



HAL
open science

Conception de système à base de traces numériques pour les environnements informatiques documentaires

Julien Laflaquière

► **To cite this version:**

Julien Laflaquière. Conception de système à base de traces numériques pour les environnements informatiques documentaires. Interface homme-machine [cs.HC]. Université de Technologie de Troyes, 2009. Français. NNT: . tel-00471975

HAL Id: tel-00471975

<https://theses.hal.science/tel-00471975>

Submitted on 9 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE L'UNIVERSITE DE TECHNOLOGIE DE TROYES

Discipline
INFORMATIQUE

Spécialité
RESEAUX CONNAISSANCES ORGANISATION

Présentée et soutenue par
JULIEN LAFLAQUIERE

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE DE TECHNOLOGIE DE TROYES

CONCEPTION DE SYSTEME A BASE DE TRACES NUMERIQUES DANS LES ENVIRONNEMENTS INFORMATIQUES DOCUMENTAIRES

Soutenu le
8 DECEMBRE 2009

Devant le Jury composé de

BRUNO BACHIMONT	ENSEIGNANT CHERCHEUR - HDR	RAPPORTEUR
SYLVIE LELEU-MERVIEL	PROFESSEUR DES UNIVERSITES	RAPPORTEUR
KHALDOUN ZREIK	PROFESSEUR DES UNIVERSITES	PRESIDENT
NATHALIE AUSSENAC-GILLES	CHERCHEUR CR1 CNRS - HDR	EXAMINATEUR
ALAIN MILLE	PROFESSEUR DES UNIVERSITES	EXAMINATEUR (INVITE)
MANUEL ZACKLAD	PROFESSEUR DES UNIVERSITES	CO-DIRECTEUR
YANNICK PRIE	MAITRE DE CONFERENCE	CO-DIRECTEUR

École Doctorale SSTO - Université de Technologie de Troyes
Avec un financement de la Région Champagne-Ardenne

Remerciements

À Émilie

Je remercie en premier lieu mes directeurs de thèse, Manuel Zacklad qui m'a accueilli dans son équipe et qui m'a laissé une grande liberté pour mener mon travail, et Yannick Prié qui m'a soutenu depuis mon DEA et à travers les (nombreuses) péripéties qui ont parsemé mon parcours de thésard. Je remercie sincèrement tous les membres de mon jury de thèse, qui ont accepté, malgré les difficultés matérielles, de se pencher sur mon travail. Je remercie la région pour le financement sans lequel cette thèse n'aurait pas été possible ainsi que l'École Doctorale qui a suivi mon travail et répondu à mes questions et mes remarques.

Je remercie chaleureusement l'équipe Tech-cico dans laquelle j'ai été intégré durant les trois années passés dans les murs de l'UTT, et tout spécialement Nada Matta pour son soutien et ses bons conseils. Je remercie mille fois toute l'équipe du LIRIS, en particulier Alain Mille avec qui les discussions ont été passionnées et passionnantes. Un merci particulier à Brigitte qui a toujours su me tirer des mauvais pas (à distance en plus), ainsi qu'à Christelle, Laure et Stéphanie mais également Paula pour leur gentillesse et leur prévenance. Un grand merci aussi à tous mes amis (ex)doctorants Yann, Caroline, Olivier, Manuel et Goritsa, qui ont réussi à me supporter pendant si longtemps au bureau...

Je remercie tous ceux qui ont accepté d'être mis à contribution durant mes travaux sur le terrain, notamment Christine, Alain et Nadia pour eLcyée et je remercie chaleureusement René, Marc et Pascal, qui m'ont consacré leur temps malgré un emploi du temps chargé. Je remercie tout particulièrement Romain, Yann et Emmanuel avec qui j'ai adoré travailler sur le terrain.

Je remercie tous ceux qui m'ont aidé à arriver jusqu'ici, avant et pendant la thèse : Yves et Odile, Dana, Sylvie et aussi Nicolas. Un grand merci à Alexandra pour le travail que nous avons pu faire ensemble et pour ses conseils et questions toujours les bienvenus. Merci à tous mes proches, famille et amis, qui ont accepté de me voir moins souvent ces derniers mois et qui m'ont toujours soutenu. Et enfin un immense merci à Émilie pour sa patience et sa compréhension, son soutien et son amour. Cet aboutissement est le nôtre.

Sommaire

Sommaire	1
Introduction	5
De la piste à la trace réutilisée	5
Organisation du manuscrit	6
Chapitre 1 Contexte et problématique	9
1.1 La notion de trace d'activité numérique	10
1.1.1 Caractéristiques de la notion de trace.....	10
1.1.2 La trace comme indice, la trace comme signe	11
1.1.3 La trace enregistrée	13
1.1.4 La trace numérique, une trace comme les autres ?.....	13
1.1.5 Bilan	15
1.2 Les activités numériques tracées	16
1.2.1 Activité Documentaire Instrumentée	16
1.2.2 La Recherche d'Information	16
1.2.3 L'apprentissage (collaboratif) à distance	18
1.2.4 Création de contenus de formation	19
1.3 Problématique	21
Chapitre 2 État de l'art des systèmes traçants	23
2.1 Introduction	24
2.1.1 Systèmes traçants et traces numériques d'interactions	24
2.1.2 Articulation du chapitre.....	25
2.2 Les Traces Numériques d'Interaction comme support d'analyse	26
2.2.2 Analyse des environnements numériques	28
2.2.3 Analyse des comportements d'utilisation et d'usage	32
2.2.4 Analyse des utilisateurs et de leurs activités	36
2.2.5 Bilan	44
2.3 Les Traces Numériques d'Interaction comme support à l'activité	44
2.3.1 Adapter automatiquement l'environnement numérique.....	45
2.3.2 Enrichir l'interaction	48
2.3.3 Assister l'utilisateur dans son activité.....	56
2.4 Conclusion	63
2.4.1 Les trois composantes d'un système traçant	63
2.4.2 Des traces de l'interaction à celles de l'activité	68
Chapitre 3 Les Systèmes à Base de Traces Modélisées	71
3.1 Introduction	72
3.2 L'approche MUSETTE	72
3.2.1 Les principaux concepts de l'approche <i>Musette</i>	73
3.2.2 Applications de l'approche MUSETTE	75
3.3 L'approche des Systèmes à Base de Traces Modélisées	76
3.3.1 Définitions formelle du concept de trace modélisée	77

3.3.2	Les Systèmes à Base de Traces modélisées	79
3.3.3	Implémentations	82
3.4	Conclusion.....	83
3.4.1	Un « méta-modèle »	84
3.4.2	Retour à la problématique	85
Chapitre 4	Cadre d'analyse et méthodes de modélisation des activités instrumentées	87
4.1	Introduction.....	88
4.2	Apports du domaine des IHM.....	88
4.2.1	Concepts clefs de la modélisation de tâche	89
4.2.2	Processus de modélisation des tâches	90
4.2.3	Formalismes	92
4.2.4	Bilan	94
4.3	Apports de l'Ingénierie des Connaissances	95
4.3.1	L'Ingénierie des Connaissances comme domaine de recherche.....	95
4.3.2	Concepts clefs de l'Ingénierie des Connaissances	96
4.3.3	Méthodes de modélisation des connaissances.....	100
4.3.4	Bilan	104
4.4	Apports du CSCW.....	105
4.4.1	Présentation du CSCW	105
4.4.2	Approche centrale : la Théorie de l'Activité.....	108
4.4.3	Approches complémentaires	112
4.5	Conclusion.....	124
Chapitre 5	Vers une méthodologie de modélisation des traces numériques	125
5.1	Introduction.....	126
5.2	Cadre théorique.....	126
5.2.1	L'activité caractérisée par ses dimensions	127
5.2.2	L'activité effective caractérisée par sa structure singulière	129
5.2.3	Les traces modélisées comme inscription de connaissance	130
5.3	Processus de modélisation générique	132
5.3.1	Phase d'analyse descriptive de l'activité.....	133
5.3.2	Phase de création du modèle de trace	137
5.3.3	Commentaires	140
5.4	Spécialisation du processus de modélisation	142
5.4.1	Traces modélisées pour l'analyse de l'activité.....	142
5.4.2	Traces modélisées pour la réflexivité de l'activité.....	144
5.5	Dynamique de mise en place d'un SBTm	146
5.6	Conclusion.....	147
Chapitre 6	Recherche action pour une méthodologie de modélisation des traces numériques	149
6.1	Introduction.....	150
6.2	Traces modélisées d'une activité pédagogique dans un EIAH	150
6.2.1	Contexte de modélisation.....	151
6.2.2	Travail d'analyse et de modélisation.....	153
6.2.3	Création d'un modèle de trace et visualisation	156
6.2.4	Bilan.....	159
6.3	Traces modélisées d'une activité de création de contenus.....	161
6.3.1	Contexte de modélisation.....	161
6.3.2	Phase d'analyse descriptive de l'activité.....	164

6.3.3	Phase de création du modèle de trace	173
6.3.4	Expérimentation : première Partie	175
6.3.5	Création de la trace modélisée	177
6.3.6	Expérimentation : seconde partie	180
6.4	Conclusion.....	184
Chapitre 7	Traces d'activité conjointes	187
7.1	Introduction.....	188
7.2	Systèmes à Base de Traces modélisées conjointes.....	189
7.2.1	Les activités conjointes	189
7.2.2	Les traces modélisées d'activités conjointes.....	191
7.3	Les potentiels usages conjoints des traces conjointes	195
7.3.1	Traces modélisées comme support d'analyse de l'activité conjointe	195
7.3.2	Traces modélisées comme support de l'activité conjointe.....	196
7.3.3	Finalités en situation	198
7.4	Conclusion.....	200
Chapitre 8	Conclusion.....	201
8.1	Bilan et apports de la thèse.....	201
8.1.1	Apports théoriques et bibliographiques	201
8.1.2	Apport méthodologique	202
8.1.3	Apport expérimental et instrumental.....	204
8.1.4	Apport conceptuel	204
8.2	Limites et commentaires.....	205
8.2.1	Critique du travail réalisé.....	205
8.2.2	Commentaires sur l'approche proposée.....	206
8.3	Perspectives.....	208
8.3.1	Prolongement du travail de terrain.....	208
8.3.2	Développement de la méthodologie de modélisation des traces.....	209
8.3.3	Nouvelles applications et nouveaux usages des traces modélisées.....	211
	Références.....	213
	Table des illustrations	231
	Annexes.....	235

Introduction

*« Suppose all interaction history is recorded, structured and indexed for later use.
In such a world, every menu choice, every document edit would be available.
But to whom and for what purposes ? »*
(Hill et Hollan, 1992, p.7)

De la piste à la trace réutilisée

Les « traces numériques » font de plus en plus souvent leur apparition dans l'actualité. Lorsqu'elles sont évoquées il s'agit souvent de désigner des moyens de « traçabilité » ou encore de « pistage » d'usager. Il s'agit parfois d'un sujet brûlant avec des problèmes touchant aux libertés fondamentales (cas notamment de la loi *HADOPI*¹ et bientôt le cas avec la *LOPPSI-2*²). Les traces laissées par certaines activités humaines permettraient ainsi d'incriminer certains utilisateurs peu scrupuleux, notamment internautes. Mais la tendance à conserver des « traces numériques » de l'activité des individus ne s'arrête pas aux seuls internautes. La société piste de plus en plus ses membres, consommateurs, simples usagers ou pirates d'autoroute numérique.

C'est ainsi notamment que les « *pass* » pour les abonnements de transport en commun (locaux comme nationaux) ou d'autoroutes, tout comme les multiples badges d'accès à des lieux « protégés » (publics ou privés), permettent l'enregistrement informatique de données relatives à la vie de tout un chacun, données qui parfois ont une durée de conservation illimitée. Ce type de procédé ne risque pas de disparaître, étant donné les derniers « progrès » industriels en la matière : notamment les tags *RFID*³ présentés comme une solution miracle pour lutter contre le vol ou bien optimiser la gestion de la production industrielle ou de ses stocks, mais surtout pour adapter un environnement au consommateur. Il s'agit là semble-t-il d'un marché porteur, puisque les industriels prévoient déjà de nombreuses applications « innovantes », comme par exemple (ce n'est qu'un exemple) le projet « *Sorting Door* »⁴ qui vise à créer un réseau qui analyserait tous les tags *RFID* actifs aux sorties des magasins, de manière à décorifier et recenser automatiquement les comportements des clients, du simple vol à l'étalage au profilage et à la personnalisation du « panier de la ménagère ». La liste des dispositifs et applications qui permettent (et permettront bientôt) de tracer nos faits et gestes de la vie quotidienne pourrait être sérieusement allongée. Les capacités techniques des machines informatiques autorisent aujourd'hui la conservation d'un volume de données tel que seule l'éthique semble devoir limiter cette tendance. Peut-être devons-nous vivre dans une société qui n'oublie rien. Selon Bruno Latour ces dispositifs

¹ Loi de création de la haute autorité de la protection des œuvres et de la propriété intellectuelle, faisant suite à la loi Davisi sur le téléchargement et la diffusion illégale d'œuvres sur Internet, voir : http://fr.wikipedia.org/wiki/Loi_Hadopi (consulté le 12/06/09).

² Loi d'orientation et de programmation pour la performance de la sécurité intérieure (LOPPSI 2) intégrée au projet de loi de finance 2009. Voir : <http://www.senat.fr/rap/108-099-327/108-099-3275.html> (consulté le 12/06/09).

³ *Radio Frequency Identification* : méthode utilisée pour stocker et récupérer des données à distance en utilisant de mini-balises métalliques.

⁴ Voir <http://www.sortingdoor.com/> (Consulté le 18/06/09).

atténuent la frontière entre vie privée et vie publique, bouleversant de fait ce qui fait l'objet d'étude des Sciences Humaines et Sociales : « *It is as if the inner workings of private worlds have been pried open because their inputs and outputs have become thoroughly traceable* » (Latour, 2007).

D'un autre côté, dans de multiples activités médiées par les outils informatiques c'est bien l'oubli justement qui pose problème. L'entreprise cherche par tous les moyens à capitaliser ses connaissances, à ne pas laisser se perdre un savoir-faire, à faire fructifier les quantités d'informations qu'elle sauvegarde par ailleurs sans les utiliser vraiment. Même pour le simple utilisateur d'un navigateur *Web* des questions simples se posent : « Comment ai-je fait pour trouver... ? Où ai-je trouvé cette information... ? Quand ai-je rempli ce formulaire ? ». Les machines sont capables d'enregistrer une multitude d'informations concernant nos faits et gestes, mais l'informatique ne semble pas avoir trouvé le moyen de tirer pleinement profit de cette capacité pour la mettre au service de nos activités pour en être une trace si ce n'est une mémoire. Les chercheurs de différents courants au sein de l'Informatique et au-delà, se sont penchés sur l'enregistrement des interactions utilisateurs et sur les possibilités techniques de les repérer, les enregistrer, les conserver. Des fonctionnalités d'historique ont fait leur apparition dans la quasi-totalité des applications informatiques courantes, et les navigateurs *Web* en particulier enregistrent durant la navigation un grand nombre d'informations qui seront réutilisées par le navigateur lui-même par la suite : formulaires, mots de passe, préférences, *cookies*.

Dans les activités documentaires aujourd'hui fortement médiées par l'informatique, en particulier professionnelles, comment obtenir des *traces* d'activités numériques *réutilisables* et *réutilisées*, et quel serait l'intérêt, le but de cette *réutilisation* ? Voilà, de manière simple, la question à laquelle cette thèse souhaite contribuer à répondre.

Organisation du manuscrit

Conservant l'Informatique comme discipline centrale, ce travail de thèse mobilise et concerne plusieurs domaines de recherche. Sur le fond, ce travail associe en effet les domaines de l'Ingénierie des Connaissances (IC), du *Computer Supported Cooperative Work (CSCW)* et des Interfaces Homme Machine (IHM). Le caractère pluridisciplinaire de ce travail est renforcé par le fait qu'il implique des domaines d'applications également variés tels que les Environnements Informatiques d'Apprentissage Humain (EIAH) ou les environnements de production de contenus, et mobilise des principes issus d'autres disciplines notamment les Sciences Cognitives et l'Ergonomie. La pluridisciplinarité est une force à condition d'être maîtrisée. C'est ce que nous avons tenté de faire en associant au cours de cette thèse un travail de recherche théorique à un travail de recherche appliquée. L'ensemble du manuscrit est divisé en huit chapitres.

Chapitre 1. Le premier chapitre sera consacré à la description du contexte de cette thèse et à la mise en place d'une problématique générale. Après la présentation de la notion de « trace numérique » que nous souhaitons adopter et la description des activités pour lesquelles nous souhaitons obtenir des traces, nous présenterons les questions auxquelles cette thèse s'attachera à répondre : comment obtenir d'une activité numérique donnée des traces réutilisables, et dans quel but.

Chapitre 2. Le second chapitre est consacré à un état de l'art des systèmes informatiques utilisant ou créant des traces numériques que nous désignons de manière générale comme des « systèmes tra-

çants ». De nombreux systèmes informatiques, dans de multiples domaines d'application, ont été conçus pour enregistrer tout ou partie des interactions d'un utilisateur avec son environnement numérique. L'état de l'art sur ces systèmes traçants a pour premier objectif de rendre compte de l'étendue des finalités réservées aux traces numériques sans être restreint à un domaine d'application particulier, ce qui est habituellement le cas des travaux bibliographiques sur ce type de systèmes. Son second objectif est de proposer une description synthétique des moyens par lesquels les dispositifs informatiques sont instrumentés pour créer concrètement des traces numériques, puis les réutiliser.

Chapitre 3. Le troisième chapitre sera consacré à la présentation de l'approche des Systèmes à Base de Traces. Cette approche, à laquelle nous souscrivons pleinement, propose une manière de considérer les traces numériques en en faisant un objet informatique à part entière, manipulable en tant que telle. Cette vision, qui constitue un cas particulier dans le paysage formé de l'ensemble des travaux sur les systèmes traçants, se concrétise dans le cadre conceptuel des Systèmes à Base de Traces modélisées. L'idée est qu'une trace numérique doit être en partie modélisée *a priori* pour dépasser les limites sémantiques habituellement caractéristiques des enregistrements automatiques d'interaction. Pour cela le cadre conceptuel fournit une formalisation des traces modélisées, se présentant comme le support à l'implémentation concrète de tout Système à Base de Trace modélisée. Si la solidité du cadre conceptuel semble acquise, il manque encore à l'approche une méthodologie de modélisation propre aux traces modélisées.

Chapitre 4. Le quatrième chapitre présentera un travail de recherche bibliographique visant à établir un cadre théorique pour la modélisation des traces numériques. Il s'agit de déterminer quels outils et quelles approches théoriques de l'activité empruntés à d'autres domaines de modélisation, peuvent être utilisés pour construire notre propre méthodologie de modélisation des traces numériques. En l'occurrence, trois domaines sont ainsi mobilisés : le domaine des Interfaces Homme Machine, celui de l'Ingénierie des Connaissances et enfin celui du *CSCW*. Ces domaines sont visés autant pour leur expertise méthodologique en termes de modélisation, que pour leurs apports théoriques sur l'activité humaine instrumentée dont nous avons besoin pour décrire et analyser les activités que nous avons à tracer.

Chapitre 5. Le cinquième chapitre sera consacré à la présentation de la méthodologie de modélisation des traces numériques que nous proposons. Sur la base des apports du chapitre précédent, nous proposons une méthodologie composée de quatre éléments : un cadre théorique d'analyse des activités instrumentées, un processus générique de modélisation, une spécialisation de ce processus relativement au contexte d'usage des traces et une démarche d'intégration de la modélisation à une démarche de conception et d'implantation d'un Système à Base de Traces modélisées en situation.

Chapitre 6. Le sixième chapitre sera consacré à la présentation du travail de recherche empirique réalisé sur le terrain pour la mise au point de la méthodologie de modélisation proposée. En effet cette dernière est le fruit d'un travail de recherche associant théorie et mise en pratique. Ainsi avons-nous directement mis à l'épreuve nos propositions sur deux terrains expérimentaux, le premier concernant une classe virtuelle et l'apprentissage collaboratif à distance, le second un atelier de production de contenu pédagogiques et la conception de formations. La présentation de ce travail de terrain montrera comment la mise en pratique a influencé la construction de la méthodologie proposée. On y trouvera également la description d'une expérimentation constituant un premier pas vers l'usage des traces modéli-

sées sur le terrain et une validation partielle de notre démarche. Ce travail de terrain aura également été l'occasion de formuler de nouvelles questions, et de mettre en perspective l'approche des Systèmes à Base de Traces modélisées dans le cadre d'activités collectives.

Chapitre 7. Le septième et dernier chapitre présente une discussion autour d'une extension de la problématique de départ en considérant les usages collectifs et coopératifs des traces modélisées. En effet, dans les situations de travail collectif, de multiples usages possibles des traces modélisées sont envisageables, remettant en perspective l'approche initiale. Dans ce contexte, l'approche des Systèmes à Base de Traces modélisées pourrait être étendue, afin de tenir compte de ces usages et offrir un cadre conceptuel et une formalisation adaptée. Nous proposons le concept de « trace conjointe » pour désigner cette extension, qui offre aux traces modélisées une nouvelle dimension en tant qu'objet de recherche et esquisse les pistes à suivre pour de futurs travaux, tant sur le plan conceptuel et technique que sur le plan méthodologique, et qui pourrait à terme donner dans la pratique une nouvelle signification aux traces numériques dans les environnements informatiques documentaires.

Chapitre 8. le huitième et dernier chapitre présente la conclusion de ce travail. Nous y présentons tout d'abord le bilan global de nos contributions, puis une discussion de leurs limites et enfin une liste de perspectives découlant de ce travail.

Chapitre 1

Contexte et problématique

Résumé du chapitre

L'expression « trace numérique » qui fait de plus souvent son apparition dans le champ de la recherche en Informatique et en déborde même parfois dans les titres de la presse grand public, désigne des réalités bien différentes. Notre thèse porte sur la question des « traces numériques » entendues comme les « traces *d'activités* numériques » ce qui est finalement une acception parmi d'autre du terme en question. C'est la raison pour laquelle ce premier chapitre est consacré en partie à donner une véritable définition à la notion de « trace numérique ». L'absence dans la littérature d'une telle définition laisse entendre qu'une trace numérique pourrait être un simple cas particulier de trace laissée par un acteur dans le sillage de son activité. Or, la notion d'une trace d'activité « numérique » diffère par nature de celle de trace d'activité « ordinaire », elle n'est pas une « marque laissée par ce qui agit sur quelque chose ». Contrairement à la notion de trace ordinaire, la trace numérique n'émerge pas de la confrontation d'une action et de l'environnement qui lui sert de support. La trace numérique *n'est pas donnée* par les propriétés d'un environnement déjà présent. Elle est *construite* de toute pièce et sa composition est nécessairement déterminée *a priori*, par la programmation des enregistrements qui seront conservés. Fondamentalement une trace d'activité numérique est le résultat d'un enregistrement systématique des interactions utilisateur qui constituent partiellement la réalisation d'une activité. À partir de là notre problématique de départ est de savoir comment *créer de telles traces* et comment les rendre *concrètement réutilisables*. Cette question se pose en particulier dans cette thèse dans le cadre d'une famille d'activités que nous appelons les *activités documentaires instrumentées* (ADI), et dont nous souhaitons pouvoir tracer la réalisation. Dans cette famille d'activité nous avons retenus trois exemples concrets. Le premier est celui de la Recherche d'Information que nous utilisons comme une activité type pour nous aider à conduire nos réflexions sur les traces, les deux autres sont liés aux terrains d'applications qui sont mobilisés dans cette thèse, à savoir une activité d'apprentissage collaboratif à distance dans une classe virtuelle et une activité de production de contenus de formations professionnelle dans l'industrie.

1.1 La notion de trace d'activité numérique

Les quelques exemples cités en introduction de ce manuscrit montrent que le terme de « trace numérique » est aujourd'hui utilisé pour désigner des objets très différents. A tel point, que n'apparaît pas clairement ce qui pourrait constituer une définition générale de la notion de « trace numérique ». La question se pose notamment pour nous qui souhaitons désigner spécifiquement par ce terme la « trace d'une activité numérique ». Curieusement, la question semble largement absente de la littérature, en tout cas pour ce qui concerne l'Informatique (au sens large). Nous profitons de l'occasion pour entamer ce manuscrit en répondant à une question simple : qu'est-ce qu'une *trace* (d'activité) *numérique* ?

1.1.1 Caractéristiques de la notion de trace

Chercher à définir la notion de trace numérique nous ramène à la définition même d'une « trace », notion à laquelle, selon A. Serres (2002), peu de penseurs consacreront leurs réflexions. Si le terme de « trace » en lui-même est largement employé dans le langage commun, la notion correspondante se révèle très complexe. Outre des sens spécifiques, dans des domaines techniques ou en mathématique (en géométrie la trace désigne le lieu d'intersection avec le plan de projection), le terme « trace » possède aujourd'hui plusieurs grandes acceptions courantes⁵. De façon générale, il fait inmanquablement penser aux traces de pas laissées dans la neige ou le sable, et se définit alors comme une « empreinte ou une suite d'empreintes sur le sol marquant le passage de quelqu'un ou quelque chose », au sens propre comme figuré, ou bien comme « une marque laissée par une action », un événement passé, la trace d'un coup par exemple. Elle est aussi parfois utilisée pour désigner « une quantité infime », dont les traces d'ADN recherchées sur la scène d'un crime sont un exemple.

1.1.1.1 Étymologie

Ces différentes acceptions sont issues de la riche étymologie du terme « trace » qui a vu sa polysémie s'étendre au fil des siècles. Le terme⁶ provient du verbe *tracer*, issu du latin *tractiare* lui-même dérivé de *tractus* signifiant l'action de tirer, le tracé ou encore la lenteur. Le terme trace connut diverses acceptions. Vers 1120, il a signifié *une empreinte ou une suite d'empreintes laissée par le passage d'un homme ou d'un animal, puis d'une chose* (d'une voiture à partir de 1690). Vers 1190 l'empreinte apparaît au sens figuré et la trace signifie la manière d'agir de quelqu'un et deviendra « l'exemple à suivre » vers 1530 d'où naîtra l'expression « marcher sur les traces de quelqu'un ». En 1250 la notion de trace s'étend à un nouveau sens, celui de l'égratignure, de la *marque laissée par ce qui agit sur quelque chose*. Le sens figuré fait son apparition vers 1538 avec l'idée d'une trace comme la marque laissée par un événement notamment en mémoire, « l'impression qu'il reste quelque chose » ou « ce qui reste du passé ». Une évolution du sens, plus récente, fait apparaître la trace, à partir de 1847, comme *une petite quantité*, résiduelle, subsistante.

Parallèlement, vers 1439 la trace devient graphie, et on fait alors une trace pour *raturer ou effacer*. Au 16^{ème} siècle on « tracerait quelque chose » pour écrire, puis au 17^{ème} siècle pour représenter au moyen de

⁵ Selon le Dictionnaire Petit Robert (2005) : « une trace est une chose ou une suite de choses laissées par une action quelconque et relatives à un être ou un objet ; une suite d'empreintes ou de marques que laisse le passage d'un être ou d'un objet ; ce à quoi on reconnaît que quelque chose a existé ; ce qui subsiste d'une chose passée ».

⁶ Les définitions sont issues du portail du Centre National de Ressources Textuelles et Lexicales, <http://www.cnrtl.fr/> (consulté le 14/03/09).

lignes, et enfin « marquer le contour », en broderie puis dans diverses spécialités. Toujours selon A. Serres (2002), l'histoire du terme fait émerger plusieurs grandes significations. Parmi elles, nous laisserons de côté (sans les ignorer) la trace comme ligne, écriture (liée à la problématique des traces écrites et non-écrites) ; la trace comme catégorie première de la connaissance historique, liée à la problématique du « document-trace » ; la trace comme « exemple à suivre » et sa dimension morale, ou encore la trace comme « empreinte psychique », mnésique, affective, cérébrale ou corticale. Puisque notre travail porte sur les traces d'activités humaines concrètes, nous retenons en revanche la trace prise dans un « paradigme indiciaire » (Ginzburg, 1989). Pour faire simple, nous essaierons de comprendre la trace en tant qu'*indice*. Indice matériel ou matérialisé, que peuvent laisser derrière elles des activités⁷, qu'il serait alors possible de – littéralement – « retracer » (Figure 1.1).

