANA. La synchronisation des deux représentations permet d’accéder au contenu des messages postés en sélectionnant simplement une case dans la représentation graphique. Dans la visualisation graphique, les cases représentent les événements d’envoi de messages liés par la relation « réponses à » et sont colorées en fonction du participant qui les a envoyés. Cette visualisation des données nous montre que les étudiants utilisent le forum pour poser des questions à l’enseignant qui est le seul à y répondre. La collaboration entre étudiants n’est donc pas améliorée par l’utilisation du forum. Cela indique que l’enseignant doit trouver de nouveaux moyens d’encourager les élèves à collaborer (e.g. en donnant un bonus à un élève qui répond à la question d’un autre élève, en introduisant un système d’e-réputation).

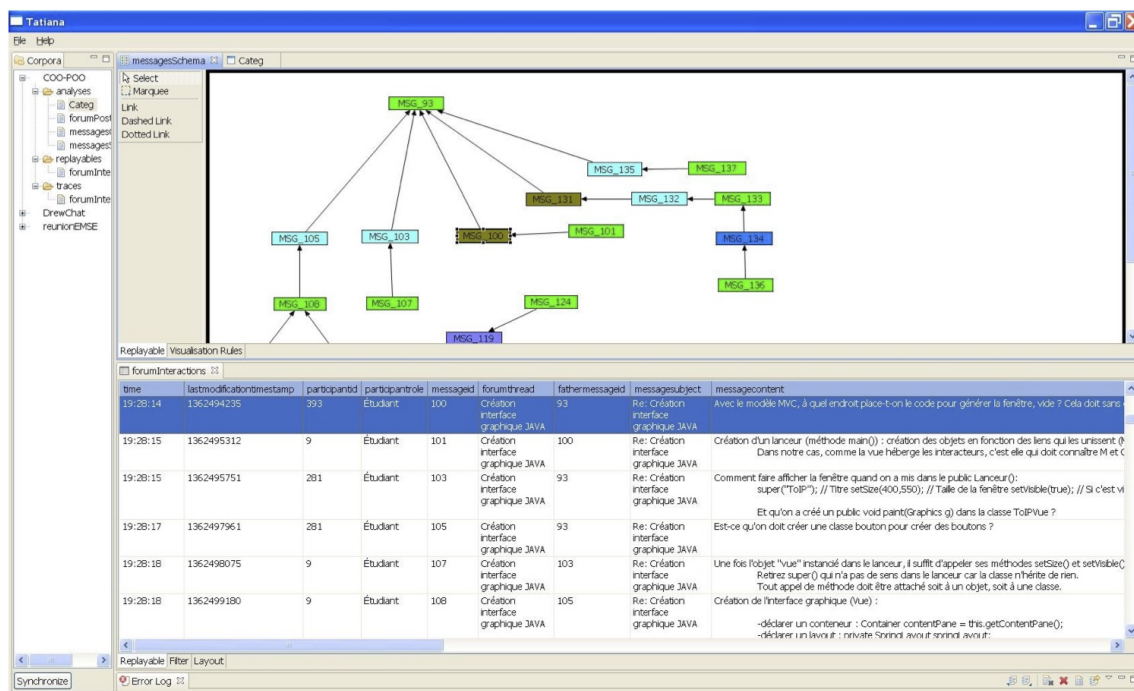


FIGURE 3.8 – Synchronisation graphe/tableau du forum dans TATIANA

Dans une seconde analyse, nous nous concentrons sur la sémantique des messages postés et nous travaillons sur une catégorisation des messages du forum. La Figure 3.9 illustre une catégorisation des types de messages échangés (e.g. initier une discussion, poser une question, répondre à une question, etc.). Les catégories définies représentent ce que les messages échangés signifient par rapport à l'activité d'apprentissage. Cette catégorisation est utilisée dans la visualisation graphique des messages échangés (cf. Figure 3.10).

The screenshot shows the TATIANA application window with the 'messagesCategories' tab selected. It displays a table with columns: Name, Description, Colour, and Appearance.

Name	Description	Colour	Appearance
Ask question	Ask a question to participants	Blue	Ask question
Answer a question - explanation	Answering a question by giving an explanation	Green	Answer a question - ...
Give information	Giving some useful information to participants	Pink	Give information
Initiating thread	Initiate a discussion thread	Orange	Initiating thread
Initiating thread, giving information	Initiate a discussion thread to give some infor...	Purple	Initiating thread, giv...
Initiating thread, asking a question	Initiate a discussion thread to answer a quest...	Yellow	Initiating thread, aski...
Asking and answering one's own question	Asking a question, then editing the post to a...	Cyan	Asking and answeri...
Answering a question, giving example	Answering a question by giving an example	Dark Green	Answering a questio...
Answering a question, explain and example	Answering a question by giving an explanatio...	Light Green	Answering a questio...
Re-asking a previous question	Re-asking a previously asked question to whi...	Dark Blue	Re-asking a previous...
Contradicting an answer	Contradicting an answer given to a question	Red	Contradicting an ans...
Commenting an answer	Commenting an answer by giving some expli...	Yellow	Commenting an ans...
Answering one's own question	A user can answer his own question when fin...	Dark Green	Answering one's ow...

FIGURE 3.9 – Configuration de la catégorisation des messages du forum dans TATIANA

Nous distinguons treize catégories : (1) poser une question ; (2) reposter une question déjà posée ; (3) répondre à une question en donnant une explication ; (4) répondre à une question en donnant un exemple ; (5) répondre à une question en donnant une explication et un exemple ; (6) contester une réponse ; (7) commenter une réponse ; (8) répondre à sa propre question ; (9) poser et répondre à une question ; (10) fournir des informations ; (11) démarrer un fil de discussion pour traiter un sujet ; (12) démarrer un fil de discussion pour donner des informations ; et (13) démarrer un fil de discussion pour poser une question.

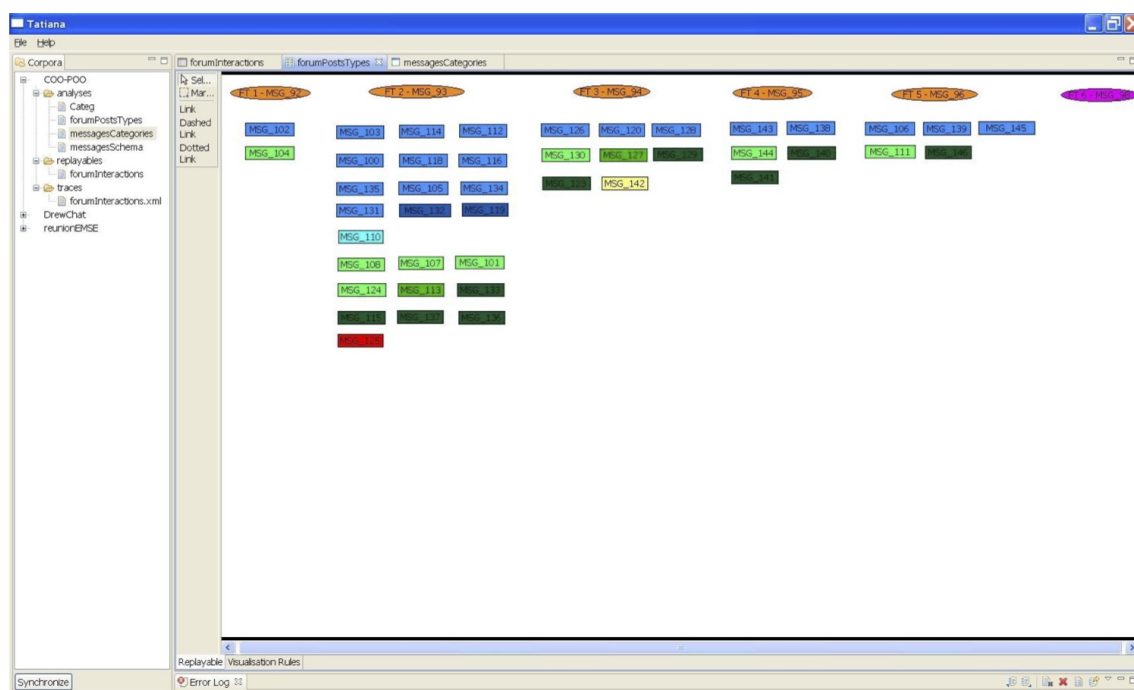


FIGURE 3.10 – Affichage de la catégorisation des messages du forum dans TATIANA

De notre point de vue, ces catégories représentent ce que l'on peut trouver dans un tel forum de discussion, mais toutes ces catégories ne se retrouvent pas pour autant dans les interactions du forum collectées. La Figure 3.10 permet d'avoir une représentation visuelle des différentes catégories de messages échangés et de leur proportion. On peut ainsi remarquer que la plupart des messages échangés correspondent à des questions (nuances de bleu) et à des réponses (nuances de vert).

Pour étudier la possibilité d'utiliser différents outils d'analyse sur un même corpus, nous utilisons deux outils de la plate-forme CALICO. La plateforme CALICO offre un ensemble d'outils qui améliorent les interactions du forum de lecture. Certains outils offrent une vision quantitative des données, quand d'autres sont destinés à effectuer des analyses approfondies du contenu des messages. La Figure 3.11 illustre l'utilisation de l'outil Authagora de la plate-forme CALICO, qui calcule quelques statistiques sur le nombre de messages envoyés pour chaque actrice ou acteur, et qui confirme qu'un utilisateur particulier (le professeur) a envoyé plus de messages que les autres participants.

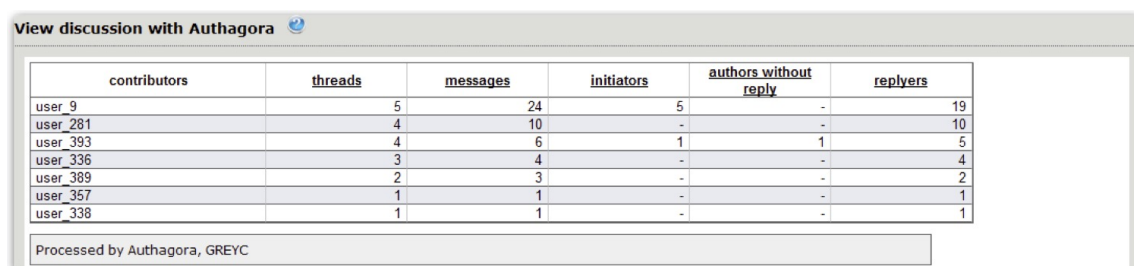


FIGURE 3.11 – Outil Authagora de CALICO

La Figure 3.12 illustre l'utilisation de l'outil Concordagora, qui recherche les occurrences de certains mots et affiche leur contexte d'utilisation. L'utilisation de cet outil nous permet de rechercher rapidement les questions à l'aide de deux mots-clés (« comment » et « est-ce que ») et d'avoir une lecture plus facile des interactions.

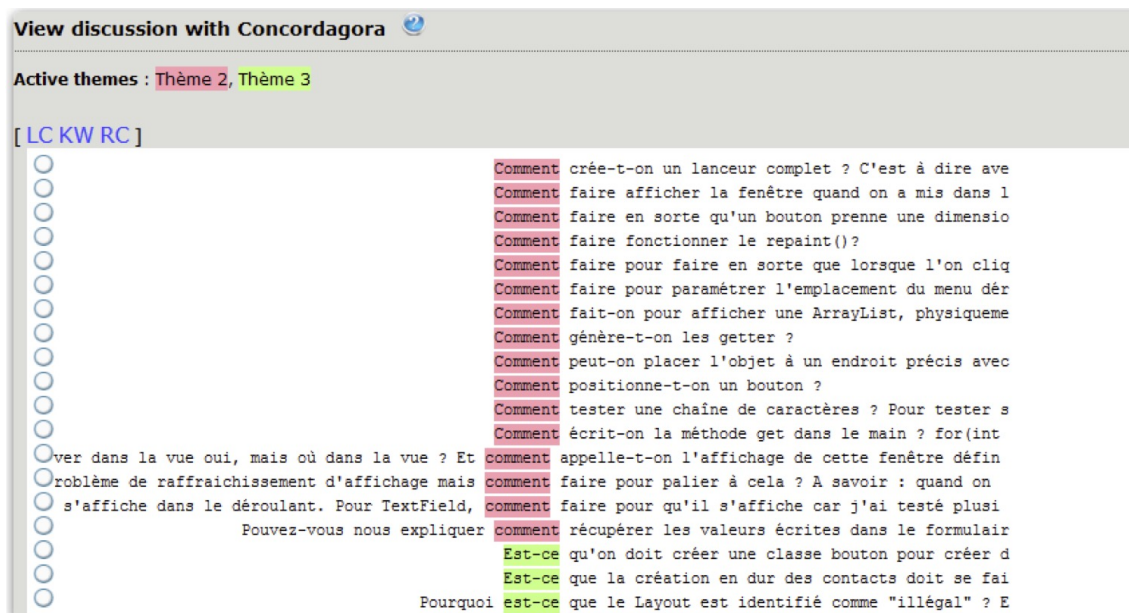


FIGURE 3.12 – Outil Concordagora de CALICO

Les analyses présentées dans cette section permettent d'étudier la question de recherche qui nous intéresse, en relation avec le rôle de la technologie dans l'amélioration de la collaboration et de la capitalisation de l'information. L'objectif d'améliorer la collaboration sur le forum n'est que partiellement atteint et il reste encore du travail à faire pour encourager les élèves à collaborer davantage (e.g. en donnant des points supplémentaires aux élèves les plus actifs sur le forum ou à ceux qui donnent une réponse correcte dans un délai donné). Le deuxième objectif, qui est de capitaliser les informations utiles pour éviter de poser plusieurs fois la même question, est atteint. Les étudiants recherchent les discussions existantes avant d'en initier de nouvelles.

La réalisation de l'expérimentation et la construction du corpus sont essentiellement prévues pour démontrer que l'approche PROXYMA est une alternative intéressante aux approches existantes concernant les questions de partage et d'analyse de corpus dans un contexte d'apprentissage assisté par la technologie. Notre hypothèse porte sur l'idée qu'éviter d'imposer une nouvelle représentation qu'il faut adopter pour pouvoir partager et analyser les données, garantit une meilleure acceptation de l'approche. Documenter le corpus avec un ensemble de métadonnées est l'effort le plus important demandé aux chercheuses et aux chercheurs. Une valeur ajoutée importante de ce travail est la visibilité accrue du corpus et la crédibilité des résultats de la recherche que les chercheuses et les chercheurs peuvent alors joindre à leurs publications scientifiques. De plus, cette approche est incrémentale et participative et il n'est pas demandé à une chercheuse qui partage un nouveau corpus de définir tous les scripts nécessaires à l'interrogation de ce corpus. Elle ne définit que les scripts nécessaires à l'analyse qu'elle souhaite effectuer. L'approche PROXYMA présente non seulement l'avantage que les chercheurs peuvent partager des données et des outils d'analyse sans avoir à investir un effort important, mais propose également une solution aux problèmes d'hétérogénéité des traces. En effet, cette approche étant incrémentale, le modèle sémantique peut être enrichi pour ajouter de nouveaux concepts nécessaires à de nouvelles analyses.

Cette expérimentation permet d'évaluer la faisabilité technique et humaine de la procédure. Nous constatons que la création de corpus initiaux partagés sans transformation des traces, prend peu de temps et ne nécessite pas de compétences particulières pour spécifier les éléments du contexte. En revanche, nous pouvons nous attendre à passer plus de temps à écrire les scripts qui exploitent ces corpus et transforment les données pour les faire correspondre aux entrées/sorties des outils d'analyse. En pratique, ces tâches s'avèrent moins difficiles, car elles se limitent aux éléments nécessaires aux analyses, au lieu de tous les éléments dans l'approche classique avec un format pivot. Une perspective intéressante est de développer un générateur de scripts qui utilise le modèle sémantique avec un dictionnaire de synonymes pour générer des scripts qui peuvent être validés ou non par les chercheuses et les chercheurs.

3.5 Conclusion

Dans ce chapitre, l'approche PROXYMA, proposée dans le cadre de la thèse d'Hajer Chebil, ouvre des perspectives aux différents systèmes à base de traces, tels que CARTE, pour le partage et l'analyse de corpus de traces d'interaction contextualisées. Notre volonté d'échapper à un format pivot pour les traces d'activité comme cœur du système à base de corpus, engendre la difficulté à trouver un compromis entre une représentation générique qui couvre différents types de traces, et une représentation spécifique qui permet d'automatiser certains processus d'analyse. Compte tenu des verrous identifiés (l'hétérogénéité et la contextualisation des traces, l'anonymat des données, la description de différents types d'analyses avec différents environnements d'analyse), l'approche PROXYMA constitue une alternative flexible et générique. Nous démontrons qu'un modèle de corpus de traces d'apprentissage contextualisées peut être créé et partagé en associant trois modèles : le modèle de corpus, le modèle sémantique et le modèle opérationnel. Dans le contexte de notre expérimentation, le modèle de corpus permet en effet aux chercheuses et aux chercheurs de conserver et de partager leurs données collectées avec leur formalisme d'origine.

L'utilisation de l'approche PROXYMA pour définir la sémantique des corpus initiaux et d'analyse est une première étape de la validation du modèle sémantique. L'interopérabilité entre plusieurs corpus est définie au moyen des scripts de la couche intermédiaire appelée « PROXY », qui permettent de décrire et d'interroger ces corpus. Les scripts d'extraction, de conversion et de filtrage permettent d'adapter les traces des corpus au format d'entrée/sortie des outils d'analyse, sans modifier leur format d'origine. Un travail d'analyse effectué sur un ou plusieurs corpus est représenté à l'aide du modèle de corpus d'analyse. Ce type de corpus permet à d'autres chercheuses ou chercheurs de récupérer les ressources d'analyse avec leur interprétation, leur documentation et les publications scientifiques correspondantes. Comme les données collectées sont toujours disponibles dans les corpus, les scientifiques peuvent mener leurs propres analyses pour vérifier ou comparer leurs résultats.

Avec ce projet « **Personnalisation des E.I.A.H.** » du pôle de recherche Cluster I.S.L.E. (Informatique, Signal, Logiciel embarqué) de la Région Rhône-Alpes, nous avons atteint nos objectifs présentés à travers les trois questions de recherche. Le partage de corpus provenant d'environnements et de contextes d'apprentissage différents devrait nous aider à valider nos modèles, notamment le modèle sémantique qui pourrait certainement être enrichi. Bien que la charge de travail du chercheur ait été diminuée et que les formats d'origine des traces aient été conservés, nous devons démontrer que ce système est adapté en pratique à différentes méthodologies d'analyse.

Ces travaux de recherche ouvrent des perspectives en termes d'exploitation des données du *World Wide Web*, pour associer d'autres ressources aux traces d'interaction en vue d'élaborer des profils de compétences. Ce travail est donc considéré, entre autres, dans le cadre du projet A.N.R. « Capture des Traces opérationnelles des acteurs d'entreprise pour construire le Capital humain et définir les Processus gagnants » (CatCap 2019-2023), où des modules d'analyse issus de l'intelligence artificielle permettront de qualifier les activités associées aux compétences effectives à partir des traces. Nous intégrons ces résultats dans notre réflexion sur la conception de systèmes à base de compétences qui sont présentés dans la partie 4 de ce document

4

Des traces aux compétences

La complexité des systèmes à base de compétences est concomitante à celle des systèmes à base de traces d'interaction ayant l'activité humaine comme dénominateur commun.

Sommaire

4.1	Introduction	100
4.2	Travaux de recherche	102
4.2.1	Les compétences	102
4.2.2	Le positionnement de notre instrumentation	107
4.2.3	Les outils de traitement de données pour les compétences	112
4.3	Expérimentation de recherche de compétences	119
4.3.1	Les objectifs de l'expérimentation	119
4.3.2	Le contexte de l'expérimentation	119
4.3.3	Les résultats de l'expérimentation	120
4.4	Expérimentation d'identification de situations d'innovation	130
4.4.1	Les objectifs de l'expérimentation	130
4.4.2	Le contexte de l'expérimentation	131
4.4.3	Les résultats de l'expérimentation	132
4.5	Conclusion	136

4.1 Introduction

Avec la transformation numérique des organisations dans le contexte de l'industrie du futur, la conception et la production industrielles sont profondément modifiées. Les grands défis de cette révolution ne sont pas seulement technologiques, mais aussi liés au capital humain (Pinzone et al. 2017). La complexité croissante des systèmes à réaliser conduit les organisations professionnelles à devoir élargir leur *pool* de compétences. La stratégie consiste donc à rechercher des compétences chez des partenaires extérieurs, ce qui préfigure la notion d'organisations étendues (ou *extended enterprises*) qui doivent se reconfigurer en permanence pour s'adapter aux besoins des projets, et ainsi préserver leur compétitivité face à la concurrence. On observe par ailleurs l'accélération de la communication entre les individus, résultant de l'utilisation des systèmes numériques essentiellement sur Internet, qui se traduit par plus de réactivité des organisations professionnelles. La mise en réseau d'individus susceptibles de participer à une activité professionnelle collaborative peut donc s'appuyer avantageusement sur ces systèmes numériques pour réaliser cette recherche. Par exemple, la réponse à un appel à projets qui nécessite l'identification dans un délai limité des partenaires potentiels issus d'un large réseau, lesquels peuvent être sélectionnés pour leur expérience commune, leurs compétences, et parfois leurs accointances.

En ce qui concerne le concept de compétence, les projets de recherche impliquant des traces numériques peuvent être classés en deux groupes :

- les projets inventoriant des compétences : les recherches sur l'industrie du futur et la gestion des ressources humaines ont conduit à la création de référentiels de compétences relativement structurés, fondés sur des déclarations de besoins ou d'offres d'emploi (Hecklau et al. 2016)(Graczyk-Kucharska et al. 2018) ;
- les projets identifiant des écarts de compétences : de nombreux travaux sur les compétences et l'exploitation des traces numériques visent à combler l'écart entre les compétences attendues par le marché de l'emploi et les curriculum vitæ des candidats (Miranda et al. 2017).

L'extraction de connaissances à partir de données numériques est un enjeu majeur pour la compétitivité des organisations (Coron 2019). Les techniques de traitement et de recherche informatiques ont considérablement évolué ces dernières années (Belmonte et al. 2020). En plus de considérer les données déposées sur le *World Wide Web*, nous nous intéressons aux données issues de la dynamique des systèmes collaboratifs utilisés au sein des organisations (i.e. traces d'interaction) et aux données qui expriment les activités « métiers » numériques ou non (e.g. *user stories* dans un Kanboard¹).

Nous partons du postulat que l'analyse des activités collaboratives fournit, entre autres résultats, des informations sur les compétences des utilisateurs dans des domaines spécifiques. Ces activités peuvent être :

- observées en dehors de tout système numérique (i.e. observation humaine, entretien) ;
- formalisées explicitement (i.e. *explicit trace sources*) dans des logiciels spécialisés (e.g. gestionnaire de tâches) ;
- stockées dans des historiques de logiciels pour des analyses à posteriori ;
- identifiées à travers des fonctionnalités de niveau « métier » dans certains logiciels ;
- déduites de l'analyse des traces d'interactions réalisées dans les systèmes numériques, à l'aide d'une instrumentation et d'un système à base de traces.

1. <https://kanboard.org/> est un outil support à la gestion de projet basé sur la méthode Kanban

Certains environnements informatiques collectent les traces d'interaction des utilisatrices et des utilisateurs dans leur base de données interne (e.g. Moodle², Alfresco³), et fournissent leurs propres indicateurs pour analyser les activités au moyen de filtres configurables (e.g. les rapports dans Moodle). Ces outils de suivi permettent généralement d'évaluer l'utilisation d'un document sur une période donnée, ou de contrôler la participation et de fournir diverses données statistiques. C'est le cas notamment d'un module de *Learning Analytics* intégré dans la plate-forme Moodle que nous avons expérimenté dans le cadre de notre participation cette année au livrable L10 du WP3 « API Moodle interuniversitaire, *Learning Analytics* et offre de formation mutualisée », du projet HyPE-13⁴.

Dans nos travaux antérieurs, l'étude croisée des traces de bas niveau d'abstraction associées aux actions réalisées dans différents environnements logiciels, permettait d'identifier des activités « métiers » par construction de traces de plus haut niveau d'abstraction (i.e. les activités). Cette construction implique l'utilisation d'un système à base de traces pour transformer les éléments de traces d'interaction en prenant en compte l'hétérogénéité de leur format.

Nous avons reconsidéré cette approche pour l'étude des compétences à cause essentiellement des difficultés d'opérationnalisation avec les systèmes utilisés, liées aux restrictions en termes d'accès pour l'instrumentation de certains outils logiciels et à la configuration fastidieuse des modèles d'utilisation associés aux traces pour chaque source. Comme nous l'avons indiqué dans la partie 1 de ce document, les deux **thématiques** « Observation des activités humaines » et « Reconnaissance des compétences individuelles et collectives » ont en commun l'activité humaine à partir de laquelle nous collectons des traces que nous exploitons en vue d'explicitier nos observations.

Pour l'observation des traces dans la première thématique, nous avons proposé l'approche originale PROXYMA (*PROXY for Multiple Analyses*) présentée dans la partie 3. Quant à la deuxième thématique, la plupart des outils logiciels étant accessibles via Internet, nous élargissons notre instrumentation aux techniques du *World Wide Web*, comme le *Web scraping*, que nous pouvons coupler à des modules d'analyse de contenu textuel de l'intelligence artificielle. Le niveau d'abstraction des informations que nous pouvons recueillir à l'aide de cette instrumentation est alors celui des outils logiciels utilisés. Selon la nature de ces outils (e.g. Kanboard⁵), l'instrumentation permet potentiellement de recueillir directement des traces de haut niveau d'abstraction (e.g. des activités).

Dans nos travaux actuels, nous étudions les compétences mobilisées à travers l'activité réalisée. Nous expliquons dans cette partie les limites précitées de notre approche d'analyse par les traces d'interaction sur les plans technique et humain. En d'autres termes, si les traces d'interaction de bas niveau d'abstraction reflètent bien les actions effectives, comme indiqué précédemment, leur exploitation pour en déduire les activités implique des contraintes techniques lourdes en termes d'instrumentation et de traitement (i.e. configuration, transformation de séquences, filtrage, etc.). Quant à l'implication humaine, des modèles d'utilisation des outils logiciels tracés doivent être spécifiés pour apporter une sémantique aux traces. Nous couplons donc délibérément les problématiques des compétences et celles des traces de haut niveau d'abstraction, c'est-à-dire placées directement au niveau de l'activité explicitée (e.g. *user story* des méthodes Agiles). Nous exposons nos choix pragmatiques pour contourner les contraintes du bas niveau, que nous exposons sur des terrains d'expérimentation réels à travers un modèle de compétence que nous présentons ci-après.

2. <https://moodle.org/>

3. <https://www.alfresco.com/fr/>

4. <https://hype13.fr/livrables/> pour le développement d'une A.P.I. Moodle pour un outil de *Learning Analytics* à déployer par les établissements

5. <https://kanboard.org/> est un outil support à la gestion de projet basé sur la méthode Kanban

Dans cette partie, nous présentons deux **expérimentations** qui illustrent les deux approches présentées précédemment, à savoir celle basée sur une instrumentation de type « bas niveau » avec la construction d'un corpus de traces gardées dans leur format initial et celle basée sur une instrumentation de type « haut niveau » en utilisant des outils logiciels qui manipulent des informations sur les activités elles-mêmes.

En d'autres termes :

- la première expérimentation consiste à valider la capacité via l'approche PROXYMA à récupérer des traces d'activités associées à un travail collaboratif assisté par ordinateur, pour identifier des compétences à partir de la reconnaissance de l'implication des individus dans certains domaines de compétences ;
- la deuxième expérimentation s'appuie sur l'architecture SBS-SEARCH appliquée à quatre contextes expérimentaux liés à des activités d'innovation, en vue de caractériser les compétences associées à des activités d'exploitation ou d'exploration.

D'autres expérimentations ont été menées sur l'identification des compétences dans le cadre du projet A.N.R. CaTCaP avec l'entreprise partenaire MANASLU, mais à défaut d'être présentées dans ce document (travaux en cours), nous nous focaliserons sur les voies potentielles d'analyse des données qu'elles ont ouvertes, notamment en termes de représentation graphique des compétences mobilisées en milieu industriel. Ces expérimentations préfigurent les systèmes de recommandation de compétences, dont les fonctionnalités s'appuieront sur les métriques issues de la théorie des graphes (liens forts et faibles, trous et ponts structurels, coefficient de *clustering*, etc.), dont nous donnons ci-dessous un aperçu de leur potentiel en termes d'analyse. Par exemple, un *cluster* de tâches similaires permet d'introduire la notion de famille de tâches utilisée pour leur référencement. Pour mesurer les similarités entre les tâches, nous avons besoin de définir des distances entre celles-ci et la simple utilisation d'une taxonomie des noms de tâches ne suffirait pas à reconnaître des tâches similaires, car le contexte de leur exécution ne serait pas pris en compte. En effet, les tâches de « programmation » diffèrent selon qu'elles sont effectuées dans le domaine de l'informatique ou celui de l'événementiel. Selon la nature des traces, il est possible de s'appuyer sur leurs méta-données (notamment la description ou l'outil utilisé) pour extraire ces informations quand elles sont disponibles. L'idéal serait donc d'analyser tous les éléments contextuels des traces afférentes et d'utiliser les techniques de l'intelligence artificielle pour faciliter les opérations généralement fastidieuses, en appliquant par exemple des algorithmes de *clustering*, comme « *K-Mean* », « *Affinity Propagation* » ou « *Hierarchical Clustering* » (Li et al. 2015) (Xu and Tian 2015) (Refianti, Mutiara, and Syamsudduha 2016).

4.2 Travaux de recherche

4.2.1 Les compétences

Le concept de compétence a fait l'objet de nombreuses définitions depuis les années 90 qui ne permettent pas jusqu'à présent d'extraire une perception commune de ce qu'est une compétence. Comme le souligne Winterton, « *it is impossible [...] to arrive at a definition capable of accommodating and reconciling all the different ways that the term is used* » (Le Deist and Winterton 2005). Le terme « compétence » a été introduit par McClelland en 1973 pour désigner des habitudes personnelles qui conduisent à une performance professionnelle supérieure (Benbrahim et al. 2018), avant d'introduire une méthodologie d'évaluation des compétences dans une société de conseil

devenue Hay McBer ([Kou, Jia, and Wang 2013](#)). Dans la littérature scientifique, les Anglo-Saxons distinguent les termes « *competency* » et « *competence* ». La « *competency* » indique la relation de cause à effet entre les caractéristiques d'un individu exécutant des activités et sa performance professionnelle, en considérant la réussite et l'efficacité. Quant à la « *competence* », elle met en évidence le lien entre un ensemble complexe de connaissances, d'aptitudes, de motivations, de comportements, et une adéquation au poste dans un contexte professionnel spécifique ([Boumane et al. 2006](#)) ([Belkadi, Bonjour, and Dulmet 2007](#)) ([Miranda et al. 2017](#)). En résumé, la « *competence* » signifie ce que les individus peuvent faire et la « *competency* » signifie comment ils le font ([Sanghi 2016](#)).