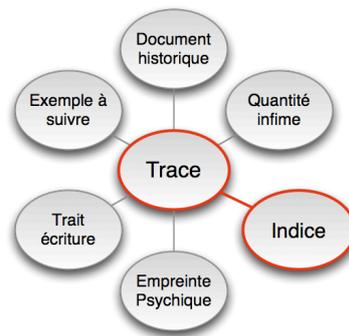


Figure 1.1 : Parmi les multiples significations de la trace nous retenons la trace comme *indice*.

1.1.2 La trace comme indice, la trace comme signe

En adoptant ce point de vue, que peut-on attribuer comme caractéristiques à une « trace » ? Tout d'abord, une trace *ne se définit pas par elle-même*, mais toujours par rapport au phénomène dont elle est la trace, une trace est toujours la trace *de* quelque chose, ce que souligne A. Serres : « la trace se caractérise par son génitif intrinsèque, si l'on peut dire, *i.e.* son caractère d'appartenance, au sens où la trace est toujours trace de quelque chose ; elle ne se définit pas par elle-même, elle n'a pas d'existence propre, autonome, au plan ontologique du moins, elle n'existe que par rapport à autre chose (un événement, un être, un phénomène quelconque) » (Serres, 2002, p.1).

En ce qui nous concerne ici, la trace n'a d'existence que relativement à l'*activité* dont elle est la trace. Si donc nous devons considérer les traces d'une activité donnée, elles seront liées à l'*environnement* de celle-ci. La trace d'activité apparaît comme une *modification* plus ou moins durable de cet *environnement*, engendrée par la *réalisation d'une activité*. La trace prend place dans la relation entre l'*environnement* (support) et l'*activité* (scripteur) qui s'y déroule. Comme l'indice, la trace est toujours susceptible de disparaître, ou simplement d'échapper à l'observateur. Une trace n'acquiert son statut de trace qu'*a posteriori* et c'est l'observateur lui-même, ou plutôt son observation qui fera d'une trace ce qu'elle est. Ainsi, elle ne prendra son sens que « *sous le regard qui la déchiffre* » (Serres, 2002, p.1) et tout produit, tout être ou objet, tout élément de l'environnement pourra être considéré à ce titre comme trace d'un phénomène ou d'une action potentiels.

⁷ Dans cette partie nous parlons bien de la notion de trace que nous qualifions de « matérielle », pour la distinguer d'une trace « numérique ».

1.1.2.1 Trace comme signe nécessaire ou non-nécessaire

La notion de trace partage donc certaines caractéristiques avec la notion d'indice. Leur rapprochement nous conduit cependant à opérer un autre rapprochement, avec la notion de « signe » cette fois. Les deux termes, « trace » et « signe », furent en effet utilisés indifféremment par le passé⁸ (Démonet, 1994). Le *signe* est ce qui est significatif de quelque chose au-delà de lui-même, ou pour Augustin « ce qui, se présentant en tant que tel à la perception sensible, présente aussi quelque chose à la perception intellectuelle » (Démonet, 1994).

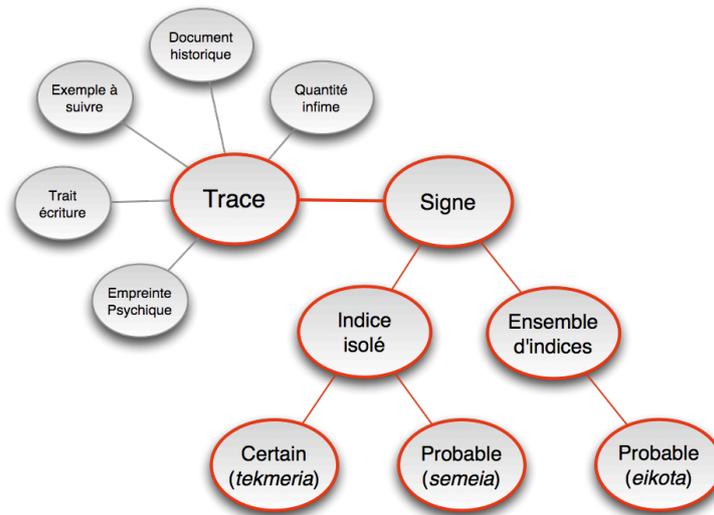


Figure 1.2 : La trace comme signe, isolée ou composée toujours interprétée.

Le signe a lui aussi cette particularité de dépendre totalement de son interprétation. Si l'on rapproche la notion de trace de celle de signe il faut distinguer deux types de signes : les *signes nécessaires* ou *indices certains* (que Platon dénommait *tekmeria*), et les *signes non-nécessaires*, ou *indices seulement probables*, voire *douteux* (les *semeia*). Une trace entendue comme indice relève donc du *signe nécessaire* (ne laisser aucun doute quant au phénomène dont elle est la trace) ou du *signe non-nécessaire* (et laisser une part de doute dans son interprétation). Le schéma ci-dessus présente de façon hiérarchique les différentes notions que nous venons d'évoquer (Figure 1.2).

1.1.2.2 Caractère non intentionnel de la trace

Autre caractéristique importante, la trace – comme le signe – désigne indifféremment un *indice isolé* ou un *ensemble d'indices*, et ce à n'importe quel niveau de précision, le phénomène pouvant se constater à toutes les échelles⁹. Quelle que soit la nature des indices (certaine ou seulement probable) considérés un par un, ils ne constitueront ensemble un *faisceau d'indices* que par l'interprétation qu'en donnera un observateur. Un peu à la manière dont un ensemble d'indices de l'enquête criminelle peuvent constituer la preuve de la culpabilité ou de l'innocence d'un accusé, le faisceau d'indices (*eikota*), est donc également un signe interprété par un *observateur sujet à l'erreur*.

⁸ Notamment au 17^{ème} siècle où on désignait ainsi les *symptômes* d'une maladie par exemple.

⁹ A condition de rester dans les limites de la perception humaine, au-delà desquelles une trace de toute façon n'existerait pas.

Du point de vue des traces, cela signifie qu'une trace « composée » de multiples « indices » d'une activité qui a eu lieu ne permet qu'une reconstitution *hypothétique* concernant un autre niveau d'interprétation¹⁰. On retrouve ce « passage sémiotique » entre l'interprétation d'une trace isolée telle l'empreinte laissée dans la neige dont on peut dire qu'elle est celle d'une patte de lièvre, et la trace composée d'un ensemble de ces empreintes que le chasseur peut interpréter comme la trace de son *déplacement* dans telle ou telle direction, et ainsi *agir* pour le suivre. Pour en revenir aux traces d'activités humaines qui nous intéressent, nous pouvons également dire d'elles qu'elles *ne sont pas nécessairement volontaires, intentionnelles* au sens où elles ne sont pas forcément le *produit* d'une action *i.e.* l'*objet* de l'activité réalisée. C'est également le propre des indices « qui produisent de la signification de manière involontaire » (Ferrer, 1997), et du signe (Leleu-Merviel, 2004). Dans cet esprit nous différencions la trace, du cas particulier de la *marque laissée volontairement* par un individu, et faisons la différence entre laisser des *traces* de pas dans la neige et former ainsi involontairement une piste que d'autres pourront suivre, et le fait de *marquer* volontairement son chemin à la manière du petit poucet pour être capable (soi ou autrui) de le suivre à nouveau.

1.1.3 La trace enregistrée

Jusqu'ici nous avons considéré la trace comme une inscription *directe*, si l'on peut dire, que peut générer la réalisation d'une activité dans son environnement. Or, les avancées technologiques ont permis d'étendre « artificiellement » l'environnement d'inscription de l'activité, en offrant la possibilité d'*enregistrer* la réalisation de celle-ci sur un support qui ne fait pas directement partie de l'environnement de réalisation de l'activité, débouchant sur une trace *indirecte* (puisque elle ne s'inscrit pas dans l'environnement de réalisation de l'activité). C'est le cas notamment d'enregistrements audio-vidéo d'une activité, filmée dans le but avoué de son analyse. Considéré dans ce cas particulier comme la *trace* de l'activité, cet enregistrement va être mobilisé *a posteriori*, par exemple pour une étude interactionnelle fine, ou encore pour provoquer des situations d'auto-confrontation (Theureau, 2009 ; Clot et Faïta, 2001).

On peut considérer que cet enregistrement est effectivement trace de l'activité puisqu'il est utilisé comme tel et que les acteurs dans leur activité n'ont pas nécessairement laissé intentionnellement des traces de leurs gestes dans le film¹¹. En revanche, cet enregistrement introduit un changement parce que la trace elle-même, en tant qu'ensemble d'indices, sera inscrite sur un *autre support* que celui de l'activité dont elle est la trace. De plus, ce support, même s'il paraît très riche, reste un point de vue partiel (et partial) sur l'activité. Le cameraman a lui-même choisi l'angle selon lequel le film serait pris, et déterminé *a priori*, d'une manière ou d'une autre, consciemment ou non, les indices et donc la trace composée qu'il pourra en tirer *a posteriori*. Comme nous allons le voir, il s'agit là d'une remarque qui s'applique également à la notion de « trace numérique » que nous allons tenter de caractériser.

1.1.4 La trace numérique, une trace comme les autres ?

Ce long préambule consacré à la notion de « trace » nous a permis de prendre la mesure de la complexité d'une notion qu'il s'agit de transposer maintenant au contexte d'une activité réalisée dans un

¹⁰ Capacité qui a rendu célèbre Sherlock Holmes et les Princes de Sérandipe (Demonet, 1994).

¹¹ Même s'il faut admettre qu'à partir du moment où ils se savent filmés les acteurs changent sans doute leur comportement.

environnement numérique. Intuitivement, on aurait tendance à vouloir parler d'une « trace numérique » comme d'une « trace comme les autres ». Cependant, le passage d'un environnement matériel à un environnement numérique et virtuel modifie les conditions d'inscription de traces telles que nous les avons vues jusqu'ici.

1.1.4.1 Support d'inscription des traces

Pour tenter de raisonner clairement sur ces modifications, reprenons la distinction entre la trace isolée et la trace composée. Pour une trace isolée, tout d'abord. D'ordinaire, lorsqu'une activité se déroule dans un environnement matériel, la matérialité de cet environnement offre, implique même, l'existence de traces laissées par l'action : l'environnement matériel constitue *en permanence* un support à l'inscription de traces laissées par les activités qui s'y déroulent. Cette matérialité autorise l'existence de traces que toute activité peut y laisser, qu'elle soit propre à un objet (une page cornée ou abîmée d'un livre), ou bien se rapportant à sa relation au reste de l'environnement (une feuille restée sur le dessus de la pile et qui sera vue/lue en premier). De plus, dans ces conditions chaque trace est un indice, nécessaire ou non, mais *unique, non prédéterminé, est lié à la contingence* d'une réalité matérielle à un instant « t ».

La particularité d'un environnement numérique, c'est sa *virtualité*, la *dématérialisation* des supports, et avec elle celle des contenus et des documents, remettant en cause jusqu'à la notion même de document (Pédaque, 2005 ; Zacklad, 2005a ; Chabin 2004). Cette dématérialisation fait disparaître certaines des *contraintes* liées à la nature physique des supports ordinaires (Leleu-Merviel, 2004), mais elle fait également disparaître du même coup les *conditions d'émergence* et d'existence de traces d'activité. Dans le produit d'une activité documentaire numérique, bien peu de traces de l'activité de réalisation persistent. Prenons l'exemple de la production d'un texte dans un éditeur : le fruit des actions finalisées de l'auteur reste, mais ni ratures, ni hésitations dans le trait, ni brouillon ne subsistent, chaque nouvelle opération écrasant les précédentes dans une nouvelle version finalisée (Ferrer, 1997).

Pour obtenir une trace numérique, il est nécessaire de créer de toutes pièces ses conditions d'existence, de créer artificiellement une possibilité d'inscription. Autrement dit, il est impératif que le concepteur du système ait déterminé *a priori*, ce que sera cette trace. Une trace numérique dans ce cas n'est pas un « donné », mais un « construit », une information enregistrée à un instant donné par la machine. Même si on peut argumenter que, du point de vue de l'utilisateur du système, un tel enregistrement ne constitue pas plus une trace intentionnelle qu'elle ne l'aurait été dans un environnement matériel, aucun de ces indices ne peut apparaître en dehors du système d'enregistrement mis en place, il n'y a que des interactions qui ont été choisies pour être enregistrées. À ce titre une trace numérique isolée est un indice nécessaire (*tekmeria*) : si enregistrement il y a, alors l'interaction correspondante a *nécessairement* eu lieu. Au-delà de la certitude de l'interaction unitaire, le « doute interprétatif » renaît lorsqu'il s'agit de considérer la trace *composée* d'un ensemble de ces enregistrements.

1.1.4.2 Caractéristiques d'une trace numérique composée

Suivant ce que nous venons de dire des traces numériques isolées, une *trace numérique composée* est constituée d'enregistrements réalisés au cours d'une activité. En tant qu'ensemble « d'indices certains » (*tekmeria*), elle se prêtera à une interprétation similaire à celle des traces composées matérielles que nous avons précédemment évoquées, en incluant le passage sémiotique dans lequel subsiste tou-

jours une part de doute (*eikota*). Ce n'est d'ailleurs peut-être qu'en tant que trace composée que l'on pourrait évoquer une notion de « trace numérique » comparable à la notion de trace en général. Des différences notables sont à signaler cependant.

Tout d'abord, la linéarité des interactions d'un utilisateur avec son système *structure temporellement* les enregistrements, et donc la trace numérique résultante. Ainsi donc une trace numérique composée d'activité sera composée d'enregistrements d'interactions successives, formant une « histoire interactionnelle ». On parle d'ailleurs « d'historiques » pour désigner l'enregistrement des opérations réalisées par un utilisateur (ou une machine) avec un outil informatique (historique de navigation Web, historique des connexions à un serveur, *etc.*). Il serait peut-être plus juste de parler de chronologie plutôt que d'historique, car le seul critère de construction est la temporalité, qui ne suffit à elle seule à raconter une « histoire ».

Autre caractéristique spécifique, toujours liée au temps : les traces numériques composées le sont d'enregistrements conservés aussi longtemps que nécessaire, du moins à l'échelle temporelle des activités auxquelles nous nous intéressons. Les traces numériques n'apparaissent que parce qu'elles ont été prédéterminées volontairement, elles ne disparaissent que dans les mêmes conditions et peuvent s'accumuler pour former de larges collections de données. Il est intéressant par exemple de rappeler que les hypertextes ont été pensés dans le contexte d'un volume toujours croissant d'information, non seulement sur l'idée de fournir un accès plus intuitif mais également et *explicitement* sur l'idée de rendre possible *l'accès réitéré* à une information déjà lue en suivant ses propres traces : « *a concept to revisit information read before by following self-created trails* »¹² (Obendorf, 2007, p.597).

1.1.5 Bilan

Ce travail définitoire a montré qu'une différence fondamentale existe entre la notion générale de « trace » et celle de « trace numérique ». Cette différence n'est généralement pas mise en lumière, ce qui explique peut-être la confusion qui règne autour de l'utilisation du terme de « trace numérique ». Nous retiendrons pour le moment que dans le contexte d'une activité réalisée dans un environnement numérique standard (un ordinateur muni de logiciels classiques) une trace numérique peut être vue comme un *ensemble d'enregistrements de données dont l'existence est provoquée par des interactions utilisateur dans le cadre de la réalisation de son activité instrumentée*. Ajoutons que dans ce cas on ne peut parler d'une trace numérique *que relativement* à une *activité réelle et concrètement réalisée* (Hawkey, 2006).

¹² Il est fait référence ici au travail de V. Bush (1945) sur *Memex*.

1.2 Les activités numériques tracées

Nous avons entamé ce premier chapitre en évoquant les traces d'une *activité* numérique, sans préciser de quelle activité il s'agissait. Or, il s'agit dans cette thèse d'une activité ou plutôt d'une famille d'activités particulières qui vont constituer le *contexte* dans lequel nous nous posons la question des traces numériques. De façon générale nous ferons ainsi référence dans notre travail à des activités que nous avons baptisées *Activités Documentaires Instrumentées* (ADI).

1.2.1 Activité Documentaire Instrumentée

De manière générale, nous entendons par Activité Documentaire Instrumentée (ADI) une activité intellectuelle qui s'appuie et se manifeste par la manipulation, l'exploitation, la création, et la gestion *individuelle* et *partagée* de documents. En ce qui nous concerne ici l'instrumentation en question est bien entendu *informatique*, c'est-à-dire que l'activité se déroule pour tout ou partie au sein d'un environnement *numérique*. Dans ce manuscrit il sera question de trois activités en particulier, que nous avons exploitées soit comme *terrains de réflexion* pour construire nos idées sur les traces dans un contexte réaliste, soit comme *terrains expérimentaux* au sein desquels nous avons concrètement travaillé. Les trois activités en question sont décrites dans les sections à suivre : la *recherche d'information* (section 1.2.2), l'*apprentissage collaboratif à distance* (section 1.2.3), et la *création collaborative de contenus pédagogiques* (section 1.2.4).

1.2.2 La Recherche d'Information

La Recherche d'Information a été retenue pour sa représentativité des Activités Documentaires Instrumentées. À ce titre nous l'avons utilisée comme un support de réflexion, et non comme support de recherches expérimentales. Notons toutefois que la Recherche d'Information est bien sûr présente dans toutes les activités documentaires médiées, y compris les activités auxquelles nous nous intéressons de près et qui seront décrites plus loin. Selon le domaine dans lequel elle est considérée, la Recherche d'Information, ne renvoie pas tout à fait à la même notion. En Informatique par exemple, la Recherche d'Information renvoie plutôt aux *Systèmes* de Recherche d'Information (SRI), leur conception, l'évaluation de leurs performances (Ihadjadene, 2004). En Sciences Cognitives, en Ergonomie ou en Psychologie Cognitive, la Recherche d'Information désigne plutôt les *processus cognitifs* et les *stratégies*, que l'on décrit pour aboutir à des modèles cognitifs du comportement de recherche engagé par des sujets (Hsieh-Yee, 2001). Enfin, les Sciences de l'Information considèrent l'objet « Recherche d'Information » de manière plus globale, en intégrant ses différentes dimensions : cognitive, comportementale, d'usage et de pratique, ou encore socio-organisationnelle et technologique. Finalement, la Recherche d'Information pourrait ne pas apparaître comme une activité *stricto sensu* : elle ne peut être considérée dans l'absolu et se trouve généralement intégrée à une situation d'activité plus ample qui lui donne sa *motivation* et ses *raisons*. Aussi, lorsque nous ferons appel dans la suite de ce manuscrit à la Recherche d'Information en tant qu'ADI, elle sera considérée *dans le contexte* d'une activité de *veille informationnelle*, en particulier sur le *Web*¹³.

¹³ C'est un domaine dans lequel nous avons accumulé de l'expérience au cours des années précédant ce travail de thèse.

La veille informationnelle représente un contexte un peu particulier qui fait appel à la recherche d'information sous toutes ses formes. De manière académique la *veille informationnelle* est définie comme une *pratique* de surveillance de l'environnement d'une organisation. Elle comprend « *des opérations de recherche, de collecte, de diffusion, de traitement, de validation et d'utilisation de l'information à caractère stratégique* » (Dartois, 2001). Cette pratique se révèle utile voire vitale, pour tout type d'organisation, dont le fonctionnement consomme et produit de l'information. Qu'il s'agisse de saisir des opportunités ou bien d'éviter des écueils, l'organisation doit dégager du flot d'informations accessibles l'information *stratégique*, génératrice de *décisions* et *d'actions* sur l'environnement (Timini, 2000).

Ces dernières années, l'explosion du *Web*, des sources d'information qu'il propose, ainsi que l'accessibilité nouvelle à l'information qu'il autorise, ont redistribué les cartes de cette pratique. La plupart des étapes du processus de veille s'en trouve touchées (Chartrand, 2003). Dans le contexte de la veille, la Recherche d'Information est une activité centrale, qui se doit d'être efficace, itérative et méthodique, orientée vers des besoins informationnels qu'elle participe elle-même à définir¹⁴. Elle s'inscrit de plus en plus dans l'exploitation d'environnement numérique, qu'il s'agisse d'outils (plateformes de veille, outils de communication, de GED¹⁵, etc.), ou bien de ressources (*Web* en général, bases de données en ligne, etc.).

Du point de vue de l'usage des environnements numériques exploités, la Recherche d'Information peut-être entendue, selon le contexte, comme deux objets de recherche distincts. D'une part l'*Information Retrieval*, pour lequel il s'agit de concevoir et évaluer des systèmes (algorithmiques) dont l'objectif est de retrouver le plus efficacement et rapidement possible une information déterminée dans un ensemble de documents donnés (Gaussier et Stéfanini, 2003¹⁶ ; Järvelin et Wilson, 2003). D'autre part, l'*Information Seeking* qui correspond plutôt à l'activité humaine de « découverte d'information », qui peut s'apparenter à une *enquête* comme le suggère M. Zacklad et H. Zaher, qui parlent de *Recherche Ouverte d'Information* (Zacklad, 2007a ; Zaher et al., 2006, 2007). Dans ce cas, on ne sait pas nécessairement ce qu'on doit s'attendre à trouver, et une bonne part de l'activité relève de la construction de sens à partir des informations découvertes. Ce que nous souhaitons considérer comme activité documentaire instrumentée est dans ce contexte plutôt l'*Information Seeking*, une activité dans laquelle un utilisateur va construire un « espace informationnel », y manipuler des documents et en construire une interprétation qu'il est souvent nécessaire de partager (Komlodi, 2004).

La Recherche d'Information constitue à nos yeux une bonne illustration du type d'activités que nous cherchons à instrumenter, de plus elle fait l'objet de nombreuses recherches dans plusieurs disciplines, ce qui nous permet de compter sur un large corpus bibliographique. De plus, comme nous serons amenés à le voir par la suite, il existe un lien fort entre les travaux sur la Recherche d'Information sur le *Web* et ceux concernant les traces d'interactions, autour des historiques bien sûr, mais plus généralement sur la réutilisation des informations et des procédures de recherche.

¹⁴ Dans la suite de ce manuscrit, nous parlerons de manière générale de Recherche d'Information en adoptant ce contexte particulier.

¹⁵ GED : Gestion Électronique de Documents.

¹⁶ On y apprend que l'expression « *information retrieval* » fut utilisée la première fois en 1948 par Calvin Mooers.

1.2.3 L'apprentissage (collaboratif) à distance

La seconde activité retenue comme activité potentiellement tracée est *l'apprentissage collaboratif à distance*. Nous entendons par apprentissage une *activité structurée d'acquisition de connaissances* par un apprenant (et non sa capacité cognitive à le faire), qui dans le cas présent se déroule à *distance* par le biais d'un logiciel dédié. L'apprentissage collaboratif à distance est une des situations¹⁷ d'apprentissage que rendent possible les Environnements Informatiques d'Apprentissage Humain (EIAH) et qui forment le *e-learning* au sens large.

Dans le cas qui va nous intéresser, l'apprentissage collaboratif à distance prend place dans l'usage d'un dispositif de « classe virtuelle », proposée par la *start-up eLycée campus*TM officiant aux Etats-Unis¹⁸ qui a lancé en 2006 une offre d'enseignement du français à distance. Cette offre a d'abord été conçue pour des enfants (de 8 à 14 ans) francophones mais ne pouvant pas pratiquer le français au sein du système scolaire américain. La « classe virtuelle » en question permet à plusieurs élèves (jusqu'à une dizaine) et à leur tuteur de se connecter au même moment à un même espace partagé pour y réaliser de manière synchrone les activités pédagogiques qui leur sont proposées. En complément de ces séances communes des séances individuelles (impliquant un apprenant seul avec le tuteur) sont également organisées. Techniquement la plateforme est constituée de trois outils (Figure 1.3) :

- Un outil de visioconférence, *Marratech*¹⁹, qui offre un espace de travail collaboratif à toute la classe (vidéo, chat public ou privé, et tableaux blancs partagés permettant d'écrire, de dessiner, de manipuler des images, *etc.*).
- Un site Web, qui regroupe les informations relatives à l'organisation des séances mais également un ensemble de liens utiles (pointant vers un dictionnaire, un glossaire, *etc.*), un ensemble de ressources exploitables, et un espace personnel permettant de gérer ses propres documents.
- Un outil de navigation conjointe dit de « co-navigation », *eMediathèque*TM, développé tout spécialement pour permettre aux apprenants d'explorer pages *Web* et ressources multimédia de façon synchrone et coordonnée²⁰.

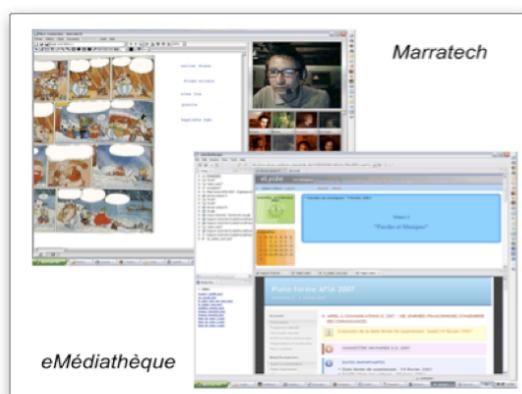


Figure 1.3 : Plateforme proposée par *eLycée*, composée d'un outil de visioconférence (*Marratech*), d'un navigateur *Web* permettant la co-navigation (*eMediathèque*) et d'un site Web dédié ici affiché dans le navigateur.

¹⁷ On y retrouve également la formation à distance, l'autoformation ou formation assistée par ordinateur, les formations en ligne, *etc.*

¹⁸ Voir <http://www.elycee.com/> (consulté le 10/06/09).

¹⁹ Voir <http://www.marratech.fr/> (consulté le 10/06/09).

²⁰ Un utilisateur peut « pousser » la page sur laquelle il se trouve aux personnes avec qui il travaille.

L'objectif de l'équipe pédagogique est double : il s'agit de développer une connaissance de la langue française par la pratique d'une part et de maintenir un contact avec la culture française d'autre part. La richesse et la diversité des technologies intégrées par cette plateforme permettent de mettre en place une démarche pédagogique s'éloignant sensiblement du schéma classique de l'apprentissage de type leçon / exercice, pour se rapprocher d'une vision constructiviste de l'apprentissage. L'idée est ici de considérer l'apprentissage *dans et par* la pratique et *dans l'interaction* avec autrui.

Dans une approche constructiviste (Gabriel, 2004), très grossièrement, ce n'est pas la rétention d'une règle ou d'un fait qui est visée en tant que telle, mais le développement de la *capacité* à comprendre et retenir cette règle ou ce fait : « *Meaning making, according to constructivists is the goal of learning processes; it requires articulation and reflection on what we know* » (Jonassen *et al.*, 1995, p.11). D'autre part, la *dimension sociale de l'apprentissage* est privilégiée. Elle est notamment présente dans plusieurs grands courants de pensée, au premier rang desquels on retrouve, entre autres, l'approche Vygotskienne de l'apprentissage (Clot, 1999 ; Ollagnier, 2006) ou encore l'approche du *Situated Learning* (Lave et Wenger, 1991). Les activités pédagogiques proposées dans ce contexte sont donc choisies pour qu'elles nécessitent et mobilisent de façon indirecte un savoir faire linguistique pour répondre à un réel besoin d'expression, tendu vers un but utile : celui de la production collaborative à travers laquelle se réalise l'apprentissage. Concrètement, les activités que vont mener les apprenants s'articulent autour de documents numériques, qu'ils vont d'une part exploiter (supports pédagogiques multimédia) et d'autre part produire (comptes-rendus, productions écrites, blog, *etc.*).