([Belkadi, Bonjour, and Dulmet 2007](#)) proposent un modèle global de compétence basé sur la modélisation des situations de travail. Dans leur étude de cas, les compétences sont définies en analysant à la fois les activités cognitives des concepteurs et les situations de conception collaborative. Au-delà de la constitution d'une équipe de projet, cette approche vise à construire une mémoire de projet orientée vers la caractérisation des compétences. ([Ley and Albert 2003](#)) caractérisent les compétences d'un utilisateur par ses connaissances, ses aptitudes et ses capacités relatives à une situation spécifique. ([Castillo et al. 2016](#)) définissent des profils de compétences dans le domaine de la santé, qui sont spécifiques au contexte, et où chaque énoncé de compétence représente une responsabilité ou une activité clé pour une catégorie donnée de professionnels de la santé.

En accord avec le postulat de la plupart des auteurs, nous considérons que les compétences dépendent des domaines d'application et sont donc associées aux connaissances correspondantes. Habituellement, les profils de compétences sont caractérisés par les données personnelles (domaines d'expertise, responsabilités, etc.) et les ressources (rapports, entretiens, articles scientifiques, etc.). Ces données sont généralement déclarées et statiques, jusqu'à ce que les utilisateurs les mettent à jour eux-mêmes. Nous pouvons ainsi observer des abus, tels que l'auto-proclamation en tant qu'expert dans un domaine. Nous posons donc comme hypothèse que ces profils pourraient être enrichis automatiquement en tenant compte des activités des utilisateurs, telles que leur participation à des forums ou à des systèmes de recommandation.

Conformément à notre approche scientifique, la construction des compétences dans une activité professionnelle implique l'identification des tâches réalisées et la construction des ressources cognitives nécessaires à leur réalisation, à leur mobilisation et leur combinaison en situations authentiques, comme le définit ([Tardif 2017](#)). Si de nombreux modèles conceptuels de compétence réalisés « à vide », i.e. en dehors des situations réelles donc sans définition du contexte, ne répondent pas à notre approche scientifique basée sur l'activité effective, nous les considérons en tant qu'ils décrivent les composants et les principes fondamentaux qui les unissent.

La thèse de Kamel-Eddin Bemmami intitulée « Modélisation d'un système à base de compétences individuelles et collectives » adresse des problématiques liées à l'industrie du futur caractérisée, entre autres, par le capital humain. Avec la transformation numérique des organisations dans l'ère de l'industrie 4.0, l'inter-connectivité est généralisée avec une invisibilité croissante d'une partie importante de l'activité, qui est supportée par des systèmes automatisés et auto-régulés. Une telle évolution demande aux acteurs de l'entreprise de nombreux apprentissages et la redéfinition de nombreuses pratiques formelles et informelles du travail. Tous ces apprentissages reposent sur des compétences individuelles et collectives, implicites et explicites, existantes et à inventer. Cette thèse s'inscrit dans le projet ANR CaTCaP qui vise à proposer des instruments pour d'une part identifier et extraire ces compétences à partir des situations singulières où elles se déploient et d'autre part, les gérer en lien avec les processus métiers, et favoriser leur diffusion et réutilisation. Le travail de cette thèse porte sur la formalisation et la pérennité des compétences extraites à partir des traces opérationnelles. Par formalisation, nous entendons caractérisation des compétences

et représentation à l'aide d'un formalisme graphique facilitant leur exploitation. Un premier modèle des compétences individuelles est élaboré dans le cadre de cette thèse (cf. Figure 4.1) pour prendre en compte l'activité et sa dynamique dans la construction des compétences (Bemmami et al. 2021).

Le modèle part d'un objectif défini comme spécifique, mesurable, atteignable, réalisable, réaliste et limité dans le temps (i.e. S.M.A.R.T.). Cette spécification de l'objectif tend à fournir un cadre formel et structuré pour comprendre les résultats attendus et les activités qui en découlent.

Un exemple d'objectif SMART issu d'une expérimentation avec l'entreprise partenaire MANASLU du projet ANR CaTCaP serait : simuler la performance énergétique d'au moins 50% des bâtiments du projet « X » dans les quatre prochains mois. Une fois l'objectif défini, un ensemble d'activités à réaliser est défini. Chacune de ces activités vise à répondre à un besoin bien précis et à contribuer à la réalisation de l'objectif initial.

En considérant l'objectif donné en exemple, nous pouvons définir les trois activités suivantes :

- Activité 1 : modélisation des bâtiments / Résultat attendu : modèles numériques pour au moins 50% des bâtiments reproduisant le plan architectural et respectant toutes les mesures définies ;
- Activité 2 : simulation énergétique / Résultat attendu : taux de déperdition de chaleur et demande d'énergie sur les 4 saisons ;
- Activité 3 : rédaction du rapport / Résultat attendu : un document comprenant les résultats de l'étude énergétique d'au moins 50% des bâtiments.

Chacune de ces activités est réalisée par un acteur selon un schème⁶ spécifique. Le schème est guidé par les exigences de l'activité (e.g. les procédures d'activité, les spécifications, les normes techniques) en termes de ressources environnementales et personnelles à mobiliser. Par exemple, en réalisant l'activité 1, l'acteur exécute une séquence d'actions, comme la reproduction du plan en deux dimensions, le passage à la conception 3D et la définition des matériaux. À travers ces actions, l'acteur mobilise, d'une part, des connaissances (e.g. lire un plan d'architecte), ses aptitudes (e.g. concevoir à partir d'un plan technique) et certaines attitudes (e.g. la rigueur et l'autonomie dans le travail). De plus, il mobilise des ressources logicielles (e.g. Design-BuilderTM, Adobe ReaderTM) et du temps de travail.

À la fin du schème, nous vérifions si l'activité est réalisée, ce qui nous amène à deux cas possibles.

- Dans le premier cas, l'activité est entièrement réalisée avec des résultats réels (e.g. 60% de tous les bâtiments sont modélisés). Ces résultats sont ensuite utilisés pour évaluer la performance en fonction des résultats attendus (e.g. moins de 30% des bâtiments sont modélisés = non performant ; entre 30% et 50% des bâtiments sont modélisés = assez performant ; 50% des bâtiments sont modélisés = performant ; entre 50% et 80% des bâtiments sont modélisés = très performant ; plus de 80% des bâtiments sont modélisés = extrêmement performant). En fonction de la performance, nous avons soit une spécification d'amélioration, soit un processus d'analyse de la performance passée. Dans la première situation, les performances sont inférieures au seuil d'acceptabilité (e.g. meilleure gestion du temps de travail, formation envisagée pour la maîtrise du logiciel de CAO). Dans la seconde situation, le processus vise à déterminer si cette performance a été régulièrement reproduite dans d'autres contextes. Si c'est le cas, cette performance est considérée comme une preuve de compétence (e.g. dans l'utilisation d'un logiciel de CAO, la conception à partir d'un dessin technique) (Nagels 2009). Sinon, aucun jugement ne peut être donné sur la compétence.

6. enchaînement d'actions pour réaliser une activité

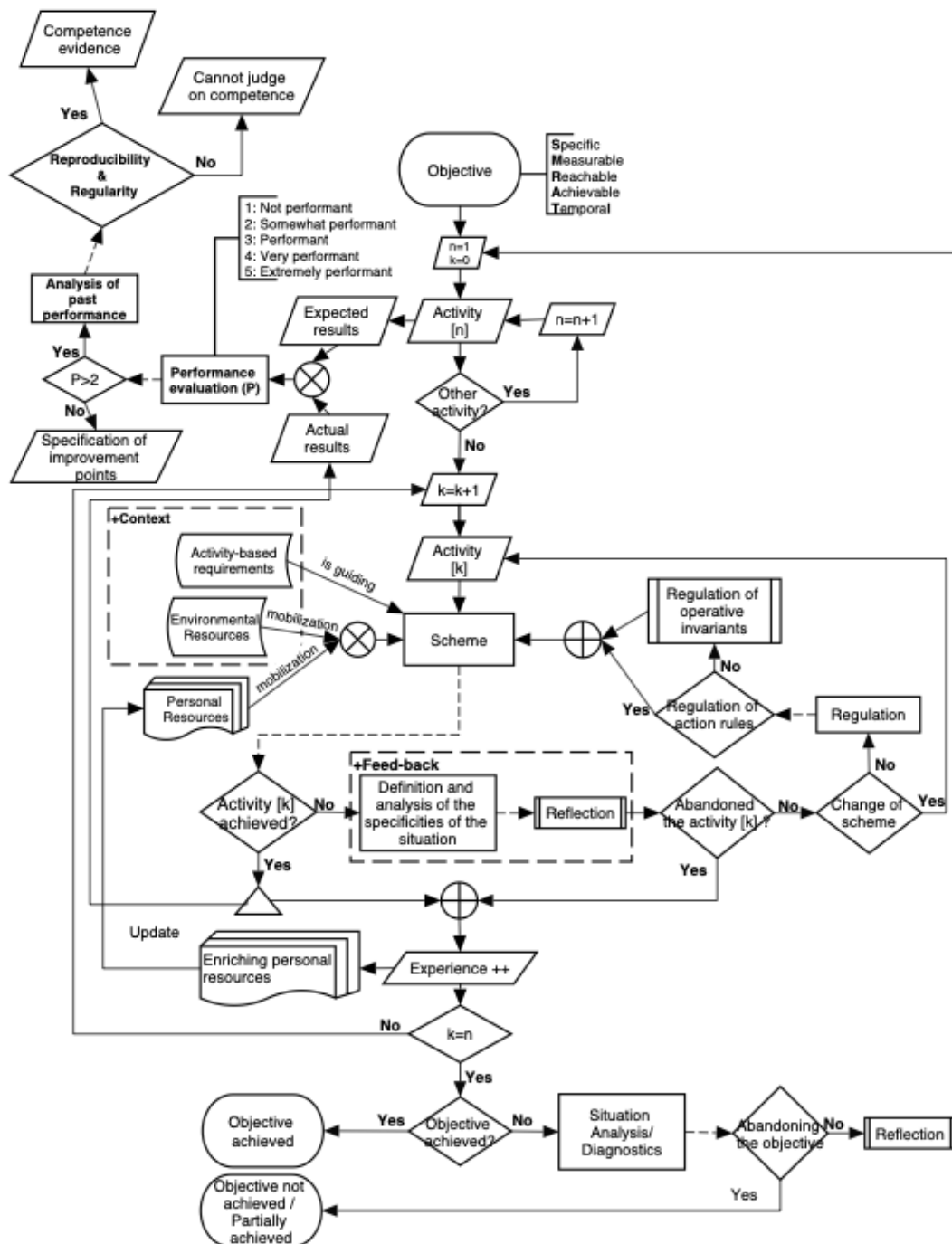


FIGURE 4.1 – modèle de compétence basé sur la dynamique de l'activité

- Dans le second cas, l'activité n'est réalisée que partiellement ou pas du tout. Dans ce cas, un processus est lancé comprenant la définition et l'analyse des spécificités de la situation de travail (contexte), afin d'identifier les causes potentielles. Dans le cas de l'activité 1, cette analyse peut indiquer, par exemple, que certains devis ne sont pas inclus dans le plan initial, ou que les volumes ont une grande complexité géométrique. Les résultats de ce processus conduisent à un sous-processus de réflexion qui permettra de déterminer si l'activité sera abandonnée ou non. Si nous poursuivons l'activité, nous devons soit modifier le schème actuel, soit le réguler pour surmonter les contraintes liées aux situations de travail. Par exemple, au cours de l'activité 1, si certaines cotes sont manquantes dans le document initial, le schème sera régulé pour ajouter une action de « complétion manuelle des cotes manquantes » dans les règles d'action. À la fin de l'activité, qu'elle soit réalisée ou non, nous avons une expérience acquise permettant d'enrichir les ressources personnelles de l'acteur (e.g. des compétences concernant la conception de géométries complexes, ou des connaissances liées à l'utilisation de logiciels de CAO). À la fin de l'exécution de toutes les activités, nous vérifions si l'objectif est atteint. Si ce n'est pas le cas, un processus d'analyse et de diagnostic de la situation de travail est effectué pour déterminer si l'objectif est encore atteignable. Si c'est le cas, un sous-processus de réflexion est lancé pour redéfinir les activités en fonction de la situation de travail actuelle. Cette analyse peut montrer, par exemple, que dans 20% des bâtiments modélisés, le calcul des pertes/gains thermiques en hiver/été n'a pas été effectué. L'une des raisons serait, par exemple, le manque de données sur les types de matériaux isolants ou de béton utilisés (e.g. béton armé, béton à haute performance, etc.). La réflexion menée par la suite permettra de redéfinir les activités pour inclure les échanges avec le bureau d'études pour compléter les données sur les matériaux utilisés ou la finalisation de la maquette numérique.

En résumé, la compétence mobilisée par l'acteur est identifiée en fonction de l'activité réalisée, de la performance observée et jugée, et du contexte d'application. Le choix de la modélisation par diagramme de processus pour ce modèle permet de représenter la construction de la compétence à travers un processus de perception et d'interprétation de la situation en cours.

Les livrables attendus dans le projet CaTCaP sont la définition d'une méthode permettant de cartographier les compétences à l'aide d'un format facilement exploitable. La cartographie ainsi obtenue permet de représenter le capital immatériel de l'entreprise à tout moment. Cette cartographie est dynamique : la méthode permettant de l'élaborer prend en compte les évolutions survenant dans les activités observées et tracées. Le modèle présenté sur la Figure 4.1 intègre la dimension dynamique permettant la création de cartographies dynamiques. Un travail est en cours pour intégrer la dimension collective dans la modélisation des compétences, afin de répondre à la question de recherche sur les caractéristiques des compétences individuelles et collectives et leur représentation pour mettre en exergue les liens qui les unissent.

Dans le cadre de la thèse intitulée « Des compétences individuelles et collectives tracées pendant les activités d'exploitation et d'exploration à la performance » d'Olivier Pouydebat, la proposition d'une architecture nommée SBS-SEARCH⁷ pour opérationnaliser un système à base de compétence (S.B.C.) est présentée sur la Figure 4.2. Cette représentation technique d'un S.B.C. est synthétisée ci-dessous, sachant qu'elle est détaillée dans (Pouydebat n.d.).

Elle comprend les éléments suivants :

- les nomenclatures de tâches (improprement appelées « référentiels de compétences ») indiquant les composants K.S.A. (*Knowledge, Skills and Attitudes* ou Savoirs, Aptitudes et Attitudes) des compétences dans des formats propriétaires ;

7. *Skill-Based System - Skills Extract, Analyze, Regroup, Consult and Historize*

- les nomenclatures consolidées dans un format unifié ;
- un annuaire pour gérer les identités et les profils d'individus ;
- le groupe de cadres expérimentaux pour isoler du système global les outils instrumentés utilisés pendant les expérimentations ;
- les services communs pour les traitements au sein des autres sous-systèmes du système global (e.g. les connecteurs E.T.L. entre les bases de données ; les filtres lexicaux pour traiter les caractères spéciaux ; les modules d'analyse basés sur les techniques de l'intelligence artificielle) ;
- le sous-système d'agrégation d'informations permettant l'agrégation des données inter-contextes permettant de calculer des indicateurs transversaux de plus haut niveau ;
- le sous-système d'analyse et d'extraction de connaissances permettant le calcul d'indicateurs de haut niveau pour des analyses statistiques en suivant le processus K.D.D. (*Knowledge Discovery in Databases*) (Fayyad, Piatetsky-Shapiro, and Smyth 1996) ;
- le sous-système de représentation et de valorisation des connaissances comprenant des tableaux de bord (e.g. outil « Tableau public »⁸) ou des représentations graphiques (e.g. « Neo4J »⁹).

Les éléments de l'architecture ayant fait l'objet d'une implémentation sont représentés en rouge sur la Figure 4.2.

4.2.2 Le positionnement de notre instrumentation

Pour la compréhension de la suite des travaux présentés, nous faisons ci-dessous un rapide rappel des considérations liées à l'instrumentation présentées dans le présent document. Nous montrons dans la partie 2 que les environnements collaboratifs assistés par ordinateur intègrent généralement des modules de collecte de traces, qui recueillent les traces générées pendant les activités des utilisatrices et des utilisateurs. Pour rappel, l'interprétation de ces traces implique la définition de la sémantique d'un modèle d'utilisation associé, c'est-à-dire un modèle décrivant l'usage de l'outil logiciel considéré. Cette association entre une trace et le modèle d'utilisation correspondant est appelée MTrace. Il existe des spécifications standard pour le partage des données, telles que le modèle SCORM¹⁰ (*Sharable Content Object Reference Model*), les métadonnées DCMI¹¹ (*Dublin Core Metadata Initiative*), les métadonnées LOM¹² (*Learning Object Metadata*), et plus récemment xAPI (spécification succédant à SCORM (Bakhouyi et al. 2020)) pour décrire les ressources respectivement issues du réseau et du domaine éducatif.

Nos travaux de recherche antérieurs ont conduit à la modélisation d'un système à base de traces et à une proposition de format appelé « séquence » CARTE (*Collection, activities Analysis and Regulation based on Traces Enriched*) (Courtin 2010), qui représente des éléments de traces soit composites, c'est-à-dire construits à partir d'éléments de traces atomiques, que nous appelons des « signaux », soit atomiques lorsqu'ils sont fournis directement par les outils de façon explicite (e.g. dans un outil de gestion de tâches). Cependant, nous avons souligné la complexité de mise en œuvre d'une instrumentation des outils pour recueillir des traces de bas niveau d'abstraction (i.e. des signaux), et de leurs transformations en vue d'atteindre un haut niveau d'abstraction (i.e.

8. <https://public.tableau.com>

9. <https://neo4j.com/v2/>

10. <https://scorm.com/>

11. <http://dublincore.org/>

12. <https://www.ieeeltsc.org/working-groups/wg12LOM/lomDescription/>

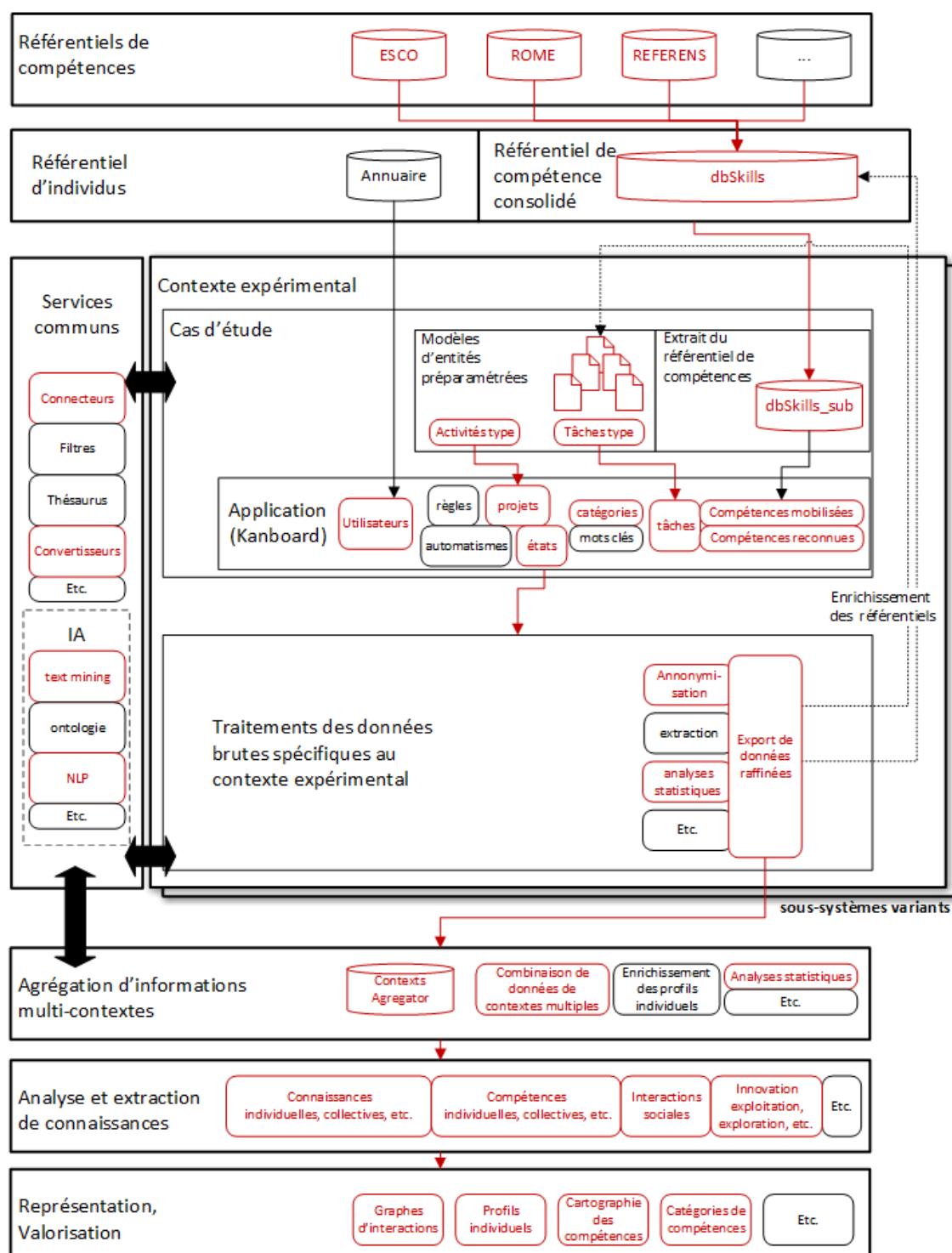


FIGURE 4.2 – Architecture SBS-SEARCH pour l'extraction des compétences

des séquences) qui nécessitent la définition de modèles sémantiques pour être interprétées par des observatrices ou des observateurs. De plus, nous observons que la plupart des systèmes logiciels ne respectent pas les spécifications qu'elles soient standards ou non, et ont leur propre format de traces. Par conséquent, les entrées des outils d'analyse spécifiques dépendent du format des traces des environnements collaboratifs assistés par ordinateur, ce qui rend difficile le partage des traces.

Les activités formant un trait d'union entre leurs traces et des compétences effectives (cf. Figure 4.3), nous identifions alors deux situations : la première où les traces recueillies sont intrinsèquement de bas niveau (e.g. *logs files*) et la seconde où les outils logiciels fournissent directement des informations sur la nature des activités (e.g. *user stories*).

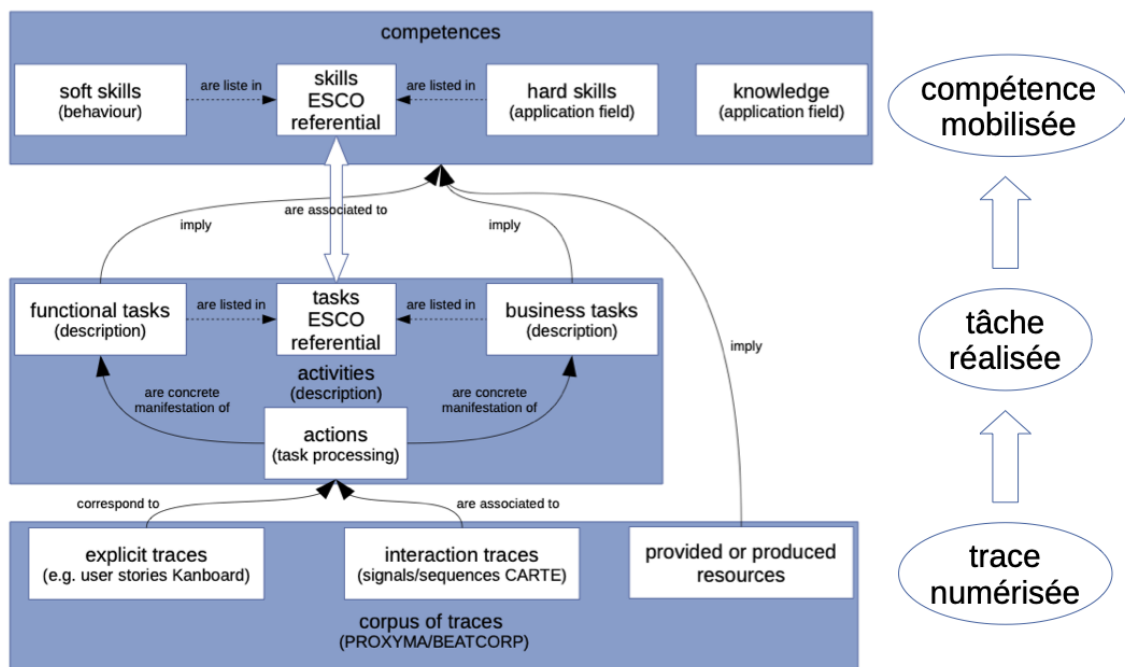


FIGURE 4.3 – les compétences et les traces d'activité

Dans le premier cas, alors que la plupart des approches classiques proposent une représentation pivot à suivre pour rendre possible le partage de traces, l'approche PROXYMA détaillée dans la partie 3 présente les caractéristiques d'un modèle ouvert à tout type de format et donc également aux plus récents, puisqu'il consiste à partager les données dans leur représentation d'origine, et à définir une couche intermédiaire appelée PROXY, qui permet de décrire et d'interroger des corpus de traces.

Pour nos travaux actuels, nous privilégions le second cas où les outils fournissent des traces explicites, assimilées à des tâches fonctionnelles ou « métier », voire à des activités (e.g. *user stories*). Les éléments de traces relatifs aux *user stories* (i.e. libellé de la tâche, description, dates, etc.) de l'outil KBskill sont stockés dans une base de données interne. À défaut de disposer d'un format d'extraction des traces, nous nous appuyons sur le format CARTE pour décrire les tâches correspondantes directement sous la forme de séquences de haut niveau d'abstraction construites à partir de ces éléments de traces numérisées, sans qu'il soit nécessaire de réaliser des transformations via un système à base de traces de bas niveau (cf. Figure 4.4). Dans notre instrumentation des outils logiciels, nous utilisons des nomenclatures de tâches afin d'uniformiser leur intitulé (e.g. le répertoire français opérationnel des métiers et des emplois ROME, la classification européenne des

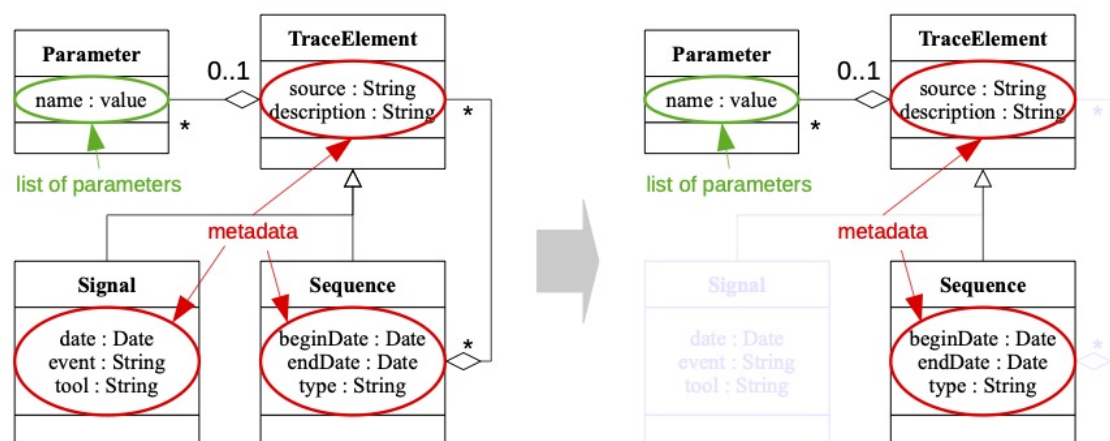


FIGURE 4.4 – séquence directement de haut niveau (à droite)

aptitudes/compétences, certifications et professions ESCO, le référentiel des professions O*NET pour les U.S.A., le référentiel français de compétences des métiers du professorat et de l'éducation REFERENS). Ces nomenclatures permettent d'associer aux tâches des composants du triptyque K.S.A. (*Knowledge, Skills and Attitudes* ou Savoirs, Aptitudes et Attitudes) des compétences. Nous illustrons notre utilisation de ces nomenclatures de tâches sur la Figure 4.5 qui présente un extrait de *user story* telle qu'elle est définie dans le logiciel KBskill¹³. Ces composants peuvent alors permettre d'inférer des compétences mobilisées avec les autres informations décrites dans le modèle de la Figure 4.1.

FIGURE 4.5 – les composants K.S.A. associés aux user stories

L'outil KBskill a été utilisé dans les travaux de thèses soutenue et en cours sur les compétences, respectivement d'Olivier Pouydebat (Pouydebat et al. 2021) et de Kamel-Eddin Bemmami (Bem-

13. outil de suivi des tâches développé en interne sur la base de l'outil *open source* KANBOARD <https://kanboard.org/>

mami et al. 2021), avant d'être abandonné par l'entreprise MANASLU au profit de l'outil propriétaire de planification et pilotage des travaux BeesBusy¹⁴ non *open source*. Le verrou technique lié au code non accessible qu'imposait ce changement d'outil a été contourné par la mise en place d'une nouvelle instrumentation basée sur le *Web scraping*. Les fonctionnalités de connexion aux nomenclatures de tâches n'étant pas disponibles dans BeesBusy, nous avons introduit des modules de fouille de texte appliqués aux descriptions des tâches et aux documents de travail pour identifier les composants K.S.A. des compétences. Le projet CaTCaP prévoit l'embauche de deux personnes informaticiennes intégratrices Web à partir d'octobre 2022 pour développer une plateforme Web, dont l'architecture sera basée sur l'approche PROXYMA pour intégrer, entre autres, les modules d'analyse précités et des modules de visualisation de cartographies des compétences avec les technologies présentées ci-après.

La question de l'instrumentation des outils logiciels s'est posée dans le travail de thèse de Kamel-Eddin Bemmami, sur la modélisation d'un système à base de compétences individuelles et collectives. Conformément aux objectifs du projet CaTCaP de fournir une méthode permettant de cartographier les compétences extraites à partir de l'analyse des traces opérationnelles, une première étape consiste à définir les moyens permettant d'obtenir ces traces. Toutes les sources potentielles de traces sont considérées dans le projet, telles que les E.N.T.¹⁵, les réseaux sociaux, mais aussi les groupes de travail « traditionnels » à l'aide d'observations humaines, d'entretiens et de transcriptions numériques (e.g. de données audio ou vidéo). L'objectif du projet est de développer une nouvelle approche de gestion des compétences basée sur la trace des activités, permettant ainsi de rompre avec les approches prescriptives et hiérarchiques résultant des approches classiques de gestion du travail. Ces approches prescriptives définies le plus souvent par des *managers*, mènent souvent à un contournement des pratiques par les acteurs, car elles ne correspondent pas à leur façon de faire ou à leurs compétences. Si le défaut de prescription libère le processus de découverte des compétences en vue d'établir des profils de compétences, nous nous appuyons également sur des informations complémentaires, comme les compétences statutaires associées aux fonctions ou les compétences déclarées (e.g. LinkedIn, curriculum vitae), pour établir un cadre de référence (e.g. référentiels de compétences « métiers ») pour notre étude. Sur cette base, les traces issues de pratiques présentent plusieurs avantages :

- l'identification des acteurs les plus à même pour exécuter des tâches données selon les contextes qui se présentent ;
- l'émergence de pratiques pertinentes directement issues des acteurs opérationnels ;
- la réutilisation intelligente des compétences capitalisées ;
- la mesure de l'écart entre la réalité et l'affichage des compétences.

La multiplication des sources de traçage offre une certaine garantie d'exhaustivité dans les données traitées. La puissance des machines capables d'appliquer des algorithmes de traitements des données ouvre des perspectives en termes de résultats. Pour autant, les scientifiques ont une responsabilité quant à l'utilisation du *big data* que ce soit en termes de vie privée ou *privacy* (i.e. défaut du respect du R.G.P.D.¹⁶ ; cf. Annexe C), en termes d'éthique (e.g. atteinte à l'intégrité des personnes, gestion de l'illectronisme), en termes d'écologie (i.e. dégradation de la sobriété numérique pour le stockage, la transmission et le traitement des données), ou en termes de stratégie (i.e. sécurité versus compétitivité). Sur le plan opérationnel, les techniques « classiques » d'analyse croisées s'appuient sur des données tabulaires gérées par des S.G.B.D.R.¹⁷. Avec des

14. <https://www.beesbusy.com/>

15. Environnements Numériques de Travail

16. Règlement Général européen pour la Protection des Données

17. Systèmes de Gestion de Bases de Données Relationnelles

données tabulaires, des matrices d'adjacence sont utilisées en croisant les lignes et les colonnes pour identifier les relations entre les données. Aujourd'hui, les *data scientists* disposent d'environnements de développement intégré complets, tels que Jupyter Notebook¹⁸ qui est ouvert et libre, qui supporte plusieurs langages de programmation, qui offre une connexion à une base de données et des possibilités d'affichage graphique de types *bar graph*, *scatter plot* ou *pie chart*. Ces I.D.E.¹⁹ facilitent l'écriture et le partage de codes informatiques indispensables, mais présentent aussi des faiblesses, notamment avec les bases de données relationnelles où l'ajout de nouveaux types de données peut générer une augmentation exponentielle de la taille de la base et conduire potentiellement à des dégradations computationnelles. D'un point de vue traçage pour limiter les données, se pose alors la question du filtrage des données « utiles » et donc du risque qui en résulte de perte de certaines relations entre les données.

4.2.3 Les outils de traitement de données pour les compétences

L'outil logiciel que devons produire dans le projet de recherche A.N.R. CaTCaP doit permettre non seulement de démontrer que l'on peut efficacement, dans le cadre d'une utilisation industrielle, exploiter les traces d'activités pour identifier et cartographier les compétences, ce qui représente un intérêt pour gérer le capital humain de l'entreprise et le valoriser, mais aussi accompagner le pilotage de l'entreprise en termes de renforcement de tel ou tel champ de compétences par exemple.

Sur le plan de la représentation des compétences formalisées, nous travaillons sur la spécification d'une cartographie pour représenter le capital humain des organisations. Dans le but d'opérationnaliser une première représentation graphique des composants K.S.A. (*Knowledge, Skills and Attitude*) des compétences, nous avons développé une interface Web (voir Figure 4.6), dans le cadre du projet « HumaNeTrust »²⁰ (H.N.T.) limitée aux composants « K », déclinée d'un précédent prototype nommé « HumaNetSkills » (H.N.S.) intégrant les deux composants « K » et « A » et implémentant l'algorithme de spatialisation « Force Atlas 2 » (voir Figure 4.7), que nous avons spécifié et développé sur la base des modèles de « PROXYMA »²¹ au sein de la plate-forme « BEATCORP »²² (voir Figure 4.8). L'étude consistait à exploiter les liens socio-professionnels des chercheuses et des chercheurs à travers leurs activités scientifiques et notamment leurs publications sur HAL²³. Dans le cadre de ce premier travail, notre panel d'utilisateurs était circonscrit en exploitant des données publiques des quelques 70 chercheurs du laboratoire SYMME. Le projet H.N.T. a été financé à hauteur de 9000€ pour investir dans deux grands écrans tactiles muraux à la disposition des personnels du laboratoire SYMME, dans des lieux propices à l'activité observée (discussions au coin « pause café » ; voir Figure 4.9) dans les deux bâtiments (la Maison de la Mécatronique à Annecy-le-Vieux et le bâtiment 2 du Bourget-du-Lac). Il s'agissait pour ces personnels de disposer d'informations sur des éléments de compétences au sein du laboratoire, supports aux discussions généralement informelles de par la nature du lieu. Dans ce prototype, les graphes sont construits à partir des données chargées dans une base relationnelle alimentée par des fichiers Json générés via des appels à l'A.P.I. fournie par HAL. Un algorithme permet de repérer des mots-clefs associés à des domaines de connaissances, et d'établir les liens du graphe à partir de calculs d'occurrences de ces mots-clefs au sein des articles en croisant avec les co-auteurs. Pour des raisons de confidentialité, les graphes sont centrés sur les composants « K » et non sur les individus. Une version centrée sur les individus qui permettrait d'identifier des ressources potentielles

18. <https://jupyter.org/>

19. *Integrated Development Environments*

20. AAP Recherche USMB en 2019

21. *PROXY for Multiple Analyses*

22. *BEenchmarking platform for Analysis of Trace CORPora*

23. Archive ouverte « Hyper Articles en Ligne » hal.archives-ouvertes.fr

pour envisager une collaboration, nécessiterait la mise en place d'un mécanisme d'anonymisation ou de pseudonymisation avec demande d'autorisation de contact conforme au R.G.P.D. Cette configuration préfigurerait un système de recommandations de ressources, qui pourraient être des individus sur la base de leur profil de compétences.

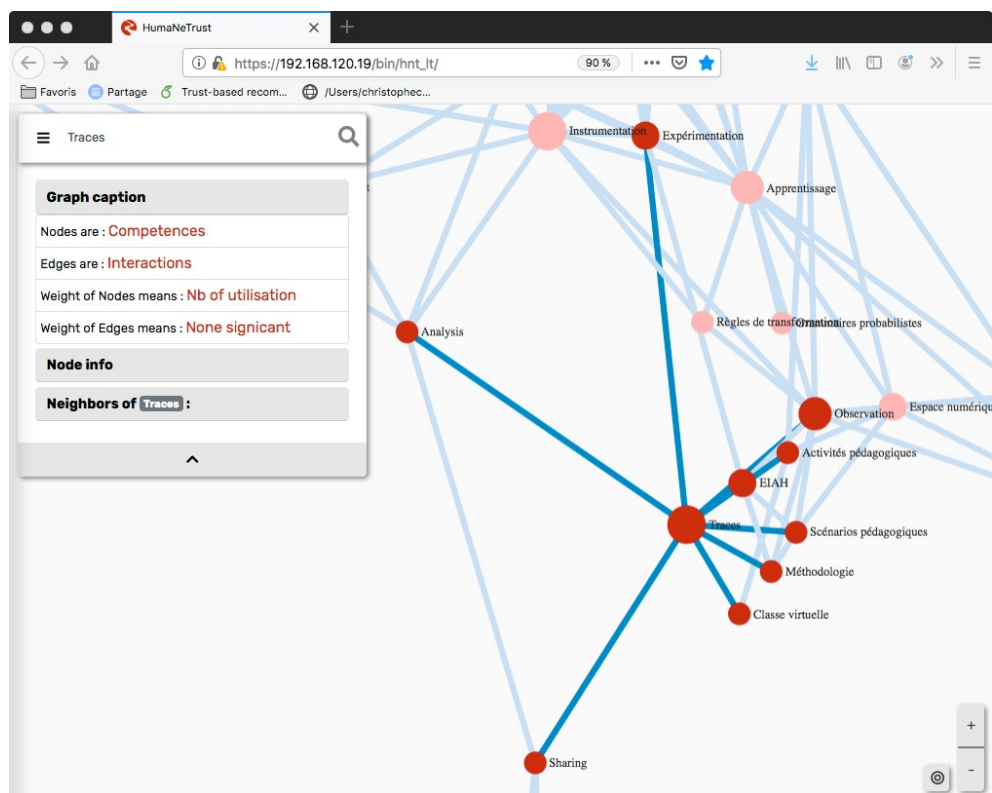


FIGURE 4.6 – prototype du visualisateur H.N.T. avec des données de H.A.L.

Pour tester l'utilité du système dans une organisation avec un objectif de pilotage stratégique, nous avons construit un réseau socio-professionnel résultant de l'analyse de ces données de publication, en vue d'avoir un indicateur de collaborations interdisciplinaires au sein du laboratoire SYMME qui compte cinq sections CNU (voir Figure 4.10). Force est de constater que les collaborations sont inégales, mais à la fois conformes aux effectifs afférents et reflètent la construction historique du laboratoire.



FIGURE 4.9 – interface *H.N.T.* interactive avec des données de *H.A.L.*

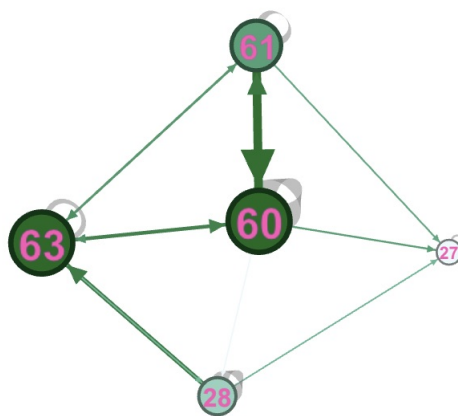


FIGURE 4.10 – *collaborations interdisciplinaires*

Dans notre publication intitulée « *From digital traces to competences* », présentée en juin 2022 à la « *10th IFAC Conference On Manufacturing Modelling, Management And Control* »²⁴, nous exposons l'architecture d'un système d'identification des compétences mobilisées, c'est-à-dire dont les composants sont issus des traces numériques des activités réalisées (cf. Figure 4.11). L'architecture comprend quatre parties : le générateur de traces, le traitement des données afférentes, le calculateur de performances et le visualisateur de cartographies de compétences.

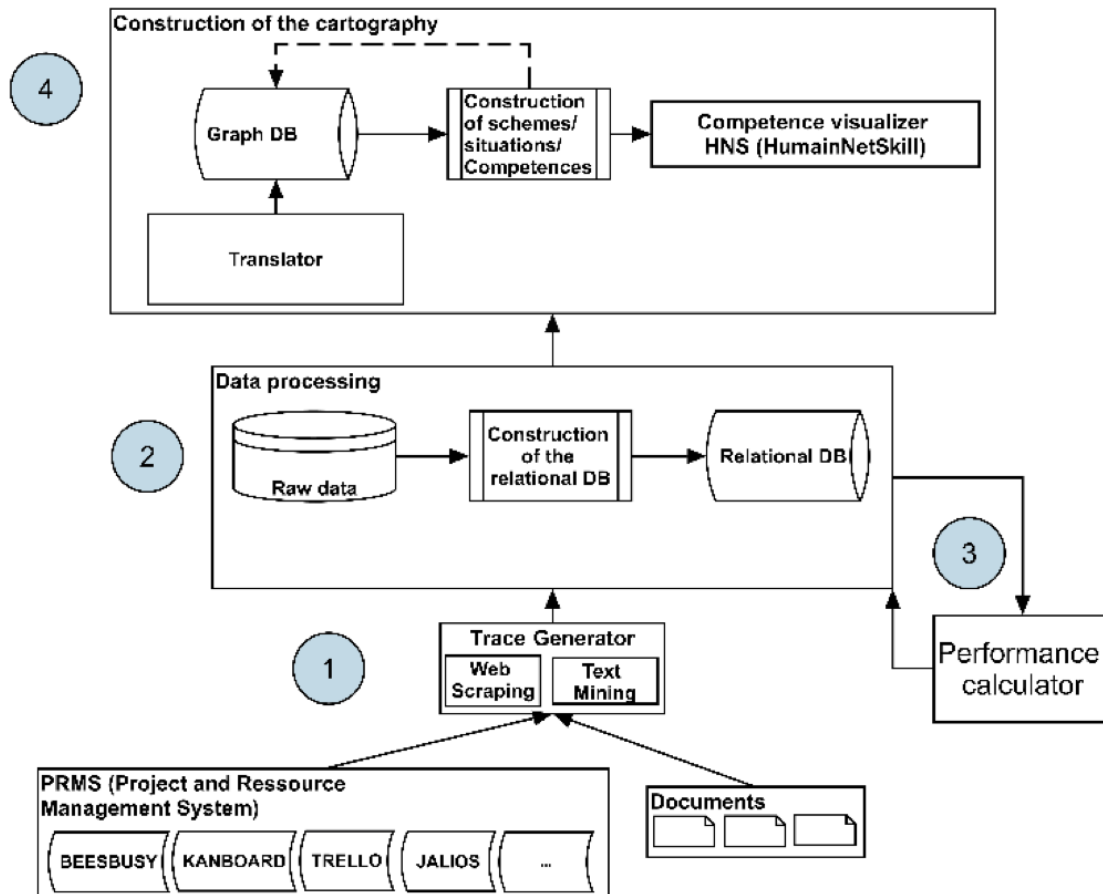


FIGURE 4.11 – architecture d'un système d'identification des compétences

Étant donnée la diversité des sources de traçage liée à l'utilisation de différents logiciels, nous utilisons plusieurs techniques de traçage, comme le *Web scraping* (cf. zone n°1 de la Figure 4.11), que nous qualifions d'« exotechnique » en ce qu'elle est extérieure aux applications, pour extraire les données des systèmes propriétaires de gestion de projets et de ressources (e.g. l'état d'une activité, la durée d'exécution estimée, la date limite de livraison). Dans ce type de système, les traces d'activité numérique peuvent prendre deux formes : les *logs* et les *tags* qui correspondent respectivement à la documentation des événements générés automatiquement et horodatés, et aux métadonnées, telles qu'un *bookmark Web*, un fichier multimédia, un enregistrement de base de données ou un fichier informatique, qui décrivent un élément en vue de le retrouver ultérieurement. L'exploitation des *logs* nécessite un accès complet au code de l'outil, mais la plupart des entreprises utilisent des outils propriétaires qui limitent cet accès. De plus, le contenu des *logs* est

24. <https://hub.imt-atlantique.fr/mim2022/mim2022/> est une conférence internationale portant sur l'ingénierie industrielle et de fabrication, la gestion de la chaîne d'approvisionnement et des opérations, l'informatique, la robotique et l'automatisation, l'intelligence artificielle, la science décisionnelle et la recherche opérationnelle. Référencement de l'article en cours.

le plus souvent très limité en termes d'informations pertinentes sur les activités, les ressources ou les acteurs. Pour toutes ces raisons, nous avons choisi dans le projet CaTCaP de baser le système d'extraction des compétences sur des documents et des *tags*. Pour analyser ces éléments textuels (e.g. les rapports, les dossiers de spécification, etc.), le générateur de traces utilise des algorithmes de fouille de texte pour récupérer les éléments qui décrivent les compétences (e.g. les ressources, les activités, les actions, etc.). Dans notre approche, nous considérons que les données générées par les techniques de *Web scraping* et les algorithmes de fouille de texte sont complémentaires. Nous retrouvons les *tags* et les documents dans le système de gestion de projets et ressources²⁵ BeesBusy utilisé par l'entreprise partenaire MANASLU, mais ils existent également dans l'outil *open source* Kanboard.

Nous utilisons un stockage en base de données relationnelle des éléments issus du *Web scraping* et du *text mining* pour permettre leur exploitation via la langage SQL (cf. zone n°2 de la Figure 4.11). L'intégration dans la base de données implique une étape de structuration de l'information généralement obtenue dans des fichiers CSV²⁶ par les outils de traçage et d'analyse avec les techniques précitées. Cette phase nécessite des transformations pour intégrer le modèle physique des données (M.P.D.) de la base de données relationnelle. Ces transformations peuvent être réalisées par des connecteurs d'outil E.T.L. (*Extract Transform Load*) spécifiquement conçus pour collecter et traiter des données provenant de diverses sources pour ensuite les envoyer dans un espace de stockage (entrepôt, base de données, etc.). Nous avons également utilisé ces technologies pour connecter en temps réel les nomenclatures de tâches (e.g. ROME, ESCO, etc.) à l'outil KBskill.

La performance de l'activité de chaque acteur est évaluée sur la base de deux indicateurs : l'efficacité et l'efficience et en supposant que les ressources sont pertinentes pour atteindre l'objectif (cf. zone n°3 de la Figure 4.11). L'efficacité est attestée par l'état de l'activité (e.g. « évaluée comme réalisée » indique que l'objectif est entièrement ou partiellement atteint) dans le P.R.M.S. Le niveau d'efficience quant à lui peut être calculé, par exemple, à partir de la différence entre la durée estimée pour réaliser une tâche et celle réellement enregistrée. D'autres critères pouvant intervenir dans le calcul de l'efficience, cet indicateur et sa graduation (i.e. non efficient, assez efficient, etc.) doivent être définis avec le maître d'ouvrage du projet considéré afin d'être au plus près de la réalité.

Nous avons choisi une représentation graphique des compétences pour identifier rapidement les éléments importants de celles-ci et les corrélations entre eux. L'utilisation de la théorie des graphes pour une analyse plus poussée est également un point important qui nous amène à choisir une représentation graphique pour la visualisation des compétences. Pour bénéficier de la puissance des outils de représentation graphique existants, nous avons opté pour l'utilisation d'une base de données graphique (cf. zone n°4 de la Figure 4.11). Certains modules de traduction facilitent la transition d'une base de données relationnelle vers une base de données de graphes, comme le traducteur proposé par le système de gestion de bases de données graphique Neo4j présenté ci-dessous. Le langage de requête orienté graphe Cypher²⁷ utilisé par Neo4j permet alors d'identifier et de représenter pour chaque activité les actions nécessaires et les situations dans lesquelles elles se produisent. Avec l'ensemble des éléments identifiés et représentés, nous pouvons identifier les compétences potentielles, en interagissant avec la base de données graphique. Des exemples de cartographies ont été publiés à MIM'2022 dans notre article intitulé « *From digital traces to competences* » (en cours de référencement).

Sur le plan des analyses des compétences tracées, les graphes natifs expriment intrinsèquement les liens entre les données, lesquelles sont représentées par des nœuds. Ainsi, lorsqu'on ajoute un

25. P.R.M.S. pour *Projects and Resources Management System*

26. fichiers de textes séparés par des virgules (*Comma Separated Value*)

27. <https://neo4j.com/developer/cypher/>

type de données dans un graphe, on ne multiplie pas la taille de la base de façon exponentielle. Le *data scientist* va ainsi pouvoir analyser toutes les données utiles grâce aux divers algorithmes de la théorie des graphes avec de bonnes performances du système d'analyse. Les algorithmes les plus utilisés incluent, entre autres, les calculs suivants : le plus court chemin, les chemins géodésiques, les mesures de centralité telles que *PageRank*, la centralité des vecteurs propres, la « proximité » (*closeness*), ou l'inter-positionnement (*betweenness*). Il existe aujourd'hui un panel d'outils d'analyse basés sur la théorie des graphes :

- Neo4j²⁸ offre aux développeurs et aux *data scientists* des outils Java avancés fiables en *open source* pour créer rapidement les applications intelligentes et des flux d'apprentissage automatique, disponibles sous la forme d'un service de *Cloud* entièrement géré ou auto-hébergé ;
- AllegroGraph²⁹ est une technologie (code propriétaire, format RDF-QuadStore) de graphe de connaissances entité-événement, distribuées horizontalement, multi-modèles (document et graphe), qui permet aux entreprises d'extraire des décisions sophistiquées et des analyses prédictives à partir de données complexes et distribuées qui ne peuvent être analysées avec des bases de données conventionnelles ;
- Sones GraphDB³⁰ est une base de données de graphes, développée par la société allemande Sones GmbH (code propriétaire, orienté .Net, langage de requête GraphQL), qui intègre des objets liés à l'aide d'une référence directe vers les nœuds adjacents pour exprimer les relations entre les nœuds/entités ;
- Virtuoso³¹ est un système (code propriétaire, format RDF) de gestion de base de données multi-modèles et une plate-forme *middleware* pour le déploiement de graphes de connaissances ;
- HyperGraphDB³² est un mécanisme de stockage de données sous forme, entre autres, de graphes (open-source Java, modèle propriétaire « Hypergraph ») conçu principalement pour la gestion des connaissances, l'intelligence artificielle et le *Web* sémantique ;
- JanusGraph³³ est une base de données de graphes évolutive optimisée pour le stockage et l'interrogation de graphes contenant des centaines de milliards de sommets et d'arêtes distribués sur un *cluster* multi-machines.

Sur le plan de l'exploitation des compétences capitalisées, il s'agit de fournir une représentation du capital humain qui soit à la fois partageable et exploitable par l'ensemble des acteurs. La représentation graphique des compétences que nous utilisons est inspirée des formalismes de représentation de connaissances proposées dans la littérature (représentations logiques, règles de production, réseaux et graphes sémantiques, ontologies). Nous entendons exploiter les capacités de parcours des graphes, de *clustering*, etc. Enfin, la traçabilité des activités et la formalisation ainsi que la représentation des compétences qui en sont extraites doivent tenir compte des contraintes de confidentialité et de *privacy* (règlementation CNIL et RGPD pour les données personnalisées).

Dans la suite du document, nous tentons d'illustrer les deux situations d'instrumentation présentées précédemment : la première où les traces recueillies sont intrinsèquement de bas niveau et la seconde où les outils logiciels fournissent directement des informations sur la nature des activités.

28. <https://neo4j.com/v2/>

29. <https://allegrograph.com/products/allegrograph/>

30. <https://dbdb.io/db/sones-graphdb>

31. <https://www.openlinksw.com/>

32. <http://hypergraphdb.org/>

33. <https://janusgraph.org/>

À travers la première expérimentation que nous présentons ci-après, nous tentons d'évaluer la pertinence de l'utilisation du *middleware* PROXYMA pour répondre à la contrainte d'hétérogénéité des traces accentuée par la multiplicité précitée, et d'évaluer le potentiel de ces traces à l'aide d'outil de fouille de texte pour identifier des compétences.

La deuxième expérimentation tente de valider l'étude de la caractérisation des compétences associées à des activités d'innovation d'exploitation et d'exploration, à travers l'opérationnalisation du système SBS-SEARCH³⁴. Il s'agit également de valider l'utilisation de traces de haut niveau d'abstraction fournies par l'outil KBskill, pour reconnaître des compétences à l'aide de nomenclatures de tâches contenant des composants K.S.A. (*Knowledge, Skills and Attitudes*) de compétences.

4.3 Expérimentation de recherche de compétences

4.3.1 Les objectifs de l'expérimentation

L'objectif principal de l'expérimentation consiste à évaluer la pertinence des traces d'interaction dans les processus de recherche de personnes (avec différents niveaux d'expertise : intermédiaire, avancé ou expert) ou de ressources (conseils, fichiers, sites *Web*, etc.).

Cette expérimentation doit également permettre d'analyser l'implication des étudiants dans des séances d'apprentissage à l'université, à travers l'observation des activités dans un module de communication. Notre étude de cas consiste à identifier les étudiants qui pourraient intégrer une équipe pour un projet de programmation logicielle. Cette observation pourrait être complétée par l'analyse des productions des étudiants, i.e. de leurs fichiers sources, comme cela est fait par (Vivacqua and Lieberman 2000) qui propose un éditeur de profils de compétences dans le même domaine.

4.3.2 Le contexte de l'expérimentation

Les travaux de la partie 3 sont notamment motivés par la volonté de partager des données d'expérimentation, en vue d'effectuer d'autres analyses. Construire une expérimentation d'apprentissage authentique est une tâche complexe et longue qui implique généralement différentes personnes (apprenants, tuteurs, développeurs, etc.). Nous nous inscrivons donc dans cette logique pour réaliser une nouvelle étude, tout en limitant les coûts humains engendrés par de la mise en place de nouvelles expérimentations. Cette nouvelle étude conforte notre approche PROXYMA / BEATCORP, puisque nous exploitons le corpus « OOD-OOP »³⁵ de l'expérimentation d'une situation d'apprentissage à l'université pendant l'unité d'enseignement « conception orientée objet / programmation orientée objet », décrite dans la partie 3.4.

L'expérimentation dure deux semaines et implique un groupe d'étudiantes et d'étudiants et un enseignant, qui utilisent deux outils de la plate-forme d'apprentissage en ligne Moodle³⁶ : le module de forum pour communiquer entre eux, et le module d'affectation pour déposer leurs travaux. L'expérimentation porte sur quatre séances de travaux pratiques, les étudiantes et les

34. *Skills Based System* (Système à base de compétences) "SEARCH" (*Skills Extract, Analyse, Regroup, Consult and Historise*)

35. *Object-Oriented Design/Object-Oriented Programming*

36. <https://moodle.org/>

étudiants devant envoyer leurs travaux en cours à la fin de chacune d'elles. Pendant le temps de classe, la communication in situ est assurée via le forum, ainsi qu'à travers des échanges verbaux. Entre les séances, le dépôt de questions à distance avec le forum est encouragé (avec des solutions de programmation Java à l'intérieur).

Dans le contexte de cette étude, l'enseignant prévoit de constituer l'équipe d'étudiantes et d'étudiants la plus appropriée, qui sera intégrée dans un futur projet de programmation logicielle. De manière traditionnelle, l'enseignant constitue une équipe parmi les personnes qu'il connaît, ou celles qui sont choisies sur les conseils de ses collègues. Le choix des personnes dépend de leurs propres compétences, et peut dépendre des relations de confiance : entre l'enseignant et chaque étudiante ou étudiant, et entre les étudiants eux-mêmes. Habituellement, ces paramètres, compétences et confiance, sont évalués en observant les activités et les relations humaines dans un contexte professionnel, comme lors d'un précédent projet.

4.3.3 Les résultats de l'expérimentation

Le corpus est créé lors de l'expérimentation présenté dans la partie 3.4 au moyen du *framework* de PROXYMA, caractérisé par les trois modèles afférents : i) le modèle de corpus, ii) le modèle sémantique et iii) le modèle opérationnel (cf. Figure 4.12).

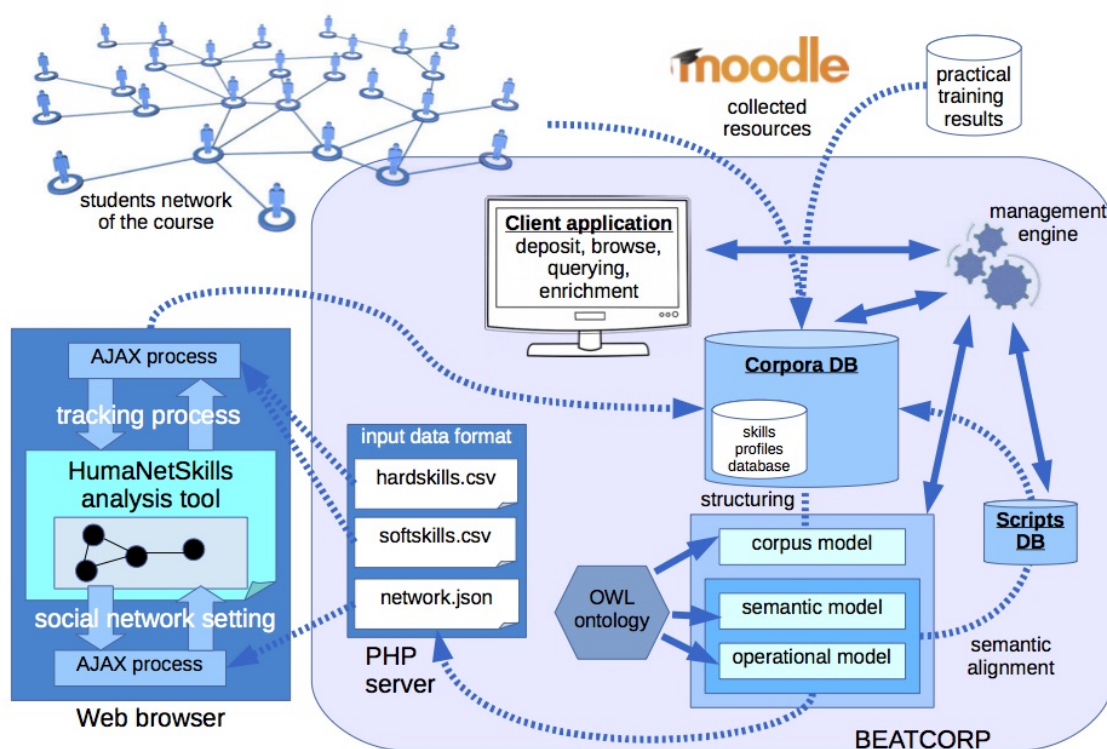


FIGURE 4.12 – L'architecture HumaNetSkills / BEATCORP

Le modèle de corpus

Le modèle de corpus présenté dans la Figure 4.13 définit les ressources et les métadonnées. Les premières contiennent le fichier pdf du sujet de l'exercice, les fichiers sources Java des élèves et les traces d'interaction avec le forum. Les secondes décrivent le corpus en termes de créateur, de description (le nombre de participants, la durée de la session d'apprentissage, etc.), de droits,

etc. Ce corpus contient des traces issues de la base de données Moodle, qui est initialement composée de 276 tables. Les traces sont collectées dans seulement une douzaine de ces tables qui sont relatives aux utilisatrices et aux utilisateurs (e.g. mdl_user, mdl_user_lastaccess), aux cours (e.g. mdl_course) et aux forums (e.g. mdl_forum, mdl_forum_discussions, mdl_forum_posts).