La situation d'apprentissage dont il est question ici relève bien des *Activités Documentaires Instrumentées*. Tout d'abord elle est réalisée entièrement dans un espace numérique, qu'il s'agisse des séances de groupe ou des séances individuelles (qui utilisent *Marratech*). De plus, qu'il s'agisse de les exploiter, de les manipuler ou d'en produire, les documents numériques sont ici au cœur des activités réalisées. L'activité d'apprentissage à distance (au sens large) a suscité de nombreux travaux de recherche dans le domaine des EIAH en général, et pour ce qui est de l'aspect collaboratif de cet apprentissage, dans le domaine de l'*Apprentissage Collaboratif Assisté par Ordinateur (ACAO)*²¹. Là encore, une abondante littérature scientifique nous donne de bonnes bases pour comprendre et caractériser cette activité, tant sur le plan cognitif de l'apprentissage humain, que sur le plan technologique des dispositifs mis à la disposition de groupes d'apprenants. Avec l'accord de l'équipe pédagogique, nous avons pu accéder à l'espace de travail et observer *in situ* (si l'on peut dire) des situations pédagogiques réelles de travail à distance. Ces activités pédagogiques ont constitué un terrain de réflexion important et ont été, partiellement, exploitées comme terrain expérimental.

1.2.4 Création de contenus de formation

La troisième et dernière activité retenue, est une activité de *création de contenus* pédagogiques. Pour être plus précis, il s'agit de l'activité d'une équipe d'ingénieurs pédagogiques au sein du département formation d'un groupe industriel français. Cette équipe est en charge de la production de contenus pédagogiques pour certaines formations internes dispensées par le groupe.

²¹ *Computer Supported Collaborative Learning (CSCL)* au niveau international.

Les contenus pédagogiques créés au sein du service le sont sur support numérique : il s'agit de présentations et d'animations utilisables directement par un formateur dans une salle de cours²². Au delà de la création *stricto sensu* de ces ressources, l'activité qui est en jeu ici est la *conception globale* des formations. Cette activité mobilise bien entendu tout au long de sa réalisation un grand nombre de documents, pour la plupart numériques. La réalisation elle-même repose sur des outils documentaires classiques (outils de présentation, de courriel, de base de donnée, *etc.*) auquel il faut ajouter l'apport récent d'un *atelier de production de contenus multimédia* (Figure 1.4) qui vient se placer au cœur du processus de production. *Emulsion*, c'est son nom, est un atelier qui a été développé et proposé par la société *Solunéa*²³ et qui regroupe les fonctionnalités nécessaires à la création de contenus multimédia, dans le respect des standards du *e-learning*, avec un rendu en format *Flash*²⁴. La conception des formations comme la réalisation des contenus nécessaires avec *Emulsion* exige la participation de plusieurs membres de l'équipe, et à plusieurs niveaux d'organisation.



Figure 1.4 : Interface d'*Emulsion*, atelier de production collaborative de contenus.

Nous aurons l'occasion de revenir en détail (chapitre 6) sur la description du processus de création de ces contenus pédagogiques. Retenons pour le moment que les membres de l'équipe sont plongés dans un espace documentaire constitué de milliers de documents sources (des supports de formation existants, des documents techniques et des documents permettant la réalisation de l'activité comme le cahier des charges) et de documents produits ou à produire, le tout à l'aide d'un ensemble d'outils informatiques documentaire.

Comme pour les deux activités précédentes, on comprendra après cette courte description en quoi l'activité de création de contenus de formation apparaît comme un bon exemple d'*activité documentaire*, et par quels dispositifs elle est *instrumentée*. Elle a de commun avec la recherche d'information son caractère ouvert, lié à la créativité demandée aux ingénieurs pédagogiques, et partage avec l'apprentissage collaboratif à distance, une dimension collaborative forte au sein de l'activité. Pour obtenir des éléments théoriques sur cette activité de création de contenus, nous pouvons bien sûr nous tourner à nouveau vers le domaine des EIAH, mais en la considérant comme une activité de conception et même de co-conception, c'est vers d'autres domaines tels l'*ergonomie*, la *psychologie cognitive* ou l'*Ingénierie des Connaissances* que nous trouverons un support théorique.

²² Il est toutefois prévu que ces supports puissent également à terme être directement utilisés dans le cadre d'auto-formations.

²³ Voir <http://www.solunea.fr/> (consulté le 10/06/09).

²⁴ Voir <http://www.adobe.com/fr/products/flash/> (consulté le 10/06/09).

Concernant la création de contenus comme objet de notre travail, elle a été un terrain de réflexion pour les traces numériques, mais a également constitué pour notre travail de thèse un important terrain expérimental. La situation d'activité au sein de l'équipe d'ingénierie pédagogique (que nous nommerons « équipe IP ») a permis un travail de terrain détaillé, mais également l'instrumentation spécifique pour des traces d'une partie importante de l'environnement de travail (l'atelier de production de contenu *Emulsion* en l'occurrence). Nous avons également pu compter sur le soutien conjoint de l'équipe et de la société *Solunea* dont les membres se sont montrés disponibles et positifs quant aux possibilités d'usages des traces numériques dans leur cadre de travail.

1.3 Problématique

Exprimé simplement, notre objectif est de *déterminer comment créer des traces d'activités numériques concrètement réutilisables dans le cadre d'activités documentaires instrumentées*, en particulier dans le cas des activités que nous venons de présenter. Pour être plus précis, nous pouvons extraire de cette problématique deux grandes questions, auxquelles nous avons à répondre dans un premier temps. Premièrement, nous venons de décrire ce que sont théoriquement les traces numériques mais il nous faut préciser comment se définissent concrètement des traces numériques en tant qu'objet informatique. Deuxièmement, nous parlons d'une trace numérique *réutilisable* ce qui implique l'existence d'un objectif de *réutilisation* de ces traces : il s'agit de se demander quel peut (ou doit) être cet objectif de réutilisation dans le contexte d'une activité documentaire instrumentée, et en particulier pour les activités que nous avons présenté comme des terrain expérimentaux.

Pour commencer à répondre à ces deux questions, nous proposons dans le chapitre suivant un état de l'art complet sur les applications informatiques utilisant ou fournissant des enregistrements d'interactions utilisateurs. Il est difficile d'affirmer pouvoir aujourd'hui fournir un état de l'art exhaustif sur la question tant le nombre d'applications concernées et de domaines de recherche impliqués est grand. Notre objectif dans ce travail bibliographique est donc de couvrir le maximum de types d'application rentrant dans ce que nous appellerons de manière générique les « systèmes traçants », en ne nous limitant pas à un domaine d'application particulier comme c'est souvent le cas. Nous souhaitons avec cette recherche, comprendre ce que sont concrètement les traces numériques dans toutes ces applications, la manière dont elles sont obtenues et comprendre également dans quel intérêt, dans quel but elles sont utilisées ou réutilisées. Ces éléments de compréhension nous sont indispensables pour *positionner* notre démarche sur un plan informatique, et pour *placer* notre travail dans une perspective de développement des traces numériques non seulement dans l'optique d'usages innovants mais également comme objet de recherche.

Chapitre 2

État de l'art des systèmes traçants

Résumé du chapitre

La mise en place et l'exploitation de dispositifs informatiques enregistrant les interactions utilisateur pour en faire des *traces d'interaction* réutilisables n'est en elle-même pas une idée nouvelle et remonte aux premiers travaux sur les Interfaces Homme-Machine. Nous avons appelé de manière générique *systèmes traçants* ces dispositifs, quels que soient les moyens d'enregistrement mis en place, la forme des traces d'interaction obtenues et les modes de réutilisation envisagés. Ce second chapitre est consacré à la présentation d'une revue complète et pluridisciplinaire de tous les types de systèmes traçants. Pour simplifier, nous avons distingué deux grandes familles de travaux en opposant deux types d'exploitations des traces numériques d'interaction : soit les traces sont exploitées *en dehors* de l'activité qui les a générées à des fins d'analyse, soit elles y sont réintroduites d'une manière ou d'une autre dans le but de soutenir et faciliter l'activité en cours. Quelle que soit leur finalité l'exploitation des traces numérique pose un double problème : le premier est de « *recupérer des données* » liées aux interactions, le second est de leur « *donner du sens* » relativement à l'activité dont elles sont issues. Deux façons d'envisager une réponse à ce problème co-existent. La première et la plus répandue repose sur l'idée que l'on peut extraire une trace qualitative de données quantitatives directement récupérées dans les systèmes informatiques (les *logs*). Les solutions techniques proposées restent principalement *ad hoc*, attachées voire intégrées aux outils utilisés et aux *logs* qu'ils peuvent fournir. On tente alors de donner du sens aux traces numériques d'activité *a posteriori*. La seconde, qui apparaît comme une vision « alternative », propose de poser le problème différemment. L'idée fondamentale est de renverser en partie l'attribution du sens, et d'augmenter *a priori* la sémantique de *ce qui sera tracé, i.e.* de générer directement au cours de l'activité une trace numérique « sémantiquement plus riche ». Dans ce cas, la trace numérique est un objet informatique à part entière, que les « systèmes à base de Traces » permettent de manipuler.

2.1 Introduction

L'idée de conserver un enregistrement, ou une inscription des opérations effectuées par un ordinateur au cours de son utilisation est loin d'être nouvelle. Il s'agit en informatique d'une pratique liée au fonctionnement même de ces dispositifs techniques. Ces machines à calculer effectuent très rapidement un grand nombre d'opérations logiques. Les informaticiens ont toujours eu besoin de pouvoir vérifier les étapes de ce calcul pour en comprendre le résultat. Ce qui était vrai pour les calculs mathématiques aux origines de l'informatique, l'est encore aujourd'hui dans le cadre du développement de logiciels. Les programmeurs ont largement recouru à des informations enregistrées concernant les opérations effectuées par leur programme, stockées dans des *logfiles* qui leur permettent de « debugger »²⁵ un code.

Ce principe, qui à l'origine était utilisé pour comprendre les opérations effectuées par la machine elle-même, a tout simplement été appliqué pour enregistrer et comprendre les interactions utilisateurs, quand celles-ci ont commencé à concerner autre chose que du calcul. Ceci correspond grosso modo à l'arrivée de l'informatique documentaire et des interfaces graphiques, lesquelles ont ouvert de nouvelles gammes d'interfaces plus abordables que la ligne de commande, et multiplié les interactions possibles. Il s'agit dès lors de comprendre non seulement les opérations de la machine mais au-delà, l'interaction qui a lieu entre l'utilisateur et sa machine. C'est donc sans surprise qu'on trouve dans les travaux d'IHM datant d'une vingtaine d'années, les premiers enregistrements systématiques d'interactions réalisés en ce sens. Ces enregistrements permettent en effet de prendre le temps d'analyser en détail ces interactions au-delà de l'immédiateté spatiale et temporelle d'une observation *in situ*. Par la suite une même démarche d'enregistrement des interactions, facilitée par les progrès techniques réalisés sur les environnements numériques (rapidité, capacité de stockage, *etc.*), a été appliquée dans de multiples domaines de recherche : Informatique bien sûr, mais également Ergonomie, Psychologie Cognitive, Sciences de l'Information, et même Sociologie et Anthropologie.

2.1.1 Systèmes traçants et traces numériques d'interactions

Dans cet état de l'art nous allons présenter un grand nombre de dispositifs enregistrant systématiquement les interactions utilisateur et allons nous intéresser à la *finalité* de ces enregistrements. Nous avons baptisé ces dispositifs « systèmes traçants » lorsque ces enregistrements sont réalisés dans un but précis et ne sont pas le fait du simple fonctionnement des logiciels utilisés. Il existe des dispositifs très divers, des outils standards que l'on a modifiés pour qu'ils enregistrent leur propre utilisation à des applications indépendantes et dédiées à ces enregistrements (intégrées ou non à d'autres applications) en permettant leur traitement et leur exploitation. Nous parlerons également de « traces numériques d'interaction » pour désigner de manière générale les enregistrements, les « historiques interactionnels » ainsi exploités.

La diversité des instrumentations en question, ajoutée à la celles des exploitations possibles et des domaines de recherche impliqués, rend difficile toute tentative de classification des travaux sur les traces d'interactions. Il est même difficile aujourd'hui de mesurer la place et le rôle qu'occupent réel-

²⁵ Trouver le moment dans l'enchaînement logique des opérations qui pose problème, identifier la source de ce problème et modifier le code pour y remédier.

lement les systèmes traçants et leurs traces d'interactions dans l'ensemble des domaines scientifiques concernés. Aucun état de l'art ne semble d'ailleurs couvrir la question des traces numériques d'interaction dans sa globalité. Si le dénominateur commun de ces travaux reste évidemment l'informatique, la plupart des états de l'art disponibles aujourd'hui sur les systèmes traçants, sont thématiquement rattachés à une discipline ou à un domaine d'application particulier : Interface Homme-Machine (IHM), Sciences de l'Information et de la Communication (SIC), Environnement Informatique d'Apprentissage Humain (EIAH), *etc.*

Lorsqu'il arrive qu'un travail bibliographique aborde de manière transversale la question des systèmes traçants, c'est en se focalisant sur des aspects purement techniques du traitement automatique des données temporelles, sans rendre compte complètement de ce que représentent les traces numériques d'interaction à la fois comme ressource et comme outil pour les domaines qui les exploitent concrètement. En tentant de rassembler les éléments d'une littérature fragmentée à travers plusieurs domaines de recherche en une vision globale des systèmes traçants et des traces numériques d'interaction qu'ils engendrent, cet état de l'art se veut une première contribution de cette thèse aux connaissances des différents domaines impliqués sur les systèmes traçants qu'ils exploitent.

2.1.2 Articulation du chapitre

Pour être à la hauteur de son ambition, notre état de l'art va devoir couvrir une large bibliographie : dans le *temps* tout d'abord, puisque les travaux sur les traces numériques d'interaction sont présents dans la littérature depuis vingt ans ; dans les *disciplines* ensuite, puisque les champs de recherche impliqués sont très variés ; dans les technologies informatiques, puisque sont utilisés des instrumentations allant de simples *plugin* jusqu'à des plateformes intégrées dédiées à l'exploitation de traces multiples. Afin de maintenir un fil conducteur dans l'exposé de ce qui pourrait apparaître comme une liste hétéroclite de travaux, cet état de l'art présentera les approches et travaux mettant en œuvre des systèmes traçants *en les distinguant selon leur finalité*, c'est-à-dire selon la *posture adoptée relativement aux traces numériques d'interactions* par ceux qui les exploitent.

Dans cet état d'esprit, une distinction simple se fait entre *deux postures* adoptées vis-à-vis des traces numériques d'interaction qui débouchent sur *deux grandes finalités* attribuables aux systèmes traçants (Wexelblat, 1998). La première est une *posture d'analyse* et désigne des travaux dans lesquels les traces numériques d'interaction d'un utilisateur sont *exploitées en dehors et indépendamment* de l'activité de celui-ci (et de l'environnement numérique dans lequel elle se déroule) et sont support d'un travail d'analyse. La seconde est une *posture de réutilisation* et désigne des travaux dans lesquels les traces numériques d'interaction sont *réintroduites et exploitées au sein même de l'activité* dont elles sont issues et en deviennent, de fait, un support plus ou moins explicite. L'opposition entre ces deux postures n'est bien sûr une réalité que dans le contexte d'un état de l'art se focalisant sur l'exploitation des systèmes traçants. La plupart des travaux d'analyse de traces numériques conduisent ainsi, à terme, à la modification de la situation dans laquelle l'activité se tient : les traces numériques y sont donc réintroduites indirectement.

Le présent chapitre sera donc articulé afin de mettre en évidence dans ses deux premières parties les travaux sur les traces numériques d'interaction comme *support d'analyse* d'une part (section 2.2), et les travaux les exploitant comme un *support de l'activité* d'autre part (section 2.3). En complément,

nous reviendrons en conclusion (section 2.4) sur ce que ces divers travaux sur les systèmes traçants peuvent avoir de *problématiques communes*, conceptuelles, techniques ou méthodologiques, concernant la constitution des traces numériques d'interaction.

2.2 Les Traces Numériques d'Interaction comme support d'analyse

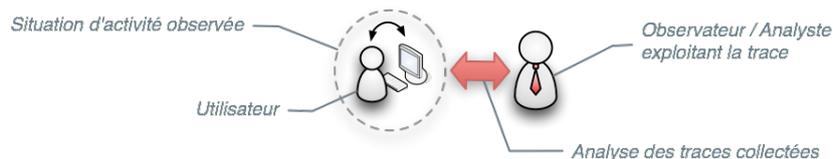


Figure 2.1 : Exploitation des traces comme support d'analyse.

Cette section de l'état de l'art est consacrée à la présentation des travaux de recherches ayant mobilisé des systèmes traçants dont les traces sont exploitées en tant que *support d'analyse*. Typiquement, nous aurons affaire à des situations d'activité dans lesquelles un *observateur*, un *analyste de l'activité*, cherche à comprendre ou rendre compte d'une activité numérique²⁶ sur la base d'une *analyse des traces* d'interaction engendrées par l'activité en question. Dans ce cas de figure, nous pourrions considérer que ces traces *n'ont pas* vocation (*a priori* tout du moins) à modifier la situation d'activité observée, et sauf exception, leur analyse se déroule *a posteriori* des interactions qui sont effectivement tracées .

2.2.1.1 L'analyse de la trace

Lorsqu'il est dit que des traces d'interaction sont *support d'analyse*, cela signifie que les traces d'interaction sont bien analysées mais que les *résultats de l'analyse* portent quant à eux sur *l'objet ou le phénomène étudié* dont ces traces sont la trace (si l'on peut dire). Ainsi le travail d'analyse aura pour but de faire émerger d'une trace des *caractéristiques remarquables* qui ne prennent sens que parce qu'on les attribue non à la trace en elle-même, mais à l'activité interactionnelle qui a eu lieu. Étant donné la nature des traces analysées (souvent de bas niveau d'abstraction), c'est classiquement une analyse *quantitative* et *statistique* des traces qui sera effectuée. Le développement de techniques statistiques spécifiques à ce type d'analyse est même devenu un objectif de recherche en soi et fédère les travaux d'un groupe de chercheurs sur l'analyse exploratoire des données temporelles ou « *Exploratory Sequential Data Analysis* » (ESDA). Leurs travaux restent focalisés sur les techniques statistiques, les traces numériques d'interaction étant réduites à leur expression la plus simple (des ensembles de données séquentielles). Ils ont contribué à définir et développer « [...] *a set of techniques for sampling, coding, and reducing sequential temporal data [...] methods for transcribing and encoding events, and analyzing the encoded data using statistical routines* » (Cuomo, 1994).

Deux types de résultats sont attendus : la *caractérisation* ou l'*évaluation* de l'objet étudié. La *caractérisation* de l'objet étudié vise à rendre compte d'un phénomène en étant capable d'en formaliser la description, voire d'en donner un modèle. Cette caractérisation s'appuiera soit sur le seul travail

²⁶ Entendue comme une activité *instrumentée* par un environnement numérique.

d'observation, soit sur des postulats théoriques. L'évaluation de l'objet étudié (parfois sa *validation*), repose sur l'existence de *critères*, permettant de comparer et de juger les objets étudiés relativement à un modèle idéal ou des valeurs prescrites²⁷. Caractérisation et évaluation sont deux démarches que l'ESDA évoque en termes de « *exploratory data analysis* » d'une part, et de « *confirmatory data analysis* » d'autre part (Sanderson et Fisher, 1994).

2.2.1.2 L'objet de l'analyse des traces

Si le travail d'analyse a pour matériau des traces d'interaction généralement similaires, les objets caractérisés ou évalués varient quant à eux sensiblement d'une recherche à l'autre. Dans le domaine de l'ESDA, on distingue ainsi différents types d'analyses dites « *cognitives* », « *comportementales* » ou « *sociales* », selon que l'objet caractérisé (ou évalué) est l'utilisateur en tant qu'agent cognitif, l'utilisation en tant que comportement ou l'utilisation partagée en tant que comportement social. Nous pouvons reprendre le principe d'une telle distinction en y ajoutant le fait que l'environnement informatique (interfaces et contenus) peut lui-même être évalué sur la base d'une analyse de traces numériques d'interaction. Cette partie de notre état de l'art sera donc articulée autour des différentes finalités possibles de l'analyse des traces. Grossièrement, il y a trois orientations possibles (Figure 2.2) et l'analyse peut permettre de caractériser (ou évaluer), soit l'environnement (section 2.2.2), soit les comportements d'utilisation ou d'usage dudit environnement (section 2.2.3), soit l'utilisateur lui-même à travers ses activités individuelles (section 2.2.4) ou collectives (section 2.2.4.3).

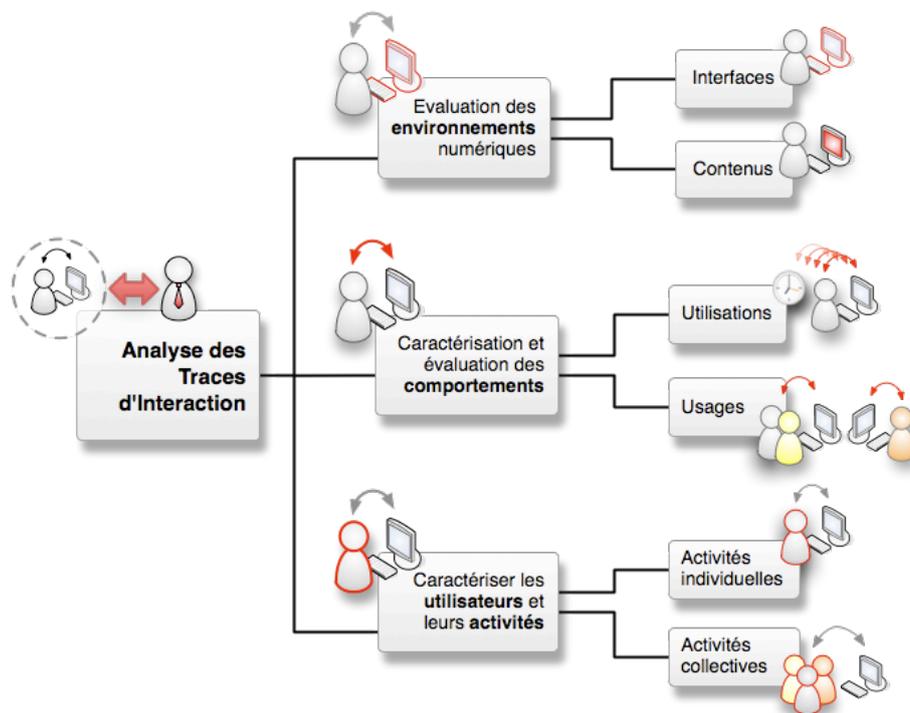


Figure 2.2 : Les finalités des traces d'interactions supports d'analyse.

²⁷ Ce qui exige que ces objets soient, un tant soit peu, caractérisés au préalable.

2.2.2 Analyse des environnements numériques

Nous allons évoquer pour commencer les traces d'interactions utilisées comme support d'analyse pour l'évaluation des « environnements » dans lesquels nous comptons les interfaces (profondes et de surface) ainsi que les contenus fonctionnels que ces interfaces permettent de mobiliser (on pense par exemple aux hypertextes et autres pages *Web*).

2.2.2.1 Les premiers Systèmes Traçants

Historiquement les premiers travaux ayant exploité explicitement des traces numériques d'interaction étaient consacrés à *l'évaluation d'interfaces utilisateur*. Avec l'arrivée dans les années 1980-1990 d'outils informatiques destinés à un large public et de nouvelles interfaces « à fenêtres », les besoins en termes d'évaluation sont devenus de plus en plus importants. Le fonctionnement des interfaces graphiques débouchait sur des problèmes inédits d'*utilité* et d'*utilisabilité*, deux critères qui servent classiquement à évaluer les interfaces utilisateurs de dispositifs informatiques²⁸. La possibilité de travailler sur des traces numériques d'interaction s'est rapidement imposée comme une alternative sérieuse aux approches classiques d'évaluation (plus intrusives et plus coûteuses)²⁹.

Les chercheurs en IHM ont donc développé les approches utilisant l'enregistrement des interactions pour les évaluer *a posteriori* de l'action les interactions utilisateur (Hanson et al., 1984). Par la suite, les outils logiciels ont facilité la récupération de « *user data* » et autres « *interaction data* » (Hilbert et Redmiles, 2000 ; Brinkman et al., 2006). Ces données sont généralement d'un assez bas niveau, que les ergonomes dénomment « *Keystroke level* » : actions sur les touches, sur le dispositif de pointage, clics, que l'on peut *associer* dans le cas d'interfaces graphiques à l'activation des menus, boutons, barres de défilement, etc. Les informations ainsi recueillies permettent *a posteriori* une analyse détaillée des suites d'opérations *réellement* effectuées par l'utilisateur. Peut alors commencer à proprement parler le travail d'évaluation, consistant à déterminer les qualités et défauts d'une interface, parfois relativement à un *modèle d'interaction* « idéal » prévu par des concepteurs, plus classiquement, relativement à des *critères d'évaluation* tels que des mesures du temps de réalisation, du nombre d'opérations nécessaires pour compléter une action, des répétitions d'opérations, etc.

Parmi les tous premiers systèmes, *Playback* (Neal et Simons, 1983) a retenu notre attention. Le système comportait un dispositif de traçage des interactions qui avait été mis en place pour permettre de « rejouer » *a posteriori* l'ensemble des interactions menées par un utilisateur durant une session d'utilisation. Ce « rejeu » pouvait être exécuté « pas à pas », à la façon dont on peut exécuter un programme informatique en mode « *debuggage* ». Cette fonctionnalité devait permettre à un analyste de disposer de temps pour faire une analyse extrêmement fine des interactions. Commodité supplémentaire pour l'analyste, *Playback* permettait également l'inclusion de notes durant la session d'observation ou bien plus tard, lors du rejeu. Les traces numériques d'interaction relevant du *Keystroke level* étant trop nombreuses et trop détaillées pour être exploitables directement, *Playback* proposait d'utiliser les

²⁸ L'*utilité* d'un système se définit comme la possibilité d'atteindre un but grâce à ce système. L'*utilisabilité* d'un système quant à elle se définit comme la possibilité qu'il offre de mettre en œuvre les moyens d'atteindre ce but (Tricot, 2007, p.179).

²⁹ Les méthodes classiques d'évaluation reposaient soit sur une *observation in situ* (qui ne porte que sur une très courte période de temps d'utilisation du système), soit sur une *expérimentation contrôlée* en laboratoire (ce qui décontextualise l'utilisation de l'outil évalué), n'étaient ni économiques, ni neutres quant à l'objet de l'évaluation : « *Both the formal experimental techniques and field research techniques are invasive as well, although in different ways* » (Siochi et Ehrich, 1991, p.310).

enregistrements réalisés d'une part pour *rejouer les interactions*, et d'autre part pour être *associés* à enregistrement audio-vidéo de l'utilisateur durant la session étudiée, permettant à l'analyste une navigation croisée dans les deux sources à sa disposition.

```

Subject 12      Session 1      Condition 4
0:20:30 CLEAR
0:20:55 A = (22/7) * R**2
0:21:20 ENTER
                                Interval: 50 sec.
----- OBSERVATIONS -----
0:20:38 P19K
0:22:02 RTC      Searching for topic in Table of Contents

-----
PF1: Next keystroke  PF2: Next function  PF3: Next interrupt
PF4: Real time      PF6: Set time      <--: Backup      CLEAR: Abort
0:21:20 E5          Syntax error because referring to wrong page in
                                Programmer's Guide

```

Figure 2.3 : Copie d'écran de *Playback* - Tiré de (Neal et Simons, 1983, p.81)

Playback n'était pas un simple outil de traitement de données enregistrées à partir d'un logiciel mais une véritable *plateforme intégrée d'évaluation* de l'utilisabilité des systèmes, disposant de nombreux modules et fonctionnalités (d'enregistrements *keystroke level*, de synchronisation et d'indexation croisée des différentes sources) mis à disposition des analystes. Ce type de plateforme était destiné à l'évaluation d'interface de dispositifs très larges, notamment des systèmes d'exploitation³⁰. Dans cette catégorie, aux côtés de *Playback* nous pouvons ranger d'autres outils comme *DRUM*³¹ (MacLeod et Renger, 1993), un environnement d'évaluation basé sur des enregistrements vidéo qui proposait des fonctionnalités de *dénombrement* et de prédéfinition de schémas d'observation. De même pour *I-Observe*³² (Badre *et al.*, 1995), un ensemble d'outils permettant, outre *l'intégration* de différentes sources, l'interrogation de la trace numérique d'interaction par requêtes complexes portant sur les valeurs des attributs de chaque événement enregistré, ou encore pour *MacSHAPA* (Sander-son *et al.*, 1994), qui fût une des plateformes de ce type les plus populaires. Développée pour l'environnement MacIntosh™, elle regroupe un certain nombre d'outils spécifiquement conçus pour traiter des traces d'interaction et plus largement tout type de données temporelles. Elle permet par exemple la gestion de la *synchronisation* d'un enregistrement vidéo et des mouvements de souris d'un utilisateur, la lecture par sélection de tranches de temps, ou encore l'application des opérations standards de traitement statistique de ces données (ré-écriture, encodage, filtrage, *etc.*).