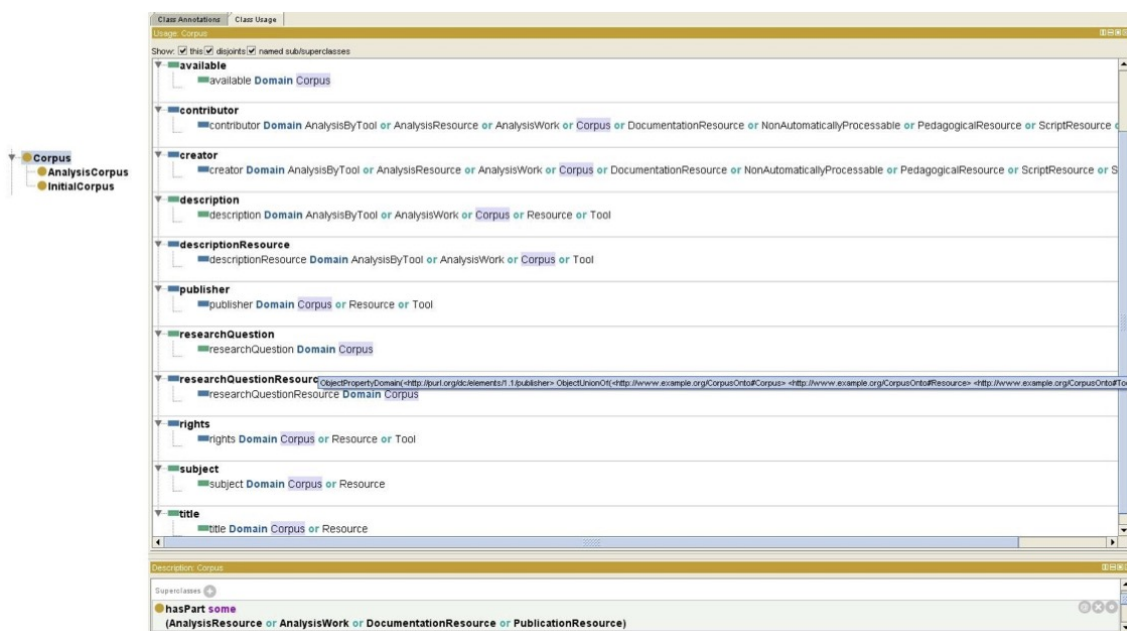


FIGURE 4.13 – Extrait du modèle de corpus (éditeur Protégé ©)

Le modèle sémantique

Avec l'approche PROXYMA, les observatrices et les observateurs peuvent conserver les modèles et les formats de leurs données collectées et les données d'entrée de leurs outils d'analyse. Le modèle sémantique, présenté dans la Figure 4.14, définit une taxonomie de concepts, faisant partie de l'ontologie globale, qui sont utiles pour nos analyses, et qui peuvent être trouvés dans les traces collectées du forum. Ce modèle fournit une base sémantique commune aux observatrices et aux observateurs qui partagent des corpus de traces et des outils d'analyse. Concrètement, un outil de rapprochement sémantique, appelé PROXY, leur permet d'aligner les données contenues dans le corpus avec les entrées des outils d'analyse.

Le modèle opérationnel

Le modèle opérationnel, présenté à la Figure 4.15, contient des scripts pour interroger et extraire les traces du corpus, et des scripts pour les convertir si nécessaire, afin d'aligner le contenu du corpus avec les concepts définis dans le modèle sémantique et les entrées des outils d'analyse. En d'autres termes, les scripts de la couche intermédiaire, nommée PROXY, assurent l'interopérabilité entre les corpus et les outils d'analyse. Pour le corpus existant, nous réutilisons les scripts d'extraction et de filtrage des données qui sont définis dans l'expérimentation dans la partie 3.4. Pour l'étude présentée dans cet article, nous utilisons deux outils d'analyse externes avec des formats de traces différents : Gephi ©³⁷, qui est un logiciel interactif ouvert et libre de visualisation de réseaux (arêtes, rangs, hubs, ...) et TXM³⁸. Le logiciel Gephi © d'exploration des données est

37. <https://gephi.org/>

38. <http://textometrie.ens-lyon.fr/spip.php?rubrique96>

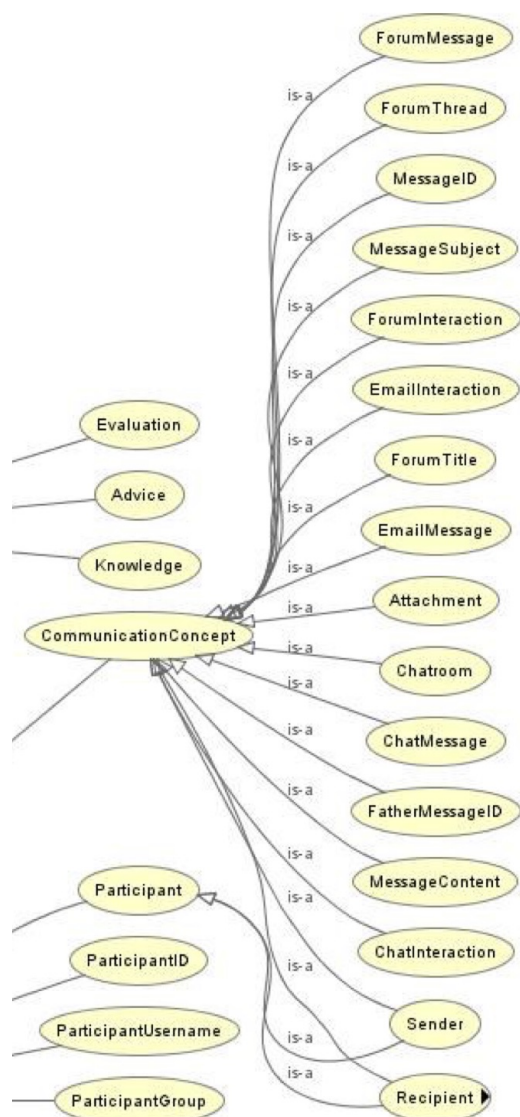


FIGURE 4.14 – Extrait de l'ontologie du modèle sémantique (OWL)

souvent utilisé pour les analyses visuelles dans le domaine de la recherche. Les traces sont formatées pour correspondre aux données d'entrée de l'outil. Par exemple, cela consiste à nommer différemment deux attributs des traces afin de s'adapter à la définition des nœuds et des arêtes du graphe représentant le réseau étudié.

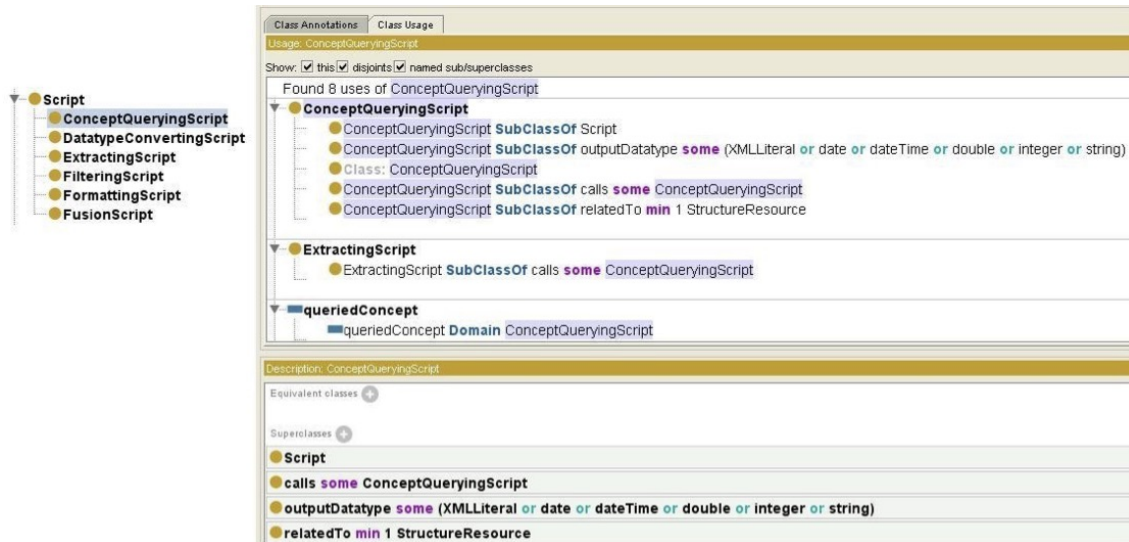


FIGURE 4.15 – Extrait du modèle opérationnel (éditeur Protégé ©)

L'analyse des traces d'interaction

La Figure 4.16 illustre une visualisation graphique des interactions du forum dans l'outil d'analyse Gephi ©. Le format de trace utilisé dépend des entrées attendues de l'outil Gephi ©. Par exemple, avec le CSV (*comma separated value*), les données représentent la structure de la matrice (liste des arêtes et leur poids). Alors qu'avec le format GraphML³⁹, les données contiennent, entre autres, la structure X.M.L.⁴⁰, le poids des arêtes et les attributs. Dans le réseau de la Figure 4.16, les nœuds représentent les messages envoyés avec leur numéro, et les étiquettes des bords nomment leur auteur. Cet outil nous permet de faire une analyse orientée vers l'intuition en manipulant le réseau pour révéler la structure sous-jacente des échanges dans le forum. Sur la Figure 4.17, l'arborescence vue du haut peut être représentée comme une étoile à plusieurs niveaux. Selon l'utilisation du forum dans cette expérimentation, le premier niveau (cercle rouge) représente les différents sujets de discussion, le deuxième niveau (cercle bleu) représente les questions, et le troisième niveau (cercle vert) devrait représenter les réponses ou les explications, mais il pourrait aussi représenter des questions supplémentaires. Quant aux autres niveaux, ils peuvent représenter soit d'autres explications, soit d'autres questions. Nous nous concentrons donc sur le troisième niveau (cercle vert), qui concerne essentiellement les explications dans cette expérimentation (cf. Tableau 4.1). La couleur des nœuds dans les Figures 4.17 et 4.18 représente le score de centralité du vecteur propre⁴¹, qui dépend du score des nœuds auxquels ils sont connectés. Ce score mesure l'influence d'un nœud dans le réseau, et permet d'identifier les messages de la Figure 4.17 dans la zone verte qui devraient être des ressources intéressantes.

En ce qui concerne le forum, il est censé encourager l'entraide entre les élèves, mais les résultats montrent que les étudiantes et les étudiants l'utilisent pour demander l'aide du professeur (auteur nommé « Christophe »), car il est presque le seul à leur répondre comme le montrent les

39. <http://graphml.graphdrawing.org/>

40. *Extensible Markup Language*

41. <http://www.analytictech.com/essex/Lectures/centrality.pdf>

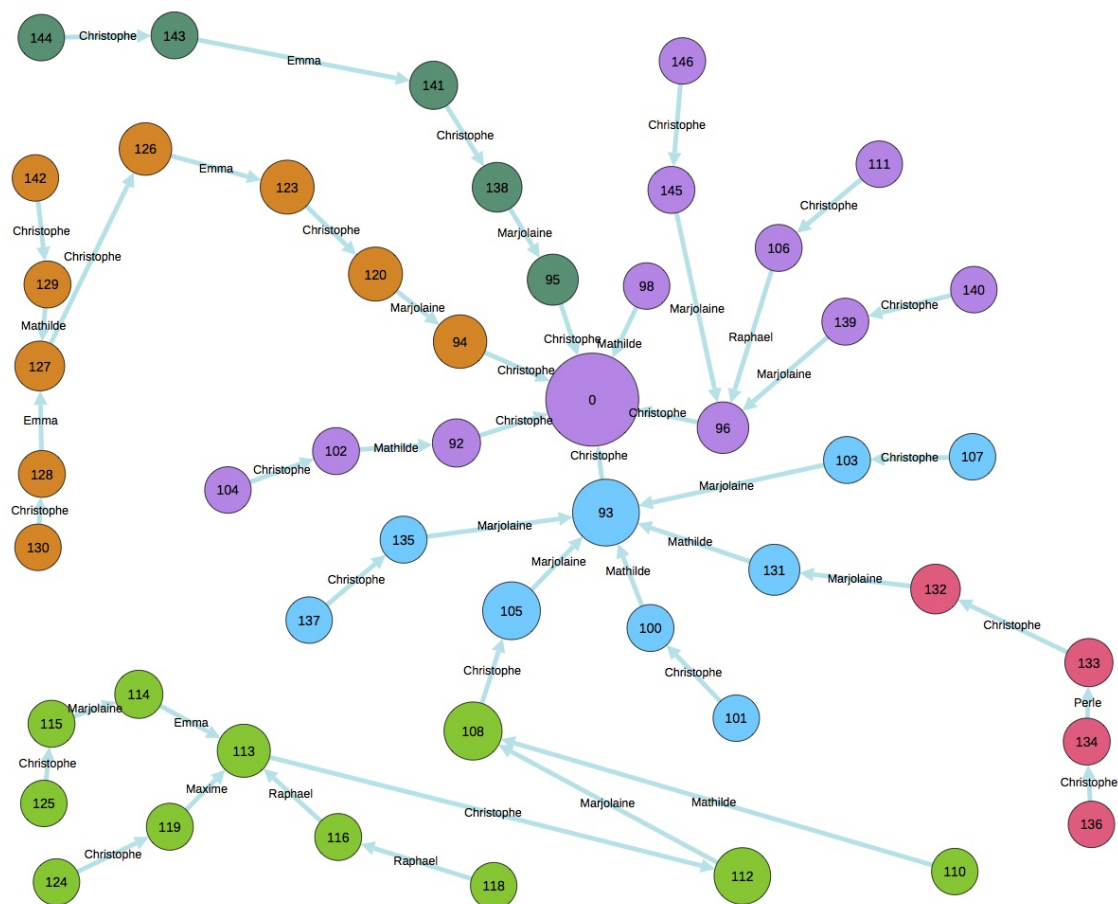


FIGURE 4.16 – Réseau des posts avec l'outil Gephi © (50 nœuds, 49 liens))

TABEAU 4.1 – Distribution des posts (sujets en gris, questions en bleu, explications en vert)

level	posts numbers	answer rate
1	92, 93, 94, 95, 96, 98 (6 posts = 6 subjects of discussion)	0%
2	100, 102, 103, 105, 106, 120, 131, 135, 138, 139, 145 (11 posts)	0%
3	101, 104, 107, 108, 111, 123, 132, 137, 140, 141, 146 (11 posts)	91%
other	110, 112, 113, 114, 115, 116, 118, 119, 124, 125, 126, 127, 128, 129, 130, 133, 134, 136, 142, 143, 144 (21 posts)	52%

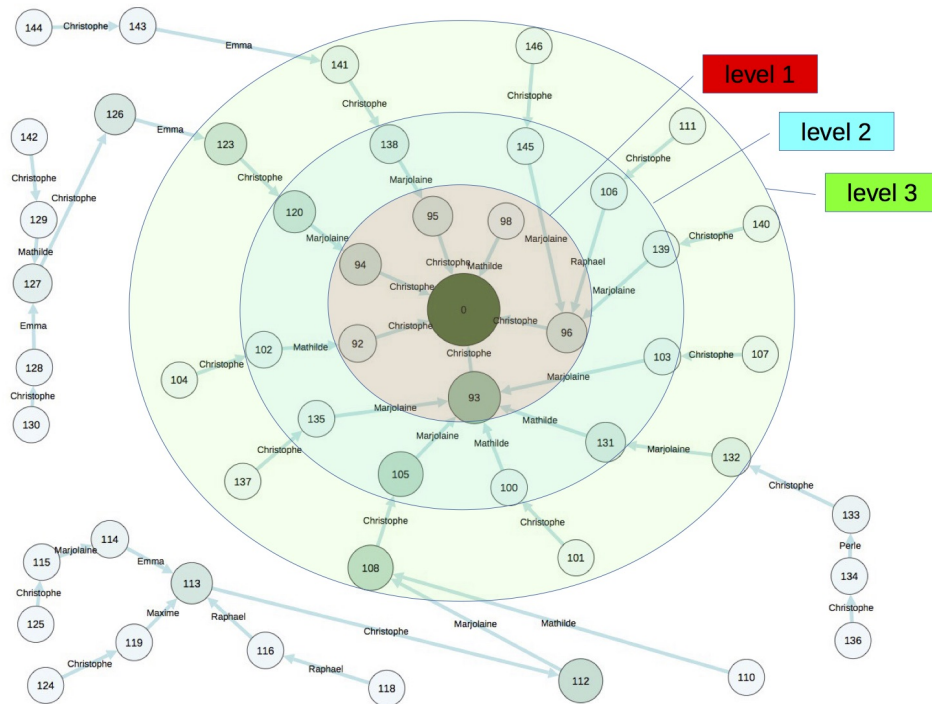


FIGURE 4.17 – Réseau des posts réorganisé

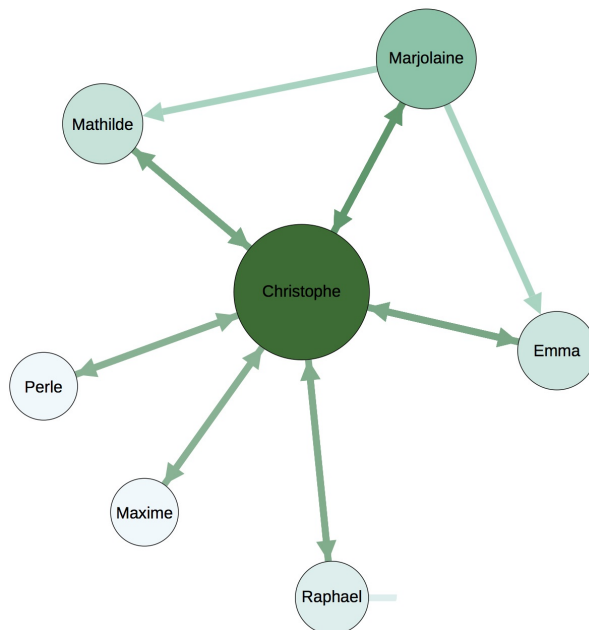


FIGURE 4.18 – Réseau des participants avec l'outil Gephi ©

Figures 4.16 et 4.18. Cette situation illustre des relations verticales : élève-enseignant. Comme la plupart du temps les élèves sont sur place, le forum est utilisé pour capitaliser les explications plutôt que pour communiquer. En outre, les élèves communiquent au moyen de ce forum générique dans l'environnement Moodle⁴², et ils travaillent dans l'IDE Eclipse⁴³, qui est un autre environnement. (George and Labas 2005) précise que l'utilisation d'un outil de communication qui n'est pas intégré dans l'environnement de travail n'incite pas les élèves à l'utiliser pour discuter des activités en cours de réalisation. Malgré ces difficultés, nous pouvons constater la pertinence des traces collectées pour identifier les compétences potentielles. Dans cette expérimentation, les analyses des traces (cf. Figures 4.16) soulignent l'implication et la réactivité des élèves pour répondre (zone verte) à des sujets de discussion spécifiques. Ainsi, l'observation des activités des élèves contribue à mettre en évidence leur participation à des domaines de compétences spécifiques, ce qui permettrait d'enrichir leur propre profil. Comme il est précisé ci-dessus, les postes de niveau 3 sont en grande majorité des explications. Par conséquent, ils peuvent être des candidats pour un système qui recherche des personnes (les auteurs des *posts*) ou des ressources (les *posts* et les éventuels liens de sites *Web* qu'ils contiennent). Quant aux postes des niveaux supérieurs, il y a autant de questions que d'explications (cf. Tableau 4.1). Il faut donc analyser les traces plus en profondeur afin de distinguer les deux types de messages.

D'un point de vue structurel, on peut penser que la longueur des postes en mots pourrait être un indicateur. Dans notre expérimentation, la longueur moyenne des explications (60 mots) est environ le double de celle des questions (37 mots), mais la longueur de chaque message n'est pas un critère discriminant pour distinguer son type, en raison de l'écart type considérable de la longueur des messages (environ 43 pour les questions et 53 pour les explications), comme nous pouvons l'observer sur la Figure 4.19.

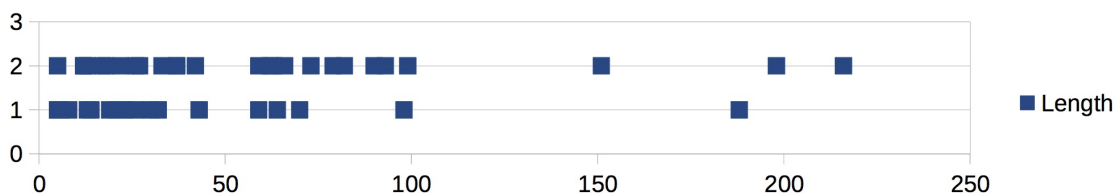
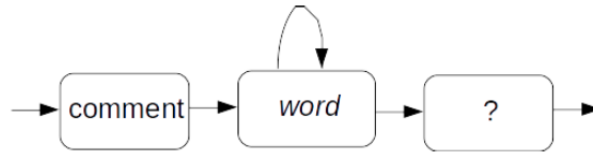


FIGURE 4.19 – Longueur des posts (question au niveau 1 et explications au niveau2)

Par conséquent, nous analysons les messages avec un autre outil d'analyse, afin de reconnaître plus précisément les types de messages. Comme le montre le Tableau 4.1, il existe trois types de messages : les sujets de discussion, les questions et les explications. Dans l'analyse précédente, nous montrons que les niveaux supérieurs de la Figure 4.16 contiennent soit des questions, soit des explications. Comme il semble plus aisé de reconnaître les questions que les explications, la nouvelle analyse consiste à détecter une forme interrogative. Par la suite, il sera possible de déduire la participation des membres aux domaines de compétences à partir de leurs *posts* d'« explication ». Cette forme interrogative peut être représentée par un diagramme de transition comme le montre la Figure 4.20. Par exemple, le *post* 112 a la forme « Comment peut-on placer l'objet à un endroit précis avec le *springLayout* ? ». Notre schéma final est beaucoup plus complexe, puisque nous devons prendre en compte toutes les formes interrogatives utilisées dans notre expérience. Ces formes peuvent commencer par « à quoi », « est-ce que », « comment », « pourquoi », « peut-on », « doit-on », « faut-il », « est-il », etc.

42. <https://moodle.org/>

43. Integrated Development Environment, <https://www.eclipse.org/>

FIGURE 4.20 – *Forme interrogative*

Nous utilisons *TXM* (Heiden 2010), un logiciel *open source* *TeXtoMetry* pour l'analyse de texte. Afin d'être aussi général que possible, *TXM* applique d'abord une analyse lexicale qui est réalisée par l'outil externe *TreeTagger*⁴⁴ (Schmid 1999). *TreeTagger* permet d'annoter un texte avec des informations sur les parties du discours (types de mots : noms, verbes, infinitifs et particules) et fournit des informations sur les lemmes⁴⁵. *TXM* offre une gamme d'outils (concordances textuelles, recherche de collocations, listes de fréquences, etc.) basés sur le moteur de recherche plein texte *CQP*⁴⁶ (*Corpus Query Language*). Nous introduisons quelques notions de *CQP* pour expliquer notre approche. La notation *CQP* de base est constituée de crochets. Chaque paire de crochets correspond à une spécification de jeton (ou *token*) d'entrée. Par exemple, `[word="wonderful"]` signifie qu'il faut rechercher « wonderful » au niveau du mot. Si nous voulons tester un *token* au niveau du lemme, nous devons utiliser des crochets comme `[frlemma="be"]` (i.e. des mots comme « is », « are » qui appartiennent au lemme « be »). Les crochets vides (`[]`) représentent un jeton quelconque. Si nous recherchons une séquence de jetons, nous avons besoin de crochets pour chaque jeton. Par exemple, la requête `[frlemma="be"] [word="wonderful"]` reconnaît une entrée de deux *tokens* tels que « is wonderful ». Enfin, *CQP* utilise des opérateurs similaires à ceux des expressions régulières (`|`, `*`, `?`, ...). Par exemple, `<|>` est l'opérateur alternatif. La requête `[frlemma="be"] [word="wonderful" | word="marvelous"]` reconnaît des entrées telles que « are wonderful » ou « is marvelous ».

Revenons à notre objectif précédent qui est la reconnaissance d'une forme interrogative. Considérons par exemple la forme « Comment placer l'objet à un endroit précis avec le *springLayout* ? », qui peut être reconnue par la requête portant sur un jeton appartenant au lemme « comment », suivi de 1 à 40 jetons quelconques, suivis d'un jeton appartenant au lemme « ? » (cf. requête suivante) :

```
[frlemma="comment"] []1,40 [frlemma=""]
```

L'interrogation finale qui est utilisée pour tous les types de formes interrogatives dans notre analyse est la suivante :

```
(([frlemma="à"] [frlemma="quoi"] ) |
[word="est"%c] "-ce" [frlemma="que"] ) |
[frlemma="comment"] | [frlemma="pourquoi"] |
([frlemma="pouvoir" | [frlemma="devoir" |
[frlemma="falloir" | [frlemma="être"]])
[frpos="PRO:PER"] )) []1,40 [frlemma="\?"]
```

Cette requête exécutée par *TXM* produit les résultats du Tableau 4.2. Notre requête retient 22 *posts* comme des questions (le *post* 134 contient deux formes interrogatives). La colonne « Référence » comprend trois champs : le numéro du billet (la source), le nom de l'auteur et un booléen *isQuestion*. Les deux premiers champs sont obtenus directement à partir des traces, tandis

44. <https://www.cis.lmu.de/~schmid/tools/TreeTagger/>

45. Nous utilisons une forme graphique contenant les termes désignant les unités sémantiques autonomes du lexique de la langue française dans notre cas.

46. http://userpage.fu-berlin.de/~flach/wp-content/uploads/CQP_A_practical_guide.pdf

que le champ *isQuestion* est défini manuellement après une lecture attentive du message. Si nous observons la valeur du champ *isQuestion*, trois messages ne sont pas réellement des questions (*isQuestion* est faux) : 92, 120 et 146. Pour comprendre cette classification erronée, il faut étudier l'ensemble du *post*. Prenons par exemple le *post* 92 (les autres sont similaires), le texte complet du poste est : « Êtes- vous sûrs d'avoir bien compris toutes les consignes relatives à la création d'un projet Java dans Eclipse ? N'hésitez pas à poser des questions MAIS après avoir bien lu toutes les informations pratiques à votre disposition dans ce cours. ». Nous pouvons remarquer l'utilisation d'une forme interrogative, mais uniquement pour encourager la participation à la discussion ou pour expliquer quelque chose. Notre approche n'est pas en mesure de traiter correctement ce type de message. Cependant, la détection d'une forme interrogative est une approche correcte dans la plupart des cas, puisque 19 messages sur 22 (86%) sont effectivement des questions.

Par contre, il y a un *post* « question » (*isQuestion* is true) qui n'apparaît pas dans les résultats. Le contenu de ce *post* est : « Moi aussi je voulais poser cette question ! :-D ». En fait, il s'agit d'une question sans forme interrogative. Une fois de plus, nous pouvons constater la difficulté de reconnaître de tels *posts*, mais ils sont extrêmement rares (1/49 dans notre expérimentation).

Par déduction, il y a 27 (49-22) *posts* « explication ». Le Tableau 4.3 montre les auteurs et le nombre de ces *posts*. Nous pouvons constater que très peu de membres (le professeur et deux élèves) fournissent des explications et que leur nombre est inégalement réparti.

En conclusion, cette expérimentation cherche donc à valider la pertinence des traces issues des activités collaboratives assistées par ordinateur, pour reconnaître la participation des utilisatrices et des utilisateurs dans des domaines de compétences. Notre démarche scientifique consiste à réaliser des analyses orientées vers l'intuition, pour mettre en évidence des éléments dans les traces d'interaction, qui sont associés à l'implication des personnes dans un forum. Un outil d'analyse graphique nous permet de manipuler les réseaux d'utilisatrices ou d'utilisateurs ou leurs messages pour révéler la structure sous-jacente des échanges dans le forum. Nous utilisons des outils d'analyse textuelle pour dépasser les limites de l'outil d'analyse graphique, afin d'analyser les *posts* plus en profondeur. Cette étude devrait renforcer la spécification de la plate-forme de *benchmarking* BEATCORP, qui repose sur l'exploitation de traces d'interaction partageables, au moyen d'outils d'analyse.

Bien que les résultats aient démontré la capacité de notre système à atteindre les objectifs de l'expérimentation, on constate la complexité croissante des analyses de traces de bas niveau d'abstraction, quand les individus utilisent plusieurs outils pour réaliser leurs tâches de communication, de coordination et de production. Dans la deuxième expérimentation que nous décrivons ci-dessous, nous montrons l'intérêt de privilégier la collecte de traces de haut niveau d'abstraction, et d'intégrer des modules de traitement de données issus de l'intelligence artificielle pour faciliter les analyses croisées.

TABLEAU 4.2 – *Posts identifiés comme étant des questions*

Référence	Pivot
92, Christophe, <i>false</i>	Êtes-vous sûrs d'avoir bien compris toutes les consignes relatives à la création d'un projet ?
110, Mathilde, <i>true</i>	Pourquoi est-ce que le Layout est identifié comme " " illégal " " ?
112, Marjolaine, <i>true</i>	Comment peut-on placer l'objet à un endroit précis avec le springLayout ?
114, Emma, <i>true</i>	Comment positionne-t-on un bouton ?
116, Raphael, <i>true</i>	Comment faire en sorte qu'un bouton prenne une dimension en fonction de son label, et...
118, Raphael, <i>true</i>	Comment faire fonctionner le repaint () ?
119, Maxime, <i>true</i>	doit-on mettre le repaint () ; ?
102, Mathilde, <i>true</i>	Comment crée-t-on un lanceur complet ?
131, Mathilde, <i>true</i>	Comment fait-on pour afficher une ArrayList, physiquement ?
134, Perle, <i>true</i>	Faut-il donc créer une collection dans le modèle ?
134, Perle, <i>true</i>	comment faire pour qu'il s'affiche car j'ai testé plusieurs choses mais rien ne marche. Il...
135, Marjolaine, <i>true</i>	Comment faire pour paramétrer l'emplacement du menu déroulant ?
120, Marjolaine, <i>false</i>	Est-ce que la création en dur des contacts doit se faire dans le ToIPModele ?
126, Emma, <i>true</i>	Comment écrit-on la méthode get dans le main ?
128, Emma, <i>true</i>	Comment génère-t-on les getter ?
138, Marjolaine, <i>true</i>	Comment faire pour faire en sorte que lorsque l'on clique sur un bouton (" " appeler " " ...
143, Emma, <i>true</i>	Pouvez-vous nous expliquer comment récupérer les valeurs écrites dans le formulaire...
106, Raphael, <i>true</i>	A quoi correspond cette erreur ?
139, Marjolaine, <i>true</i>	comment faire pour palier à cela ?
145, Marjolaine, <i>true</i>	Comment tester une chaîne de caractères ?
146, Christophe, <i>false</i>	peut s'écrire bien plus simplement, non ?
100, Mathilde, <i>true</i>	comment appelle-t-on l'affichage de cette fenêtre définie ?
105, Marjolaine, <i>true</i>	Est-ce qu'on doit créer une classe bouton pour créer des boutons ?

TABLEAU 4.3 – *Posts identifiés comme étant des explications*

Auteur	Nombre de <i>posts</i> d'explication
Christophe	22
Marjolaine	3
Mathilde	2
Total	27

4.4 Expérimentation d'identification de situations d'innovation

4.4.1 Les objectifs de l'expérimentation

Ce travail adresse des problématiques associées aux sciences de gestion et à la représentation informatique des concepts du système complexe à construire. La recherche de la performance au sein des organisations a fait l'objet de nombreux travaux de recherche qui ont conduit, entre autres, à des méthodes de gestion et d'organisation telles que le *LEAN manufacturing* et le *LEAN management*. Au-delà des considérations liées aux activités d'exploitation, qui visent à assurer les opérations quotidiennes en s'appuyant sur des compétences existantes, la performance de l'organisation tient aussi à la capacité à innover en menant des activités d'exploration et de recherche de nouvelles compétences. Des travaux de recherche ont montré l'importance de l'ambidextrie organisationnelle, qui combine exploitation et exploration, ainsi que l'importance de la nature et des conditions du partage du temps de travail entre ces deux types d'activité. Les besoins d'agilité et de flexibilité dans la configuration collective du travail, et notamment dans la recherche de compétences pour répondre aux besoins en production et en innovation, remettent en question la cohabitation des structures hiérarchique (pouvoir formel) et en réseau formée par les individus qui interagissent entre eux, la nature des liens pouvant être liée au degré de couplage dans leurs activités (co-rédaction, co-construction, ...) ou aux relations interpersonnelles (confiance, ...).

L'objectif ici est d'évaluer la capacité des données d'interaction pour l'identification des compétences dans le domaine de l'innovation et de vérifier le degré de conformité des compétences déduites avec la réalité. Les problématiques scientifiques afférentes concernent l'identification des composantes des compétences et des modèles numériques existants, ainsi que les solutions techniques réalistes pour tracer les compétences individuelles et collectives au sein d'une organisation. Le travail de recherche sur la caractérisation des compétences, selon qu'elles sont associées à des activités d'exploitation ou d'exploration, s'est heurté à l'absence de cadre théorique sur la dynamique des activités d'innovation. Pour appliquer la méthode abductive basée sur des expérimentations pour construire ce cadre théorique, nous avons développé l'outil KBskill sur la base de l'outil de gestion de projet Kanboard⁴⁷, que nous avons couplé via des connecteurs E.T.L. (*Extract Transform Load*) à des nomenclatures de tâches reconnues : le répertoire français opérationnel des métiers et des emplois ROME ; la classification européenne des aptitudes/compétences, certifications et professions ESCO ; le référentiel des professions O*NET pour les U.S.A. ; le référentiel français des métiers du professorat et de l'éducation REFERENS).

Dans cette expérimentation associée au travail de thèse d'Olivier Pouydebat sur les activités d'innovation⁴⁸ (Pouydebat n.d.), nous avons posé la question de recherche suivante :

Quelles sont les traces d'interaction numérique qui permettront de formaliser, à l'aide de transformation des traces, des compétences relatives aux activités d'innovation d'exploitation et d'exploration ?

Dans ce document, nous synthétisons l'expérimentation, dont tous les éléments détaillés sont disponibles dans (Pouydebat n.d.). Les analyses présentées dans cette expérimentation portent sur des sources de traces explicites, dont le haut niveau d'abstraction considéré dans le système à base de traces BEATCORP permet de s'affranchir des transformations de la plate-forme (*Explicit Trace Sources* de la Figure 2.6 de la partie 2 du document). Nous nous appuyons sur le processus issu de l'architecture SBS-SEARCH d'extraction des compétences, à partir des activités individuelles

47. <https://kanboard.org/> est un outil support à la gestion de projet basé sur la méthode Kanban

48. thèse intitulée « Des compétences individuelles et collectives tracées pendant les activités d'exploitation et d'exploration à la performance »

inscrites dans l'outil KBskill et reconnues dans les nomenclatures de tâches (ROME, ESCO, etc.). Notre démarche scientifique consiste donc à analyser ces traces de haut niveau, afin de mettre en évidence leurs caractéristiques structurelles et/ou sémantiques, qui sont liées à des compétences d'innovation d'exploitation ou d'exploration.

4.4.2 Le contexte de l'expérimentation

Cette expérimentation se déroule dans les deux domaines d'application que nous étudions : l'éducation et l'entreprise. Elle se décompose donc en deux groupes, dont le premier compte trois parties qui se sont déroulées dans le cadre de projets encadrés par des enseignants durant moins d'une semaine avec des groupes de 110 à 138 étudiants d'école de management et le deuxième qui a été effectué dans le bureau d'étude d'ingénierie du bâtiment de l'entreprise MANASLU, partenaire du projet CaTCaP, sur une durée de plus de sept semaines dans le cadre de deux projets réels. L'outil *open source* KBskill a été utilisé dans les deux groupes d'expérimentations, ce qui a simplifié les analyses et les comparaisons entre les différentes situations. Nous distinguons deux contextes d'usage de l'outil KBskill : i) l'usage « libre », c'est-à-dire sans aucune directive quant aux actions à réaliser ; ii) l'usage « guidé », c'est-à-dire avec des directives à suivre relatives aux actions à effectuer. Nous distinguons également deux types de situation d'innovation : i) l'innovation d'exploitation visant à améliorer des processus existants ; ii) l'innovation d'exploration visant à imaginer des nouveaux produits ou de nouveaux concepts.

	Usage libre de KBskill	Usage guidé de KBskill
Contexte d'innovation d'exploration	Groupe 1 (partie 1) du 4 au 5 janvier 2021	Groupe 1 (partie 2) du 11 au 15 janvier 2021
Contexte d'innovation d'exploitation	Groupe 2 (entreprise) du 25 Aout au 15 Octobre 2020	Groupe 1 (partie 3) du 18 au 21 janvier 2021

TABLEAU 4.4 – Contextes d'expérimentation

Les contextes et les caractéristiques de l'expérimentation sont synthétisés dans les tableaux 4.4 et 4.5. (Pouydebat n.d.).

	Durée (jours)	Type usage	Type innov.	Nb. util. déclarés	Nb util. actifs	Nb. projets	Nb. tâches créées	Nb. tâches validées
Groupe 1 (partie 1)	2	libre	explor.	112	5	4	103	73
Groupe 1 (partie 2)	5	guidé	explor.	110	84	20	619	506
Groupe 1 (partie 3)	4	guidé	exploit.	138	14	4	144	138
Groupe 2	51	libre	exploit.	7	7	2	133	78

TABLEAU 4.5 – Synthèse des caractéristiques de l'expérimentation

L'outil KBskill est configuré de manière identique pour les quatre parties de l'expérimentation, à savoir avec les quatre colonnes « à faire », « en cours », « en attente », « fait », et « validé ». Les tâches listées dans la dernière colonne sont considérées comme étant réussies, parce qu'elles attestent d'un niveau de performance suffisant et permet ainsi la reconnaissance des compétences

afférentes. C'est ce qu'on appelle la validation sociale et seules les tâches qui auront atteint cette colonne seront considérées pour cette reconnaissance. On remarque que les pourcentages de tâches validées sont plus importants dans les situations avec usage guidé du KBskill, comme en atteste le tableau 4.5.

Les sujets traités par quatre équipes d'étudiants dans chacune des trois parties du premier groupe sont respectivement :

- concevoir un bâtiment innovant, dit « inclusif », dont l'enjeu pour les urbanistes et les architectes est de proposer de nouvelles formes de bâtiments reliés au tissu urbain dans lequel ils s'insèrent ;
- imaginer et réinventer la montagne durable du futur afin de modifier l'offre touristique en rupture avec les pratiques actuelles connues ;
- optimiser les processus existant de production de gaufres au sein de l'entreprise, en améliorant les cadences, en réduisant les coûts, etc.

Les quatre équipes d'étudiants disposaient d'une instance de KBskill pour organiser leurs activités de *design thinking*. Un membre par équipe était identifié comme étant référent « éditeur » des *user stories* (ou tâches) dans KBskill. Dans les parties guidées de l'expérimentation, les étudiants étaient accompagnés pour les aider à organiser leur travail et à l'inscrire dans KBskill.

Le deuxième « groupe », qui est constitué d'une seule partie, concerne un effectif très réduit, mais sur une durée plus longue et dans un contexte professionnel réel. Les activités considérées dans cette expérimentation concernent les mesures et les calculs de l'efficacité énergétique des bâtiments. Les livrables sont des expertises techniques et méthodologiques pour des clients extérieurs. D'autres activités internes visent à proposer des techniques innovantes pour améliorer la performance des procédés actuels de mesure. Il s'agit donc de processus d'innovation d'exploitation, où les processus existants sont optimisés pour permettre à l'entreprise de rester concurrentielle. L'application KBskill a donc été utilisée dans le cadre de deux projets, l'un étant interne et l'autre externe.

4.4.3 Les résultats de l'expérimentation

Les traces d'activité

Les tâches (i.e. *user stories*) et les composants K.S.A. (*Knowledge, Skills and Attitudes*) de compétences issus des nomenclatures de tâches au sein de l'outil KBskill, sont stockés dans deux bases de données distinctes. La collecte de ces données montre que l'usage guidé de l'outil augmente considérablement le volume d'informations rentrées (plus d'interactions avec l'outil, plus de tâches déclarées, meilleur taux de remplissage des *user stories*, etc.). Les composants K.S.A. étant sélectionnés au moment de la configuration des *user stories* (cf. Figure 4.5), les associations tâches/composants K.S.A. peuvent être facilement récupérées dans la base de données des tâches. À noter qu'en cas d'absence de composants K.S.A. dans les nomenclatures de tâches, l'utilisateur peut ajouter des composants qui seront stockés dans une autre base de données pour les distinguer des composants « standards ». Un traitement ultérieur pour rechercher les alignements sémantiques éventuels, via une taxonomie par exemple, permettra de valider ces nouveaux composants et de les intégrer dans la base principale des composants K.S.A. de KBskill.

La préparation des données à analyser

Si l'interface graphique de KBskill permet de choisir les libellés exacts des composants K.S.A. à l'aide d'une fonctionnalité d'auto-complétion, il n'en est pas de même avec les libellés des *user*

stories dans la version actuelle de l'outil. Pour les besoins de l'analyse, il est donc primordial d'effectuer des alignements sémantiques, mais le nombre important de tâches rend cette opération très fastidieuse. Nous avons donc réalisé ce travail d'association tâches/composants K.S.A. manuellement pour l'une des tâches puis, à l'aide de l'intelligence artificielle, nous avons automatiser ce travail d'association pour toutes les tâches similaires. Le module que nous avons utilisé est un développement interne d'un outil en PHP, qui se base sur l'algorithme de Levenstein ([Schimke and Vielhauer 2007](#)) pour proposer les tâches considérées comme proches. L'opération d'association tâches/composants K.S.A. issus des nomenclatures ROME et ESCO a ensuite été automatisée à l'aide de la *library open-source* Python SpaCy⁴⁹ spécialisée dans le traitement de corpus de textes en langage naturel (N.L.P.). Nous avons néanmoins pu constater que l'automatisation gagnerait à intégrer une phase d'apprentissage pour être plus performante, car certaines associations ont dû être corrigées manuellement.

Conformément à notre démarche scientifique, nous recherchons des compétences effectives et non des compétences déclarées. Seules les associations tâches/composants K.S.A. sont déclarées et la validation de ces derniers est subordonnée à la validation des tâches afférentes. Quant aux compétences effectives, leur construction nécessite la prise en compte de tous les éléments du modèle de compétence structurel proposé par ([Pouydebat n.d.](#)) (i.e. le contexte, les activités, les composants K.S.A., etc.) issus de la collecte des traces pendant l'expérimentation.

L'analyse des données

Les alignements sémantiques réalisés pour les tâches et les composants K.S.A. nous permettent de réaliser des analyses et de comparer les résultats entre les différentes sessions expérimentales. Les données récoltées dans les bases de données de KBskill ont permis de calculer des indicateurs dans le but d'identifier et d'isoler différents groupes de composants K.S.A. correspondant à des contextes d'innovations spécifiques.

Les premiers indicateurs statistiques qui ont été calculés sont les nombres d'occurrences de chaque composant K.S.A. dans chacune des quatre parties d'expérimentation. Une matrice d'occurrences (4 colonnes pour les 4 parties et 258 lignes pour les 258 composants K.S.A.) a été créée pour les quatre parties de l'expérimentation. Pour permettre la comparaison des informations d'une partie à l'autre, deux autres matrices ont été créées pour calculer d'une part, les fréquences de composants K.S.A. par partie d'expérimentation et d'autre part, les fréquences des parties par composant K.S.A.

Un indicateur binaire est venu compléter les fréquences précitées, à savoir la présence (valeur 1) ou non (valeur 0) d'un composant K.S.A. dans chacune des parties de l'expérimentation. Cet indicateur peut alors être utilisé par les filtres des outils de visualisation des données relatives aux compétences. Une combinaison binaire classique sur quatre digits représentant les quatre parties permet d'identifier 2^4 combinaisons de présence des composants (e.g. 0000 correspond à l'absence du composant dans toutes les parties, alors que 1100 signifie qu'il est présent dans les deux premières parties).

La combinaison des deux indicateurs permet d'isoler des composants K.S.A. selon la typologie des parties de l'expérimentation (e.g. les composants associés à de l'innovation d'exploitation avec la valeur 0011 versus d'exploration avec la valeur 1100). Les résultats ont été présentés sous deux formes graphiques, l'une sous forme d'histogrammes statiques avec l'outil « Tableau public » (cf. Figure 4.22) et l'autre à l'aide de graphes avec l'outil « Neo4J » (cf. Figure 4.21). Les histogrammes de l'outil « Tableau public » apportent une vision statistique des résultats,

49. <https://spacy.io/>

c'est-à-dire sans qu'il soit possible d'interagir avec l'interface graphique, contrairement au graphe interactif de l'outil « Neo4J ». Les représentations des résultats d'analyse s'inscrivent dans la partie basse de l'architecture SBS-SEARCH. Dans les travaux que nous menons actuellement dans le projet CaTCaP, nous avons choisi de nous appuyer sur l'architecture représentée sur la Figure 4.11 pour migrer vers la base de données graphique « Neo4J » et bénéficier de son langage de requête orienté graphe « Cypher », plus adapté que les requêtes SQL.

Les exemples ci-dessous attestent du potentiel de ces représentations pour répondre à la question de recherche sur la capacité des traces de haut niveau d'abstraction à identifier des compétences dans des contextes d'innovation d'exploitation et d'exploration. Cette expérimentation permet de valider l'utilisation d'un outil P.R.M.S. (*Projects and Resources Management System*) pour nos futures analyses, tel que BeesBusy utilisé désormais par l'entreprise partenaire MANASLU. La perspective de ce travail à partir de ces résultats pourrait être de créer les conditions favorables aux activités d'innovation d'exploitation ou d'exploration au sein d'une entreprise selon la stratégie qui aura été définie en son sein. Il s'agirait par exemple de mobiliser des ressources humaines ayant des compétences propices à telle ou telle forme d'innovation. Bien entendu, ce ne sont pas les seuls paramètres à prendre en compte, mais de plus amples expérimentations pourraient mettre en exergue des invariants qui permettraient de tendre vers ces conditions favorables.

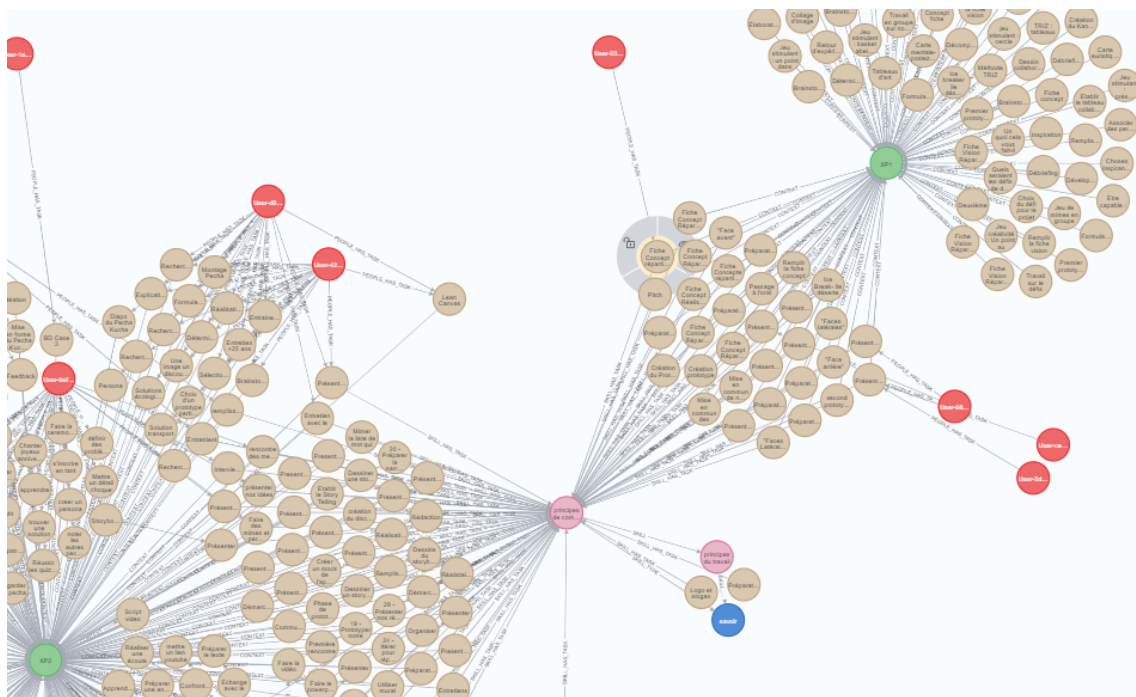


FIGURE 4.21 – Résultats d'analyse sous forme d'un graphe interactif avec « Neo4j »

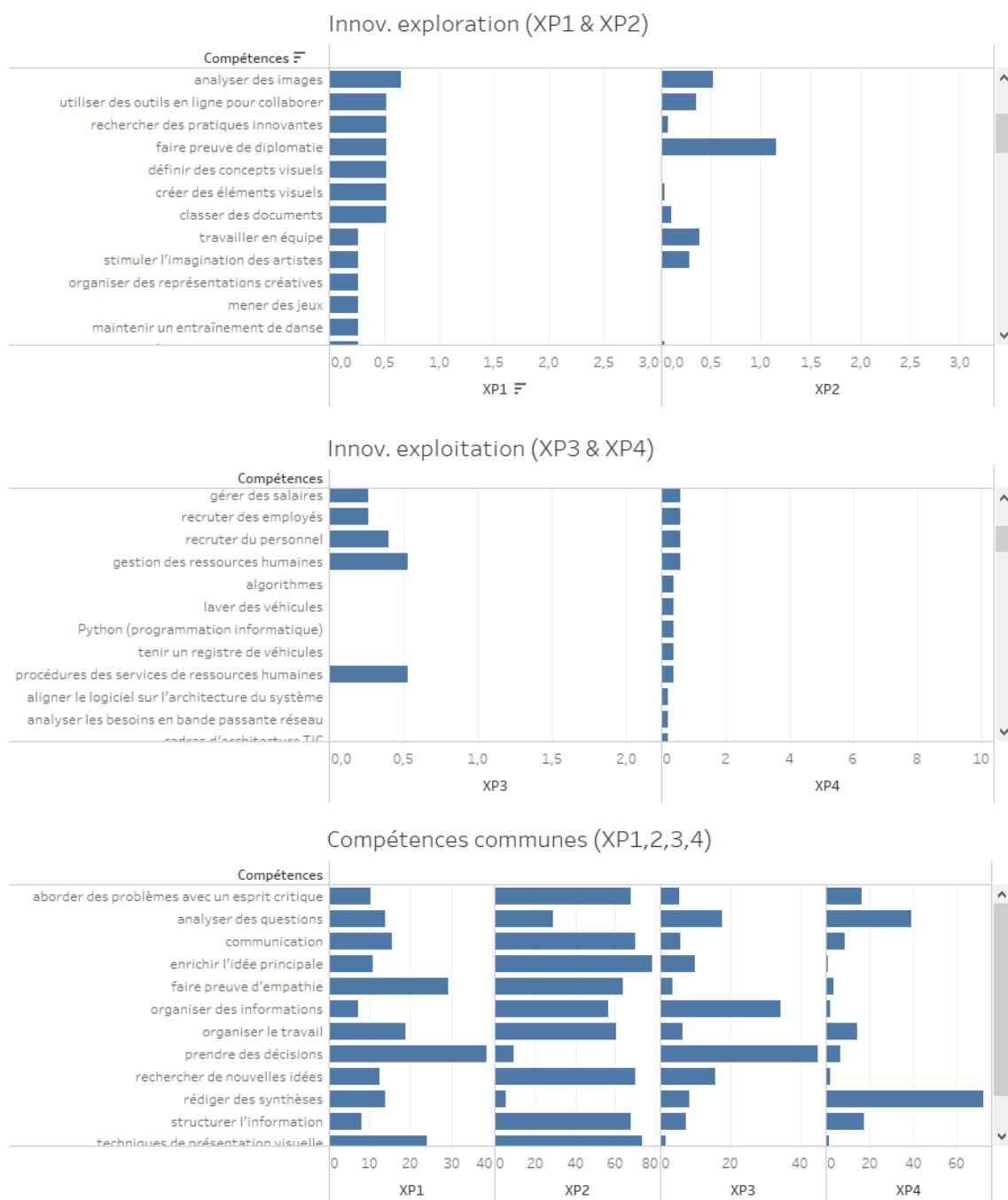


FIGURE 4.22 – Résultats d'analyse sous forme d'un tableau de bord statique avec « Tableau public »

4.5 Conclusion

Dans cette dernière partie, nous exposons le lien entre les traces d'interaction et les compétences mobilisées, c'est-à-dire qui s'expriment à travers l'activité effective. Contrairement aux activités qui ont un ancrage dans les systèmes numériques avec les traces d'interaction, les compétences sont déduites d'indicateurs, comme la performance, et identifiables à travers les activités. Par conséquent, ce que nous recherchons, ce sont avant tout les activités menées avec un objectif défini et un niveau de performance attendu.

Nous avons donc privilégié les traces de haut niveau d'abstraction (i.e. les séquences), que l'on peut collecter dans des outils P.R.M.S. (*Projects and Resources Management System*) sous la forme de *user stories*. Ce choix pragmatique est motivé à la fois par la simplification des traitements de données (i.e. transformations limitées des traces) et par la rapidité de l'opérationnalisation qui en découle. Nous procédons ensuite à des analyses couplées avec des nomenclatures de tâches pour établir le lien avec les compétences, lesquelles sont qualifiées par rapport à l'atteinte de l'objectif initial et la qualité du résultat. Pour mettre en exergue les items et leurs liens qui ont conduit à l'identification des compétences, nous choisissons une représentation graphique pour visualiser et explorer ces éléments sous la forme de cartographies.

L'idée que nous avons en perspective, est d'apporter une dimension dynamique aux cartographies en introduisant la diachronie, c'est-à-dire des photographies à des instants différents, et ainsi constater les évolutions des compétences, voire en déduire leur cycle de vie. Les algorithmes de spatialisation et les métriques de la théorie des graphes (e.g. liens forts, liens faibles) permettront de construire des indicateurs à partir de ces cartographies, qui seront utiles pour établir une stratégie de développement cohérente avec les besoins des individus et des entreprises.

Ces résultats nourriront bien sûr les futurs travaux sur l'articulation entre les processus managériaux pour la gestion des compétences dans l'entreprise et les processus de reconfiguration continue des formations professionnelles. Dans le projet P.I.A. « Territoires d'innovation pédagogique » / « Campus des métiers et des qualifications », nous traiterons la question des indicateurs caractérisant l'adaptabilité de l'entreprise dans un contexte donné pour formaliser, à l'aide d'un mécanisme de régulation, les transformations nécessaires dans l'offre de formation.

Conclusion générale et perspectives

Ce document a une fonction de recueil pour la transmission des travaux antérieurs, que nous nous efforçons d'expliquer lors de nos encadrements de thèses et de Master, et lors de nos collaborations avec d'autres chercheuses et chercheurs. C'est aussi et surtout l'occasion de prendre du recul sur les voies suivies dans les communautés de recherche, et de définir des orientations pour les travaux à venir. Ces orientations doivent être davantage motivées par les besoins des utilisatrices et utilisateurs et leurs usages du numérique, mais aussi par l'impact du numérique sur les structures sociales (analyse de « niveau macro ») et les conséquences en terme d'éthique dont il est responsable, plutôt que sur les outils « à la mode » dictés par des technologies passantes. Il convient de noter que ce recueil de travaux adresse les deux facettes du métier d'enseignant-chercheur, de par le lien étroit qui unit nos activités pédagogiques et les problématiques de recherche dans notre cas.

Dans ce document, nous présentons les travaux de recherche conduits depuis près de deux décennies, pour répondre à des problématiques qui alimentent de manière incrémentale et itérative une chaîne cinématique conceptuelle, dont nous présentons la vision synoptique Figure 1.5 au début de ce document (page 19). Les problématiques traitées au cours de ces années sont issues d'un constat d'insuffisance liée à l'introduction des systèmes numériques collaboratifs dans l'activité humaine (e.g. les notions de *group awareness* pointant l'appauvrissement du contexte des échanges médiés par les outils informatiques). Nous avons donc proposé d'exploiter les traces d'interaction numériques pour recouvrer les conditions du travail présentiel, voire les enrichir en ajoutant des indicateurs (e.g. des *feedbacks*) et des traitements spécifiques (e.g. le mécanisme de rétroaction).

Nos premiers travaux inscrits dans la thématique « Observation des activités humaines » ont conduit à la proposition du format de traces CARTE (Courtin 2010) présenté dans la partie 2 pour répondre à la problématique d'hétérogénéité des formats en l'absence de format standard. Si ce format s'adapte bien à la représentation de traces issues d'outils logiciels très variés (e.g. communication, production, coordination, etc.), son adoption au sein d'une plate-forme d'outils existants d'analyse de traces nécessiterait l'adaptation de leur format d'entrée.

L'évolution vers des corpus de traces centrés sur un format pivot, comme les travaux MULCE (Chanier and Ciekanski 2010) ou PSLC-DataShop (Stamper et al. 2010), ont montré leurs limites, car ils nécessitent des transformations des traces pour permettre des analyses croisées de corpus d'origines diverses. L'architecture BEATCORP (Courtin and Tomasena 2016) présentée dans la partie 3, basée sur les trois modèles (de corpus, sémantique et opérationnel) de l'approche PROXYMA (Chebil, Courtin, and Girardot 2015), et dont la spécificité est le traitement générique des formats de traces hétérogènes, se démarque par l'absence de format pivot. À noter que la charge de travail relative aux transformations de formats de traces des systèmes basés sur un format pivot, s'est reportée sur l'initialisation des opérations d'alignement sémantique de l'approche PROXYMA à l'arrivée de nouveaux formats. Une fois cette phase d'initialisation passée, les opérations sont disponibles pour les analyses suivantes sans autres transformations. Le partage d'outils d'analyse est donc rendu possible, ce qui contribue à l'adaptation de notre système BEATCORP à une plus large palette d'outils.

Notre travail sur la modélisation des corpus de traces d'apprentissage contextualisées nécessite néanmoins des adaptations pour que nos modèles, et notamment le modèle de corpus, soient reproductibles dans le domaine de l'industrie pour prendre en compte les données spécifiques. Le principe du *middleware* (i.e. le PROXY), ainsi que les modèles sémantique (i.e. la taxonomie) et opérationnel (i.e. les six opérateurs) quant à eux sont reproductibles. En rehaussant le niveau d'abstraction des traces collectées, l'approche PROXYMA ouvre des perspectives très prometteuses pour l'instrumentation des outils numériques fortement ancrées au sein des organisations. L'exploitation des traces s'appuie sur la taxonomie du modèle sémantique de l'approche PROXYMA pour effectuer un alignement sémantique entre les traces issues des différents outils du même type (e.g. P.R.M.S. « *Projects and Resources Management System* ») utilisés dans le domaine industriel. De là, une large palette d'applications potentielles destinées à répondre aux enjeux socio-économiques actuels ouvre à d'autres problématiques, vers lesquelles nous orientons nos travaux actuels et futurs. Ces problématiques couvrent, entre autres, les recommandations de ressources et la régulation de l'activité dans des organisations étendues (i.e. décloisonnées). La démarche scientifique privilégiée passe par l'expérimentation soit en situation écologiquement fondée (i.e. qui s'apparente à une situation authentique), soit directement en immersion sur le terrain (i.e. en situation professionnelle). Pour aborder les problématiques liées à l'observation de l'activité humaine qui est intrinsèquement ancrée dans la réalité, et notamment dans des domaines d'application accessibles tels que l'éducation ou l'industrie, nous privilégions une logique abductive alliant pratique et théorie pour réaliser nos tâches de modélisation.

Dans les travaux inscrits dans la thématique « Reconnaissance des compétences individuelles et collectives » sur la recherche de compétences au sein des organisations, les principales contributions de nos recherches concernent la prise en compte des activités humaines réelles et des relations interpersonnelles dans nos modèles de compétence, à partir d'items numériques dont les traces et les productions. La mise à l'épreuve de nos modèles dans d'autres domaines que l'Éducation, moyennant des évolutions pour tendre vers plus de généricité, auront démontré l'adaptation de nos travaux à différents contextes pour répondre à de nouveaux objectifs.