2.2.2.2 Évolution de l'évaluation d'interface à base de traces

Des travaux très divers marquent l'évolution de l'exploitation des traces numériques d'interaction dans ce contexte. C'est le cas par exemple de l'éditeur de commande modifié par (Siochi et Ehrich, 1991) qui permettait de générer une *transcription automatique (transcripts)*, constituée de l'enregistrement temporalisé des différentes commandes tapées par un utilisateur et des réponses fournies par la machine. Notons que l'enregistrement *ne se limitait pas* à la liste des commandes valides, mais incluait tout ce que l'utilisateur avait *effectivement* tapé, pour être sûr de prendre en compte l'ensemble des opéra-

³⁰ En effet, tous les grands créateurs d'interfaces graphiques, notamment les concepteurs de systèmes d'exploitation (Microsoft™, Apple™ et Sunsoft™ notamment), se sont dotés de plateformes d'évaluation similaires à celles que nous venons d'évoquer.

³¹ *Diagnostic Recorder for Usability Measurement*.

³² *Interface Observation, Evaluation, Recording and Visualization Environment*

tions de l'utilisateur, dans l'évaluation de l'*utilisabilité* du système de commande. Les auteurs proposent ensuite un algorithme de détection de répétition de patterns, la répétition étant selon leur analyse un marqueur d'un problème d'interaction homme-machine.

De façon générale les analyses de trace d'interaction de très bas niveau ne sont plus que rarement nécessaires. Quelques exceptions persistent, comme par exemple lorsqu'il s'agit d'évaluer l'utilisation concrète d'un logiciel dont les développements successifs ne permettent plus d'interpréter globalement le code et donc son fonctionnement exact. Dans l'exemple proposé par (El-Ramly *et al.*, 2001) les traces utilisées sont constituées d'une suite d'états et transitions élaborées à partir de listes de *captures d'écran* de l'application (les états) et d'opérations de l'utilisateur (les transitions). L'évaluation d'interface sur la base de *critères quantitatifs* est quant à elle toujours d'actualité. Il n'est cependant plus nécessaire aujourd'hui de faire appel à de grosses plateformes d'évaluation. L'analyse quantitative des interactions peut se baser sur des outils simples tel *Experiscope* (Figure 2.4) qui permet la visualisation et le traitement de *logs* d'interaction de bas niveau, et fournit un support pour le travail d'évaluation de l'analyste (Guimbretière *et al.*, 2007)³³.

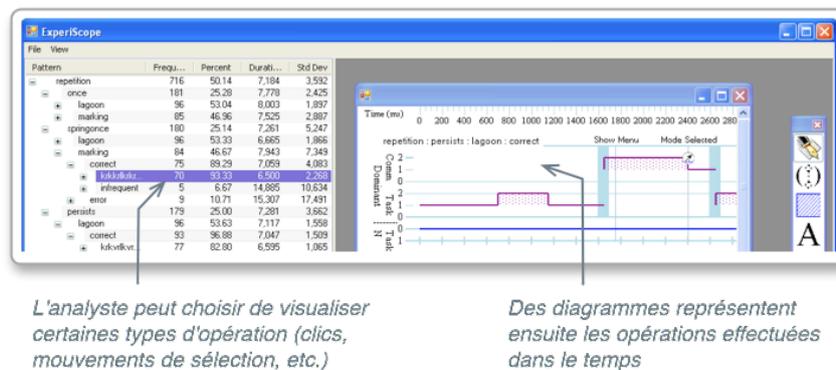


Figure 2.4 : Interface d'*Experiscope* - Tiré de (Guimbretière *et al.*, 2007, p.1333).

2.2.2.3 Évaluation des « interfaces Web » à base de traces de navigation

Une autre catégorie d'interfaces particulières a suscité beaucoup de travaux d'évaluation, y compris et surtout sur la base de l'analyse de traces d'interaction. Ce sont les « interfaces Web », à savoir les interfaces de tout type de contenu accessible par le *Web*. Du point de vue de l'évaluation ces interfaces sont un cas particulier, constituant à elles seules un environnement composite à évaluer : le navigateur autant que la structure ou la mise en forme du contenu, sont susceptibles de modifier les conditions de l'environnement dans lequel évolue un internaute³⁴. Du point de vue de l'évaluation sur la base d'analyse de traces d'interaction, il s'agit également d'un cas à part, car la nature des informations enregistrées est différente. Il s'agit plus de développer des moyens de capter des interactions de bas niveau, mais uniquement de se baser sur des informations disponibles dans le système qui font office de source de collecte, en l'occurrence les *logfiles* (ou *logs*) de navigation, qui sont *générés* nativement par un navigateur³⁵.

³³ Voir aussi <http://www.cs.umd.edu/hcil/experiscope/> (consulté le 30/04/2009).

³⁴ Situation délicate pour l'évaluation puisqu'on ne sait finalement pas exactement ce qui est évalué : le site, sa mise en forme, ou le navigateur lui-même. Cette difficulté est pourtant peu relevée dans la littérature concernée.

³⁵ Le fonctionnement de l'architecture client-serveur génère des fichiers de *logs*, *i.e.* des fichiers listant les opérations effectuées, côté client comme côté serveur. Nous aurons l'occasion d'évoquer cette question plus en détail ultérieurement.

Ces fichiers, souvent temporaires, ne sont pas conçus pour être collectés de la sorte, mais ils permettent d'enregistrer des informations sur les interactions de l'utilisateur avec des *objets de l'environnement Web* : URL¹, fichiers, répertoires, pages, sites, métadonnées, adresses de courriel, etc. Les traces numériques d'interaction ainsi constituées sont mobilisées, entre autres, comme support d'analyse destiné à l'évaluation ergonomique de sites *Web* (Dubois *et al.*, 2000). Une particularité des travaux sur les « *interfaces Web* » est la tentative de mettre au point des systèmes d'*analyse automatique* des traces d'interaction (des *logs* de navigation en l'occurrence) pour l'évaluation des sites *Web*. Dans (Pirolli, 2002) par exemple, les auteurs utilisent les *logs* d'un navigateur (instrumenté à cet effet), pour caractériser les internautes afin d'en créer des modèles implémentés dans un « simulateur » d'utilisateur qui lui-même permettra d'évaluer et de *comparer automatiquement* plusieurs interfaces alternatives pour un site donné. Il s'agit là d'un premier pas vers une sorte de « Graal de l'évaluation », que constituerait son automatisation complète. On trouvera un état de l'art spécifiquement consacré à ce type de systèmes dans (Ivory et Hearst, 2001).

2.2.2.4 De l'évaluation à la réingénierie des environnements

Au-delà de l'évaluation des interfaces, les systèmes traçants sont aujourd'hui mobilisés au cœur même de l'ingénierie de *conception* et de *re-conception* d'environnements numériques. Cette possibilité est même devenu un centre d'intérêt particulier pour certains chercheurs, comme en témoigne des *Workshops* comme celui organisé par W.P. Brinkman en 2006³⁶ et dont l'objectif est précisément le rôle des traces d'interactions au service de la conception et/ou re-conception des systèmes.

On peut même dire que la réingénierie en question concerne toutes les dimensions qui peuvent composer les environnements concernés : non seulement l'interface de surface, comme on vient de le voir, mais également le modèle d'interaction sous-jacent, ou encore le contenu actif qui fait parfois partie intégrante de cet environnement. C'est le cas par exemple dans les EIAH, dans lesquels les *scenarii* pédagogiques font l'objet d'une ingénierie (Tchounikine, 2002) qui compte les traces numériques d'interactions parmi les outils de réingénierie des environnements pédagogiques proposés aux apprenants. Les *scenarii* pédagogiques en question sont parfois développés à la main, en dehors de l'environnement pédagogique, mais la tendance (liée à la poussée du *e-learning* et des standards³⁷ qui l'accompagne) est plutôt à l'intégration, et nombre d'EIAH proposent de scénariser *a priori* les activités pédagogiques au sein même du système (Koper *et al.*, 2003 ; Ferraris *et al.*, 2007). Dans ce cas précis donc, les traces numériques d'interaction sont mobilisées pour suivre le déroulement *effectif* du *scenario* lors d'une session réelle d'utilisation de l'EIAH, et de le comparer avec le déroulement *prévu* au départ (Marty *et al.*, 2004, 2007). Certains EIAH intègrent même directement cette possibilité en donnant accès aux logs de l'EIAH lui-même, comme c'est le cas par exemple avec *REDIM* (Choquet et Iskal, 2007). Les exemples de travaux sur l'évaluation des environnements numériques sont relativement nombreux, et leurs problématiques se voient renouvelées régulièrement, à l'occasion de changement technologiques ou d'usages. Il est assez difficile de trouver des travaux de synthèse couvrant sur ce domaine particulier. Le lecteur intéressé pourra toutefois se référer à un état de l'art riche proposé par D. Hilbert et D. Redmiles (2000).

³⁶ The 2006 Workshop on Computer Assisted Recording, Pre-Processing, and Analysis of User Interaction Data, HCI 2006 conference, London, UK, Tuesday 12th September 2006.

³⁷ Voir les normes SCORM / LOM par exemple.

2.2.3 Analyse des comportements d'utilisation et d'usage

Le deuxième objet visé par l'analyse des traces numériques est le *comportement* des utilisateurs. Il faut comprendre par ce terme que l'objectif est l'étude des *comportements d'utilisations et d'usages* des environnements dont on étudiait précédemment les interactions. Les travaux les plus avancés sur le sujet sont développés dans le domaine du *Web*, avec en particulier la communauté du « *Web Usage Mining* » (Srivastava *et al.*, 2000). Le domaine du *Web* comporte une particularité importante du point de vue des traces numériques : la navigation d'un Internaute engendre nativement un certain nombre d'informations stockées dans des logs de navigation qui seront largement et quasiment exclusivement exploités.

Au-delà de cette particularité, ce qui change est le fait qu'on ne parlera plus de traces d'interaction (ce qu'elles restent fondamentalement) mais directement de traces de *comportement*, de « *trace of behavior* » de l'utilisateur (Pirolli, 2002, p.3), et plus précisément du comportement au centre de l'attention de ces chercheurs c'est-à-dire le « *Web browsing behavior* » (Hawkey *et al.*, 2005). La notion de « comportement » ou de « comportement de navigation » reste un peu vague. Chercher à caractériser ou évaluer « dans l'absolu » ce comportement, n'aurait à vrai dire pas beaucoup de sens. Un comportement n'a de sens que relativement à une activité. En fait il semble qu'une majorité de chercheurs dans ce domaine, considère (plus ou moins explicitement) que l'activité « par défaut » est la *Recherche d'Information*³⁸. Parmi les nombreux travaux disponibles, nous avons choisi de présenter des exemples impliquant une analyse des traces d'interactions pour caractériser l'utilisation d'un navigateur d'une part, et pour caractériser des usages du *Web* d'autre part.

2.2.3.1 Caractériser l'utilisation du navigateur *Web*

La première façon d'entendre la caractérisation du « comportement » d'un internaute est d'analyser l'utilisation de son navigateur. C'est en tout cas de cette façon que bon nombre de chercheurs abordent la question. En se basant sur les logs de navigation produits par les navigateurs Hawkey *et al.* (2005) proposent par exemple une caractérisation du « comportement de navigation » fondée sur des mesures du nombre de pages visitées, du nombre de fenêtres de navigateur utilisées, de la vitesse de navigation, de la durée des sessions, *etc.* Ces données sont ensuite traitées et représentées graphiquement de façon à mettre en évidence des *patterns temporels* remarquables susceptibles d'être caractéristiques du « comportement » de navigation observé. Dans la figure ci-dessous (Figure 2.5) on peut voir par exemple comment, graphiquement, les auteurs présentent l'utilisation différenciée des fenêtres actives.

Toujours dans l'optique de caractériser l'utilisation du navigateur, Weinreich *et al.* (2006) proposent une des rares études longitudinales réalisées sur la « navigation *Web* » depuis l'apparition des navigateurs grand public, les précédentes étant celles de Catledge et Pitkow (1995), Tausher et Greenberg, (1997), puis de Cockburn et ses collègues (2002). Au cœur de ce travail, une observation longitudinale de 25 internautes durant une centaine de jours (en moyenne) s'est basé sur une instrumentation du navigateur *Firefox™ 1.0* permettant d'enregistrer l'ensemble des *logs* de navigation (côté client). Ces

³⁸ À tel point d'ailleurs que pour certains, comportement, navigation *Web* et Recherche d'Information ne font qu'un.

logs sont nettoyés pour ne comporter que des « *actions de navigation* », *i.e.* toute opération volontaire de l'utilisateur qui déclenche l'affichage d'une nouvelle page dans le navigateur³⁹.

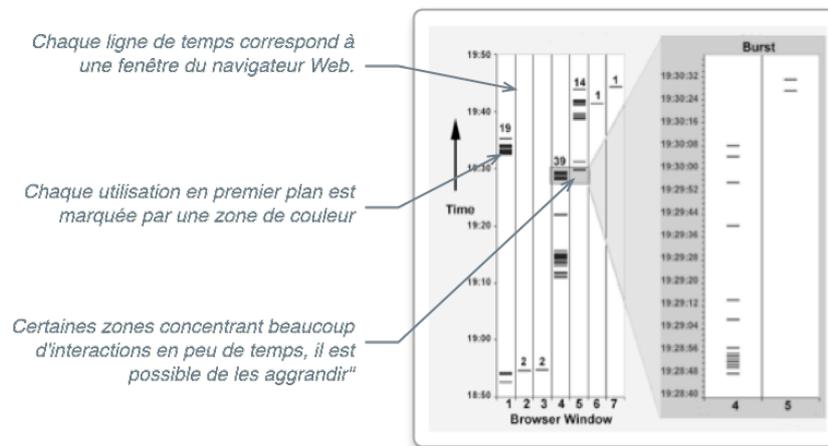


Figure 2.5 : Exemple de visualisation de l'utilisation des fenêtres actives d'un navigateur, permettant la mise en évidence de « *patterns of web browsing* » - Tiré de (Hawkey *et al.*, 2005 p.1444)

On notera au passage le fait, amusant, que 40% des pages affichées lors d'une navigation *ne relèvent pas* d'un acte *volontaire* de l'internaute mais principalement de pages publicitaires (Figure 2.6). Du point de vue des traces numériques de « navigation » cela pose évidemment un problème de bruit récurrent, qui exige un *filtrage* des logs récupérés. Sur les traces débarrassées de ce bruit Weinreich *et al.* (2006) appliquent de manière systématique des algorithmes de *dénombrements*, *statistiques*, *mesures de corrélation* entre des variables choisies dans les logs, *etc.*

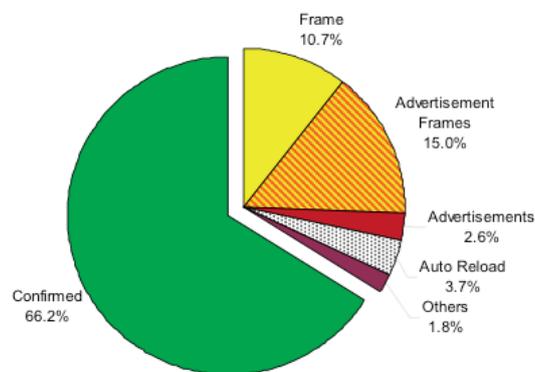


Figure 2.6 : Répartition des actions de navigation - Tiré de (Weinreich *et al.*, 2006, p.135)

Ces diverses mesures sont sensées *caractériser*, rappelons-le, le comportement de navigation. Prenons l'exemple de l'utilisation du bouton « retour »⁴⁰. Cette étude nous apprend qu'il n'est à l'origine que de 14,3% des actions de navigation, ce qui constitue d'ailleurs selon les chiffres avancés par les auteurs, une baisse significative par rapport aux études antérieures d'une dizaine d'année⁴¹. Pour tenter

³⁹ La définition d'une action de navigation devient très vite complexe, notamment à cause des changements automatiques dans les pages, obligeant les auteurs à exclure *frame*, *pop-up*, *etc.*

⁴⁰ Qui anime beaucoup le travail des chercheurs en conception de site *Web* depuis longtemps (Tauscher et Greenberg, 1997 ; Greenberg et Cockburn, 1999 ; Moyle et Cockburn, 2003 ; Milic-Frailing, 2004)

⁴¹ L'étude de Tauscher et Greenberg (1997) montrait que le bouton « retour » était une des actions de navigation principale dans les premiers temps d'internet, loin devant d'autres fonctionnalités comme l'historique.

d'expliquer ce résultat, les auteurs pointent une corrélation entre le niveau d'utilisation du bouton « retour » et la fréquence d'utilisation des *onglets* (fonctionnalité disponible dans *Firefox 1.0*), et des *nouvelles fenêtres*, qui offre la possibilité de naviguer « en parallèle » sur plusieurs pages, « délinéarisant » la navigation. Par conséquent le bouton « retour » est moins utilisé au profit d'un parcours entre les différentes fenêtres ou onglets ouverts (Kaasten et Greenberg, 2001). L'hypothèse défendue est alors que le comportement de l'utilisateur est de maintenir ouvertes, dans un onglet ou une fenêtre, les pages sur lesquels il pense revenir. Le même type de démarche a été affiné et appliqué à d'autres mesures comme le *temps passé* sur chaque page, que les auteurs cherchent à mettre en lien avec la « *scannability* »⁴² des pages en question.

2.2.3.2 Caractériser des parcours de navigation *Web* (usages)

Une autre façon de voir le comportement de navigation est de considérer les parcours des internautes sur le *Web* en général ou sur un domaine en particulier. Dans ce cas, il s'agit de ne plus se limiter aux opérations de manipulation de l'outil comme dans le cas précédent, mais de tenir compte des contenus qui sont parcourus. D'un point de vue technique pour les traces, cela signifie (en général) un changement de source des *logs* de navigation, un passage du *client* au *serveur*, pour récupérer les données d'usage, *i.e.* « *data that describes the pattern of usage of Web pages, such as IP address, pages references, and the date and time of accesses* » (Srivastava *et al.*, 2000, p.13). C'est le cas dans la thèse de T. Beauvisage (2004) sur la « sémantique des parcours *Web* » où le dispositif de traçage est constitué d'une sonde placée sur les machines d'internautes volontaires donnant accès aux *logs* de navigation tels qu'ils sont échangés entre client et serveur.

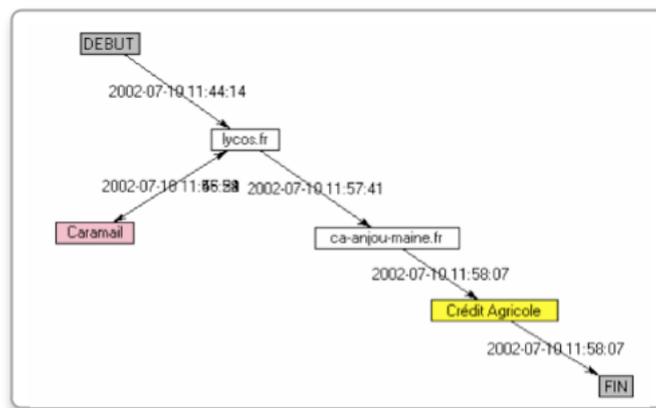


Figure 2.7 : Visualisation sous forme de graphe de parcours *Web*.
Tiré de (Beauvisage, 2004, p.144).

L'auteur se fixe pour objectif de « décrire et visualiser la dynamique des parcours » en définissant lui-même une topologie originale (Beauvisage, 2004, p.165). Après la toujours indispensable phase de *filtrage* des *logs* de navigation, des traitements statistiques simples sur la *forme*, la *temporalité* et le *rythme* des actions de navigation ont été utilisés pour « reconstituer » des parcours. L'originalité vient ici d'un *enrichissement a posteriori des parcours*, enrichissement qui consiste à *typer les URLs* en fonction d'une classification similaire à celle proposée par les annuaires *Web* (*Catservices* en

⁴² Que l'on pourrait traduire par « propension d'une page à être parcourue rapidement du regard ».

l'occurrence). L'analyse des parcours proprement dite, qui va donc *tenir compte du contenu* est ensuite effectuée à l'aide de techniques de visualisation graphique (Figure 2.7).

De nombreux travaux d'analyse des usages du *Web*, ou d'un site *Web*, s'appuient sur ce genre de technique de visualisation (Benford *et al.*, 1999 ; Cugini et Scholtz, 1999 ; Cadez *et al.*, 2000 ; Chi, 2002 ; Chen *et al.*, 2004). En mettant au point la bonne combinaison entre statistique et visualisation appliquée les chercheurs tentent de faire émerger les points critiques du site étudié : partie isolée, peu ou pas assez fréquentée, *etc.*). Une fois les parcours ou les « cheminement » les plus prégnants mis en évidence, la structure du site peut être modifiée en fonction des objectifs propres à chaque *Webmaster*. Dans le travail de F. Rossi *et al.* (2005)⁴³, les auteurs proposent une méthode de visualisation des parcours qui se *dégage de la structure du site* lui même (graphe des liens hypertextes). L'idée consiste à opérer des *rapprochements* et des *mis en relations* de pages qui ne sont ni proches, ni en relation dans l'architecture du site, mais qui sont « proches » dans les parcours des internautes.

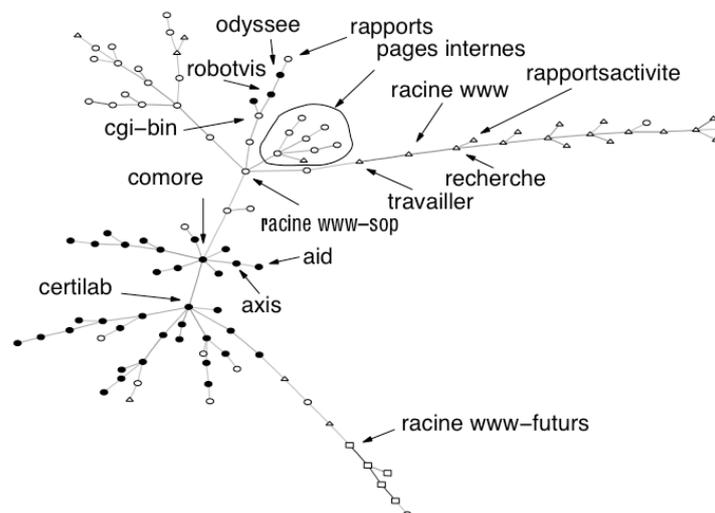


Figure 2.8 : Arbre couvrant minimal du site de l'INRIA - Tiré de (Rossi *et al.*, 2005, p.573).

Les données récupérés dans ce cas sont les *logs serveur* d'un site (celui de l'INRIA⁴⁴ en l'occurrence), donnant une séquence ordonnée des pages demandées au serveur par un utilisateur⁴⁵ qui constitue au final le « cheminement » de ce dernier. Chaque suite de requête devient un vecteur (d'URLs). L'ensemble des vecteurs permet de construire un espace (l'espace des navigations), dans lequel on peut finalement calculer des dissimilarités entre des groupes d'URLs, et donc recréer des « distances » entre des parties du site qui ne sont plus calculées à partir de la structure du site. Le résultat est finalement visualisé à l'aide d'un arbre couvrant minimal (Figure 2.8). Les auteurs affirment ainsi permettre de « *confronter la vision éditoriale avec celle des internautes et comprendre les modes d'utilisations du site* » (Rossi *et al.*, 2005, p.574).

De façon générale la caractérisation de l'utilisation d'un navigateur ou de l'usage du *Web* issue de l'analyse des traces d'interactions n'est pas produite pour elle-même, et est utilisée à d'autres fins : prédire automatiquement les actions de l'utilisateur par exemple (Lettkeman *et al.*, 2006), ou bien

⁴³ F. Rossi a publié d'autres travaux similaires dont les références sont disponibles en ligne sur : <http://apiacooa.free.fr> (consulté le 07/03/09).

⁴⁴ Institut National de la Recherche en Informatique et en Automatique (<http://www.inria.fr/>).

⁴⁵ Ne sont conservés que des suites de requêtes envoyées par un client identifiable, avec pour seule contrainte de ne pas laisser d'écart de temps de plus de 30 minutes entre deux requêtes (seuil au-delà duquel les auteurs considèrent que l'utilisateur a interrompu sa navigation).

compléter l'évaluation de l'interface *Web* d'un site (cf. section 2.2.2). Du côté des traces d'interaction, les *logs* de navigation qui les constituent dans les exemples précédents, restent difficilement exploitables sans autre information, ce que reconnaissent les auteurs eux-mêmes : « *This makes their contextual interpretation inherently difficult and additional qualitative information is needed to support a detailed task-related evaluation of the data* » (Weinreich, 2006, p.141). Il est à ce sujet intéressant de voir que lorsqu'on demande à un internaute la raison de son utilisation des nouvelles fenêtres ou onglets, la réponse est la suivante : « *new tabs for closely related tasks and new Windows for parallel tasks* », ce qui relève bien de l'activité en situation, constitue un comportement, mais ne peut être déduit à partir des seuls *logs* d'utilisation de l'outil. Effectivement il y a là une limite à ce genre de démarche, « *aims and tasks of the users often stay below the surface* » (Weinreich, 2006, p.141). C'est pourtant ce niveau que d'autres chercheurs se donnent comme objectif d'atteindre avec l'analyse de traces, pour caractériser ou évaluer les *utilisateurs* et leurs *activités*.

2.2.4 Analyse des utilisateurs et de leurs activités

Le troisième niveau de caractérisation et d'évaluation sur la base de l'analyse de traces numériques d'interactions que nous considérons est celui des utilisateurs et de leurs activités. Concernant les utilisateurs, il s'agit d'approches qui travaillent sur des *modèles d'utilisateur*, *cognitifs* et *comportementaux*, ou de manière moins formelle sur des *profils utilisateurs*. Concernant les activités de ces utilisateurs, il s'agit de travaux se fondant sur l'analyse des traces pour *qualifier*, voire *évaluer* relativement à des critères de performance, la *réalisation d'activités* par un ou plusieurs utilisateurs. Contrairement aux travaux précédents, l'activité en question est nécessairement identifiée.