Cette adaptation a commencé par les travaux de thèse d'Olivier Pouydebat (Pouydebat n.d.) sur la recherche de compétences à travers l'activité qui a abouti à la définition de l'architecture SBS-SEARCH⁵⁰ basée sur celle de BEATCORP⁵¹ (Courtin and Tomasena 2016). Pour permettre l'extraction des compétences à partir de l'activité, nous avons expérimenté la *library* d'intelligence artificielle SpaCy⁵² pour faciliter l'automatisation de l'association « tâche/compétence »⁵³ ayant des libellés proches de ceux des nomenclatures de tâches incluant des composants K.S.A. (*Knowledge, Skills and Attitudes*). Dans le système SBS-SEARCH, le modèle conceptuel des données pour les compétences est de type relationnel et les informations sont stockées dans une base de données MariaDB⁵⁴. Dans le modèle de compétence structurel proposé par (Pouydebat n.d.), les éléments de contexte de l'activité sont pris en compte, ainsi que les composants K.S.A. des compétences, mais la dimension temporelle de l'activité pour représenter l'enchaînement des actions afférentes n'est pas intégrée, comme en atteste le formalisme UML utilisé (représentation statique du modèle de compétence sous forme de diagramme de classes). Ce type de modèle associé à l'architecture SBS-SEARCH a permis, à l'aide d'expérimentations, de répondre à la question de recherche portant sur la caractérisation des compétences relatives aux activités d'innovation d'exploitation et d'exploration et formalisées à partir des traces d'interaction numérique. Cependant,

50. *Skill-Based System - Skills Extract, Analyze, Regroup, Consult and Historize*

51. *BEenchmarking Platform for Analysis of Traces CORPora*

52. *library open-source* de traitement naturel du langage, <https://spacy.io/>

53. via l'algorithme de Levenshtein (Schimke and Vielhauer 2007)

54. <https://mariadb.org>

nous avons pu mesurer les limites de ce type de représentation structurelle et tabulaire pour une exploitation dynamique des compétences sous la forme de cartographies.

Dans le modèle de (Bemmami et al. 2021), la compétence est mobilisée par l'activité et se construit dans un schème⁵⁵ mobilisant et combinant les ressources environnementales et personnelles en intégrant la dynamique de construction et d'évolution de la compétence. Ce travail s'inscrit dans le projet CaTCaP⁵⁶ visant à proposer des instruments pour d'une part identifier et extraire des compétences à partir de situations singulières où elles se déploient, d'où la nécessité de représenter la dimension temporelle de l'activité, et d'autre part, les gérer en lien avec les processus métiers, et favoriser leur diffusion et réutilisation. Dans les travaux du projet, nous affichons l'idée de découverte de compétences non encore répertoriées. Nous avons donc intégré des nomenclatures de tâches/compétences dans notre système pour repérer les compétences connues et ainsi circonscrire celles qui pourraient être nouvelles, modulo un défaut de reconnaissance par notre système. Le nombre de tâches pouvant être conséquent, nous nous appuyons sur des modules d'IA (e.g. SpaCy) pour faciliter cette reconnaissance. Pour répondre aux objectifs du projet, nous avons choisi de caractériser et de représenter les compétences à l'aide d'un formalisme graphique facilitant leur exploitation. Un premier modèle de compétences individuelles est élaboré pour prendre en compte l'activité et sa dynamique dans la construction des compétences (Bemmami et al. 2021). La représentation graphique est en cours de développement via le système de gestion de bases de données graphiques natives Neo4J, pour stocker et gérer les données et leurs relations dans leur état « naturel ». Les bases de données orientées « graphe » peuvent permettre de répondre aux problématiques des données fortement connectées (*LinkedData*). Ce système offre aux développeurs et aux *data scientists* des outils avancés fiables en *open source* qui permettent des requêtes rapides, la définition de contextes pour les analyses et un modèle de données modifiable en un minimum d'efforts.

Avec le démarrage en 2022 du projet P.I.A. « Territoires d'innovation pédagogique » / « Campus des métiers et des qualifications » qui adresse des problématiques au croisement de nos précédents travaux dans les domaines des E.I.A.H. et de l'industrie du futur, nous encadrons une nouvelle thèse intitulée « Adaptation continue des formations professionnelles à l'évolution des besoins en compétences dans les organisations ». Les processus de reconfiguration de la formation professionnelle qu'elle soit initiale ou alternante doivent s'inscrire dans un continuum avec les processus managériaux et métiers des organisations pour répondre aux enjeux économiques. Si les formations évoluent régulièrement pour prendre en compte les besoins des organisations, la temporalité des évolutions n'est pas toujours adaptée à la réalité économique. Il conviendrait de synchroniser la réactivité d'adaptation de la formation avec celle de l'activité professionnelle à l'aide de mécanismes de régulation. Ces mécanismes pourraient s'appuyer sur des indicateurs relatifs à l'activité au sein des organisations pour réguler les activités de formation. Les besoins d'agilité et de flexibilité dans la configuration collective du travail, et notamment dans la recherche de compétences pour répondre aux besoins des organisations, engendrent des besoins d'agilité sur les processus de reconfiguration de la formation professionnelle. La problématique majeure de ce projet concerne l'articulation entre les processus managériaux pour la gestion des compétences dans l'entreprise et les processus de reconfiguration continue des formations professionnelles. Le travail de recherche proposé dans le cadre de cette thèse concerne la caractérisation de la régulation assistée de l'offre de formation, voire des activités de formation, en fonction de l'évolution des besoins en compétences mobilisées dans les activités professionnelles. La question de recherche que nous adressons dans ce travail porte donc sur la définition d'indicateurs caractérisant l'adaptabilité

55. enchaînement d'actions pour réaliser une activité

56. « Capture des Traces opérationnelles des acteurs d'entreprise pour construire le Capital humain et définir les Processus gagnants »

de l'entreprise dans un contexte donné pour formaliser, à l'aide d'un mécanisme de régulation, les transformations nécessaires dans l'offre de formation.

Nous présentons sur la Figure 4.23 les notions traitées dans les travaux antérieurs, les notions en cours de traitement dans les projets courants et les notions à traiter dans les futurs projets. Les résultats en cours du projet A.N.R. CaTCaP sur les caractéristiques d'une cartographie dynamique et interactive des compétences, constituent un socle solide pour aborder efficacement les notions relatives d'une part, aux systèmes de recommandations de compétences dans des contextes de pilotages stratégiques au sein des organisations et d'autre part, aux mécanismes de régulation des activités dans les programmes de formation continue au vu des activités « métier » menées dans les entreprises.

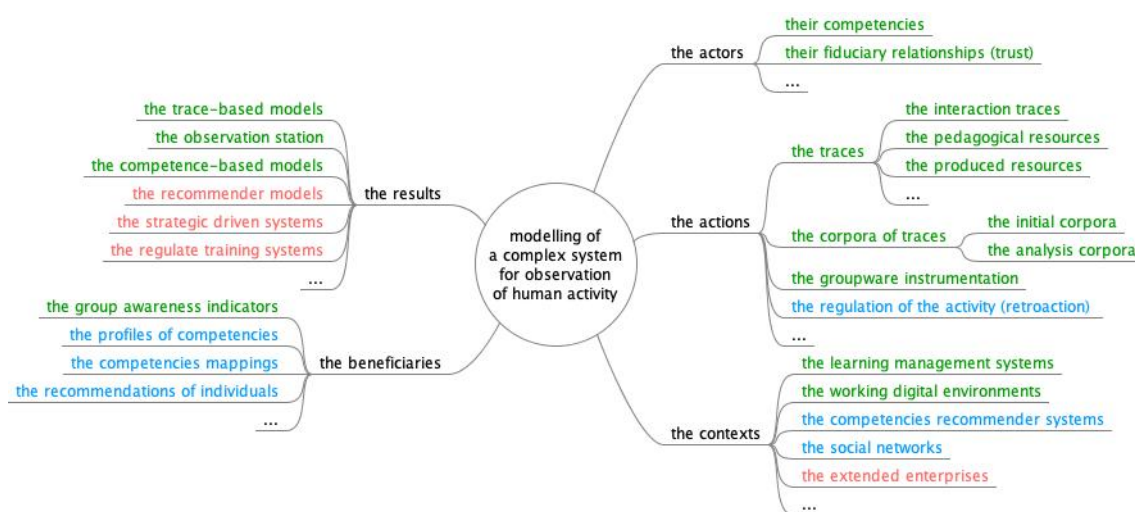


FIGURE 4.23 – couverture des notions d'un système complexe d'observation de l'activité humaine

Des techniques d'analyse avancées pour l'automatisation

Nous avons conscience du paradoxe apparent de notre approche *user-centered*, avec un contrôle continu par l'humain, face à notre volonté d'automatisation des processus (e.g. collecte, analyse, transformation, rétroaction, visualisation, etc.) pour traiter de plus grands volumes de données en un temps très réduit. Nous connaissons l'apport des techniques d'apprentissage automatique (i.e. *machine learning*) et d'apprentissage en profondeur (i.e. *deep learning*) de l'intelligence artificielle, notamment en termes de puissance de calcul, et nous mesurons les limites qualitatives des traces sur lesquelles sont réalisés les traitements. Le changement des comportements humains dû à l'évolution des usages du numérique, souligne les limites d'une approche purement informatique d'analyse de traces et le besoin d'analyses multi-disciplinaires.

En perspective, nous identifions deux cadres de réflexion : celui de l'apprentissage non supervisé orienté vers la découverte de classes de compétences et celui du *learning analytics* utilisant les traces pour mesurer l'écart par rapport aux objectifs fixés. Nous élargirons les analyses sémantiques de ces traces à l'aide d'outils de Traitements Automatiques du Langage Naturel (T.A.L.N.⁵⁷) pour extraire des informations plus pertinentes et ainsi améliorer la qualité des

57. ou *N.L.P.* pour *Natural Language Processing*

résultats fournis par les outils d'exploitation de ces traces (systèmes de recommandations de compétences, tableaux de bord d'indicateurs pour le pilotage stratégique, mécanisme de régulation d'activité de formation basé sur l'adaptation des compétences, etc.). Nous sommes confortés dans l'intérêt de l'intelligence artificielle à laquelle nous avons eu recours dans les travaux de la thèse d'Olivier Pouydebat. Nous avons par exemple utilisé l'algorithme de Levenshtein (Schimke, Vielhauer, and Dittmann 2004) (Schimke and Vielhauer 2007) (Yujian and Bo 2007) pour assister les opérations d'association tâches-compétences au sein de l'outil KBskill, support à la gestion de projet basé sur la méthode Kanban. Nous avons étendu ces opérations d'automatisation à l'aide de la bibliothèque SpaCy⁵⁸ dans une expérimentation où nous avons introduit des nomenclatures de tâches/compétences. Cette recherche ayant fait l'objet d'un plan de gestion des données de la recherche, il sera possible d'entraîner un module d'intelligence artificielle avec les données que nous mettons à disposition de la communauté scientifique (Pouydebat n.d.).

Pour répondre aux problématiques des systèmes de recommandations de compétences, et notamment celles relatives à la prédiction, nous introduirons les techniques de l'intermédiation (transformation, stockage, partage, exploitation) très exploitées dans le secteur économique, pour renforcer la connaissance du passé nécessaire à l'application efficace des techniques d'apprentissage automatique. Le système pourrait ainsi transformer la connaissance de l'activité passée d'individus, comprenant potentiellement l'utilisation antérieure d'un service de recommandation, en informations à valeur augmentée pour leur proposer des futures recommandations pertinentes. Les travaux relatifs aux projets A.N.R. CaTCap et P.I.A. « Campus des métiers » relatifs aux outils de représentation de cartographies interactives des compétences devront être prolongés pour exploiter les analyses topologiques des graphes, en vue de créer des indicateurs pertinents à partir de la détection de motifs caractéristiques. Par ailleurs, la construction de tableaux de bord à l'aide de techniques de type *data-viz* fournira des indicateurs chiffrés et objectifs qui seront utiles à la conduite stratégique des organisations. Compte tenu d'une production croissante des données (i.e. *big data*), les problématiques relatives aux stockages massifs, aux manipulations multiples parfois fallacieuses (e.g. manipulation d'opinion), aux prédictions conditionnées (e.g. filtrage collaboratif, théorie de la longue traîne) devront être considérées.

De la recherche à l'application sur le terrain

Que ce soit dans le domaine de l'éducation ou celui de l'industrie du futur, les réponses aux besoins actuels en termes de compétences n'en sont qu'aux prémices. Les travaux scientifiques relatifs au numérique seront d'autant plus incontournables qu'ils contribueront à l'évolution vers l'industrie 5.0, qui est caractérisée par l'exploitation des données générées par les systèmes communicants (e.g. *Internet of Things*) et le *big data* pour piloter des machines intelligentes. Les perspectives de nouveaux travaux sont donc très larges et les résultats obtenus jusqu'à aujourd'hui encourageants. Les principales actions à venir s'inscrivent dans la partie haute de la chaîne cinématique présentée dans le document, qui adresse notamment les systèmes de recommandations de ressources diverses basés sur l'activité humaine. Le décroisement des domaines de recherche résultant de l'« hyper-technologie » omniprésente nous conduit à investiguer davantage dans les analyse instrumentées à l'aide de l'intelligence artificielle. C'est dans cette perspective d'élargissement des domaines d'investigation que s'inscrivent, entre autres, nos futurs travaux de construction de systèmes complexes d'observation de l'activité humaine.

58. *Opensource library* de traitement naturel du langage <https://spacy.io/>

Je suis actuellement impliqué en tant qu'enseignant-chercheur dans des actions concrètes, pour la mise en place de l'« Approche Par Compétences » (A.P.C.) pour le B.U.T.⁵⁹ au sein du réseau national des référents de l'A.D.I.U.T.⁶⁰, pour la construction de référentiels de compétences afférents dans les A.C.D.⁶¹, et pour la participation à des projets de recherche et développement d'applications pour l'Éducation, à travers des expérimentations en situation réelle d'apprentissage. Nous avons notamment mené une expérimentation dans le projet ANR HyPE-13 « Hybridation et partage des enseignements » qui, à partir d'une A.P.I. Moodle pour opérationnaliser un outil de *Learning Analytics*, produit un ensemble de ressources à destination des enseignants pour les accompagner dans la transformation pédagogique et le passage à l'hybridation. Les outils KBskill⁶² développé au sein de notre projet CaTCaP et KAPC+⁶³ développé dans le projet @SPIRE⁶⁴ sont utilisés pour l'évaluation des compétences des étudiants au sein de certaines formations universitaires depuis janvier 2022. Une phase d'analyse des résultats issus de cette opérationnalisation est prévue à l'automne 2022, avec comme objectif la validation des modèles sous-jacents.

Un système complexe modulaire

Le système complexe pour l'observation de l'activité humaine que nous nous efforçons de réaliser est le résultat d'une construction itérative et incrémentale de modules ayant pour objectif de répondre à diverses problématiques. La vision d'un système globale s'est nourrie progressivement des avancées tant en termes de résultats de recherche que de capacités des systèmes, lesquelles ont profondément modifié les usages du numérique. En ce qui concerne les systèmes à base de traces, où les formats de traces dits standards peinent à faire consensus, nous avons pris le parti de nous en abstraire en créant un modèle générique de corpus de traces PROXYMA⁶⁵. Si le modèle de trace CARTE⁶⁶, présenté dans le chapitre 1, a été conçu comme une réponse au besoin de genericité, laquelle a été validée par des expérimentations en milieux écologiques (Courtin and Talbot 2009)⁶⁷, sa généralisation impliquant une standardisation que des formats plus répandus ont obtenu, supposerait une volonté non exprimée d'interopérabilité de tous les outils d'analyse de traces. Le système PROXYMA prend en charge l'interopérabilité des systèmes de collecte, de construction, d'analyse ou de visualisation des traces, ce qui garantit l'indépendance de ces systèmes qui peuvent être remplacés par d'autres selon les sources de traces considérées. Notre système d'observation est donc constitué de modules faiblement couplés, offrant sur le long terme une ouverture vers des technologies émergentes, produisant de nouvelles sources de données.

Nos travaux de modélisation des compétences intègrent les activités associées en tant que preuves objectives de la mobilisation desdites compétences, sans y intégrer d'éléments d'identification desdites activités, lesquels pouvant être les traces d'interaction au travers d'un système à base de traces ou les informations sur la validation de l'exécution des activités avec un système de gestion de projet utilisant la méthode Kanban⁶⁹ (Sugimori et al. 1977). Nous exprimons une fois

59. Bachelor Universitaire de Technologie

60. Assemblée des Directeurs d'Institut Universitaire de Technologie

61. Assemblées de Chefs de Département

62. outil de suivi des tâches développé en interne sur la base de l'outil *open source* KANBOARD <https://kanboard.org/>

63. outil KARUTA <https://karuta-france-portfolio.fr/> adapté à l'approche par compétences

64. « Accompagnement, Spécialisation Progressive et Individualisation pour la Réussite de tous les Étudiants »

65. *PROXY for Multiple Analyses*

66. *Collection, activity Analysis and Regulation based on Traces Enriched*

67. disponible sur https://rd.springer.com/chapter/10.1007/978-3-642-04636-0_52 classée A+ selon l'A.T.I.E.F.⁶⁸ <http://www.atief.fr/ressources/classement/conferences-internationales>

69. La méthode Kanban a été inventée par l'ingénieur Taiichi Ōno dans les années 1960 dans les usines Toyota.

de plus l'indépendance des systèmes à base de traces et ceux à base de compétences, d'autant plus motivée par le fait que les compétences peuvent s'appuyer sur d'autres systèmes, et donc d'autres sources d'information, pour leur identification.

En perspective, nous démarrons un projet P.I.A.⁷⁰ alliant l'analyse des compétences mobilisées au sein des entreprises et les mécanismes de régulation de l'activité de construction d'un programme de formation. Ce travail pourra s'appuyer, entre autres, sur la cartographie des compétences que nous développons dans le projet A.N.R. CaTCaP, mais intégrera très probablement des modules d'analyse issus de l'IA pour élaborer un module de régulation dynamique de l'activité d'adaptation de la formation professionnelle.

Une démarche scientifique abductive

Au cours de nos recherches sur les compétences tracées, nous suivons une démarche scientifique de découverte plutôt que d'application de théories, pour nous affranchir de toute démarche prescriptive, comme par exemple la recherche de compétences à partir de profils de compétences prédéfinis (e.g. fiches de poste). Nous privilégions les études de terrains qui constituent des milieux écologiquement fondés, pour opérationnaliser les éléments mesurables que nous exprimons dans nos questions de recherche et qui garantissent l'objectivité de nos résultats. Nous avons conscience que l'omniprésence de la technologie dans les activités professionnelles induit des considérations dans le champ des sciences humaines et sociales, lesquelles produisent des résultats qualitatifs difficilement réifiables de par les liens étroits avec les spécificités individuelles et sociétales. La dimension sociologique qui est apparue avec l'extension du monde virtuel, au point de supplanter le monde réel, introduit une part de subjectivité qu'il nous faudra considérer dans nos futures analyses. Le choix d'une approche systémique pour l'observation de l'activité humaine, et celui d'exploiter les données numériques associées, tiennent à nos exigences scientifiques dans le domaine de l'informatique, comme celle d'exprimer nos questions de recherche à partir de termes mesurables objectivement. Nous nous appuyons sur les données empiriques glanées dans des expérimentations pour construire notre connaissance scientifique et nos hypothèses. Ces observations nous permettent d'établir des corrélations objectives entre les données, à partir desquelles nous construisons nos modèles théoriques (volonté de généralité) qui permettent de valider ou d'invalider nos hypothèses.

Nos travaux initiaux s'inscrivent principalement dans la communauté E.I.A.H., laquelle a bénéficié progressivement du décloisonnement des communautés de recherche induit par l'intérêt croissant des résultats de recherche et des applications dans des domaines plus larges. L'ancrage industriel associé à la formation professionnelle (projet P.I.A. « Campus des métiers et des qualifications »), et les orientations en termes d'analyse des données de nos travaux (ouverture à l'intelligence artificielle) nous conduisent à accentuer nos futures investigations dans des disciplines spécifiques, telles que le *Learning Analytics* et le *Data Analytics* par exemple. Les questions relatives à l'humain étant considéré comme des secteurs sensibles et notamment dans la communauté du génie industriel très intéressée par les résultats de l'intelligence artificielle, il nous paraît important de préserver la dimension humaine dans les problématiques que nous adressons. Nos futurs travaux dans le projet P.I.A. nous conduiront à explorer les problématiques de l'extraction des données éducatives, en vue de développer des méthodes d'exploration de celles-ci de plus en plus volumineuses provenant des environnements éducatifs. Les travaux publiés dans les conférences telles que « *Education Data Mining* » et « *Learning Analytics and Knowledge* » permettront de positionner le curseur du contrôle humain dans nos futurs travaux.

70. Programme d'Investissement d'Avenir « Campus des métiers et des qualifications »

Dans le cadre du projet CaTCaP, nous explorons les compétences sur la base des traces numériques laissées à travers les processus « métier ». L'exploitation des données d'interaction pendant les processus ouvre le champ du *process mining* (exploration de processus) qui nous semble être une piste intéressante pour l'élaboration du système de régulation du projet « Campus des métiers et des qualifications ». Nous mesurons les enjeux d'une démarche *process mining* (P.M.) combinée au *trace mining* (T.M.) qui remet l'individu au centre du système.

La Figure 4.24 qui fournit une photographie actuelle de nos cibles en termes de publications devrait évoluer avec l'extension de certains domaines, comme les réseaux sociaux en tant que source croissante d'informations utiles et les sciences de l'Homme et de la société concernées par les effets bienfaisants et délétères du monde virtuel.

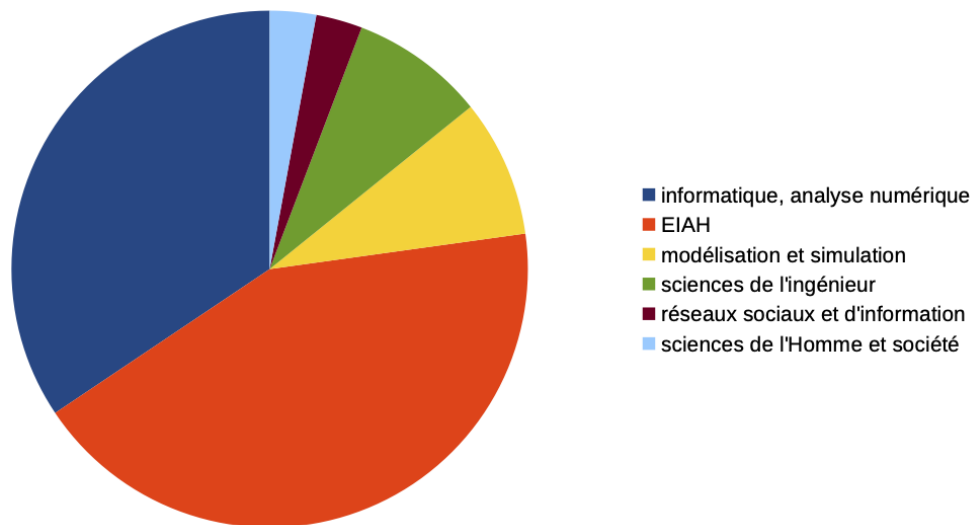


FIGURE 4.24 – domaines de publication

En perspective, nous continuerons d'explorer des solutions techniques sur des terrains éducatifs (projets HyPE-13, @SPIRE et « Campus des métiers et des qualifications ») et industriels (projets CaTCaP et « Campus des métiers et des qualifications ») en vue d'élaborer des modèles théoriques et opérationnels à partir d'analyses des données d'activités collaboratives issues des environnements informatiques. Nos travaux de recherche montrent la complexité des systèmes capables d'exploiter l'activité humaine pour assister les personnes dans leurs tâches, pour adapter les interfaces humain-machine, ou pour automatiser des tâches « métier » récurrentes.

L'objectif principal de cette habilitation à diriger des recherches est de créer une synergie en fédérant des moyens autour d'équipes de chercheuses et de chercheurs de plusieurs disciplines scientifiques dans une topologie d'organisation étendue (i.e. plusieurs organismes organisés en réseau), afin de produire des systèmes complexes basés sur l'observation de l'activité humaine.

Bibliographie

- Abdelmalek, Ali Aït. 2004. "Edgar Morin, sociologue et théoricien de la complexité :." *Societes* no 86 (4) : 99–117. Bibliographie_available : 1 Cairndomain : www.cairn.info Cite Par_available : 1 Publisher : De Boeck Supérieur, Accessed 2021-07-14. <https://www.cairn.info/revue-societes-2004-4-page-99.htm>.
- Bakhouyi, Abdellah, Rachid Dehbi, Mouad Banane, and Mohamed Talea. 2020. "A Semantic Web Solution for Enhancing the Interoperability of E-Learning Systems by Using Next Generation of SCORM Specifications." In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019)*, edited by Mostafa Ezziyyani, Cham, 56–67. Springer International Publishing.
- Belkadi, F., E. Bonjour, and M. Dulmet. 2007. "Competency characterisation by means of work situation modelling." *Computers in Industry* 58 (2) : 164–178. Accessed 2021-08-02. <https://linkinghub.elsevier.com/retrieve/pii/S0166361506001552>.
- Belmonte, Jesús López, A. Segura-Robles, Antonio José Moreno Guerrero, and M. Parra-González. 2020. "Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science." *undefined* Accessed 2022-04-01. <https://www.semanticscholar.org/paper/Machine-Learning-and-Big-Data-in-the-Impact-A-with-Belmonte-Segura-Robles/9112db841ba3735c07616af0333caa0dd6d4df73>.
- Bemmami, Kamel Eddin, Jean-Luc Maire, Lilia Gzara, Christophe Courtin, and Olivier Pouydebat. 2021. "Toward A New Model Of Competences In Work Situations." *IFAC-PapersOnLine* 54 (1) : 1150 – 1155. Publisher : Elsevier, Accessed 2022-01-14. <https://hal.archives-ouvertes.fr/hal-03430160>.
- Benbrahim, Chahinaze Fikri, Elboudhidi Jaber, Bakkali Hajira, N. Cheikhrouhou, N. Sefiani, and K. Reklouli. 2018. "Towards an ontological approach to company competences management." *2018 International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA)*.
- Boucher, Xavier, and Patrick Burlat. 2003. "Vers l'intégration des compétences dans le pilotage des performances de l'entreprise." *Journal Européen des Systèmes Automatisés (JESA)* Vol. 37 (3) : pp. 363. Accessed 2022-07-13. <https://hal-emse.ccsd.cnrs.fr/emse-00755899>.
- Bouhineau, Denis, Vanda Luengo, Nadine Mandran, Michael Ortega, Claire Wajeman, and Ben-Manson Toussaint. 2013. "Open platform to model and capture experimental data in Technology enhanced learning systems." Accessed 2021-07-28. <https://hal.inria.fr/hal-00948795>.
- Boumane, Abderrazak, Abdennebi Talbi, Christian Tahon, and Driss Bouami. 2006. "Contribution à la modélisation de la compétence." Apr.
- Butoianu, Valentin, Philippe Vidal, Katrien Verbert, Erik Duval, and Julien Broisin. 2010. "User context and personalized learning :A federation of contextualized attention metadata." *Journal of Universal Computer Science* 16 (16) : 2252–2271.
- Büttner, Ricardo. 2018. "Development of an Efficient Europe-wide e-Recruiting System (European Recruiting 2020)." In *German-Turkish Perspectives on IT and Innovation Management*, edited by Fehim Bakırcı, Thomas Heupel, Orhan Kocagöz, and Üstün Özen, FOM-Edition, Wiesbaden, 267–274. Springer Fachmedien.
- Castillo, Joseph, Carmel J. Caruana, Paul Simon Morgan, Catherine Westbrook, and Adrian Mizzi. 2016. "Validation of a competence profile for MR radiographers using a formal research process." Accepted : 2017-04-02T19 :31 :18Z Publisher : European Society of Radiology, Accessed 2020-12-14. <https://www.um.edu.mt/library/oar/handle/123456789/18043>.

- Chabert, G., J. Ch. Marty, B. Caron, T. Carron, L. Vignollet, and C. Ferraris. 2006. "The Electronic Schoolbag, a CSCW workspace : presentation and evaluation." *AI & SOCIETY* 20 (3) : 403–419. Accessed 2021-07-15. <https://doi.org/10.1007/s00146-005-0026-1>.
- Chanier, Thierry, and Maud Ciekanski. 2010. "Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage : un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage." *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication* (Vol. 13). Number : Vol. 13 Publisher : Adalsic, Accessed 2021-07-28. <https://journals.openedition.org/alsic/1666>.
- Chebil, Hajer, Christophe Courtin, and Jean-Jacques Girardot. 2012a. "An ontology-based approach for sharing and analyzing learning trace corpora." In *6th IEEE International Conference on Semantic Computing*, Palermo, Italy, Sep., 101–108. Accessed 2019-02-25. <https://hal-emse.ccsd.cnrs.fr/emse-00742148>.
- Chebil, Hajer, Christophe Courtin, and Jean-Jacques Girardot. 2012b. "The proxy model : a new approach to sharing and analyzing learning traces corpora." *International Journal of Information and Education Technology* 2 (4) : 208–211. Accessed 2019-02-25. <https://hal-emse.ccsd.cnrs.fr/emse-00741929>.
- Chebil, Hajer, Christophe Courtin, and Jean-Jacques Girardot. 2014. "Building and Analyzing a Corpus of Contextualized Traces Collected during a Technology Enhanced Teaching Module." In *International Conference on Advanced Learning Technologies - ICALT*, Athens, Greece, Jul. Accessed 2019-02-25. <https://hal.archives-ouvertes.fr/hal-01186561>.
- Chebil, Hajer, Christophe Courtin, and Jean-Jacques Girardot. 2015. "A case study to validate the PROXYMA approach to share and analyse contextualised interaction trace corpora in a TEL environment." *International Journal of Learning Technology* 10 (4) : p.291–312. Accessed 2019-02-25. <https://hal.archives-ouvertes.fr/hal-01270899>.
- Choquet, Christophe, and Sébastien Iksal. 2007. "Modeling Tracks for the Model Driven Re-engineering of a TEL System." *Journal of Interactive Learning Research* 18 (2) : 161–184. Publisher : Association for the Advancement of Computing in Education (AACE), Accessed 2021-07-27. <https://www.learntechlib.org/primary/p/22967/>.
- Corbel, Annie, Jean-Jacques Girardot, and Kristine Lund. 2006. "A method for capitalizing upon and synthesizing analyses of human interactions." Oct., 38. Accessed 2021-07-26. <https://halshs.archives-ouvertes.fr/halshs-01493921>.
- Coron, Clotilde. 2019. "Big Data et pratiques de GRH." *Management & Data Science* 3 (1). Accessed 2022-04-01. <https://halshs.archives-ouvertes.fr/halshs-01961214>.
- Courtin, C., and S. Talbot. 2006. "Trace Analysis in Instrumented Collaborative Learning Environments." In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, Jul., 1036–1038.
- Courtin, Christophe. 2010. "CARTE An Observation Station to Regulate Activity in a Learning Context." In *Learning and Instruction in the Digital Age*, 207–223. Springer US. Accessed 2018-11-23. <https://hal.archives-ouvertes.fr/hal-00804910>.
- Courtin, Christophe, and Stéphane Talbot. 2005. "AN ARCHITECTURE TO RECORD TRACES IN INSTRUMENTED COLLABORATIVE LEARNING ENVIRONMENTS." In *IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA)*, Porto, Portugal, Dec. Accessed 2021-07-22. <https://hal.archives-ouvertes.fr/hal-02885190>.
- Courtin, Christophe, and Stéphane Talbot. 2007. "Une Station d'Observation pour des Situations d'Apprentissage Collaboratif Instrumenté." In *Actes de la conférence EIAH 2007*, Jun. INRP. Accessed 2018-11-23. <https://hal.archives-ouvertes.fr/hal-00161613>.
- Courtin, Christophe, and Stéphane Talbot. 2009. "Automatic Analysis Assistant for Studies of Computer-Supported Human Interactions." In *Learning in the Synergy of Multiple Disciplines*, edited by Ulrike Cress, Vania Dimitrova, and Marcus Specht, Lecture Notes in Computer Science, Berlin, Heidelberg, 572–583. Springer.