2.2.4.1 Modèles cognitifs et comportementaux d'utilisateurs

Les modèles cognitifs et comportementaux sont des concepts issus de la Psychologie et des Sciences Cognitives, mais sont régulièrement utilisés dans d'autres domaines de recherches, notamment IHM et CSCW. En principe ces modèles rendent compte des processus mentaux ou comportementaux d'un individu en donnant une représentation formalisée de leur architecture et de leur fonctionnement. Ces modèles peuvent être parfaitement descriptifs, mais dans le cadre de l'informatique il s'agit plutôt de modèles opérationnels, permettant d'expliquer ou de qualifier l'activité de l'utilisateur, voire de prédire celle-ci (OBrien et *al.*, 2006). Pour mettre au point ou valider ces modèles les chercheurs ont besoin de données d'observation de l'activité des utilisateurs obtenues par observation directe, ou bien – ce qui nous intéresse ici – par observation indirecte, c'est-à-dire à partir de traces d'interaction. Les traces d'interactions sont utilisées pour mettre à l'épreuve des modèles théoriques ou pour les affiner, ou même pour en définir de nouveaux (Halvey et *al.*, 2005). Les chercheurs adoptent pour cela une démarche itérative qui consiste à faire des allers-retours entre le modèle et les données réelles. Une formalisation en a été proposée, par Ritter et Latkin, (1994), sous la forme du « *Task-Based Protocol Analysis* ». Ce protocole repose sur les trois étapes (Ritter et Latkin, 1994, p.346) :

- (a) *using an initial model's sequential predictions (a trace) to find structure in the sequential data,*
- (b) *summarizing the model's successes and failures in accounting for the data, and*
- (c) *using the results this summary to improve the model in an iterative cycle.*

Dans le travail présenté par (Pirolli *et al.*, 2002) la modélisation de l'utilisateur cumule un modèle cognitif *et* un modèle comportemental, le tout pour tenter de simuler un utilisateur navigant sur le *Web*. La mise au point des modèles repose principalement sur l'analyse des traces d'interaction ou plutôt de navigation (dans une tâche prédéfinie plutôt simpliste). Les *logs* de navigation ne sont cette fois pas la seule source de traçage : les *logs* sont en effet synchronisés et indexés avec un enregistrement vidéo et celui d'un *eye-tracker*, permettant de situer à tout instant le point de fixation du regard d'un utilisateur sur l'écran. Au final, la trace exploitée sera constituée « *d'actions de navigation* », qu'un *analyste* aura dû *transcrire à la main* en respectant un « *Protocol Transcript* ». Le document final prend la forme d'un tableau listant selon les colonnes, les différents évènements en fonction de la source de traçage (Figure 2.9).

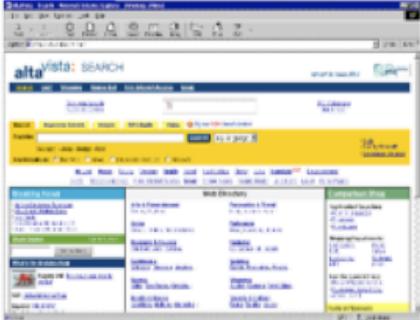
	SYSTEM	OBSERVED ACTIONS AND TRANSCRIPT	MODEL INTERPRETATION
4:25:00		(Task #2:ANTZ) (Reads Question) Ok, Um.	(O*READ-QUESTION)
		Let's go to Altavista because I like Altavista.	(G*GO-TO SITE "Altavista")
4:49:22		Click: www.altavista.com in pull-down menu. Um, and let's look up Antz. And see what's generally available.	(O*GO-TO SITE "www.altavista.com") (G*SEARCH WEB "www.altavista.com "Antz")
		Type search: Antz	(O*SEARCH WEB "www.altavista.com" "Antz")
4:55:00	http://www.altavista.com/cgi-bin/query?psqp&sc=on&hl=on&q=antz&kl=XX&stype=stxt	"The official Antz web ring," that sounds appropriate.	(O*EVAL LINK "The official Antz web ring" hi)

Figure 2.9 : *Protocol transcript* - Tiré de (Pirolli, 2002, p.5).

Les auteurs proposent de créer un modèle utilisateur fondé *théoriquement* sur (a) un *modèle cognitif* rendant compte de phénomènes perceptifs (*SNIF-ACT*) et (b) un *modèle comportemental* de recherche d'information (*Information Scent*). Ces modèles sont implémentés de façon à créer un « simulateur d'internaute ». Une fois ce simulateur en place, la démarche consiste à utiliser les traces d'interactions de navigation *effective* pour affiner les modèles. Celles-ci permettent d'évaluer la pertinence du simulateur, relativement à un utilisateur réel qui navigue sur les mêmes pages pour réaliser une même tâche⁴⁶, le but étant de donner un modèle utilisateur le « plus proche possible de la réalité »⁴⁷. En ce qui concerne spécifiquement le rôle des traces numériques dans ce travail, nous retiendrons qu'elles sont ici exploitées soit comme *source d'information* pour *caractériser* le comportement d'un utilisateur, soit comme *élément de comparaison* permettant d'évaluer le modèle utilisateur.

⁴⁶ Les tâches en questions sont un peu restrictives pour qu'on puisse envisager que l'on parle ici de navigation dans l'absolu.

⁴⁷ Dans ce travail, les auteurs vont bien au-delà de l'amélioration d'un modèle, cognitif ou comportemental : une fois les ajustements réalisés, le même dispositif sera utilisé comme un « simulateur d'utilisateur ».

Les activités de navigation ou de recherche d'information sur le *Web* ne sont pas les seules activités cognitives que les chercheurs tentent de modéliser. Dans le travail de M. Rauterberg et M. Fjeld (1998) par exemple, c'est dans le contexte d'une activité de *résolution de problème* que des traces numériques d'interaction sont mobilisées. Défendant l'originalité de leur approche qui « *suggest automatic logging of behavioural data* » (Rauterberg et Fjeld, 1998, p.153), les auteurs proposent de générer des *modèles de tâches réalisées* à l'aide de *résumés* de traces numériques d'interaction. L'objectif est de reconstituer, à partir de ces traces, le « *mental task model* » d'un utilisateur lors d'une activité. Ici, la reconstruction du modèle en question est réalisée à l'aide d'un *réseau de Pétri*⁴⁸ qui permet de rassembler de manière cohérente l'ensemble des éléments collectés (Figure 2.10). Nous pourrions multiplier les exemples si nous sortions du cadre des activités directement instrumentée par des artefacts informatiques. C'est par exemple un modèle cognitif de conduite automobile qui a motivé le travail de O. Georgeon (Georgeon *et al.*, 2006a ; 2006b ; Georgeon, 2008) et du système *Abstract (Analysis of Behaviour and Situation for menTal Representation Assessment and Cognitive acTivity modelling)* dont nous reparlerons (Chapitre 3).

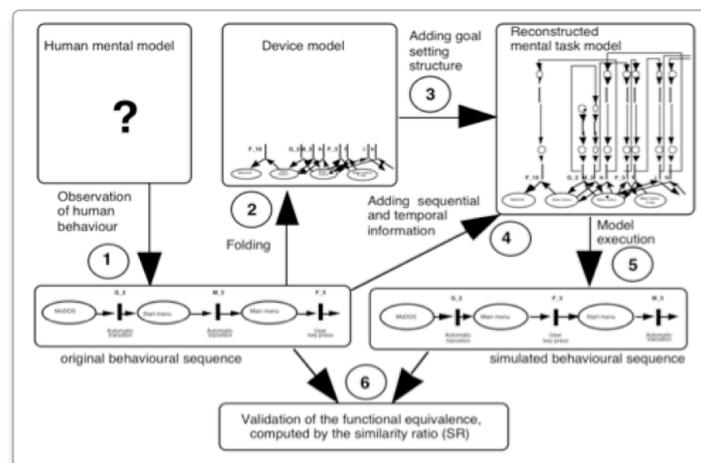


Figure 2.10 : Architecture en six étapes de l'approche de modélisation de comportement proposée par M. Rauterberg et M. Fjeld - Tiré de (Rauterberg et Fjeld, 1998, p.155)

2.2.4.2 Profils utilisateurs

Des approches plus pragmatiques que les précédentes, dont l'ambition n'est pas forcément de caractériser précisément des modèles cognitifs de l'utilisateur proposent d'en établir des *profils*. Partant de l'idée qu'on peut caractériser un *comportement* à partir de traces numériques d'interaction, des *comportements types* sont déterminés et les utilisateurs catégorisés en fonction de leur type de comportement. En dehors des travaux de psychologie cognitive sur le sujet (Chevalier et Ciaccia, 2009), cette pratique se développe principalement dans le *Web Usage Mining* (Brinkman *et al.*, 2006 ; Grčar, 2004) où de nombreux travaux se consacrent au « profilage » les internautes⁴⁹. La demande est souvent portée par les besoins des sites marchands pour le ciblage publicitaire. Dans ce cas la limite entre « *User*

⁴⁸ Composé de *places* et de *transitions*, un *réseau de Petri* est un réseau *causal*, qui évolue en fonction de son *marquage*, i.e. des valeurs qui sont attribuées à chaque place, les transitions faisant varier ces valeurs.

⁴⁹ Voir les « *Workshop on Web Usage Analysis and User Profiling* » ou la revue *Web Usage Analysis and User Profiling* (Springer).

Profile » et « *Customer profile* » a tendance à s'estomper⁵⁰. Dans le domaine des EIAH il est également parfois question de « profil d'apprenant », et il s'agit dans ce cas plutôt de déterminer le *profil psychologique* de celui-ci, plutôt que de le classer dans un profil type d'utilisateur. En fait la caractérisation (comme l'évaluation) des utilisateurs se fait à travers la caractérisation (ou évaluation) de leurs activités, ce qui nous amène au point suivant.

2.2.4.3 Caractériser et évaluer une activité

Parler de traces d'activité, alors que nous n'avons jusqu'ici parlé que d'interaction, d'utilisation ou d'usages, implique un saut qualitatif de la démarche de recherche : on ne vise plus uniquement à déterminer des régularités ou des caractéristiques remarquables dans leur répétition ou leur régularité d'apparition, mais on cherche à déterminer la *qualité singulière* d'une activité donnée, que cela soit pour l'évaluer ou la caractériser. Dans cet exercice, la seule différence fondamentale qui distingue l'évaluation de la caractérisation d'une activité est l'existence de critères précis, de valeurs de références, d'un *modèle de performance*. Dans le cas général d'une activité complexe comme celle qui nous intéresse il est difficile, voire impossible, de fixer totalement *a priori* ce modèle de performance. Les activités pédagogiques constituent toutefois, comme nous allons le voir, une exception (section 3.2.5.1).

Pour *évaluer* une activité, il faut disposer d'un modèle de performance auquel comparer ladite activité. Or, il est très difficile de disposer *a priori* de tels point de référence pour des activités numériques complexes, comme la recherche d'information sur le *Web* par exemple (Iqbal et Bailey, 2007). Même dans le cas où on admet un modèle de référence simpliste, comme par exemple le fait que l'utilisateur est performant lorsqu'il « *maximize information gained per unit of time* » (Pirolli, 2002, p.1), il reste une grande part de subjectivité dans l'évaluation de la « quantité d'information » recueillie, et dans ce qu'elle est censée représenter. En l'absence de tels critères, et pour une activité en général, on cherchera la mise en place d'*indicateurs* (section 3.2.5.2) permettant la caractérisation de l'activité et qu'il appartiendra à l'analyste d'interpréter. Avant de reparler des indicateurs, notons qu'il existe en principe une exception, c'est-à-dire des activités dont on a en règle général un modèle de performance : les activités pédagogiques dans les EIAH.

De telles activités constituent une exception car les critères d'évaluation peuvent être définis en fonction des objectifs pédagogiques visés par le concepteur de l'activité, et ils sont définissables *a priori*. Ceci ouvre la voie à la conception de systèmes d'évaluation automatique gérée par les EIAH eux-mêmes. Ceux-ci ne vont cependant guère plus loin que l'enregistrement des scores aux exercices ou aux tests inclus dans l'activité pédagogique. Pour aller plus loin dans l'évaluation de l'activité seul un enseignant-tuteur directement présent au sein de l'EIAH est à même de réaliser une évaluation qualitative. Dans ce cas, les traces des apprenants sont plutôt taitées de façon à être présentée et analysées par l'enseignant lui-même : « *The capacity to easily record activities could also benefit learners who use computers. An entire session could be captured, so that peers or mentors could analyze a student's work* » (Plaisant *et al.*, 1999). Dans les EIAH instrumentés à cet effet le tuteur peut accéder de différentes façons aux traces de l'activité des apprenants.

⁵⁰ *User profile* est défini pour l'occasion comme « *Data that provides demographic information about users of the Web site. This includes registration data and Customer profile information* » (Srivastava *et al.*, 2000, p.13)

Le premier est l'utilisation de dispositifs traçants permettant le « rejeu » complet de la session d'un apprenant, comme c'est le cas dans des EIAH comme pour *DREW*⁵¹ (Baker *et al.*, 2003), pour *Combien?*⁵² (Le Calvez *et al.*, 2003) ou encore pour *SimPLE*⁵³ (Plaisant *et al.*, 1999) ou *ColAT* (Avouris *et al.*, 2007). L'avantage pour le tuteur est évidemment d'accéder à distance à une information très riche sur ce qui s'est passé (qu'il n'aurait peut-être même pas eu en présentiel), comme les hésitations, les étapes par lesquels l'apprenant passe, les outils qu'il mobilise à un instant donné, *etc.*

Le second consiste à créer des moyens de *visualiser le parcours* de l'apprenant dans l'activité pédagogique qui lui est proposée. Bien entendu un tel dispositif ne peut être mis en place qu'à la condition que l'environnement et l'activité s'y prêtent, *i.e.* que les « étapes » qui constituent le parcours soient *a priori* définies et représentables en tant que telles. C'est ce que proposent par exemple des outils comme *REFLET* (Desprès *et al.*, 2004) ou *FORMID* (Guéraud *et al.*, 2004) qui affichent la progression du parcours des apprenants. C'est également, dans une certaine mesure le cas d'outils comme *Collablogger* (Morse et Steves, 2000) ou *VLE* (Hardy *et al.*, 2004) qui donnent accès à la progression des apprenants, ou encore des outils comme *CourseVis* (Mazza et Dimitrova, 2003) qui proposent des fonctionnalités de *visualisation* des actions d'apprenants lors de l'utilisation de la plateforme d'apprentissage *WebCT*, ou plus récemment avec le système *GISMO* avec *Moodle*⁵⁴ (Mazza et Milani, 2005).

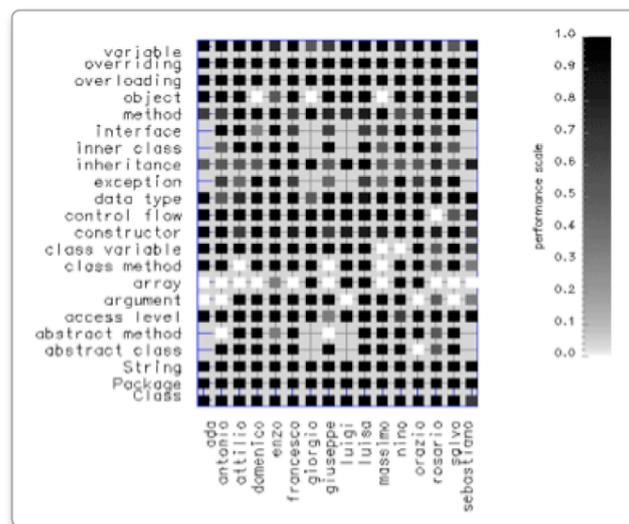


Figure 2.11 : Exemple de visualisation proposée par *CourseVis* : il s'agit ici de montrer pour chaque élève quels sont les concepts maîtrisés ou simplement abordés dans le cours de leur activité⁵⁵.

Même si certaines activités pédagogiques sont entièrement déterminables *a priori*, cela ne constitue pas la totalité des cas, loin de là. On se retrouve donc bien souvent dans le cas général, à savoir une activité que l'on souhaite évaluer, ou caractériser, par exemple en tant que tuteur dans un EIAH, sans avoir le temps d'accéder directement à un enregistrement complet de l'activité, ni disposer d'un outil permettant la reconstitution d'un parcours adéquat. Le souci d'instrumentation des tuteurs dans ce

⁽⁵¹⁾ *Dialogical Reasoning Educational Web tool*, pour plus d'information voir <http://lead.emse.fr/>

⁽⁵²⁾ Pour plus d'information voir <http://combien.lip6.fr/>

⁽⁵³⁾ *Simulated Process in a Learning Environment*.

⁽⁵⁴⁾ Pour *WebCT* voir : <http://www.blackboard.com/>, pour *Moodle* voir : <http://moodle.org/> (consulté le 02/05/2009).

⁽⁵⁵⁾ Exemple tiré d'une présentation en ligne : <http://www.comp.leeds.ac.uk/vania/umuas/coursevis.html> (consulté le 02/05/2009).

cadre est très présent dans le domaine des EIAH, comme le montre par exemple le travail présenté dans (Laperroussaz, 2006) autour du système *TACSI* dont c'est explicitement le but. La solution privilégiée, et très largement présente également dans la littérature est donc celle des *indicateurs*.

Caractériser une activité avec des indicateurs

Les *indicateurs* sont des variables, au sens mathématique, auxquelles sont attribuées une série de caractéristiques (Bratitsis et Dimitracopoulou, 2006). Les indicateurs auxquels nous nous intéressons ici sont calculés sur les traces numériques générées par l'utilisation d'un système informatique. Ces indicateurs peuvent dépendre d'autres variables tout comme ils peuvent eux-mêmes être à leur tour utilisés dans le calcul d'autres indicateurs. Leur objectif est de réaliser en quelque sorte, une partie du traitement exhaustif des traces, que ne peut faire un analyste ou un tuteur. Ce dernier se voit mettre à disposition des indicateurs dont il doit être capable de tirer une caractérisation de l'activité qu'il analyse. Encore une fois, c'est très largement dans le domaine des EIAH que se trouvent les travaux les plus riches et les plus nombreux sur le sujet.

Dans le travail de (René, 2000), des indicateurs sont issus de statistiques de dénombrement plus ou moins simples : nombre d'accès à une ressource en ligne, nombre d'essais pour chaque réponse, temps passé pour la réalisation de chaque opération, *etc.* L'environnement *Combien?*⁵⁶ est un autre exemple. Celui-ci fournit des indicateurs statistiques sur le temps mis pour trouver une solution, les exercices achevés, le nombre d'erreurs (Le Calvez, 2003). L'environnement d'apprentissage *ESSAIM* calcule et fournit à l'observateur le temps passé par un apprenant sur chaque étape de l'activité pédagogique (Desprès et Leroux, 2003). Et dans *Virtuoso* c'est par courriel que les « scores » des apprenants, toujours calculés sur le même type d'information, sont communiqués au tuteur concerné à la fin d'une session pédagogique (Stephanov et Stephanova, 2005).

De tels indicateurs montrent toutefois leurs limites quant à ce qu'ils permettent de dire d'une activité. Dans le cas des EIAH notamment, la répétition des actions et/ou le pourcentage de réussite aux exercices automatiques ne donnent pas nécessairement d'information sur la façon d'apprendre (Lomic-ka, 1998), information qui *in fine* est celle que recherche un tuteur pour juger de l'activité de son apprenant. En effet, ce que doivent « indiquer » les indicateurs à un observateur extérieur comme le tuteur, ce sont les moyens de comprendre les stratégies de l'apprenant, ses difficultés et leur raisons, de déterminer ce qu'il a réellement compris et est capable de mobiliser à nouveau.

Le domaine des EIAH n'est pas tout à fait le seul à utiliser des indicateurs pour caractériser une activité pédagogique dont il récupère les traces numériques. L'activité de recherche d'information⁵⁷ sur le *Web* fait également l'objet de ce type de caractérisation. Nous ne pouvons ici présenter dans le détail tous les exemples, mais concernant spécifiquement la recherche d'information, il est possible d'en trouver par exemple dans l'état de l'art de (Jansen et Pooch, 2000). Il est cependant un point sur lequel nous allons nous arrêter un instant : la *caractérisation des activités collectives*.

⁵⁶ Voir : <http://combien.lip6.fr/> (consulté le 02/05/2009).

⁵⁷ En tant que telle et non comme l'activité « par défaut » de tout internaute.

Caractériser des activités collectives

Les activités collectives ayant fait une entrée en force dans les situations de travail informatisées, (tout particulièrement dans le cadre des EIAH où l'apprentissage *collaboratif* constitue une véritable pédagogie), la question de leur caractérisation et de leur évaluation a bien entendu été abordée par les chercheurs depuis plusieurs années. Dès le début de l'utilisation des traces numériques d'interaction, on note un certain flou entre ceux qui adoptent une démarche similaire aux précédentes et ceux qui soulignent qu'il s'agit d'une caractérisation trop particulière qui nécessite un autre point de vue. D'un côté donc des chercheurs en *ESDA* (section 2.2.1.1) ont considéré qu'il s'agissait de tracer des activités collectives et ont choisi de les aborder de la même manière que les activités individuelles. Ainsi, le travail de Vortac *et al.* (1994) dans lequel était proposé de caractériser l'activité de contrôleurs aériens à partir de l'analyse des *logs* du système qu'ils utilisaient. Sur les traces d'interactions récupérées étaient réalisés des traitements statistiques permettant de caractériser douze « comportements » dont onze étaient des comportements relevant d'une utilisation individuelle du système, la douzième relevant du « comportement d'équipe ». D'un autre côté, on notait dans (Sanderson et Fisher, 1994) la nécessité de procéder différemment ou pour le moins avec des bases méthodologiques différentes.

De manière générale, la mobilisation des traces numériques d'interactions pour caractériser ou évaluer l'activité collective a pris place dans des démarches similaires à celles que nous avons évoqué jusqu'ici. On relève, encore une fois dans le domaine des EIAH principalement, deux types de mobilisations des traces numériques. De la même manière que pour les activités individuelles, certains outils vont fournir à un tuteur les moyens de suivre la progression de plusieurs apprenants en parallèle, par exemple *TACSI* (Laperroussaz, 2006), *Collablogger* (Morse et Steves, 2000), *SimPLE*⁵⁸ (Plaisant *et al.*, 1999) ou *ColAT* (Avouris *et al.*, 2007). Un aspect plus intéressant est la manière dont les chercheurs utilisent les traces numériques comme support d'analyse pour la caractérisation des activités *collaboratives*, en utilisant à nouveau des *indicateurs*.

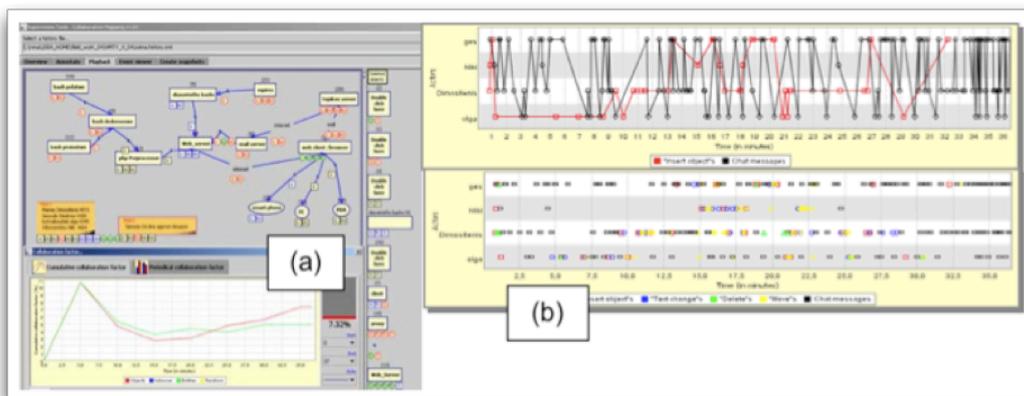


Figure 2.12 : Visualisation (a) d'un indicateur de collaboration et (b) de l'évolution du degré de participation dans *Synergo* - Tiré de (Avouris *et al.*, 2004, p.17).

Les indicateurs de la (bonne ou mauvaise) collaboration au sein d'un groupe sont généralement basés sur des données relativement simples, portant sur les dimensions *communicative* et *participative* supportées par l'environnement numérique impliqué. L'environnement d'apprentissage *ESSAIM* par

⁵⁸ *Simulated Process in a Learning Environment*.

exemple, propose au tuteur un indicateur correspondant à un « *degré d'interaction* » de ses apprenants (Desprès *et al.*, 2001). L'outil *Synergo*⁵⁹ (Avouris *et al.*, 2004) permet de visualiser (Figure 2.12) le *nombre* de messages émis considéré comme « *facteur de collaboration* » (*collaboration factor*), et le *nombre* d'objets manipulés considéré comme le « *degré de participation* » (*actor activity*).

Des indicateurs de collaboration plus ou moins similaires sont présentés dans quelques états de l'art dont celui de (Dimitracopoulou, 2005) et de (Gogoulou *et al.*, 2005). Dans les exemples que nous venons d'évoquer il s'agit plutôt d'indicateurs quantitatifs basés sur la « *participation* » des utilisateurs membres du collectif de travail, cette participation étant entendue comme le volume plus ou moins grand d'interventions dans les espaces de communication des systèmes utilisés. Pour aller plus loin que cette volumétrie, des travaux ont tenté de rentrer plus en détail dans l'analyse des traces de communication pour tenter de mieux caractériser la collaboration au sein d'un collectif. Les chercheurs se sont pour cela associés à des spécialistes de *Communication Médiatisée par Ordinateur* (CMO) pour tracer l'utilisation de *messaging*, de *forums* ou de *chats* (May *et al.*, 2008a, 2008b ; Anjewierden et Efimova, 2006).

Les indicateurs de collaboration, tels que ceux utilisés dans les travaux que nous venons de présenter sont avant tout des mesures *quantitatives* obtenues par des traitements systématiques des traces numériques. On peut douter de la capacité d'une approche qualitative à caractériser à elle seule la richesse et la singularité d'une activité humaine, d'autant plus lorsque celle-ci implique la collaboration de plusieurs utilisateurs (Soller *et al.*, 2005).



Figure 2.13 : Interface *Replay tool* - Tiré de (Rouncefield *et al.*, 2006, p.38)

En d'autres termes, il est difficile de caractériser, de manière systématique, une *collaboration* qui n'existe pas en soi mais à travers la *réalisation* d'une activité donnée. Cette particularité avait été remarquée par Sanderson et Fisher en 1994, lesquels notaient déjà que si de nombreux travaux existent sur l'analyse de « *données séquentielles* » issues d'activités collaboratives, celles-ci y étaient abordées

⁵⁹ Pour plus d'information voir : <http://hci.ece.upatras.gr/synergo/synergo.php> (consulté le 12/02/2009)

de manière sensiblement différente, par une démarche *qualitative* d'analyse. Pour l'occasion, ce sont des chercheurs d'autres domaines qui s'y employaient, des chercheurs en Sciences Sociales et particulièrement en ethnologie. Ainsi le travail présenté dans (Rouncefield *et al.*, 2006) autour de leur outil *Replay Tool* (Figure 2.13), dont ils expliquent l'intérêt qu'il représente pour une étude qualitative des activités collectives, tout particulièrement pour étudier les « phénomènes de collaboration ». Ces chercheurs en Sciences Sociales comme certains de leurs prédécesseurs (Woodruff *et al.*, 2002 ; Brown *et al.*, 2003), réclament même le développement de cette approche et espèrent des futurs systèmes qu'ils tracent « *new forms of data* » plus à même de se prêter aux *raisonnements analytiques* (Rouncefield *et al.*, 2006).

2.2.5 Bilan

Nous venons de parcourir un panorama des différentes finalités réservées aux traces numériques d'interaction lorsque celles-ci servent de *support d'analyse* pour un observateur extérieur à l'activité. Qu'elles prennent place dans des processus de conception ou de re-conception d'artefacts informatiques, pour des activités individuelles aussi bien que collectives, l'analyse de ces traces repose en grande partie sur le traitement quantitatif de grands volumes de données. Les résultats de ces analyses n'ont en principe pas d'impact direct sur l'activité, et les traces n'y sont pas nécessairement présentes pour le ou les utilisateur(s) du système. Comme nous allons le voir dans la section à suivre, la réutilisation des traces numériques d'interactions au sein même d'une activité est un sujet qui suscite de nombreux travaux de recherche.

2.3 Les Traces Numériques d'Interaction comme support à l'activité

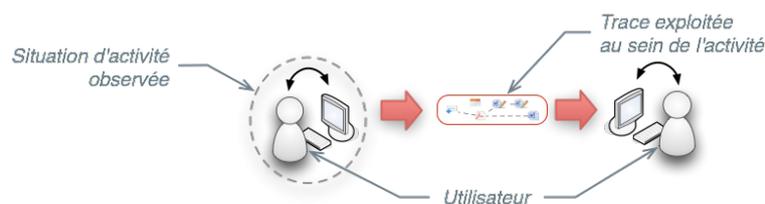


Figure 2.14 : Exploitation des traces numériques d'interaction au sein même de l'activité.

La seconde partie de cet état de l'art sur les systèmes traçants présente des travaux de la littérature dans lesquels des traces numériques d'interaction sont exploitées pour *modifier la situation* d'activité des utilisateurs, en y étant *réintroduites* sous une forme ou une autre. Présentées explicitement ou non aux utilisateurs, cette réintroduction des traces dans l'activité a pour objectif général d'en *faciliter* la réalisation : les traces numériques d'interaction deviennent ainsi, plus ou moins directement, un *support à l'activité* (Figure 2.14). Comme nous allons le voir, certaines des démarches présentées ici se situent dans le prolongement naturel des travaux d'analyse évoqués dans la première partie de l'état de l'art. Aux problèmes précédemment évoqués de *source* de traçage et de *traitements* des traces d'interaction, s'ajoutent pour les systèmes traçants que nous allons décrire le problème d'*intégration* des traces à la situation d'activité, et celui de leur *présentation* aux utilisateurs (celle-ci ne devant pas

alourdir l'interaction, ni gêner la réalisation de la tâche qu'elle est censée faciliter). Comme pour les travaux concernant les traces numériques comme support d'analyse, le travail bibliographique n'est pas simplifié par la diversité des domaines de recherche impliqués, et est même compliqué par l'utilisation de termes clefs très différents désignant, ou sous-entendant l'existence, des systèmes traçants, notamment en anglais : « *feedback* » de « *reuse* » ou « *reusing* », « *trail* », « *reflective* », *etc.*

Quoi qu'il en soit, nous articulons cette section en suivant le même principe d'organisation que la précédente, en présentant les travaux autour de *trois grandes finalités* (Figure 2.15). Celles-ci seront traitées de manière ordonnée, en fonction du caractère implicite ou explicite de la réutilisation des traces du point de vue de l'utilisateur : la première est *l'adaptation automatique* de l'environnement (section 2.3.1), la seconde *l'enrichissement* de l'interaction par *contextualisation* des objets manipulés ou par la *représentation explicite* de l'activité (section 2.3.2), et la troisième *l'assistance* à l'utilisateur dans son activité (section 2.3.4). C'est dans cette dernière section que nous ferons quelques remarques spécifiques relatives à l'assistance basée sur les traces dans le domaine des EIAH.

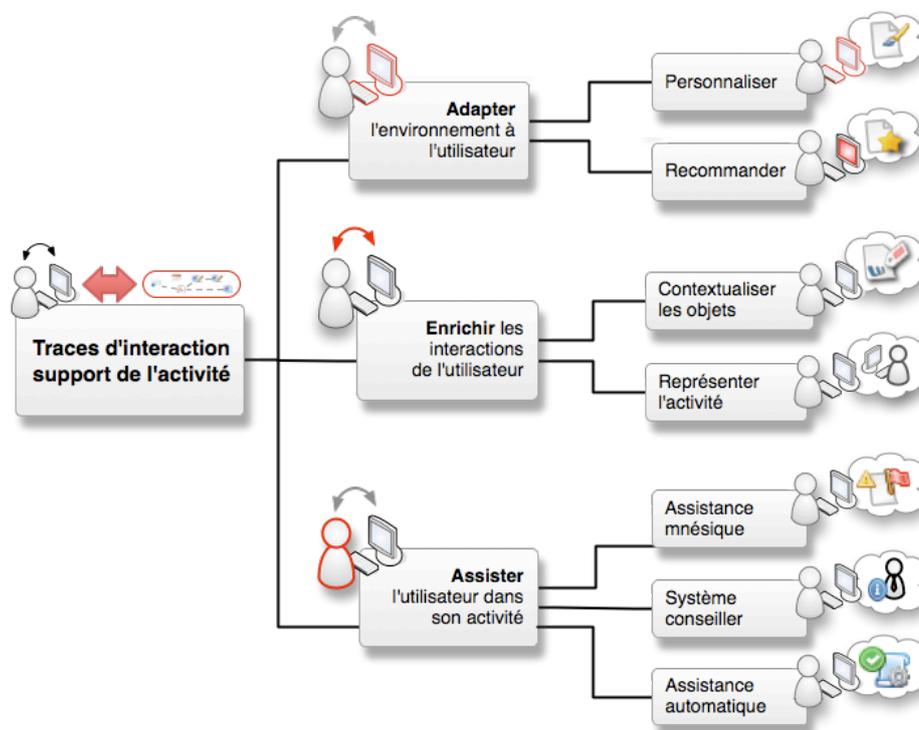


Figure 2.15 : Finalités des traces numériques utilisées comme support à l'activité.

2.3.1 Adapter automatiquement l'environnement numérique

L'impact le plus direct que les traces numériques puissent avoir, est d'être à l'origine de *l'adaptation automatique* de l'environnement à sa propre utilisation. Comme précédemment, nous entendons par environnement, à la fois les contenus et les outils pour les manipuler. Nous n'aborderons pas ici la question des applications auto-adaptatives (Bourguin et Derycke, 2005) que nous pouvons par ailleurs rapprocher de la question générale de la *réingénierie des environnements* (précédemment abordée section 2.2.2), et allons nous concentrer sur *l'adaptation des contenus* au sens large : principalement la

personnalisation des contenus, et certains processus de *recommandation* qui peuvent être considérés également comme une adaptation de contenu relativement à une utilisation ou un utilisateur.

2.3.1.1 La personnalisation des contenus

Le principe de la personnalisation est simple : pour un utilisateur donné, on va chercher à modifier son environnement pour que ce dernier corresponde au mieux à ses besoins et ses attentes *supposés*. Pour cela, il est possible d'exploiter *directement* les traces de son activité passée, ou bien de le faire *indirectement* par le biais d'un *profil utilisateur* par exemple (section 2.2.4.2). Les meilleures illustrations de ce type d'approches se trouvent du côté des publications du *Web Usage Mining* qui consacre spécifiquement une partie de ses recherches à l'adaptation automatique des *sites Web* et de leurs contenus. Il existe en effet plusieurs niveaux de personnalisation des sites *Web*. La plus répandue consiste à analyser le parcours de l'internaute sur un site, puis de conserver cette information sur la machine de celui-ci sous forme de *cookies*. À chaque visite ces données sont consultées par le serveur pour proposer à l'internaute un contenu ou une mise en forme adapté(e) aux préférences supposées de celui-ci : « *If we determine that you like to get advice from peers, you're going to see lots of advice from peers* »⁶⁰. Le « *Website morphing* » (Hauser et al., 2008) que développe l'École de Management de *Sloan* au MIT⁶¹ par exemple, est fondé sur ce principe. L'objectif affiché est toujours d'augmenter le confort de navigation pour l'internaute, mais l'objectif réel, et peut-être moins louable, est de faire du ciblage publicitaire et d'adapter à chaque profil utilisateur, le contenu des bandeaux publicitaires pour en augmenter l'efficacité. Il faut dire que « *The researchers' initial studies show that morphing a website to suit different types of visitors could increase the site's sales by about 20 percent* », ce qui ne manque pas d'intéresser les annonceurs (Chatterjee et al., 2003).

La personnalisation des contenus est également une question abordée dans le cadre de l'utilisation des EIAH. L'idée est que l'environnement se fonde sur un profil d'apprenant pour définir de manière adéquate les contenus, voire l'interface, proposée à un utilisateur. Un premier niveau de personnalisation simple peut consister à fabriquer une unique activité pédagogique, constituée de contenus différents en fonction du niveau de l'apprenant, niveau lui-même calculé à partir du parcours de celui-ci dans l'EIAH, directement ou par le biais d'un profil utilisateur, comme c'est le cas dans le projet *Ambre*⁶² que nous aurons l'occasion d'évoquer plus tard.

2.3.1.2 Les systèmes de recommandation

Une autre façon d'exploiter implicitement les traces numériques d'interaction pour modifier une situation d'activité consiste à les utiliser de manière transversale dans un collectif d'utilisateurs. Dans le cadre d'une navigation sur le *Web* il s'agira de proposer à l'internaute des sites intéressants au regard des sites déjà visités durant la navigation engagée.

⁶⁰ Voir <http://www.technologyreview.com/Biztech/20872/> (consulté le 12/02/09).

⁶¹ Voir <http://mitsloan.mit.edu/> (consulté le 12/02/09).

⁶² Voir <http://www710.univ-lyon1.fr/~nguim/ambre.html> (consulté le 20/06/09).

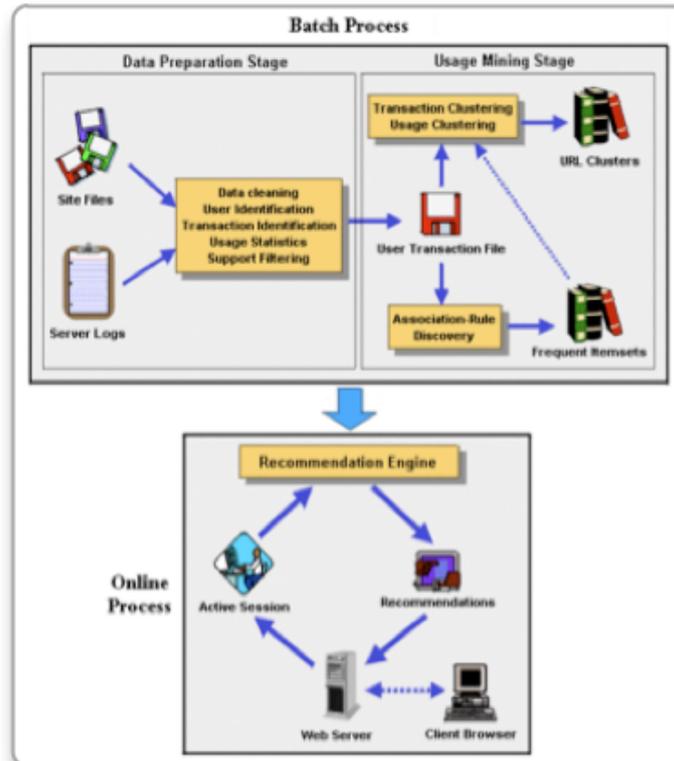


Figure 2.16 : Architecture globale du système de recommandation.
Tiré de (Mobasher et al., 2000 p.145).

Le principe est assez simple. Les sources de traçage étant le plus souvent restreintes⁶³, les traces d'interactions sont généralement constituées d'une suite d'*URLs*. De la même façon que sont construits les profils utilisateur, les chercheurs utilisent des statistiques, qu'ils appliquent à cette trace : règles d'association, classifications, *clustering*, *pattern* de séquences, modèles de dépendance, *etc.* (Srivastava et al., 2000). Le but ultime est d'être capable d'anticiper la navigation d'un internaute. Dans (Mobasher et al., 2000) par exemple, des *URLs* constituent l'essentiel de la trace et les auteurs cherchent à y mettre en évidence des « groupes d'items fréquents », ainsi que des « règles d'association » permettant par la suite de proposer automatiquement à un internaute des suites possibles à sa navigation (Figure 2.16).

Un développement intéressant de ces systèmes de recommandation est leur exploitation « croisée » : on utilise dans ce cas le rapprochement entre des utilisateurs aux profils et/ou parcours similaires pour leur proposer à l'un les actions que l'autre utilisateur a réalisées, à l'instar du site marchand *Amazon*⁶⁴, qui fonde des recommandations d'achat sur des rapprochements avec d'autres listes d'achats similaires réalisés par d'autres utilisateurs. Un exemple original de ce type de système de recommandation nous est proposé avec le système *Histview* de (Terveen et al., 2002). L'objectif du système est d'offrir des *playlists* de morceaux musicaux correspondant au mieux aux goûts d'un utilisateur donné, en se basant sur les morceaux qu'il a déjà aimé par le passé et sur ceux que d'autres utilisateurs ont sélectionnés comme faisant partie d'une même *playlist*. Dans cet exemple l'utilisateur est mis à contribu-

⁶³ Aux serveurs et *proxys*.

⁶⁴ Voir <http://www.amazon.com/> (consulté le 26/03/09)

tion puisque c'est à lui de déterminer le niveau de préférence pour un morceau donné, et l'on devine comment des systèmes de recommandation très complexes peuvent être obtenus en croisant les informations concernant de nombreux utilisateurs d'un même environnement.

On remarquera que l'on atteint rapidement les limites de la catégorie que nous souhaitons illustrer ici, à savoir la modification des contenus par une exploitation implicite des traces numériques d'interaction. En effet, à partir du moment où les préférences de l'utilisateur sont déterminées avec son concours, le caractère implicite de l'existence et de l'exploitation de la trace s'estompe et on se rapproche d'approches plus explicites par rapport à cette exploitation. De même, lorsque la recommandation reste fondée sur l'exploitation statistique d'une suite d'URLs sans modèle d'usage, et qu'elle s'exprime par modification transparente (pour l'utilisateur) des contenus de son environnement, nous restons dans le domaine de la recommandation telle que nous l'avons définie. Si en revanche, cette expression passe par une interface plus explicite, on parlera de *systèmes conseillers* (qui seront abordés dans une section ultérieure).

2.3.2 Enrichir l'interaction

Une autre finalité des traces numériques d'interaction est *l'enrichissement* de l'interaction. Il s'agit d'extraire de l'activité des informations qui seront, d'une manière ou d'une autre, portées à la connaissance de l'utilisateur. Cet enrichissement, par lequel on espère toujours faciliter l'utilisation et les usages d'un environnement, se présente sous plusieurs formes de contextualisation soit des objets manipulés (histoire interactionnelle), soit de l'utilisation d'un outil particulier (historique d'application).

2.3.2.1 Les objets enrichis de leur histoire interactionnelle

L'idée de faire porter aux objets numériques eux-mêmes la trace de leur manipulation a fait son apparition dès le début des années 1990. En l'occurrence, un premier travail marquant intitulé « *Edit wear and read wear* » (Hill *et al.*, 1992 ; Hill et Hollan, 1993), et avait pour idée centrale de : « *Record on computational objects (e.g. documents, menus, spreadsheets, images, email) the events that comprise their use, and then, on future occasions, when the objects are used again, display useful graphical abstractions of the accrued histories as parts of the objects themselves* » (Hill *et al.*, 1992, p.3). Il s'agit donc d'utiliser les traces d'interaction pour enrichir des objets manipulés par les utilisateurs. En l'occurrence, ce sont des *lignes de code* informatique⁶⁵ qui sont ici « manipulées ». Ces lignes de code sont regroupées en zones. L'éditeur *Zmacs*, modifié à cet effet, enregistre le nombre de « *spots* » d'interactions, soit de *lecture* (*read*), soit d'*écriture* (*edit*), que l'utilisateur mène sur chaque zone. Chaque « *spot* » vient s'ajouter aux précédents pour être représenté par un histogramme venant se placer dans l'ascenseur de l'éditeur, dénommé pour l'occasion « *attribute-mapped scroll bar* ». Cette barre permet de distinguer les zones plus ou moins « *éditées* » ou plus ou moins « *lues* » (Figure 2.17).

Dans ce système pionnier, la gestion des traces est, malgré les apparences, relativement complexe. En effet, il faut mettre en place un système d'enregistrement lourd, qui calcule en permanence pour chaque « *spot* » du code affiché à l'écran combien de temps il y est affiché ou édité. De plus, chaque « *spot* » enregistré se voit également affecté d'un certain nombre de « *metadata* » le caractérisant (son

⁶⁵ Il s'agit de *LISP*.

auteur, son rôle, le projet impliqué) permettant d'indexer et surtout de filtrer l'affichage de la barre par la suite. Il est intéressant de voir ici poindre un problème fondamental sur lequel nous reviendrons largement par la suite : une fois le système fonctionnel en place, les auteurs se posent la question de savoir *ce qui doit être considéré comme un interaction* avec un objet, et donc susceptible d'être effectivement pris en compte dans la trace. Jusque là, ce qui était pris en compte pour l'édition par exemple était « la ligne » de code, mais ne pourrait-on pas plutôt considérer « le caractère » ou « un groupe de lignes » ? Le problème se pose également pour la lecture, qui n'est caractérisée que par le temps d'affichage d'une ligne à l'écran à un instant donné.

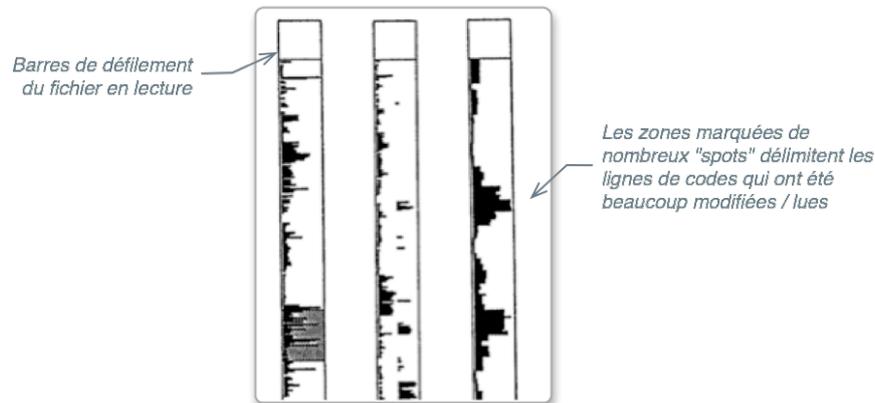


Figure 2.17 : Exemple de « *Attributed-mapped scroll bar* » dans *Zmacs*, montrant des spots de lecture et d'écriture sur certaines zone d'un code - Tiré de (Hill *et al.*, 1992, p.4).

Aujourd'hui des formes de contextualisation minimaliste, portée par les objets eux-mêmes, ont fait leur apparition dans les interfaces d'outils courants, comme les compteurs de lectures des fichiers audio par exemple.

2.3.2.2 Les historiques d'application

Contrairement au système « *read and edit wear* », qui « fragmente » l'histoire interactionnelle et « l'écrase » dans un histogramme, les historiques d'application regroupent et présentent de manière *temporelle* les *opérations effectuées* par un utilisateur dans une *application*. Dans le cas d'une application quelconque, l'existence d'un enregistrement systématique des opérations effectuées est rarement explicite. Lorsque c'est le cas cependant, les chercheurs se sont beaucoup attachés aux moyens d'en améliorer la présentation.

La présentations visuelle des historiques d'applications

L'enregistrement systématique des interactions ne débouche pas toujours sur un historique explicitement accessible pour l'utilisateur. Ces enregistrements sont pourtant présents dans la plupart des applications actuelles et sont utilisés notamment avec la fonctionnalité « *undo / redo* » pour « défaire / refaire » une série d'opérations. La plupart du temps, la présentation de l'historique de ces opérations se fait sous la forme d'une liste de texte, comme c'était déjà le cas lors de la création de ce type de fonctionnalité dans des applications comme *Gina* (Berlage, 1994) ou *Amulet* (Myers *et al.*, 1997). Certaines applications, éditeurs de texte, logiciels de traitement des photographies numériques ou de

dessin, se prêtent mal à ce genre d'historiques textuels. Aussi plusieurs travaux ont été engagés au cours des années pour offrir d'autres présentations à l'utilisateur.

Une des pistes explorée est la présentation sous la forme d'une suite de captures d'écran pour illustrer l'état du système à un moment donné. Le résultat devient une sorte de *story-board* de l'activité passée. Dans cette catégorie, nous pouvons citer les système *Chimera* (Kurlander et Feiner, 1992) et *Pusruit* (Modugno et Myers, 1994) qui utilisent une présentation proche de la bande dessinée. Il en est de même pour le tout récent travail de Nakamura et Igarashi (2008) dans lequel les auteurs proposent de manière générique une méthode de visualisation de l'historique des opérations de toute interface graphique. L'exemple choisi est celui d'un logiciel de dessin. La séquence des opérations est représentée par une suite de *vignettes*. D'une part, l'ensemble des vignettes illustre l'évolution du dessin, et d'autre part chaque vignette peut-être *annotée* avec des symboles de façon à *figurer* les opérations effectuées (Figure 2.18).

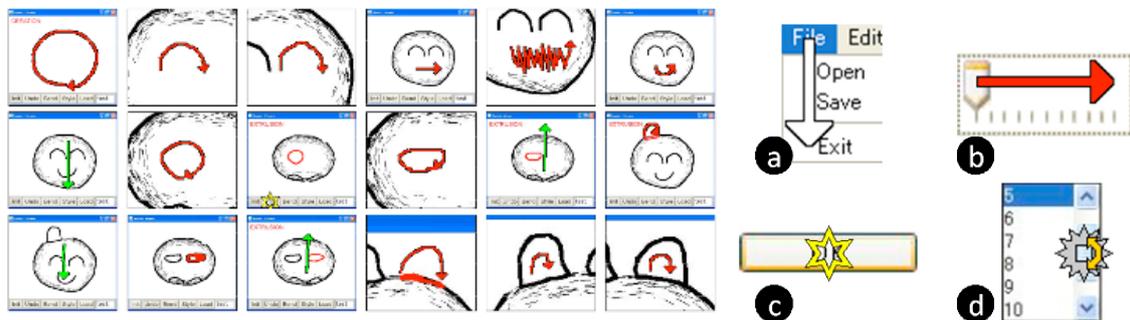


Figure 2.18 : Historique visuel et annoté (à gauche), et symboles d'annotation (à droite)
Tiré de (Nakamura et Igarashi, 2008, p.23).

Les historiques de navigation Web

S'agissant de traces numériques présentées à l'utilisateur sous forme d'historique, il est impossible d'éviter la question des *historiques de navigation Web*. La *présentation* de ces historiques est d'ailleurs peut-être un des sujets qui a le plus suscité de travaux dans le cadre de la recherche autour des interfaces de navigation sur le *Web*. Cet intérêt a été entretenu au cours du temps par les résultats de recherche concernant l'utilisation des navigateurs. Nous pouvons citer par exemple le travail de Tauscher et Greenberg en 1997, qui montrait que 58% des *URLs* consultées par des internautes l'avaient déjà été par ces mêmes internautes par le passé, et que les nouvelles pages visités ne l'étaient souvent qu'une seule fois.

Les auteurs plaidaient ainsi pour que les navigateurs intègrent une instrumentation et des fonctionnalités dédiées à la « re-visite » (que sont bien loin de proposer les historiques de navigateurs *Web* classiques). Depuis leur apparition, les navigateurs ont sensiblement évolués, mais leur interface n'a subit aucune « révolution ». On les a dotés de fonctionnalités telles que l'historique, système d'auto-complétion ou de favoris, en ayant dans l'idée que ces fonctionnalités devaient être suffisantes pour « *assist users when navigating to previously visited pages* » (Hawkey, 2006, p.1). Les dernières avancées à signaler sont en rapport avec la recherche d'information et les services en ligne de conservation des historiques, en particulier celui de *Google Web History* qui permet de conserver ses requêtes, ou de manière plus disparate, le développement de *plugins* Firefox™ qui font régulièrement leur appari-

tion pour disparaître presque aussi vite, comme *NavTracer*⁶⁶ par exemple qui fournissait un historique de navigation amélioré, prenant en compte le focus de l'utilisateur permettant ainsi d'affiner les informations enregistrées.

Globalement, et malgré leur évolution, les *historiques* de navigateur *Web* restent *sous-exploités*. Plusieurs raisons sont évoquées (cf. section 2.2.3.1), la plus triviale étant que la simple présentation d'une suite d'*URLs* ne constitue pas un historique de navigation satisfaisant. Plusieurs travaux se sont donc attachés à développer des présentations plus élaborées. Leur objectif est de redonner à la navigation représentée une *profondeur*, une *structure* plus proche de celle de l'activité menée par l'utilisateur. Pour cela, les pages dans l'historique sont regroupées en fonction de leur site, de leur proximité temporelle ou lexicale, *etc.* Le résultat est alors présenté sous forme d'arbres ou de réseaux.

Dès 1995 était proposé, sur le navigateur *Mosaic*, le système *Webmap* qui permettait de construire un *arbre* de navigation (Figure 2.19) qui était présenté soit en tant réel, soit à la fin d'une session (Domel, 1995). Dans le même ordre d'idée, nous pouvons également citer le système *Padprint*, présenté comme « *Graphical Multiscale Web Histories* » (Hightower *et al.*, 1998) et qui offre à l'utilisateur une présentation de ses pages visitées sous forme d'*arbre horizontal* dont les nœuds sont des pages visitées (Figure 2.19). Chacune d'elle est représentée par une *vignette* issue d'une capture d'image effectuée sur la machine client, alors que les *arcs* sont les *liens hypertextes* utilisés récupérés sur un serveur *proxy*.



Figure 2.19 : Les historiques de navigation dans *Webmap* (à gauche) et *Padprint* (à droite)
Tirés respectivement de (Domel, 1995, p87) et (Hightower *et al.*, 1998, p.121).

En 1997 était présenté *Footprints*, un système conçu pour donner en temps réel à l'utilisateur un *historique enrichi* de sa navigation sous la forme d'un *graphe* calculé sur les *logs* de navigation côté serveur (Wexelblat et Maes, 1997, 1999). La navigation de chaque internaute « enrichit » les pages visitées, en permettant leur *annotation* (puis leur mise en relation et leur regroupement dans un graphe). L'idée est d'utiliser les précédentes utilisations de ces pages comme faisant partie de l'environnement de l'utilisateur dans son activité. Le graphe obtenu apparaît finalement soit sous la forme d'un *arbre*, soit sous la forme d'une *carte* sur laquelle l'utilisateur aurait évolué (Figure 2.20).

⁶⁶ <http://navtracer.mozdev.org/>

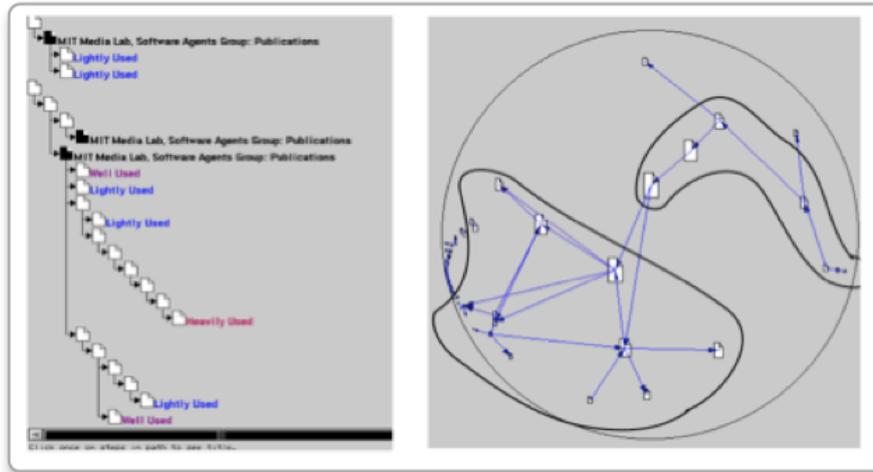


Figure 2.20 : Arbre (à gauche) et carte (à droite) de l'historique de navigation avec *Footprints*.
Tiré de (Wexelblat et Maes, 1999, p.7 et p.9).

Les auteurs opposent *histoire* et *accumulation* pour justifier leur démarche qui consiste à faire émerger une structure dans la navigation des internautes. On constatera avec intérêt, que ce système traçant a *dès le départ* été pensé pour une utilisation par un *collectif*. Les auteurs expliquent en effet qu'ils considèrent comme plus valorisant les traces de navigation lorsqu'elles sont *partagées* (Wexelblat et Maes, 1999). En l'occurrence, *Footprints* permet de partager des cartes sur lesquelles sont représentés par un graphe les seules pages visitées sur un site et offrant la possibilité d'accéder à la séquence des pages telles qu'elles ont été visitées par d'autres.