- Courtin, Christophe, Stéphane Talbot, and Mathieu Mangeot. 2010. "Activity regulation for the participative creation of data on-line." In *International Conference on Machine and Web Intelligence*, Algiers, Algeria, France, Oct., 132–139. Accessed 2019-02-25. <https://hal.archives-ouvertes.fr/hal-00968711>.
- Courtin, Christophe, and Miguel Tomasena. 2016. "A benchmarking platform for analyzing corpora of traces : the recognition of the users' involvement in fields of competencies." In *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, Shanghai, China, Dec. Accessed 2019-02-25. <https://hal.archives-ouvertes.fr/hal-01537995>.
- Desmarais, Lise, Lise Duquette, Delphine Renié, and Michel Laurier. 1997. "Evaluating Learning and Interactions in a Multimedia Environment." *Computers and the Humanities* 31 (4) : 327–349. Accessed 2021-07-23. <https://doi.org/10.1023/A:1001062407103>.
- Dimitracopoulou, A., A. Martinez, Y. Dimitriadis, A. Morch, S. Ludvigsen, A. Harrer, U. Hoppe, et al. 2004. "State of the art on interaction analysis : interaction analysis indicators. Délivrable D26.1 du JEIRP (Jointly Executed Integrated Research Project) ICALTS (Interaction and Collaboration Analysis Supporting Teachers' and Students' Self-regulation) du réseau européen Kaleidoscope." .
- Drulhe, Marcel. 2008. "Orientations épistémiques et niveaux d'analyse en sociologie." *SociologieS* Publisher : Association internationales des sociologues de langue française (AISLF), Accessed 2021-07-27. <https://journals.openedition.org/sociologies/2123>.
- Dyke, Gregory, Jean-Jacques Girardot, Kristine Lund, and A. Corbel. 2007. "Analysing face to face computer-mediated interactions." Budapest, Hungary, Aug. Accessed 2021-07-27. <https://halshs.archives-ouvertes.fr/halshs-01494954>.
- Dyke, Gregory, Kristine Lund, and Jean-Jacques Girardot. 2009. "Tatiana : an environment to support the CSCL analysis process." In *Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1*, CSCL'09, Rhodes, Greece, Jun., 58–67. International Society of the Learning Sciences.
- Ellis, Clarence, and Jacques Wainer. 1994. "A conceptual model of groupware." In *Proceedings of the 1994 ACM conference on Computer supported cooperative work, CSCW '94*, New York, NY, USA, Oct., 79–88. Association for Computing Machinery. Accessed 2021-07-15. <https://doi.org/10.1145/192844.192878>.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. "From Data Mining to Knowledge Discovery in Databases." *AI Magazine* 17 (3) : 37–37. Number : 3, Accessed 2022-07-22. <https://ojs.aaai.org/index.php/aimagazine/article/view/1230>.
- Ferraris, C., and C. Martel. 2000. "Regulation in groupware : the example of a collaborative drawing tool for young children." In *Proceedings Sixth International Workshop on Groupware. CRIWG 2000*, Oct., 119–127.
- Ferraris, Christine, Laurence Vignollet, Christian Martel, Andreas Harrer, and Yannis Dimitriadis. 2009. "Competitive challenge on adapting activities modeled by CSCL scripts." In *Proceedings of the 9th international conference on Computer supported collaborative learning - CSCL'09*, Vol. 2, Rhodes, Greece, 229–230. Association for Computational Linguistics. Accessed 2021-07-31. <http://portal.acm.org/citation.cfm?doid=1599503.1599583>.
- Galinon-Mélénec, Béatrice. 2011. "Déclinaisons du paradigme de la trace." In *L'Homme trace, Perspectives anthropologiques des traces contemporaines*, edited by GALINON-MELENEC Béatrice, CNRS Alpha, 351–371. CNRS éditions. Accessed 2022-01-03. <https://halshs.archives-ouvertes.fr/halshs-01080009>.
- George, Sébastien, and Hervé Labas. 2005. "Fostering Discussions with Forum Tools Based on E-Learning Standards." In *IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2005)*, Porto, Portugal, Dec., 60–68.

- Gertner, Abigail S., and Kurt VanLehn. 2000. "Andes : A Coached Problem Solving Environment for Physics." In *Intelligent Tutoring Systems*, edited by Gilles Gauthier, Claude Frasson, and Kurt VanLehn, Lecture Notes in Computer Science, Berlin, Heidelberg, 133–142. Springer.
- Giguet, Emmanuel, Nadine Lucas, François-Marie Blondel, and Eric Bruillard. 2009. "Share and explore discussion forum objects on the Calico website." In *Proceedings of the 9th international conference on Computer supported collaborative learning - CSCL'09*, Vol. 2, Rhodes, Greece, 174–176. Association for Computational Linguistics. Accessed 2021-07-31. <http://portal.acm.org/citation.cfm?doid=1599503.1599562>.
- Glahn, C., M. M. Specht, and E. J. R. Koper. 2007. "Processing learner profiles for smart indicators." Maastricht, The Netherlands, Oct., 454–463. Accessed 2021-07-27. <https://research.ou.nl/en/publications/processing-learner-profiles-for-smart-indicators-2>.
- Graczyk-Kucharska, Magdalena, Maciej Szafranski, Marek Golinski, Malgorzata Spychala, and Kamila Borsekova. 2018. "Model of Competency Management in the Network of Production Enterprises in Industry 4.0—Assumptions." In *Advances in Manufacturing*, edited by Adam Hamrol, Olaf Ciszak, Stanisław Legutko, and Mieczysław Jurczyk, Lecture Notes in Mechanical Engineering, Cham, 195–204. Springer International Publishing.
- Guin, Nathalie, Stéphanie Jean-Daubias, and Sandra Nogry. 2007. "Le projet AMBRE : utiliser un paradigme d'apprentissage fondé sur le RÀPC pour faire acquérir à des élèves une méthode de résolution de problèmes." In *Raisonnement à partir de cas 2 - surveillance, diagnostic et maintenance*, edited by J. Renaud, B. Chebel-Morello, B. Fuchs, and J. Lieber, 203–226. Hermès-Lavoisier. Accessed 2021-07-27. <https://hal.archives-ouvertes.fr/hal-01573251>.
- Gutwin, Carl, and Saul Greenberg. 2002. "A Descriptive Framework of Workspace Awareness for Real-Time Groupware." *Computer Supported Cooperative Work (CSCW)* 11 (3) : 411–446. Accessed 2021-07-20. <https://doi.org/10.1023/A:1021271517844>.
- Guéraud, Viviane, Jean-Michel Adam, Jean-Philippe Pernin, Gaëlle Calvary, and Jean-Pierre David. 2004. "L'exploitation d'Objets Pédagogiques Interactifs à distance : le projet FORMID." *STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation)* 11 : 46 pages. Publisher : ATIEF, Accessed 2021-07-31. <https://hal.archives-ouvertes.fr/hal-00696400>.
- Harrower, Kathryn. 2019. "Algorithmic Decision-Making in Organizations : Network Data Mining, Measuring and Monitoring Work Performance, and Managerial Control." *Psychosociological Issues in Human Resource Management* 7 (2) : 7–12. Publisher : Addleton Academic Publishers, Accessed 2022-04-10. <https://www.ceeol.com/search/article-detail?id=809353>.
- Hecklau, Fabian, Mila Galeitzke, Sebastian Flachs, and Holger Kohl. 2016. "Holistic Approach for Human Resource Management in Industry 4.0." *Procedia CIRP* 54 : 1–6. Accessed 2022-04-01. <https://www.sciencedirect.com/science/article/pii/S2212827116308629>.
- Heiden, Serge. 2010. "The TXM Platform : Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme." Vol. 2, Waseda University, Sendai, Japan., Nov., 389–398. Institute for Digital Enhancement of Cognitive Development, Waseda University. Issue : 3, Accessed 2021-08-03. <https://halshs.archives-ouvertes.fr/halshs-00549764>.
- Jaczynski, Michel, and Brigitte Trousse. 1998. "WWW assisted browsing by reusing past navigations of a group of users." In *Advances in Case-Based Reasoning*, edited by Barry Smyth and Pádraig Cunningham, Lecture Notes in Computer Science, Berlin, Heidelberg, 160–171. Springer.
- Jaillon, Philippe, Annie Corbel, and Jean-Jacques Girardot. 2002. "DREW : A Dialogical Reasoning Web Tool." In *International Conference on Information and Communication Technologies in Education (ICTE 2002)*, edited by Université Paul Sabatier, Badajoz, Spain, Nov., 5p. Accessed 2021-07-26. <https://hal-emse.ccsd.cnrs.fr/emse-00772776>.
- Kalfoglou, Yannis, and Marco Schorlemmer. 2003. "Ontology mapping : the state of the art." *The Knowledge Engineering Review* 18 (1) : 1–31. Publisher : Cambridge University Press, Accessed 2021-07-29. <https://www.cambridge.org/core/journals/knowledge-engineering-review/article/abs/ontology-mapping-the-state-of-the-art/9A273424C3873243A0DC50FD43C6AD4E>.

- Kaptelinin, Victor, and Bonnie Nardi. 2018. "Activity Theory as a Framework for Human-Technology Interaction Research." *Mind, Culture, and Activity* 25 (1) : 3–5. Publisher : Routledge _eprint : <https://doi.org/10.1080/10749039.2017.1393089>, Accessed 2021-07-26. <https://doi.org/10.1080/10749039.2017.1393089>.
- King, Gary. 2007. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." *Sociological Methods & Research* 36 (2) : 173–199. Publisher : SAGE Publications Inc, Accessed 2021-07-28. <https://doi.org/10.1177/0049124107306660>.
- Koedinger, Kenneth R., Kyle Cunningham, Alida Skogsholm, and Brett Leber. 2008. "An Open Repository and analysis tools for fine-grained, longitudinal learner data." *Enterprise Development and Microfinance* 157–166.
- Kou, Yue, Zhiyong Jia, and Yihua Wang. 2013. "A Comparative Research on Competency and Competence, Competency Model and Competence Model." In *Proceedings of the Sixth International Conference on Management Science and Engineering Management*, edited by Jiuping Xu, Masoom Yasinzaï, and Benjamin Lev, London, 681–693. Springer.
- Laflaquière, Julien, Lotfi S. Settouti, Yannick Prié, and Alain Mille. 2006. "Trace-Based Framework for Experience Management and Engineering." In *Knowledge-Based Intelligent Information and Engineering Systems*, edited by Bogdan Gabrys, Robert J. Howlett, and Lakhmi C. Jain, Lecture Notes in Computer Science, Berlin, Heidelberg, 1171–1178. Springer.
- Lallé, Sébastien, Vanda Luengo, and Nathalie Guin. 2013. "Assistance in Building Student Models Using Knowledge Representation and Machine Learning." In *Artificial Intelligence in Education*, edited by H. Chad Lane, Kalina Yacef, Jack Mostow, and Philip Pavlik, Lecture Notes in Computer Science, Berlin, Heidelberg, 754–757. Springer.
- Le Deist, Françoise Delamare, and Jonathan Winterton. 2005. "What Is Competence?" *Human Resource Development International* 8 (1) : 27–46. Accessed 2022-04-03. <http://www.tandfonline.com/doi/abs/10.1080/1367886042000338227>.
- Ley, Tobias, and Dietrich Albert. 2003. "Identifying employee competencies in dynamic work domains : Methodological considerations and a case study." *Journal of Universal Computer Science* 9 (12) : 1500–1518.
- Li, Jingran, Fei Tao, Ying Cheng, and Liangjin Zhao. 2015. "Big Data in product lifecycle management." *The International Journal of Advanced Manufacturing Technology* 81 (1) : 667–684. Accessed 2022-06-29. <https://doi.org/10.1007/s00170-015-7151-x>.
- Luengo, Vanda, Lucile Vadcard, Michel Dubois, and Dima Mufti-Alchawafa. 2006. "TEL EOS : de l'analyse de l'activité professionnelle à la formalisation des connaissances pour un environnement d'apprentissage." In *17e journées francophones d'Ingénierie des connaissances IC'2006*, edited by Mounira Harzallah, Jean Charlet, and Nathalie Aussenac-Gilles, Nantes, France, Jun., 101–110. Accessed 2021-07-29. <https://telearn.archives-ouvertes.fr/hal-00190719>.
- Mangeot, Mathieu, and Antoine Chalvin. 2006. "Dictionary Building with the Jibiki Platform : the GDEF case." In *LREC 2006*, Genova, Italy, May, 1666–1669. Accessed 2021-07-21. <https://hal.archives-ouvertes.fr/hal-00968611>.
- Mangeot, Mathieu, and Hong Thai Nguyen. 2009. "Building Lexical Resources : Towards Programmable Contributive Platforms." In *2009 IEEE-RIVF International Conference on Computing and Communication Technologies*, Jul., 1–8.
- Martel, Christian, Christine Ferraris, Bernard Caron, Thibault Carron, Ghislaine Chabert, Christophe Courtin, Laurence Gagnière, Jean-Charles Marty, and Laurence Vignollet. 2004. "A Model for CSCL Allowing Tailorability : Implementation in the "Electronic Schoolbag" Groupware." In *Groupware : Design, Implementation, and Use*, edited by Gert-Jan de Vreede, Luis A. Guerrero, and Gabriela Marín Raventós, Lecture Notes in Computer Science, 322–337. Springer Berlin Heidelberg.

- Martínez, A., P. De La Fuente, and Y. Dimitriadis. 2003. "Towards an XML-Based Representation of Collaborative Action." In *Designing for Change in Networked Learning Environments : Proceedings of the International Conference on Computer Support for Collaborative Learning 2003*, edited by Barbara Wasson, Sten Ludvigsen, and Ulrich Hoppe, Computer-Supported Collaborative Learning, 379–383. Dordrecht : Springer Netherlands. Accessed 2021-07-28. https://doi.org/10.1007/978-94-017-0195-2_46.
- Martínez-Monés, Alejandra, Andreas Harrer, Beatriz Barros, Lars Bollen, Tharrenos Bratitsis, Weiqin Chen, Ageliki Dimitracopoulou, et al. 2005. *Library of Interaction Analysis Tools*.
- May, Madeth, Sébastien George, and Patrick Prévot. 2011. "TrAVis to Enhance Online Tutoring and Learning Activities : Real Time Visualization of Students Tracking Data." *International Journal of Interactive Technology and Smart Education* 8 : 52–69. Publisher : EmeraldInsight, Accessed 2021-07-28. <https://hal.archives-ouvertes.fr/hal-01354507>.
- Mill, John Stuart. 1856. *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles, and the Methods of Scientific Investigation*. J. W. Parker. Google-Books-ID : 8hzkAAAA-MAAJ.
- Mille, Alain, Guy Caplat, and Mick Philippon. 2006. "Faciliter les activités des utilisateurs d'environnements informatiques : quoi, quand, comment ?" *Intellectica* 44 (2) : 121–143. Publisher : Persée - Portail des revues scientifiques en SHS, Accessed 2021-07-19. https://www.persee.fr/doc/intel_0769-4113_2006_num_44_2_1294.
- Miranda, Sergio, Francesco Orciuoli, Vincenzo Loia, and Demetrios Sampson. 2017. "An ontology-based model for competence management." *Data & Knowledge Engineering* 107 : 51–66. Accessed 2022-04-01. <https://www.sciencedirect.com/science/article/pii/S0169023X16303640>.
- Mostow, Jack, and Joseph E Beck. n.d. "Project LISTEN's Reading Tutor : Interactive Event Description." 3.
- Nagels, Marc. 2009. "Évaluer des compétences ou des performances ?" In *Évaluation et développement professionnel*, edited by Université catholique de Louvain-La-Neuve, Louvain-La-Neuve, Belgium, Jan., C004. Accessed 2022-07-15. <https://hal.archives-ouvertes.fr/hal-00357588>.
- Nguyen-Xuan, Anh, Jean-François Nicaud, Anne Bastide, and Emmanuel Sander. 2002. "Les expérimentations du projet APLUSIX." *Sciences et techniques éducatives* 9 (1) : 63–90. Accessed 2021-07-31. https://www.persee.fr/doc/stice_1265-1338_2002_num_9_1_1500.
- Nicolaescu, Sergiu Stefan, Adrian Florea, Claudiu Vasile Kifor, Ugo Fiore, Nicolae Cocan, Ilie Receu, and Paolo Zanetti. 2020. "Human capital evaluation in knowledge-based organizations based on big data analytics." *Future Generation Computer Systems* 111 : 654–667. Accessed 2022-04-01. <https://www.sciencedirect.com/science/article/pii/S0167739X19306351>.
- Nogry, Sandra, Stéphanie Jean-Daubias, and Nathalie Guin-Duclosson. 2012. "How to combine objectives and methods of evaluation in iterative ILE design : lessons learned from designing Ambre-add." *Interactive Learning Environments* 20 (2) : 155–175. Publisher : Taylor & Francis (Routledge), Accessed 2021-07-31. <https://hal-univ-paris8.archives-ouvertes.fr/hal-01015773>.
- Pernin, Jean-Philippe, and Anne Lejeune. 2004. "Modèle pour la réutilisation de scénarios d'apprentissage." In *TICE Méditerranée*, Nice, France, 48.
- Pinzone, Marta, Paola Fantini, Stefano Perini, Stefano Garavaglia, Marco Taisch, and Giovanni Miragliotta. 2017. "Jobs and Skills in Industry 4.0 : An Exploratory Research." Vol. AICT-513, Sep., 282. Springer International Publishing. Issue : Part I, Accessed 2022-04-01. <https://hal.inria.fr/hal-01666201>.
- Pouydebat, Olivier. n.d. "Des compétences individuelles et collectives tracées pendant les activité d'exploitation et d'exploration à la performance." Accessed 2022-04-11. <http://www.theses.fr/s210257>.

- Pouydebat, Olivier, Christophe Courtin, Kamel-Eddin Bemmami, Elodie Gardet, and Georges Habchi. 2021. "Automation of competences extraction - HumaNeTrust : an experimental computer system used in an academic laboratory." In *CIGI-Qualita21 : Conférence Internationale Génie Industriel QUALITA - Grenoble*, Grenoble, France, May. Accessed 2021-08-02. <https://hal.archives-ouvertes.fr/hal-03203592>.
- Rabardel, Pierre. 1995. *Les hommes et les technologies ; approche cognitive des instruments contemporains*. Armand Colin. Accessed 2021-07-16. <https://hal.archives-ouvertes.fr/hal-01017462>.
- Rath, Andreas S., Didier Devaurs, and Stefanie N. Lindstaedt. 2009. "UICO : an ontology-based user interaction context model for automatic task detection on the computer desktop." In *Proceedings of the 1st Workshop on Context, Information and Ontologies - CIAO '09*, Heraklion, Greece, 1–10. ACM Press. Accessed 2021-07-31. <http://portal.acm.org/citation.cfm?doid=1552262.1552270>.
- Reffay, Christophe, and Marie-Laure Betbeder. 2009. "Sharing corpora and tools to improve interaction analysis." In *EC-TEL 2009, 4th European Conference on Technology Enhanced Learning*, France, 196–210. Accessed 2021-07-31. <https://hal.archives-ouvertes.fr/hal-00662914>.
- Reffay, Christophe, Thierry Chanier, Muriel Noras, and Marie-Laure Betbeder. 2008. "Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche." *STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation)* 15 : x–x. Publisher : ATIEF, Accessed 2021-08-01. <https://edutice.archives-ouvertes.fr/edutice-00159733>.
- Refianti, Rina, Achmad Mutiara, and A.A. Syamsudduha. 2016. "Performance Evaluation of Affinity Propagation Approaches on Data Clustering." *International Journal of Advanced Computer Science and Applications* 7.
- Rey, Bernard. 2015. "La notion de compétence : usages et enjeux." *Le français aujourd'hui* 191 (4) : 15–24. Place : Paris Publisher : Armand Colin, Accessed 2022-07-13. <https://www.cairn.info/revue-le-francais-aujourd-hui-2015-4-page-15.htm>.
- Sanghi, Seema. 2016. *The Handbook of Competency Mapping : Understanding, Designing and Implementing Competency Models in Organizations*. SAGE Publications India. Google-Books-ID : tGDIDAAA-QBAJ.
- Sauermann, Leo, Gunnar Aastrand Grimnes, Malte Kiesel, Christiaan Fluit, Heiko Maus, Dominik Heim, Danish Nadeem, Benjamin Horak, and Andreas Dengel. 2006. "Semantic Desktop 2.0 : The Gnowsiss Experience." In *The Semantic Web - ISWC 2006*, edited by Isabel Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Mike Uschold, and Lora M. Aroyo, Lecture Notes in Computer Science, Berlin, Heidelberg, 887–900. Springer.
- Schimke, S., C. Vielhauer, and J. Dittmann. 2004. "Using adapted Levenshtein distance for on-line signature authentication." In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 2, Aug., 931–934 Vol.2. ISSN : 1051-4651.
- Schimke, Sascha, and Claus Vielhauer. 2007. "Similarity searching for on-line handwritten documents." *Journal on Multimodal User Interfaces* 1 (2) : 49–54. Accessed 2022-04-13. <https://doi.org/10.1007/BF02910058>.
- Schmid, H. 1999. "Improvements in Part-of-Speech Tagging with an Application to German." In *Natural Language Processing Using Very Large Corpora*, edited by Susan Armstrong, Kenneth Church, Pierre Isabelle, Sandra Manzi, Evelyne Tzoukermann, and David Yarowsky, Text, Speech and Language Technology, 13–25. Dordrecht : Springer Netherlands. Accessed 2021-08-03. https://doi.org/10.1007/978-94-017-2390-9_2.
- Settouti, Lemya, Nathalie Guin, Vanda Luengo, and Alain Mille. 2011. "Adaptable and Reusable Query Patterns for Trace-Based Learner Modelling." In *Towards Ubiquitous Learning*, edited by Carlos Delgado Kloos, Denis Gillet, Raquel M. Crespo García, Fridolin Wild, and Martin Wolpers, Lecture Notes in Computer Science, Berlin, Heidelberg, 384–397. Springer.

- Settouti, Lotfi Sofiane, Yannick Prié, Pierre-Antoine Champin, Jean-Charles Marty, and Alain Mille. 2009a. *A Trace-Based Systems Framework : Models, Languages and Semantics*. report. Accessed 2021-07-21. <https://hal.inria.fr/inria-00363260>.
- Settouti, Lotfi Sofiane, Yannick Prié, Damien Cram, Pierre-Antoine Champin, and Alain Mille. 2009b. "A Trace-Based Framework for supporting Digital Object Memories." In *1st International Workshop on Digital Object Memories (DOMe'09) in the 5th International Conference on Intelligent Environments (IE 09)*, Barcelona, Spain, Jul., 39–44. IOS Press. Accessed 2021-07-16. <https://hal.archives-ouvertes.fr/hal-01437735>.
- Smith, Raymond, and Bogdan Korel. 2001. "Slicing Event Traces of Large Software Systems." *arXiv :cs/0101005* ArXiv : cs/0101005, Accessed 2021-07-16. <http://arxiv.org/abs/cs/0101005>.
- Stamper, John, Ken Koedinger, Ryan S. J. d. Baker, Alida Skogsholm, Brett Leber, Jim Rankin, and Sandy Demi. 2010. "PSLC DataShop : A Data Analysis Service for the Learning Science Community." In *Intelligent Tutoring Systems*, edited by Vincent Aleven, Judy Kay, and Jack Mostow, Lecture Notes in Computer Science, Berlin, Heidelberg, 455–455. Springer.
- Stern, Hermann, Rene Kaiser, Philip Hofmair, Peter Kraker, Stefanie N. Lindstaedt, and Peter Scheir. 2010. "Content recommendation in APOSDLE using the associative network." *Journal of Universal Computer Science* 16 (16) : 2214–2231.
- Sugimori, Y., K. Kusunoki, F. Cho, and S. Uchikawa. 1977. "Toyota production system and Kanban system Materialization of just-in-time and respect-for-human system." *International Journal of Production Research* 15 (6) : 553–564. Publisher : Taylor & Francis _eprint : <https://doi.org/10.1080/00207547708943149>, Accessed 2022-04-09. <https://doi.org/10.1080/00207547708943149>.
- Talbot, Stéphane, and Christophe Courtin. 2008. "Trace Analysis in Instrumented Learning Groupware : an experiment in a practical class at the university." In *Seventh IASTED International Conference WEB-BASED EDUCATION (WBE)*, Innsbruck, Austria, Mar. Accessed 2021-07-21. <https://hal.archives-ouvertes.fr/hal-02884949>.
- Talbot, Stéphane, and Philippe Pernelle. 2003. "Helping in collaborative activity regulation : modeling regulation scenarii." In *Proceedings of the 15th Conference on l'Interaction Homme-Machine, IHM '03*, New York, NY, USA, Nov., 158–165. Association for Computing Machinery. Accessed 2021-07-21. <https://doi.org/10.1145/1063669.1063691>.
- Tardif, Jacques. 2017. *Chapitre 1. Des repères conceptuels à propos de la notion de compétence, de son développement et de son évaluation*. De Boeck Supérieur. Bibliographie_available : 0 Cairndomain : www.cairn.info Cite Par_available : 1 ISSN : 0777-5245 Pages : 15-37 Publication Title : Organiser la formation à partir des compétences, Accessed 2021-08-02. <https://www.cairn.info/organiser-la-formation-a-partir-des-competences--9782807307858-page-15.htm>.
- Teutsch, Philippe, Francis Bangou, and Charlotte Dejean-Thircuir. 2007. "ViCoDiLi, un outil de Visualisation de Contenus de Discussions en Ligne." In *EPAL - Echanger Pour Apprendre en Ligne*, edited by M.-N. Lamy, F. Mangenot, and E. Nissen, Actes du colloque Epal 2007 (Échanger pour apprendre en ligne), Grenoble, France, Jun. Accessed 2021-08-01. <https://hal.archives-ouvertes.fr/hal-02018124>.
- Vargas-Vera, Maria, and Miklos Nagy. 2015. "State of the Art on Ontology Alignment." *International Journal of Knowledge Society Research (IJKSR)* 6 (1) : 17–42. Publisher : IGI Global, Accessed 2021-07-29. <https://www.igi-global.com/article/state-of-the-art-on-ontology-alignment/www.igi-global.com/article/state-of-the-art-on-ontology-alignment/133138>.
- Vivacqua, Adriana, and Henry Lieberman. 2000. "Agents to assist in finding help." In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, The Hague, The Netherlands, 65–72. ACM Press. Accessed 2021-08-02. <http://portal.acm.org/citation.cfm?doid=332040.332408>.

- Wolpers, Martin, Jehad Najjar, Katrien Verbert, and Erik Duval. 2007. "Tracking Actual Usage : the Attention Metadata Approach." *Journal of Educational Technology & Society* 10 (3) : 106–121. Publisher : International Forum of Educational Technology & Society, Accessed 2021-07-28. <https://www.jstor.org/stable/jeductechsoci.10.3.106>.
- Xu, Dongkuan, and Yingjie Tian. 2015. "A Comprehensive Survey of Clustering Algorithms." *Annals of Data Science* 2 (2) : 165–193. Company : Springer Distributor : Springer Institution : Springer Label : Springer Number : 2 Publisher : Springer Berlin Heidelberg, Accessed 2022-06-29. <https://link.springer.com/article/10.1007/s40745-015-0040-1>.
- Yahiaoui, Soumaya, Christophe Courtin, Pierre Maret, and Laurent Tabourot. 2017. "Competences Network Based on Interaction Data for Recommendation and Evaluation Aims." In *Complex Network 2017*, Lyon, France, 989–1001. Accessed 2019-02-25. <https://hal.archives-ouvertes.fr/hal-01994599>.
- Yujian, Li, and Liu Bo. 2007. "A Normalized Levenshtein Distance Metric." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (6) : 1091–1095. Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence.
- Zaiane, O.R., and J. Luo. 2001. "Towards evaluating learners' behaviour in a Web-based distance learning environment." In *Proceedings IEEE International Conference on Advanced Learning Technologies*, Aug., 357–360.