Pour A. Wexelblat il s'agit là de partager l'expérience que les utilisateur font du site : « *These maps show sequence, obviously, but at a higher level they can be seen to expose the kinds of experiences which users can get from a web site* ». Afin d'exprimer le partage de ces expériences de navigation les auteurs invoquent la notion de « *social navigation* » (Wexelblat, 1998), notion qui ne manque pas de raisonner avec les évolutions actuelles du « *Web 2.0* ». Pour faire écho à ce que nous avons dit concernant les systèmes de recommandation, il s'agit là bien plus que des suggestions d'*URLs*, car c'est bien la navigation, ou une partie de la navigation en tant que telle qui est partagée.

La présentation de l'historique de navigation continue de motiver le travail des chercheurs. Parmi les travaux récents, celui de Shirai *et al.* (2006) est particulièrement original. Développant l'idée de contextualisation des pages *Web*, les auteurs proposent d'estomper la frontière qui sépare la navigation sur des pages *visitées actuellement* des pages *visitées par le passé*. Ils ont par exemple développé sur cette idée un interface qui place la page actuellement visitée à la croisée de deux axes : verticalement les différentes versions (lors des différentes visites) de la page courante et horizontalement les pages visitées dans la même séquence de navigation (Figure 2.21).

Dans ce même travail, les auteurs proposent aussi de donner accès à l'utilisateur à un historique de navigation, sous forme de *ligne de temps*, axé sur les recherches (*search queries*) et mettant en avant la séquence des résultats consultés. Ils proposent enfin de regrouper les pages présentées non seulement en fonction de leur apparition dans une séquence de navigation, mais en fonction de leur similarité *lexicale* ou de la ressemblance de leurs *URLs*.

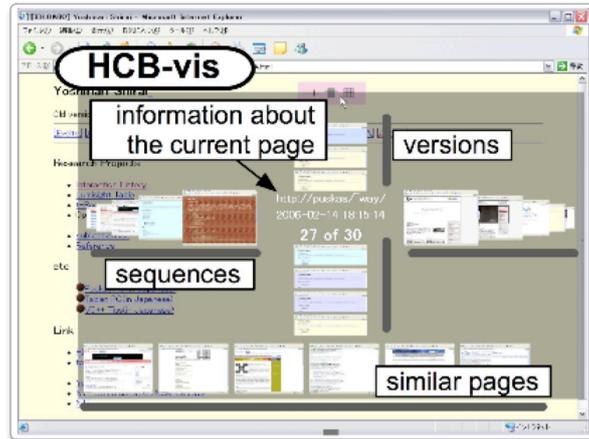


Figure 2.21 : Présentation d'une page *Web* visitée à la croisée de deux axes temporels.
Tiré de (Shirai *et al.*, 2006, p.1323).

Contexte et mémoire

On peut sans doute trouver de multiples exemples pratiques de l'intérêt d'enrichir l'interaction par la *contextualisation* des objets manipulés. L'intérêt et la pertinence de cette contextualisation pour faciliter l'activité, repose principalement sur le fonctionnement de la mémoire humaine, ou tout au moins sur ce que la psychologie et les sciences cognitives ont déjà mis en évidence il y a longtemps sur le fonctionnement de la mémoire humaine. C'est ce que nous rappelle le système traçant proposé par M. Ringel et ses collègues (2003), qui se propose de faciliter la recherche de documents au sein d'un environnement numérique personnel. Pour cela, le système présente à l'utilisateur une ligne de temps contenant un certain nombre d'objets (documents) créés ou utilisés à un instant donné. Pour se retrouver dans une liste qui pourrait s'apparenter à une simple visualisation de *logs* sans filtrages particulier, la ligne de temps comporte également des *éléments personnels particulier* qui vont fournir des *repères temporels* pour la remémoration de objets recherchés. L'idée de M. Ringel et ses collègues est de s'appuyer sur les particularités de la *mémoire épisodique* qui fonctionne grâce à un certain nombre d'*événements* mobilisés par la mémoire comme des repères identifiant des *épisodes* de la vie quotidienne. Ainsi, comme on le voit sur la figure ci-dessous, la ligne de temps proposée permet l'insertion de ce type de *repères*, par exemple des photos prises lors d'un événement particulier de la journée. L'utilisateur peut lui-même intervenir sur ces jalons, mais il est également imaginé que ce genre de contextualisation soit automatisé, dans la mesure du possible et dans les limites qu'aura lui-même fixées l'utilisateur.

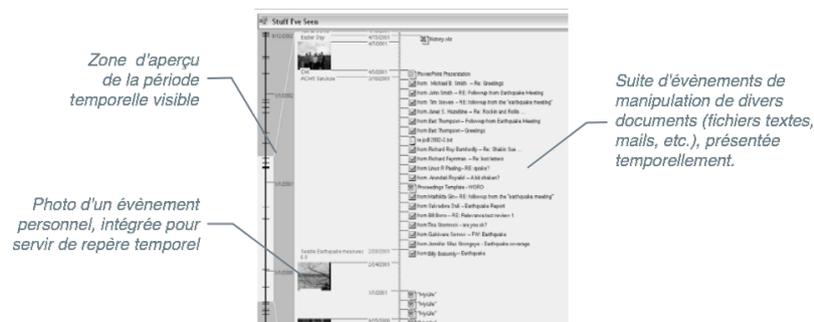


Figure 2.22 : La présentation sous forme de ligne de temps de *Stuff I've Seen*.
Tiré de (Ringel *et al.*, 2003, p.186).

Dans ces différentes tentatives de présentation structurée des historiques d'application, on perçoit un objectif sous-jacent, qui n'est pas simplement de contextualiser les objets manipulés et les manipulations uniquement par rapport à l'organisation temporelle des interactions mais par rapport à l'organisation de *l'activité* elle-même. Une première limite à la contextualisation de l'interaction par une représentation de l'activité elle-même est une contrainte technique, qui contraint de façon générale les systèmes traçants à dépendre d'une application en particulier, là où l'activité engage l'utilisation de plusieurs outils. Lorsque, comme dans le cas de l'outil de dessin présenté plus haut, toute l'activité est supportée par le même outil, il est possible de rendre compte d'une progression de l'activité. Lorsqu'au contraire l'activité se déploie sur plusieurs outils, ou que plusieurs activités peuvent être menées en parallèle sur le même outil, la difficulté apparaît.

2.3.2.3 Enrichir l'interaction avec une représentation de l'activité

Une autre façon d'enrichir l'interaction d'un utilisateur en situation d'activité, est de lui présenter explicitement une *trace numérique d'interaction qui figure son activité*, i.e. lui présenter un ensemble de ses interactions « en extension » dans une présentation temporelle. Cette fois, la confrontation de l'utilisateur avec la trace numérique de ses interactions est clairement annoncée, d'autant plus lorsque la présentation de celle-ci dépend d'une application dédiée.

La présentation d'interactions de multiples applications

L'utilisation d'une seule application, fut-elle aussi importante que le navigateur *Web*, ne constitue peut-être pas une source de traces suffisamment riche. C'est l'idée qui a donné naissance à quelques applications, plus ou moins complexes, qui se chargent de présenter un historique des interactions ayant eu lieu avec *plusieurs* applications d'un même environnement. Ainsi V. Kaptelinin (2003) entre autres, remarque que la métaphore du bureau et le cloisonnement des applications existantes ne sont pas vraiment en phase avec la manière dont se *structure réellement* l'activité d'un utilisateur. Celui-ci travaille plutôt sous la forme de *projets* qui vont mobiliser un nombre d'outils parfois impressionnant : un client messagerie, un agenda, des notes, des traitements de texte plus ou moins compliqués, des tableurs, des calculatrices, des outils d'édition d'images (*etc.*), auxquels il ne faut pas oublier de rajouter d'éventuelles applications en ligne. Il s'agit là d'un grand thème du domaine du *PIM*⁶⁷. Pour apporter une réponse à ce problème, V. Kaptelinin (2003) propose la création d'un environnement permettant de regrouper et de contextualiser les ressources utilisées dans l'activité *par les projets* dont ils relèvent : *UMEA (User-Monitoring Environment for Activities)*. Le système récupère et stocke des événements issus de MS-Office™, comme l'ouverture d'une page *Web* ou l'impression d'un document texte⁶⁸, les regroupe par projets pour finalement les présenter à l'utilisateur.

Un dispositif très similaire est actuellement proposé sous la forme d'une petite application dédiée appelée *Slife (ex-Onlife)*. Cette petite application exploite des *logs* produits par le dernier système d'exploitation sous Macintosh™ (OS X), ce qui lui permet d'enregistrer des interactions avec de nombreuses applications, y compris (bien entendu) les navigateurs *Web* comme *Firefox™* ou *Safari™*. Le

⁶⁷ *Personal Information Management* domaine qui se consacre à l'instrumentation des espaces numériques de travail (cf. articles autour de cette question à CHI-2008 : <http://pim2008.ethz.ch/acceptedpapers.php>)

⁶⁸ Ce travail est un des rares exemples où les *logs* utilisés ne sont pas des logs de navigation sur le *Web*.

système propose de présenter à l'utilisateur une ligne de temps figurant le partage de son temps de travail en fonction des applications utilisées au premier plan.

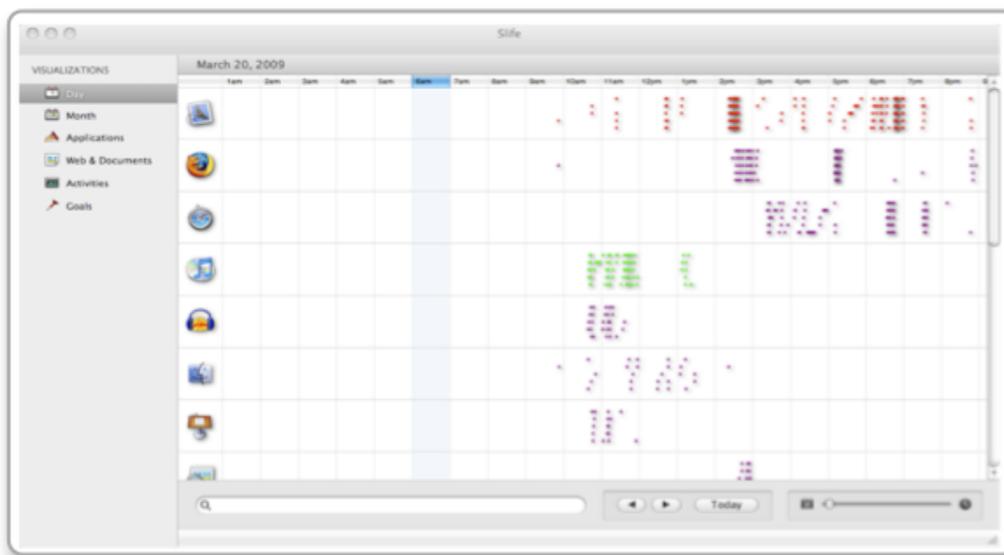


Figure 2.23 : Interface principale de *Slife*.

Comme on peut le voir (Figure 2.23) des « spots » de couleur représentent la durée passée sur tel ou tel fichier de telle ou telle application. La ligne de temps offre deux échelles, soit la journée, soit le mois. Les spots donnent une information sur le document manipulé uniquement si cette information est directement disponible (dans les métadonnées du fichier) mais n'en garde ni aperçu, ni copie. Ce n'est pas un moyen d'accès aux documents, c'est un moyen de représenter de manière très globale le temps passé sur chaque application en essayant de couvrir au maximum tout ce qui peut relever de l'activité menée par l'utilisateur. La présentation de son activité que lui propose *Slife* a pour but de l'aider à prendre conscience du temps passé sur chaque application et indirectement à réguler son activité en fonction de ses objectifs propres⁶⁹. Si l'effort de prise en compte d'une multitude d'applications par le système traçant est notable, il ne paraît cependant pas suffisant pour parler réellement d'une contextualisation de l'interaction par l'activité elle-même. L'organisation de l'activité telle qu'elle est vécue et réalisée concrètement par l'utilisateur n'est que difficilement réductible à la somme des applications mobilisées pendant un laps de temps dans l'environnement numérique. Pour se rapprocher de l'objectif, une des pistes privilégiée par les chercheurs est celle de l'implication de l'utilisateur.

L'implication de l'utilisateur

Qu'il s'agisse de contextualiser plus efficacement des objets selon l'activité qui les a mobilisés ou bien de figurer cette activité elle-même, le problème est de ne pouvoir identifier automatiquement et de manière sûre *l'activité de laquelle une action de l'utilisateur relève* à un instant donné. Sachant que l'utilisateur peut changer d'activité à tout instant, il est effectivement très difficile du point de vue de la machine de déterminer celle qui est en cours de réalisation. Plusieurs travaux se tournent donc vers la seule personne capable de trancher de manière sûre : l'utilisateur lui-même.

⁶⁹ Dans sa dernière mouture, l'utilisateur peut se fixer des objectifs de temps à passer pour telle ou telle activité / application.

C'est le cas dans les deux exemples que nous venons d'évoquer : dans *UMEA* tout comme dans *Slife*, l'utilisateur dispose d'un panel auquel il peut accéder rapidement pour indiquer, au moment où il agit, l'activité qu'il est en train de mener (Figure 2.24). Dans le cas du système *UMEA*, l'interface utilisateur permet de gérer *a posteriori* l'attribution d'une activité (appelée « projet ») de rattachement pour les événements constituant la trace d'interaction, le système rattachant par défaut tout nouvel événement à l'activité courante.

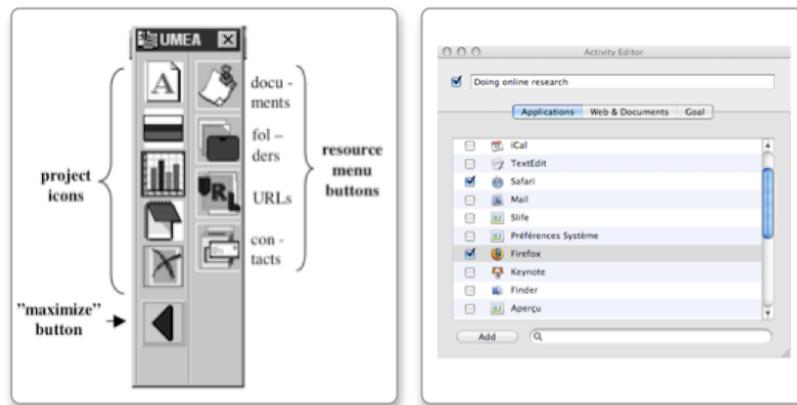


Figure 2.24 : Interface utilisateur de *UMEA* (à gauche) et le panel de *Slife* (à droite). Respectivement tirés de (Kaptelinin, 2003, p.356) et d'une capture d'écran de l'application (2009)

On retrouve également cette idée dans des approches concernant la navigation sur le *Web*. Après une tentative infructueuse de caractérisation automatique de la navigation *Web* à base de *logs*, Hawkey, et *al.* (2006) proposent à l'utilisateur un outil de « *reporting* » qui permet de *tagger* et *d'annoter* les pages qu'il estime avoir « trouvées » dans sa recherche d'information. Il déclare au passage à quelle tâche en cours cette « découverte » doit être intégrée. Ceci permet ensuite de regrouper les différentes parties des traces de navigation en fonction de la tâche en cours. On voit ici des traces numériques d'interactions qui changent de nature, et qui pour être pertinentes finissent par s'appuyer sur le concours *volontaire* de l'utilisateur lui-même, ce qui enlève à ces traces leur caractère en principe *involontaire*. Nous terminerons d'ailleurs cette section, consacrée à la contextualisation rendue possible par les systèmes traçants, en faisant quelques remarques liées à cette implication de l'utilisateur dans la constitution de sa propre trace numérique.

2.3.3 Assister l'utilisateur dans son activité

La dernière finalité à mettre en avant dans l'exploitation des traces numériques est celle de *l'assistance à l'utilisateur*. Dans les différents exemples de systèmes traçants évoqués jusqu'ici, l'objectif attribué aux traces numériques se cantonne généralement à la *facilitation* de l'utilisation des applications mobilisées : « *helps make applications easier to use* » (Nakamura et Igarashi, 2008, p.23). Dans les travaux que nous allons aborder maintenant, il ne s'agit plus simplement de facilitation de l'utilisation grâce à des fonctionnalités pratiques, mais il s'agit de manière *explicite*, pour les systèmes traçants et les traces numériques qui en sont issues, de *soutenir effectivement l'activité* engagée par l'utilisateur. Cette assistance explicite peut prendre plusieurs formes. Nous allons ici en évoquer trois, dont nous verrons qu'elles sont, en partie, le prolongement de certaines démarches déjà évoquées :

l'assistance *automatique* (section 2.3.3.1), l'assistance *mnésique* (section 2.3.3.2) et enfin les assistants *conseillers* (section 2.3.3.3).

2.3.3.1 Assistance automatique : ré-exploiter directement les traces d'interaction

De manière basique, une trace numérique en tant que séquence enregistrée d'opérations effectuées par la machine, peut très bien être *ré-exécutée*. C'est le principe du « *undo / redo* » que nous avons déjà évoqué ou du « *rejouage* » de session offert par certains outils d'analyse des interaction. À ce titre, des traces numériques d'interaction peuvent être, pour tout ou partie, ré-exécutée, dans des circonstances différentes, par un utilisateur désireux de ne pas répéter une séquence d'opérations déjà réalisée plusieurs fois. C'est ce qui se passe de manière simpliste avec les systèmes *auto-complétion* qui évitent à l'utilisateur retaper lui-même des valeurs identifiables (nom, email, date, *etc.*) dans des formulaire en ligne par exemple.

Très tôt, des chercheurs (Greenberg et Witten, 1988) ont noté que les utilisateurs de systèmes informatiques répètent certaines actions ou opérations, et ont étudié la possibilité de réutiliser des séquences répétitives de toutes sortes (télétypes, sélections graphiques, éditions, navigations dans des menus, *etc.*). Dans ce cas, la trace numérique n'est pas présentée en tant que telle à l'utilisateur. Lorsque c'est le cas, dans des historiques d'opérations par exemple (section 2.3.2.2), la « *navigation* » dans les opérations en question reste linéaire et enfermée dans la séquence complète dont elles font partie. En cassant cette linéarité, des systèmes comme *Collagen* décrit dans (Rich et Sidner, 1997), permettent de sélectionner un « *segment* » de la trace d'interaction, et de le ré-exécuter dans un autre contexte.

Petit à petit, en complexifiant les opérations prises en compte et les paramétrages possibles de la ré-exécution, on se rapproche du principe des « *macros* » disponibles dans des logiciels comme Excel™. Et en poussant encore un peu, on aboutit aux systèmes de *programmation par l'exemple* ou *PBD*⁷⁰. Dans ce type de système, l'utilisateur agit de manière à ce que le système enregistre la séquence d'opérations désirée, puis généralise les opérations en un programme. C'est le cas du système *DocWizards* (Bergman *et al.*, 2005) qui permet de créer des démonstrations ou « *follow-me* » exécutables et commentables. Pour le système *Koala* (Little *et al.*, 2007), il s'agit d'enregistrer des interactions correspondant à des « *comportements* » de navigation en ligne (section 2.2.3).

On enregistre des opérations (recherches, remplissages de formulaires, consultations de sites spécifiques) qui sont traduites en *pseudo langage naturel* afin d'être compréhensible à la fois pour l'utilisateur et pour la machine, en l'occurrence *FIREFOX*™, qui devra les ré-exécuter. Le résultat produit ressemble fort à la création assistée de *démonstration automatique de logiciels*, avec un script annoté (mise en surbrillance des boutons ou champs sélectionnés par exemple) et rejouable de sorte qu'on puisse faire comprendre à quelqu'un qui rejoue le script ce qu'il y avait à faire et comment le faire. Les auteurs imaginent ainsi pouvoir *capitaliser* de « *bonnes pratiques* » qui seraient alors *partageables* entre collaborateurs d'une équipe de travail : « *[allowing] users to document and record the specifics of performing a business process, and put the process on a wiki to share their experience and allow others to automatically execute the process for themselves* » (Little *et al.*, 2007, p.203). Comme le laisse deviner ce dernier exemple la *réutilisation* des traces numériques d'interaction pour

⁷⁰ *Programming By Demonstration.*

l'assistance ne se limite pas à sa *ré-exécution directe* par la *machine*, mais également sa remobilisation par l'utilisateur lui-même.

2.3.3.2 Assistance mnésique : ré-exploiter indirectement les traces d'interactions

La capitalisation par une trace numérique peut-être en effet entendue dans d'autres circonstances comme une forme de *réention*. Le système traçant peut ainsi être considéré comme un dispositif « mnésique » ou « rétenseur », dont l'objectif en termes d'assistance n'est plus la ré-exécution possible d'opérations telles quelles, mais plutôt de permettre à l'utilisateur de se remémorer ces opérations pour *remobiliser l'expérience associée* à une séquence d'action singulière. Dans le prolongement de ce que nous avons dit sur la *contextualisation* (section 2.3.2.2), un système traçant peut dans ce cas proposer explicitement une *assistance mnésique*, une *assistance indirecte à l'activité* de l'utilisateur.

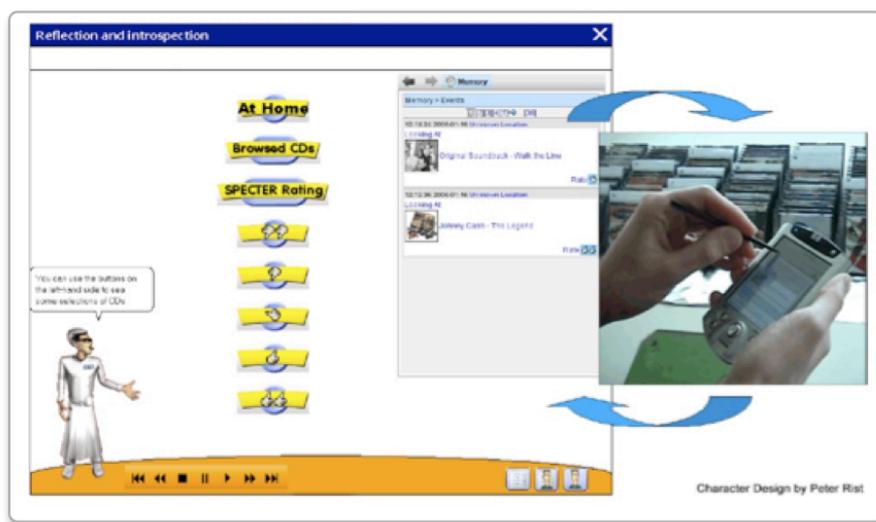


Figure 2.25 : Interfaces du système *SPECTER* - Tiré de (Kröner *et al.*, 2006, p.5).

C'est ainsi que se présente par exemple le système *SPECTER* (Schneider *et al.*, 2005 ; Kröner *et al.*, 2006) qui s'annonce clairement comme une « prothèse mnésique ». Dans cet exemple original, les interactions enregistrées qui vont constituer la trace sont récupérées grâce à des capteurs dans le monde réel⁷¹. Dans l'exemple présenté, l'activité choisie est celle de l'achat d'un CD musical dans un grand magasin. L'utilisateur peut alors accéder à la trace de son activité d'achat sur un dispositif numérique mobile d'une part, et sur son ordinateur à la maison d'autre part (Figure 2.25). Le système lui-même est conçu selon un modèle de mémoire « naturelle », et peut fournir à l'utilisateur des épisodes passés dont la similarité avec le contexte présent est grande, permettant ainsi de re-contextualiser lui-même une l'activité engagée.

Cette même idée est également exploitée par le système *SHERLOCK* (Lesgold *et al.*, 1992), destiné à l'entraînement de techniciens en avionique. Il propose des activités pédagogiques pour lesquelles il est capable de donner des explications grâce à un système de dialogue que l'utilisateur peut utiliser pour poser des questions. Au sein de ce système, Lemaire et Moore (1994) proposent que la trace numérique d'interaction soit utilisée pour *augmenter les explications* fournies à l'utilisateur. Ce dernier peut

⁷¹ Il est également fait appel à l'utilisateur qui peut contribuer en annotant certaines de ses actions que le système lui propose de rejouer.

sélectionner pour une activité en cours, une explication passée fournie par le système sur une situation similaire et lui demander de la comparer à l'explication en cours. Le système génère alors automatiquement un texte comparant les deux situations pour soutenir la compréhension de l'apprenant. Pour interagir avec les traces d'interaction, l'utilisateur peut pointer et sélectionner des portions de dialogue et poser une question à son propos. Comme nous le verrons dans le chapitre suivant, nous avons nous même exploré cette piste de recherche, que constitue ce type d'assistance par réutilisation indirecte des traces numériques, et que nous pensons prometteuse (voir l'approche Musette, Chapitre 3).

2.3.3.3 Les assistants « conseillers »

Il est un peu difficile de constituer une catégorie pour les *assistants conseillers en général*, cette approche s'étant surtout développée dans l'environnement plus favorable des EIAH (objet de la section suivante). Leur principe de base est proche de ce que nous avons précédemment présenté comme les systèmes de *recommandation*. Nous marquerons la différence avec ces derniers en deux points : premièrement, les premiers avaient pour but de suggérer des actions de manière intégrée à l'environnement, alors que les systèmes conseillers sont conçus comme des éléments tiers dans l'interaction ; deuxièmement, les systèmes conseillers font appel à des données et des calculs plus complexes pour proposer divers types d'actions ou opérations à effectuer.

Même s'il est difficile de déterminer à l'avance dans le cas de la navigation sur le *Web*, une « bonne façon » de réaliser une activité donnée (sans connaître au préalable les buts de l'utilisateur), certains chercheurs ont tenté de mettre au point des assistants capables de conseiller un utilisateur durant sa navigation sur le *Web*. C'est le cas de *Letizia*, un navigateur *Web* « conseiller » (Lieberman, 1995, 2001). Comme pour les travaux visant la *prédiction* des actions d'un utilisateur, *Letizia* se base sur les interactions de l'utilisateur (accès à des ressources, consultations des écrans, clics effectués, temps passés aux opérations, choix effectués, réponses données aux éventuelles questions, etc.) pour calculer des actions à suggérer. Les actions suggérées sont proposées à l'utilisateur qui est libre de les accepter ou de les rejeter.

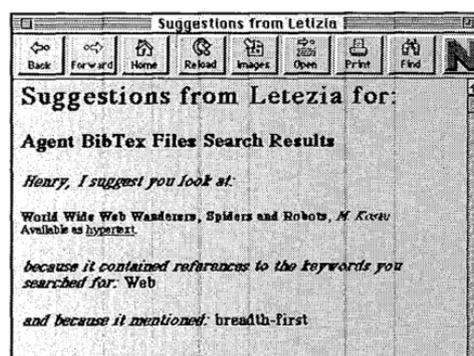


Figure 2.26 : Interface de l'assistant conseiller à la navigation du système Letizia.
Tiré de (Lieberman, 1995, p.1).

Pour tenter de correspondre au mieux à la réalité de l'activité en cours, le système *Letizia* a été amélioré pour ne pas se contenter de *donner* des conseils. Le système peut en effet lui-même tenir compte des « conseils » que l'utilisateur peut donner, et effectuer une navigation « parallèle » à celle de l'utilisateur (Lieberman, 2001). Nous noterons que la question du collectif a également été abordée par

les auteurs. Pas de manière aboutie avec le système *Letizia*, mais avec l'agent *Let's Browse* proposé par Lieberman et ses collègues en 1999. Associé à un dispositif qui reconnaît l'utilisateur (à la condition que celui-ci porte le badge adéquat) de n'importe quel machine dans un bâtiment, l'agent en question utilise les traces de navigation (traces d'interaction avec le navigateur) pour tenter soit de rapprocher des personnes ayant des intérêts communs, soit de faire profiter chacun des parcours des autres.