Table des figures

1.1	couplage faible entre CARTE et TATIANA	10
1.2	couplage fort entre CARTE et TATIANA	10
1.3	conversion entre CARTE et CMC	11
1.4	structure du projet CaTCaP	16
1.5	fil conducteur de recherche du bas vers le haut	19
1.6	frise chronologique des activités de recherche	20
1.7	observation et virtualisation	23
1.8	notions d'un système complexe d'observation de l'activité humaine	25
2.1	Structure hiérarchique de l'activité	34
2.2	Les tâches au sein de l'activité	35
2.3	Système à base de traces	37
2.4	Format générique de traces	37
2.5	Encapsulation d'un signal dans une séquence	38
2.6	Niveaux d'abstraction des traces collectées	40
2.7	Modèle du trèfle augmenté	40
2.8	Architecture de la station d'observation	50
2.9	Architecture logicielle pour l'expérimentation	53
2.10	Outil d'édition des règles	54
2.11	Outil de co-édition « Jibiki »	56
2.12	Outil de coordination	57
2.13	Outil de communication « CoffeeRoom »	58
2.14	Outil d'observation « CARTE » pour l'enseignant	59
2.15	Outil d'observation « CARTE » pour le chercheur	59
2.16	Observation avec les logs	61
2.17	Observation avec l'instrumentation	62
2.18	Reconnaissance de séquences et génération de signaux	63
3.1	Les modèles de l'approche PROXYMA	78
3.2	Les modèles de corpus de l'approche PROXYMA (1)	79
3.3	Les modèles de corpus de l'approche PROXYMA (2)	81
3.4	Le modèle sémantique de l'approche PROXYMA	83
3.5	Le méta-modèle du modèle sémantique de l'approche PROXYMA	84
3.6	Le modèle opérationnel de l'approche PROXYMA	86
3.7	L'architecture de la plate-forme de benchmarking BEATCORP	88
3.8	Synchronisation graphe/tableau du forum dans TATIANA	93
3.9	Configuration de la catégorisation des messages du forum dans TATIANA	93
3.10	Affichage de la catégorisation des messages du forum dans TATIANA	94
3.11	Outil Authagora de CALICO	94
3.12	Outil Concordagora de CALICO	95
4.1	modèle de compétence basé sur la dynamique de l'activité	105
4.2	Architecture SBS-SEARCH pour l'extraction des compétences	108

4.3	les compétences et les traces d'activité	109
4.4	séquence directement de haut niveau (à droite)	110
4.5	les composants K.S.A. associés aux user stories	110
4.6	prototype du visualisateur H.N.T. avec des données de H.A.L.	113
4.7	prototype du visualisateur H.N.S.	114
4.8	architecture de BEATCORP avec HAL	114
4.9	interface H.N.T. interactive avec des données de H.A.L.	115
4.10	collaborations interdisciplinaires	115
4.11	architecture d'un système d'identification des compétences	116
4.12	L'architecture HumaNetSkills / BEATCORP	120
4.13	Extrait du modèle de corpus (éditeur Protégé ©)	121
4.14	Extrait de l'ontologie du modèle sémantique (OWL)	122
4.15	Extrait du modèle opérationnel (éditeur Protégé ©)	123
4.16	Réseau des posts avec l'outil Gephi © (50 nœuds, 49 liens)	124
4.17	Réseau des posts réorganisé	125
4.18	Réseau des participants avec l'outil Gephi ©	125
4.19	Longueur des posts (question au niveau 1 et explications au niveau2)	126
4.20	Forme interrogative	127
4.21	Résultats d'analyse sous forme d'un graphe interactif avec « Neo4j »	134
4.22	Résultats d'analyse sous forme d'un tableau de bord statique avec « Tableau public »	135
4.23	couverture des notions d'un système complexe d'observation de l'activité humaine	140
4.24	domaines de publication	144
A.1	proximité des vues et des tâches dans des situations collaboratives	XVI
A.2	différents types de group awareness	XVI
A.3	Éditeur ZWikiCoop	XVII
A.4	programme scientifique	XIX
B.1	Exemple de règle	XXII
B.2	Description des règles	XXIII
C.1	avis du comité d'éthique pour la recherche	XXVIII

Liste des tableaux

4.1	Distribution des posts (sujets en gris, questions en bleu, explications en vert) . . .	124
4.2	Posts identifiés comme étant des questions	129
4.3	Posts identifiés comme étant des explications	129
4.4	Contextes d'expérimentation	131
4.5	Synthèse des caractéristiques de l'expérimentation	131



Annexe 1 : activité scientifique

Éditeur ZWikiCoop / Journée scientifique.

Éditeur ZWikiCoop

Notre première utilisation des traces numériques d'interaction est motivée par le besoin d'indicateurs (e.g. *feedback*) dans un éditeur partagé pour faciliter la collaboration. Nous nous basons sur une trame conceptuelle élaborée par (Gutwin and Greenberg 2002) et représentée sur la Figure A.1.

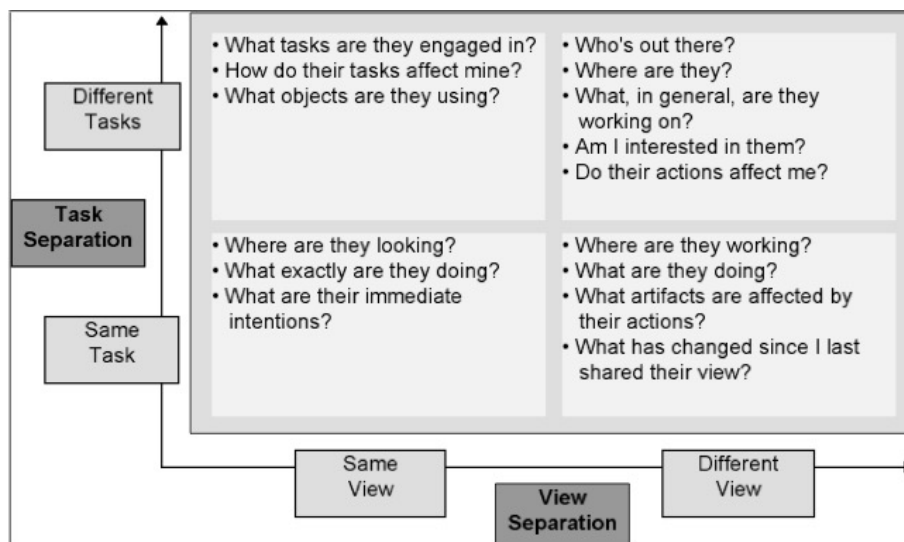


FIGURE A.1 – *proximité des vues et des tâches dans des situations collaboratives*

Nous expérimentons à l'aide de l'outil Web collaboratif « ZWikiCoop », développé en 2002 dans le cadre du **projet BQR (Bonus Qualité Recherche) n°R2001-7**, la spécification d'un éditeur de texte dans un contexte scolaire. Ce projet introduit ainsi les notions de granularité des éléments partagés, d'indicateurs de **group awareness** (ou conscience du groupe) de différents types comme décrits dans la Figure A.2

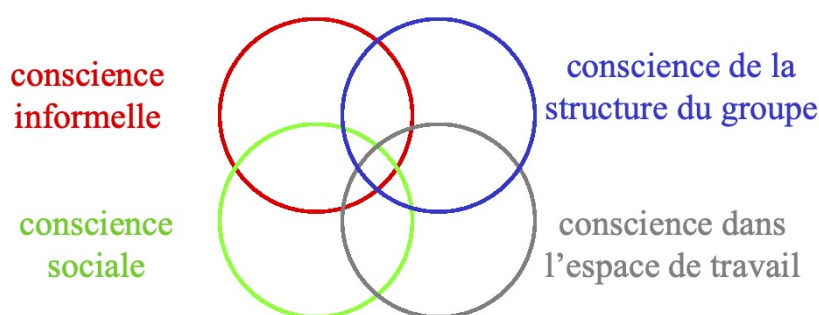


FIGURE A.2 – *différents types de group awareness*

La Figure A.3 présente l'interface graphique de l'outil Web collaboratif « ZWikiCoop » pour illustrer les concepts de granularité des éléments partagés et de *group awareness* à travers des indicateurs graphiques.

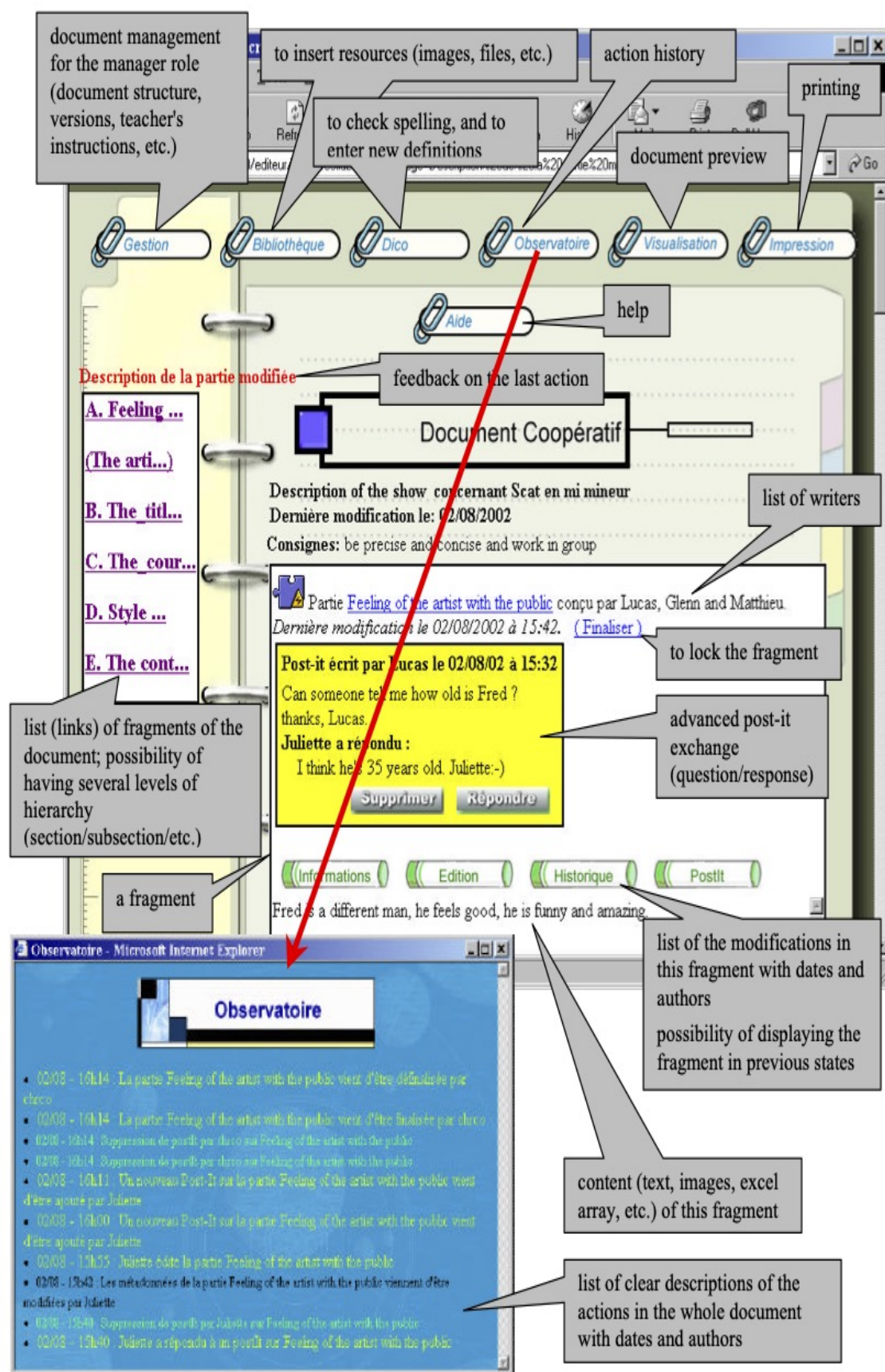


FIGURE A.3 – Éditeur ZWikiCoop

Journée scientifique

La Figure [A.4](#) présente le programme de la journée scientifique organisée par Christophe Courtin et financée par l'A.R.C.6 – « T.I.C. et usages informatiques innovants » (Région Auvergne Rhône-Alpes).

13H30 - 15H : exposés, démonstrations, discussions	• Jérémy Mary, INRIA, Université de Lille. ◦ Systèmes de recommandation et algorithmes de bandits Évaluation des approches statistiques sur les systèmes de recommandation à travers de difficiles challenges (Netflix, Adabo, Yahoo !). Les approches statistiques ont permis de résoudre le problème de recommandation de produits (films, livres, musique, etc.) avec des garanties théoriques. Une attention particulière est portée à la méthodologie d'évaluation de la performance et à la prise en compte des contraintes pour la recommandation. Quelques remarques de votre plus général issues de problématiques préoccupant les acteurs industriels du monde de la recommandation et sur les moyens techniques nécessaires. • Geoffray Bonnin et Anne Boyer, LORIA, université de Lorraine. ◦ De l'importance des facteurs humains dans la recommandation de musique : L'exemple du besoin en diversité Les systèmes de recommandation sont très utilisés dans l'industrie pour assister efficacement les utilisateurs dans leur recherche d'information sur Internet (e-commerce, intranets documentaires, services de musique en ligne, etc.). Ces systèmes reposent sur des techniques d'apprentissage automatique qui ont fait leurs preuves en terme de précision. Néanmoins, les systèmes actuels ne tiennent pas compte des facteurs humains (confiance, personnalité, facteurs culturels, humeur, etc.) ayant un impact sur le taux d'acceptation des recommandations et sur la prise de décision. De récentes études ont montré que l'apport en diversité – fourni d'autres facettes de la personnalité – est un facteur déterminant pour améliorer la satisfaction des utilisateurs. Nous présentons le cas de la génération automatique de listes de lecture musicales, ou playlists, qui peut être considérée comme une forme particulière de recommandation. Dans ce cadre, nous avons comparé la diversité de playlists créées et partagées sur les réseaux sociaux, de plusieurs plateformes musicales à celles des playlists qui génèrent les services de streaming en ligne et les algorithmes ayant une grande précision, et avons pu observer de fortes divergences. Cette observation suggère que l'utilisation exclusive des mesures habituelles, fondées sur la précision, n'est peut-être pas suffisante. Par ailleurs, il a été démontré que l'observation de l'évolution de la diversité dans les séquences de consultation des utilisateurs permet de caractériser automatiquement le contexte de recherche et d'ajuster les recommandations en conséquence. Ces recherches ont été appliquées à la recommandation de contenu vidéo (YouTube) et de contenu audio (Spotify). Les recommandations sont basées sur des réseaux sociaux, les intranets documentaires, la e-éducation, le marketing, la publicité ciblée, la recommandation de musique, de VOD ou d'événements culturels, etc.
15H - 15H15 : pause café	
15H15 - 16H45 : exposés, démonstrations, discussions	• Eka Ngere, LAMSADE, université Paris Dauphine. ◦ Systèmes de recommandations pour les villes "intelligentes" Le concept de "ville intelligente" (smart city) n'a pas encore été clairement défini. Cependant, il existe 6 caractéristiques/catégories pour classer ce type de villes et les compter : économie intelligente (smart economy), mobilité intelligente (smart mobility), environnement intelligent (smart environment), personnes intelligentes (smart people), vie intelligente (smart living), gouvernance intelligente (smart governance). En outre, l'"intelligence" est un challenge de plus en plus important pour de nombreuses villes et communautés. C'est d'un intérêt particulier dans le domaine des Technologies de l'Information et de la Communication (TIC) et pour de tels systèmes ou à l'usage des objets économiques, sociaux, ... Nous présentons comment mettre en place un système de recommandation pour les villes (et les personnes) pour améliorer la qualité de vie des citoyens et des entreprises. Nous présentons des décisions et des recommandations de différents contextes : des individus et des entreprises, des entreprises et des individus. Notre objectif est d'aider l'expérience des villes "intelligentes" pour aider des villes qui veulent améliorer leur intelligence en mettant en place les actions appropriées. • Maria Malek, LARIS, EISTI, Cogit. ◦ Analyse des réseaux complexes pour la recommandation : approche par combinaison des informations topologiques et nodales Nous décrivons dans cet exposé une approche de recommandation qui explore et combine trois types d'information à partir d'un réseau complexe. Le premier type est celui des liens de voisinage au sein du réseau, le second est celui des propriétés des nœuds du réseau (comme la centralité, la proximité, etc.) et le troisième est celui des propriétés des communautés. Le dernier type est celui des propriétés des communautés elles-mêmes, comme la structure interne du réseau. Nous utilisons par exemple la technique de l'arbre couvrant ainsi que des algorithmes de parcours en largeur ou en profondeur associés à des heuristiques. Le troisième type d'information concerne les propriétés du réseau et consiste à utiliser les mesures locales comme la centralité, l'intermédiation et le prestige pour guider les algorithmes de navigation dans le graphe. Plusieurs algorithmes basés sur cette approche sont proposés et ont été appliqués dans le domaine de e-commerce, ainsi que la recommandation d'experts dans un réseau bibliographique. Nous abordons à la fin du problème des grands graphes pour lesquels les algorithmes classiques ne sont plus applicables. Nous présentons des applications en préopératoire pour le calcul de certains indicateurs sociaux qui constituent des éléments importants dans les systèmes de recommandation.
16H45 - 17H : conclusion de la journée	

9H : café d'accueil	   
9H30 - 10H : introduction de la journée	Laurent Tabourat, directeur du laboratoire SYMME et chargé de mission S.I. à l'USMB Christophe Courtin, Christine Ferraris, Christian Martel, Miguel Tomaseva, Laurence Viennot, SYMME, université de Savoie Mont Blanc 10H - 11H30 : exposés, démonstrations, discussions • Kavé Salamandjan, LISTIC, université de Savoie Mont Blanc. ◦ Densification de réseaux sociaux par système de recommandation Les réseaux de partage de vidéos dépendent fortement de la contribution de la communauté de leurs utilisateurs et de leurs interactions sociales. Néanmoins, la mesure de ces réseaux montre que la connectivité sociale de ces réseaux est très faible comparée à ce qui est observé dans les réseaux sociaux populaires. Nous présentons une méthode de recommandation de vidéos pour densifier les réseaux sociaux. Nous utilisons, deux métriques de similarités entre utilisateurs de ces réseaux, qui déterminent la proximité entre deux personnes en regardant leurs fichiers tagués et les tags et les favoris qu'ils ont partagés sur leurs profils. Après une validation de ces deux métriques, nous utilisons ces mesures de proximité afin de recommander des relations sociales entre utilisateurs afin de densifier le réseau social. L'application au réseau de Youku montre que le système de recommandation améliore grandement l'accès des utilisateurs à des vidéos qui peuvent les intéresser. • Julien Subercarpe, LISTIC - Equipe Satin, Télécom Saint-Etienne, Université Jean Monnet, Saint-Etienne. ◦ Méthodes de hachage pour la recommandation sur Twitter Les larges volumes de données des réseaux sociaux posent de nombreux problèmes pour le passage à l'échelle des systèmes de recommandation. Nous présentons nos travaux sur des techniques de hachage permettant de représenter de manière compacte le profil d'un utilisateur (Twitter notamment) et de recommander efficacement des utilisateurs similaires. Dans une première partie nous montrons que l'utilisation de méthodes classiques de hachage permet de maintenir une précision correcte par rapport aux méthodes classiques. Dans une seconde partie nous présentons nos travaux sur de nouvelles méthodes de hachage similaires à la détermination qui permettent de dépasser le problème de l'égalité des mots-clés.
12H30 - 13H30 : repas sur place (offert par l'A.R.C.6), affichage posters	

FIGURE A.4 – programme scientifique



Annexe 2 : trace d'interaction

Description des règles d'analyse des traces d'interaction avec une DTD (Document Type Definition) dans l'outil d'édition de la station d'observation.

Description des règles

La Figure B.1 présente un exemple de règle basée sur une DTD (*Document Type Definition*) présentée sur la Figure B.2, qui spécifie l'indicateur « 5 entrées ont été créées par un contributeur en 1 mois » (entrée = définition dans l'expérimentation décrite dans la partie 2.5 page 51).

```
<rules>
  <rule>
    <description> the first rule </description>
    <conditions>
      <condition>
        <action> contribution </action>
        <numbersOfResults> 5 </numbersOfResults>
        <parameters>
          <parameter>
            <id> month </id>
            <value> 1 </value>
          </parameter>
        </parameters>
      </condition>
    </conditions>
  </rule>
</rules>
```

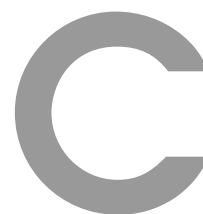
FIGURE B.1 – *Exemple de règle*

```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" version="1.0">
  <xs:element name="rules" type="listOfRules"/>
  <xs:complexType name="listOfRules"/>
    <xs:sequence>
      <xs:element minOccurs="0" name="rule" type="rule"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="rule">
    <xs:sequence>
      <xs:element minOccurs="0" name="description" type="xs:string"/>
      <xs:element minOccurs="0" name="conditions" type="listOfConditions"/>
      <xs:element minOccurs="0" name="retroActions" type="listOfRetroactions"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="listOfConditions">
    <xs:sequence>
      <xs:element minOccurs="0" name="condition" type="condition"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="condition">
    <xs:sequence>
      <xs:element maxOccurs="1" minOccurs="1" name="action" type="action"/>
      <xs:element maxOccurs="1" minOccurs="1" name="numberOfResults" type="xs:int"/>
      <xs:element maxOccurs="1" minOccurs="1" name="numberOfResultsSemantic" type="xs:string"/>
      <xs:element maxOccurs="1" minOccurs="1" name="subCondition" type="listOfConditions"/>
      <xs:element maxOccurs="1" minOccurs="1" name="parameters" type="listOfParameters"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="action">
    <xs:sequence>
      ...
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="listOfParameters">
    <xs:sequence>
      ...
    </xs:sequence>
  </xs:complexType>
  ...

```

FIGURE B.2 – *Description des règles*



Annexe 3 : gestion des données personnelles

Point juridique / Conception éthique

Point juridique

L'utilisation des traces d'interaction est un sujet très sensible en tant qu'il peut être très intrusif sans un cadrage strict et transparent. L'exploitation des traces est donc encadrée au niveau national par la loi « Informatique et Libertés » n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, et son acte d'application, le décret n°2019-536 du 29 mai 2019. Depuis 2018, le règlement général sur la protection des données (R.G.P.D. ou *G.D.P.R.* en anglais) fixe de nouvelles contraintes au niveau européen aux organisations concernant le traitement des données à caractère personnel. Le R.G.P.D. vient remplacer la directive sur la protection des données personnelles en Europe, et permet de responsabiliser les personnes qui font des traitements de ces données. Selon le R.G.P.D., la limitation du profilage des consommateurs élimine la collecte de données sans consentement, de même qu'il renforce le droit à l'oubli de l'individu. Pour les projets de type R.I.P.H. (Recherche Impliquant la Personne Humaine) du domaine biomédical ou médical majoritairement, un comité de protection des personnes (C.P.P.) est nécessaire au regard notamment de la protection des personnes se prêtant à la recherche.

Ces lois tentent d'offrir un cadre renforcé et harmonisé de la protection des données, tenant compte des récentes évolutions technologiques et de l'augmentation conséquente des volumes de données (*Big Data*, objets communicants, Intelligence Artificielle, etc.).

L'individu est placé au cœur du dispositif R.G.P.D. et voit ainsi ses droits sécurisés :

- obligations d'information de la part des organisations ;
- restrictions en termes de recueil de consentement ;
- droit à la portabilité des données ;
- droit à l'effacement.

Malgré les nombreuses méthodes visant à préserver la confidentialité et la légalité dans les systèmes de recommandation (anonymisation, randomisation, protocoles cryptographiques avec ou sans serveur), aucune confidentialité totale ne peut être offerte aux personnes. La disponibilité des informations personnelles est limitée par les paramètres de confidentialité des utilisatrices et utilisateurs, les directives de protection des données (R.G.P.D.) et les lois sur la protection de la vie privée ([Büttner 2018](#)).

La protection des données est aussi un aspect à considérer dans le cadre de nos travaux de recherche, où sont exploitées des informations issues du *Web* ou de plates-formes collaboratives. Au-delà de la réglementation, un positionnement syndical doit être considéré par rapport à la problématique de gestion des données personnelles dans le cadre de la recommandation de personnes. Les effets induits par des recommandations sur la base de l'utilisation des systèmes numériques peuvent être discriminatoires, valorisant certains profils aux dépens d'autres profils tout aussi riches de compétences, mais dans un registre moins visible.

Conception éthique

Bien que nos travaux s'apparentent à des recherches impliquant la personne humaine, le domaine des ressources humaines n'implique pas la validation des traitements de données par un comité de protection des personnes (C.P.P.), créé pour les domaines biomédical et médical. Les caractéristiques relatives à la protection des personnes dans nos expérimentations ont donc été communiquées au comité d'éthique pour la recherche (C.E.R.), qui est l'équivalent du C.P.P. pour les autres domaines (cf. Figure C.1). Dans nos études, aucune activité invasive ne vient perturber l'activité « métier », laquelle aurait pu influencer sur les résultats obtenus. Les données personnelles des participantes et participants sont utilisées pour caractériser la cohorte étudiée, mais les personnes restent pseudonymisées¹ pour les publications scientifiques. En d'autres termes, l'identité et les coordonnées des personnes ne sont pas divulguées, mais uniquement analysées et prises en compte dans le traitement des données et leur interprétation. Il est possible que les données soient demandées aux auteurs plusieurs années après la publication des résultats. Les données sont donc conservées pendant quinze ans, de manière à répondre aux mêmes exigences que celles fixées par un comité de protection des personnes. Au-delà de cette période, les données sont gardées sous forme pseudonymisée et un code est attribué à chaque individu pour protéger son identification (i.e. suppression du fichier d'indexation).

Comme les systèmes de recommandation reposent sur de nombreuses informations personnelles, les risques pour la vie privée sont inversement proportionnels à la précision des recommandations. Ces systèmes doivent trouver un compromis acceptable entre la confidentialité et la précision pour être efficaces. L'objectif du système de recommandation de profils de compétences dans un réseau social est d'aider la ou le commanditaire dans sa tâche, en augmentant sa conscience du groupe (*group awareness*). Les effets de ce système sont à la fois une augmentation des chances de sélection des membres et une intrusion dans leur vie privée. Par contre, les systèmes de recommandation peuvent être injustes, voire discriminatoires, pour les personnes qui n'utilisent pas très souvent les environnements numériques. Dans nos expériences, nous nous assurons que tous les membres utilisent les mêmes outils de la même manière, et que leurs informations personnelles ne soient pas communiquées aux autres membres à leur insu. Le domaine technologique évoluant généralement plus vite que le domaine juridique, certains services de recommandation sont plus ou moins acceptés par les utilisatrices ou utilisateurs, en raison de l'affichage de leurs informations personnelles. Dans un contexte traditionnel, les niveaux de confiance des membres envers les autres ne sont pas des informations communiquées par exemple. Dans nos expériences impliquant la confiance inter-personnelle, cette dernière est représentée par un lien non orienté entre deux nœuds dans la représentation graphique de l'outil d'analyse. Pour les besoins de notre étude, les chercheurs conservent le prénom avec un numéro pour les homonymes afin de pouvoir comparer le graphique avec la réalité. Dans la version du système basé sur la confiance, seules des étiquettes pseudonymes sont affichées, comme « étudiant1 », « étudiant2 », etc. et les personnes ne voient pas les niveaux de confiance choisis par les autres. Une version du système pour les membres du réseau eux-mêmes n'existe pas encore en raison des restrictions présentées ci-dessus.

Dans une future version, les participants devront se connecter au système de recommandation pour obtenir un graphe égocentré, où les autres nœuds seront étiquetés avec des pseudonymes. L'obtention de l'identité d'une autre personne nécessitera de la contacter avec un e-mail « aveugle », en sélectionnant le pseudonyme correspondant pour obtenir son autorisation. Si un tel système peut être utile pour aider une personne à élargir son groupe de connaissances, son utilisation pour embaucher des personnes doit être à la fois légale et éthique. Outre la conformité

1. nécessité d'avoir recours à des informations supplémentaires pour retrouver leur identité

juridique et éthique, le système de recommandation devra être accepté par les personnes elles-mêmes, et par les syndicats avant d'être utilisé dans une organisation.

CER_2020_9



Chambéry, le 18 novembre 2020

Le CER de l'Université de Savoie a examiné l'étude intitulée :

Capture des Traces opérationnelles des acteurs d'entreprise pour construire le Capital humain et définir les Processus gagnants.

Placée sous la responsabilité de :

M. Christophe Courtin, chercheur au laboratoire "Systèmes et Matériaux pour la Mécatronique" (SYMME) de l'université Savoie Mont Blanc.

cette étude a motivée un avis favorable de la part du CER.

Mention spéciale : l'avis rendu ne vaut pas caution pour les usages inappropriés qui pourraient, à l'insu des chercheurs, être fait des résultats obtenus dans le cadre de la recherche susmentionnée.

Le Président

Pascal BOUCHEZ



FIGURE C.1 – avis du comité d'éthique pour la recherche

"Des systèmes complexes pour l'observation de l'activité humaine"

Résumé

Nos recherches en science informatique s'inscrivent dans deux champs principaux que sont les Environnements Informatiques pour l'Apprentissage Humain (E.I.A.H.) et l'industrie du futur. Ce document relate les travaux entrepris pour concevoir et réaliser des systèmes complexes pour l'observation de l'activité humaine. L'approche suivie est celle du traitement des traces d'interaction numériques témoins de l'activité observée. La démarche d'investigation retenue répond à une logique abductive, considérant que les modèles complexes doivent refléter des phénomènes réels avant d'être généralisés. Les nombreuses problématiques abordées s'inscrivent dans une chaîne cinématique conceptuelle et implique des contributions pluridisciplinaires. Nos contributions majeures concernent les systèmes à base de traces d'un point de vue conceptuel, les systèmes de partage de corpus de traces et d'outils d'analyse et l'identification des compétences mobilisées. Les perspectives de ce travail impliquent l'élargissement des champs d'investigation et la réalisation concrète de systèmes complexes basés sur l'activité et répondant aux enjeux sociétaux actuels. Notre stratégie consiste à fédérer des ressources humaines issues des réseaux de la recherche dans de futurs projets nationaux et internationaux.

Mots-clés : observation, trace d'interaction, corpus, compétence, recommandation, régulation, conscience de groupe, confiance, cartographie, apprentissage, industrie 5.0

Abstract

Our research in computer science is focused on two main fields : Technology-Enhanced Learning and Teaching (T.E.L.T.) and the industry of the future. This document deals with the undertaken work to design and realize complex systems for the observation of human activity. The followed approach is that of processing the digital interaction traces that are witnesses of the observed activity. The adopted investigation method is based on an abductive logic, considering that complex models must reflect real phenomena before being generalized. The numerous issues addressed are part of a conceptual kinematic chain and involve multidisciplinary contributions. Our major contributions concern trace-based systems from a conceptual point of view, systems for sharing the corpus of traces and analysis tools and the identification of mobilized competencies. The perspectives of this work imply the widening of the fields of investigation and the concrete realization of complex systems based on the activity and answering the current societal stakes. Our strategy is to bring together human resources from research networks in future national and international projects.

Keywords : observation, interaction trace, corpus, competence, recommendation, regulation, group awareness, trust, mapping, learning, industry 5.0