Un autre exemple de mise en œuvre d'un système conseiller sur la base des traces de navigations se trouve dans les travaux de P. Pirolli et ses collègues (2000, 2002). Le système en question est sensé prédire, connaissant son parcours jusqu'à l'instant t , quelles pages un internaute est susceptible de visiter à l'instant $t + 1$. Les auteurs font alors un pas supplémentaire et affirment que ce type de système pourrait guider l'utilisateur et lui suggérer des choix directement pertinents en se basant non seulement sur l'historique *Web*, mais aussi sur les tâches engagées par l'utilisateur dans son environnement : « *This is a theory at the adaptationist level, where we predict what actions it would be beneficial for the user to take based on the task environment* » (Pirolli, 2002, p.1).

Ce que les auteurs suggèrent ici, c'est qu'un tel système serait capable d'interpréter la tâche de navigation en cours pour décider quelles actions pourraient, ou non être bénéfiques. Même dans le cadre d'un EIAH où l'activité est plus ou moins prédéterminée, ceci est loin d'être évident. Comme nous l'avons souligné, il est très difficile de représenter (section 2.3.2.3) et *a fortiori* d'assister, une activité qui n'est pas correctement caractérisée et identifiable du point de vue de la machine. Sans doute parce que ses besoins sont plus grands, peut-être aussi parce qu'ils sont théoriquement plus facilement caractérisables, c'est à l'*apprenant utilisateur d'EIAH*, que beaucoup de système traçants visent à porter assistance.

2.3.3.4 Remarque sur l'assistance dans les EIAH

L'idée d'exploiter les traces numériques d'interaction comme un support à l'activité d'un apprenant a été mise en avant et développée très tôt dans le domaine des EIAH, domaine qui est sans aucun doute aujourd'hui un des plus avancés sur la question de l'*assistance* à base de *traces numériques d'interaction* : « *Capturing and manipulating the history of user actions seems especially promising as an aid in learning environments* » (Plaisant *et al.*, 1999, p.348). C'est la raison pour laquelle nous avons choisi de traiter spécifiquement les systèmes traçants qui y sont proposés dans ce cadre, dans une section à part.

Les chercheurs du domaine ont clairement identifié deux type de systèmes assistants, basés sur les traces numériques d'interaction : les systèmes de « *guiding* » et les systèmes de « *mirroring* » (Jermann *et al.*, 2001 ; Soller *et al.*, 2005). L'idée directrice des systèmes dits de « *guiding* » est d'intervenir plus ou moins radicalement pour *orienter l'apprenant* dans son activité. Au fur et à mesure de la réalisation de l'activité pédagogique, le système compare l'état de l'interaction relativement au modèle et peut ainsi, à un instant donné orienter l'utilisateur, lui *suggérer* ou *imposer* des actions à réaliser pour poursuivre son activité. On se rapproche en cela des systèmes que nous avons qualifiés de « *conseillers* » (section 2.3.3.3). Le principe des systèmes de *mirroring*, comme leur nom l'indique, est de présenter à l'apprenant une trace de son activité afin de le *confronter* à un « *reflet* », une « *représentation* », de ce qu'il a fait dans le cadre de son activité d'apprentissage. De manière générale, cette présentation doit permettre aux apprenants de *contrôler* leur comportement, de *réfléchir* à leur

progression, et *d'acquérir* des connaissances en révisant leurs *expériences*. Au-delà de la réflexion sur une activité singulière, ce que cette confrontation à ses propres traces cherche à toucher, c'est la prise de conscience « *métacognitive* », c'est-à-dire une prise de conscience, non de l'objet de l'apprentissage, mais du *processus d'apprentissage* lui-même.

On a montré que les processus métacognitifs, non seulement ont une importance cruciale pour le succès d'un apprentissage particulier, mais sont également indispensables dans le développement des capacités d'un apprenant, notamment en terme de transfert de connaissances (Bull et Kay, 2008 ; Romero, 2004 ; Gagnière *et al.*, 2004 ; Escorcía, 2006). De manière très sommaire nous pouvons décrire la métacognition comme une « cognition qui prend la cognition pour objet ». Il s'agit d'une part des *connaissances* métacognitives qui sont les connaissances que l'on dispose sur soi-même sur la manière de se comporter, de réfléchir ou d'agir en fonction de préférences cognitives, et d'autre part de la *régulation* métacognitive qui est le contrôle exercé sur son propre raisonnement, stratégie ou activité cognitive en général (Flavell, 1977). Les traces numériques d'interaction a un rôle fondamental à jouer, en permettant de donner un support à cette activité cognitive particulière (Dimitracopoulou, 2005).

À la suite de (Carroll *et al.*, 1996) et (Guzdial *et al.*, 1996), C. Plaisant et ses collègues (1999) soulignent les rôles multiples que peuvent jouer les traces d'interactions dans le cadre d'une activité d'apprentissage : en aidant l'apprenant à comprendre ce qu'il a fait, en lui permettant de corriger ou modifier une étape de son travail, de rejouer son activité, de sauvegarder la trace de son activité pour une consultation ultérieure (seul ou avec un pair), d'annoter cette trace pour l'envoyer à un pair pour demander son avis ou son aide, mais permettant aussi de composer des démonstrations ou tutoriels jouables, de créer des macros ou de répéter des séquences ainsi définies. Le présentation de ces traces numériques peut également, et tout simplement aider un apprenant à se situer dans son activité. C'est ce que propose de faire le système *PIXED* par exemple (Héraud *et al.*, 2004). Ce système propose aux apprenants de visualiser leur *état d'avancement* au sein d'un *réseau de notions* à acquérir. Cours, activités pédagogiques et tests de connaissances forment les nœuds du réseau en question qui incite l'apprenant à évaluer par lui-même son niveau de maîtrise des notions requises pour continuer les activités. Le réseau en lui-même est annotable et sert également à accéder aux différentes ressources ou activités pédagogiques qui y sont symbolisées (Figure 2.27).

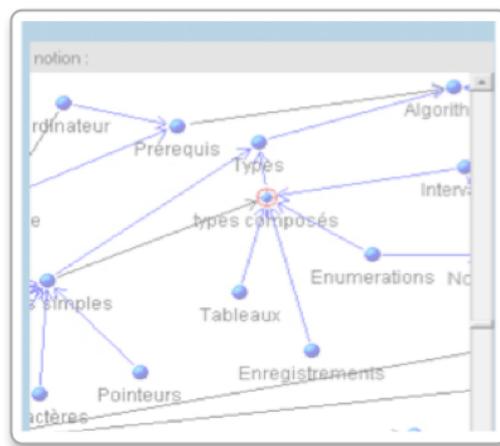


Figure 2.27 : Réseau de notion de *PIXED* - Tiré de (Héraud *et al.*, 2004)

D'autres systèmes du même type ont été proposés depuis l'apparition des EIAH. C'est le cas de *SHERLOCK*, que nous avons évoqué précédemment, de *SHERLOCK II* (Katz et al., 1992), ou encore du « journal d'apprentissage » de (Carroll et al., 1996). L'effort de recherche consenti dans le domaine des EIAH, pour développer les systèmes traçants (quel que soit leur type) est à la hauteur du rôle important que les traces numériques peuvent occuper dans des activités pédagogiques : « *Learning histories can help students and professionals make more effective use of digital library searching, word processing tasks, computer-assisted design tools, electronic performance support systems, and web navigation* » (Plaisant et al., 1999). D'un certain point de vue, on peut même dire des traces numériques elles-mêmes qu'elles représentent un véritable *objet pédagogique* (Choquet et Iksal, 2007 ; Laflaquière et al., 2007) dont tout le potentiel n'a pas encore été exploité.

Assister le groupe dans son activité collective d'apprentissage

Le domaine des EIAH est également pionnier en ce qui concerne l'assistance explicite des activités collaboratives. Au-delà des processus métacognitifs qu'elles peuvent susciter, les traces d'interaction ont un rôle à jouer dans les situations d'apprentissage collaboratif. Pour (Jermann et al., 2001 ; Jermann, 2004) en effet un apprentissage collaboratif satisfaisant est un apprentissage permettant à un apprenant de *collaborer* et à *apprendre à collaborer*. Les environnements d'apprentissage doivent donc guider, soutenir et assister l'apprentissage collaboratif en régulant, modérant et conseillant l'apprenant aussi bien dans sa tâche d'apprentissage que dans sa tâche de collaboration (Ollagnier, 2006).

C'est ce qu'affirment également les concepteurs de *SimPLE* (*Simulated Process in a Learning Environment*), système qui enregistre, conserve et surtout met à disposition des autres apprenants et du tuteur, les traces d'activité de ses utilisateurs (Plaisant et al., 1999 ; Rose et al., 2000). L'activité en question étant la conception de montage en physique, le module « historien visuel » permet, outre le partage, le rejeu, l'annotation et l'édition des morceaux de l'histoire en question (Figure 2.28).

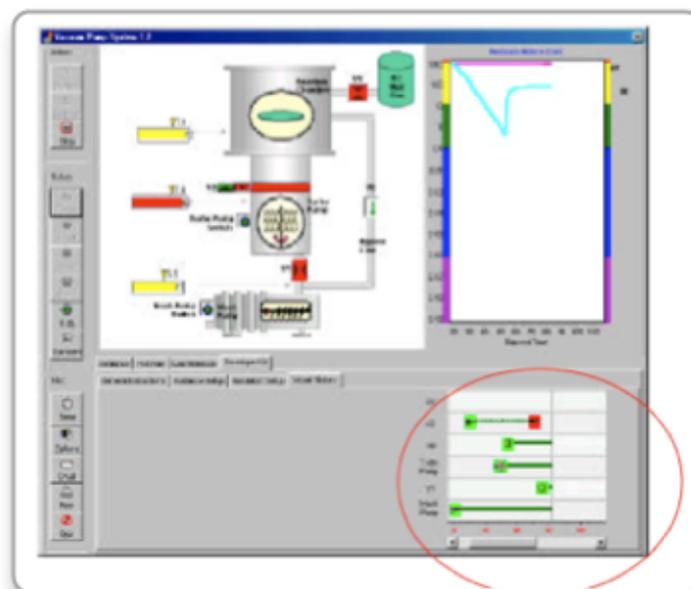


Figure 2.28 : Interface de *SimPLE* - Tiré de (Plaisant et al., 1999, p.349)

Une des pistes explorées pour renvoyer à un collectif d'apprenants une représentation exploitable de leur activité, est d'utiliser des indicateurs, utilisés habituellement dans les travaux d'analyse des activités collaboratives. C'est le cas pour le système *DIAS*⁷² (Bratitsis et Dimitracopoulou, 2006) par exemple, qui permet la visualisation d'indicateurs (degrés d'interaction, taux de contribution) par les apprenants eux-mêmes. Ces indicateurs sont censés permettre aux apprenants d'évaluer non seulement leur contribution relativement au travail de groupe, mais de prendre conscience du comportement du groupe dans son ensemble. Les apprenants seraient alors à même de modifier leurs comportements pour tenter de se rapprocher des objectifs de collaboration qui leur sont fixés.

2.4 Conclusion

Nous avons organisé la présentation de ce travail bibliographique autour des multiples finalités attribuées aux traces numériques d'interaction. Pour conclure ce chapitre nous allons faire quelques remarques pour donner un autre point de vue sur l'ensemble des systèmes traçants et travaux que nous venons d'évoquer : sur un *plan technique* puis sur un *plan conceptuel et méthodologique*.

2.4.1 Les trois composantes d'un système traçant

2.4.1.1 Collecter des traces numériques d'interaction

Historiquement, ce qui constitua le premier défi pour les chercheurs fut la mise au point de moyens techniques de récupérer des traces d'interaction. La première technique employée a été de *modifier des applications existantes*, comme ce fut le cas pour l'éditeur de texte *Zmacs*, modifié par (Hill *et al.*, 1992) pour leur « *Edit Wear* » et « *Read Wear* ». Bien que nécessaire parfois, la modification d'une application est une tâche souvent difficile, voire impossible (pour des problèmes de droit d'accès au code source), et dans tous les cas très coûteuse en termes de réalisation (codage, *reverse engineering*). Par la suite, l'évolution des environnements informatiques⁷³ a fourni de nouvelles sources de collecte. Nous avons, dans notre présentation, largement évoqué les *logsfiles*, ces fichiers utiles aux programmes pour réaliser des diagnostics de leur propre fonctionnement, mais également utilisés par les développeurs d'applications pour déboguer leurs codes ou pour trouver une solution à une situation de blocage de leurs programmes. Ces fichiers permettent de constituer plus « facilement » certaines traces numériques d'interaction. Sur le plan du développement, l'avantage est évident : nul besoin de créer ces *logs*, il suffit de développer les moyens de conservation et d'accès nécessaire à leur exploitation. C'est le cas, par exemple avec *logs* de navigation sur le *Web*, collectés depuis le *client* (Choo *et al.*, 2000) ou depuis le *serveur Web* (Spiliopoulou, 2003 ; Jansen et Pooch, 2000).

Quel que soit le dispositif, les performances actuelles des machines utilisées rendent quasiment transparent l'enregistrement continu d'une multitude de *logs* pendant l'activité, sans interférer avec celle-ci. La contrepartie de cette facilitation (car il y en a une), est que *primo* les informations recueillies sont *totale-ment dépendantes* du modèle de conception du système dont les *logs* sont issus, *deuxio* un énor-

⁷² *Discussion Interaction Analysis System*

⁷³ Notamment l'évolution que constitue la programmation objet d'une part et les composants modulaires d'interface d'autre part.

me volume de données est à traiter à la sortie de ces enregistrements : « *unfortunately this [method] produces voluminous data* » (Siochi, 1991, p.309) ; enregistrements qu'il faut de plus « nettoyer » d'un certain nombre d'enregistrements non « significatifs » (Weinreich *at al.*, 2006b), ou encore anonymiser en respectant le caractère *privé* de certaines navigations par exemple (Hawkey et Inkpen, 2005).

2.4.1.2 Transformer les traces numériques d'interaction

Comme nous venons de le dire, les traces d'interaction constituées de *logs* d'applications ne sont pas forcément exploitables en tant que telles (Kort et DePoot, 2005). Ces *logs* sont des données très, voire trop, précises et de bas niveau concernant l'interaction, et il faut trouver un « compromis entre granularité et précision » (Dragunov et *al.*, 2005) pour faire en sorte que les traces obtenues soit interprétables. De nombreuses techniques visent ainsi à *transformer*, d'une manière ou d'une autre, des traces d'interaction de bas niveau d'abstraction en traces d'interaction interprétables. C'est la deuxième composante des systèmes traçants, qui représente l'étape fondamentale qui consiste à « donner du sens » aux traces numériques. Parmi les travaux et approches que nous avons évoqués, de *plusieurs techniques* ont été explorées pour atteindre cet objectif. Voici les plus importantes d'entre elles.

La transcription manuelle

Certainement considérée en informatique comme un « dernier recours » la transcription manuelle de traces numériques d'interaction n'est pas rare dans les travaux de la littérature. Il semble même qu'elle soit la seule technique permettant d'obtenir réellement des traces d'interaction qui ont du sens à un certain niveau d'abstraction. Dans bien des cas cette contribution « humaine » au processus de traitement est minimisée, mais elle est parfois explicitement soutenue par une méthodologie (ou une instrumentation) particulière comme c'est le cas avec un « *Protocol Transcript* » dans (Pirolli et *al.*, 2002).

L'association à une autre source d'information

Une autre manière de transformer une séquence d'enregistrements bruts tels que les *logs*, est de l'associer à une autre source d'information. Cette technique est généralement réservée à une problématique d'analyse. L'exemple le plus courant est l'association d'une trace d'interaction à un enregistrement vidéo (Macleaod et Renger, 1993 ; Sanderson et Fisher, 1994). Cette association nécessite la *synchronisation temporelle* des supports et la création de *fonctionnalités* de navigation, pour que l'utilisateur de la trace puisse effectuer une exploration croisée, *i.e.* se déplacer dans la vidéo et accéder à la trace correspondante et inversement.

Les techniques de visualisation

Les techniques de visualisation partent de l'idée que si les données récoltées ne sont pas traitables sous leur forme textuelle car trop nombreuses pour être lisibles, il est possible de les *représenter graphiquement*. On compte alors sur les propriétés visuelles du résultat et sur les capacités d'interprétation de l'utilisateur de la trace pour lui « donner du sens ». L'ensemble de ces techniques fédère une petite communauté de chercheurs spécialisés dans les *techniques de visualisation* de ce type de données, notamment autour de la conférence *VISVIP* (Cugini et Scholtz, 1999). Nous devons ici avertir le lecteur que par *techniques de visualisation*, nous entendons les techniques qui permettent de visualiser

des données brutes, pour en faire « émerger visuellement du sens » et non une mise en forme pour la présentation visuelle d'une trace destinée directement à une interprétation globale de la trace.

Les techniques de réécriture automatique

Pour réduire le volume des données tout en conférant aux traces un niveau d'abstraction supérieur, bon nombre de travaux se concentrent sur des techniques permettant de *réécrire* la trace automatiquement. Un premier niveau de réécriture est celui du *filtrage*, c'est à dire des techniques qui visent à supprimer de la trace d'interaction des informations « non pertinentes » ou « redondantes ». Bien que pouvant comporter quelques pièges, ce premier niveau reste simple comparé au second qui consiste non pas à supprimer, mais à *composer* des éléments entre eux.

Ces *compositions* ont pour but de faire apparaître dans la trace des éléments de *plus haut niveau d'abstraction*. Dans la très longue liste des techniques utilisées pour réaliser ce type de composition nous pouvons citer (de manière non exhaustive) : les techniques de *clustering*, de *classification* ou de *regroupements automatiques* (Srivastava *et al.*, 2000) et les techniques (principalement empruntées au *datamining*) de détection de *séquences*, *patterns* ou *motifs temporels* qui constitueront à leur tour des éléments remarquables d'une trace numérique de plus haut niveau d'abstraction : matrice de fréquence de transition⁷⁴ (Vortac *et al.*, 1994), *Pathfinder Scaling Algorithm*⁷⁵ (Cooke *et al.*, 1996), chaînes de *Markov*, cycles de fréquences ou encore les algorithmes de détection des *Maximal Repeating Pattern*⁷⁶ (Cuomo, 1994), réseau de Petri (Rauterberg et Fjeld, 1998), *Temporal-pattern* ou *T-pattern* (Magnusson *et. al.*, 2000), *etc.*

Les techniques de dénombrement et indicateurs statistiques

Une autre piste à avoir été explorée est celle des *statistiques descriptives*. Toujours dans le but de transformer la trace afin d'en extraire son sens éventuel, les chercheurs se sont attachés à créer des *indicateurs* basés sur des mesures *quantitatives* des constituants de la trace numérique. Ces indicateurs sont ainsi censés donner une valeur *qualitative* à partir de mesures *quantitatives*. De nombreux indicateurs, parfois très complexes ont été mis au point pour tenter de décrire l'utilisation ou les usages de navigation des internautes. C'est le cas dans le travail de (Oberndorf *et al.*, 2007) pour la mise au point d'un indicateur statistique pertinent du taux de réitération des visites de pages ou de sites *Web*, ou dans le travail de (Jermann *et al.*, 2001, 2004) pour la mise au point d'indicateurs de la bonne - ou moins bonne - collaboration au sein du travail collectif.

L'implication de l'utilisateur

L'ultime manière de « donner du sens » à une trace d'interaction relativement à l'activité qui a été concrètement réalisée est encore de poser la question à l'auteur de l'activité. Cette idée est diversement exploitable. Dans (Hawkey *et al.*, 2006) par exemple, l'utilisateur doit dire ce qu'il fait au fur et à mesure de son activité en déclarant, dans un « journal de bord » de sa navigation (Figure 2.29), à quelle tâche il se consacre à chaque nouvelle opération. Dans cet exemple précis, l'utilisateur concerné peut également donner des informations complémentaires *a posteriori* d'une session, toujours à travers

⁷⁴ Entre des actions utilisateur, ou des comportements observés.

⁷⁵ Utilisé par exemple pour déterminer le chemin « le plus fort » dans un parcours sur le *Web* par exemple.

⁷⁶ Correspond à la plus longue chaîne se répétant dans cette même séquence.

son journal de navigation. Dans des situations extrêmes on peut enrichir, voire créer, une trace d'interaction sur la base des déclarations d'un utilisateur. On fera alors appel à des méthodes de *verbalisation*, d'*explicitation* (Weinreich *et al.*, 2006a)⁷⁷ et/ou d'*auto-confrontation* (Kröner *et al.*, 2006). Cette dernière solution, si on peut en attendre des résultats d'une grande richesse (puisque la trace numérique représentera vraiment ce que l'utilisateur a fait ou voulu faire), a en revanche un gros défaut : la trace d'interaction y perd son aspect non-invasif.

Window ID	Date / Time	Page Title	URL	Privacy Level
263892	3/15/2005 00:08:43.25	Google	http://www.google.ca/	Don't Save
264016	3/15/2005 00:08:43.99	Google	http://www.google.ca/	Don't Save
264454	3/15/2005 00:08:42.15	Google Search: a	http://www.google.com/search?source=webclient	Public
329560	3/15/2005 00:08:38.80	Google	http://www.google.ca/	Don't Save
329560	3/15/2005 00:09:02.82	Google Search: b	http://www.google.com/search?source=webclient	Public
461094	3/15/2005 00:08:42.83	Google	http://www.google.ca/	Don't Save
461094	3/15/2005 00:09:12.79	Google Search: d	http://www.google.com/search?source=webclient	Public
876548	3/15/2005 00:08:41.79	Google	http://www.google.ca/	Don't Save
3406780	3/15/2005 00:14:17.67	Google	http://www.google.ca/	Don't Save
3406780	3/15/2005 00:14:24.34	Google Search: d	http://www.google.com/search?source=webclient	Public
3539624	3/15/2005 00:24:13.74	Google	http://www.google.ca/	Don't Save
3539624	3/15/2005 00:24:19.26	Google Search: atoney scott	http://www.google.com/search?source=webclient	See Public
3539624	3/15/2005 00:24:31.06	Google Search: atoney scott defiance	http://www.google.com/search?source=webclient	See Public
3539624	3/15/2005 00:24:38.47	Google Search: atoney scott defiance	http://www.google.com/search?source=webclient	See Public
3539624	3/15/2005 00:24:51.58	Google Search: atoney scott defiance calgary	http://www.google.com/search?source=webclient	See Public
132134	3/15/2005 00:38:45.40	Google	http://www.google.ca/	(null)
332134	3/15/2005 00:40:34.62	22-Searched:02	22-search for medical info:02	Private
332134	3/15/2005 00:40:43.08	22-Searched:02	22-search for medical info:02	Private
332134	3/15/2005 00:40:25.83	22-Searched:02	22-search for medical info:02	Private
332134	3/15/2005 00:40:37.02	22-Searched:02	22-search for medical info:02	Private
332134	3/15/2005 00:40:54.95	22-Searched:02	22-search for medical info:02	Private
332134	3/15/2005 00:41:14.76	22-Searched:02	22-search for medical info:02	Private
332134	3/15/2005 00:44:23.24	22-Searched:02	22-search for medical info:02	Private
197892	3/15/2005 09:27:52.95	Google	http://www.google.ca/	(null)
197892	3/15/2005 09:28:00.92	Canada411	http://www.canada411.com	(null)
197892	3/15/2005 09:28:03.39	http://canada411.yellowpages.ca/	http://canada411.yellowpages.ca/	(null)
197892	3/15/2005 09:28:03.68	Canada411	http://canada411.yellowpages.ca/search/business	(null)
197892	3/15/2005 09:28:20.39	Canada411	http://canada411.yellowpages.ca/search/business	(null)
197892	3/15/2005 09:28:22.43	Canada411	http://canada411.yellowpages.ca/search/business	(null)
1705624	3/15/2005 11:56:27.34	Google	http://www.google.ca/	(null)
1705624	3/15/2005 11:59:12.58	http://www.google.ca/search?source=backup	http://www.google.com/search?source=backup	(null)
1705624	3/15/2005 11:59:22.63	Backing up the Windows registry	http://support.microsoft.com/SUPPORT/registry	(null)
1705624	3/15/2005 11:59:10.99	Google Search: backup registry	http://www.google.ca/search?source=backup	(null)
1705624	3/15/2005 11:59:24.80	Google Search: windows registry regy	http://www.google.ca/search?source=backup	(null)
1705624	3/15/2005 11:59:30.07	Windows Registry help	http://www.computerpage.com/registry.htm	(null)

Figure 2.29 : Journal de navigation – Tiré de (Hawkey, 2006, p.2).

2.4.1.3 Présenter les traces numériques d'interaction

Une fois le résultat du traitement des traces d'interaction obtenu, reste à en donner accès à leurs utilisateurs, ce qui constitue la troisième composante des systèmes traçants. L'interprétation des traces dépendra en effet autant de ce qu'elles contiennent que de la manière dont ce contenu est accessible, ce qui exige « *a well-designed user interface to present this information* » (Shneiderman, 1998). En fonction de leur *finalité* et du *contexte* d'exploitation, mais également de la nature de l'activité, de l'expertise de l'utilisateur de la trace, *etc.*, les modalités d'accès aux traces en question peuvent varier fortement. Cette variabilité repose sur les *multiples dimensions* de la présentation des traces à leurs utilisateurs, dont nous proposons ci-après une description des principales.

Synoptique (histoire interactionnelle vs objets enrichis)

Lorsque les traces numériques d'interaction sont présentées à un utilisateur c'est globalement pour lui permettre de *se la représenter*. Comme nous l'avons vu dans les sections précédentes, il y a deux façons d'y parvenir. De manière *implicite* tout d'abord, avec des traces qui viennent enrichir des objets de l'interface ou bien modifier l'interface elle-même (Hill *et al.*, 1992 ; Wexelblat et Maes, 1997, 1999). De manière *explicite* ensuite, avec des traces présentées dans leur extension temporelle

⁷⁷ Les sessions d'observations sont encadrées par deux entretiens de 90 minutes censées permettre d'enrichir les traces d'informations plus qualitatives. Les auteurs notent toutefois que ces informations ne sont pas suffisantes pour remplir le rôle qu'on leur avait attribuer.

de manière à « figurer une activité ». C'est le cas par exemple avec l'activité d'apprentissage représentée sous forme de réseaux ou diagrammes dans *Pixed* (Héraud, 2004) ou *DREW* (Baker *et al.*, 2003) ou avec l'activité de navigation représentée sous forme de carte dans *Padprint* (Hightower *et al.*, 1998) et *Footprint* (Wexelblat *et al.*, 1999), de graphe dans *Abstract* (Georgeon *et al.*, 2006) et *Lifeline* (Plaisant *et al.*, 1996, 1998), ou encore sous forme de table avec *ColAT* (Avouris *et al.*, 2007).

Échelle temporelle

Un des facteurs primordiaux pour la présentation d'une trace numérique est la *temporalité* : sa temporalité propre d'une part, celle de son exploitation d'autre part. Concernant leur temporalité propre, les traces numériques peuvent représenter une plus ou moins grande période de temps, à l'échelle de l'activité tracée : de quelques minutes à quelques heures pour des sessions de navigation *Web* (Wexelblat et Maes, 1997, 1999), de quelques heures à quelques semaines pour des activités plus complexes (Héroult *et al.*, 2004), voir à l'échelle de toute une vie (Plaisant *et al.*, 1996). Concernant cette fois la temporalité de leur présentation, les traces numériques d'interactions peuvent être présentées en *temps réel* durant la réalisation de l'activité lorsqu'il s'agit de *supporter* celle-ci (Héroult *et al.*, 2004, Plaisant *et al.*, 1996). Au contraire, dans la plupart des approches concernant l'*analyse* à base de traces, la présentation se fait *a posteriori*, quand l'activité sera terminée (Avouris *et al.*, 2004, 2005).

Partage et accès multiples

Dans de multiples situations d'exploitation, y compris en temps réel, les traces numériques d'interaction peuvent exiger d'être présentées de manière partagée. C'est le cas par exemple pour la trace qu'un apprenant aurait à partager en temps réel avec son tuteur, aussi bien que des traces partagées *a posteriori* entre plusieurs analystes. Nous pouvons donc distinguer deux cas de figures ne comportant pas tout à fait les mêmes difficultés : le cas de la présentation à un utilisateur individuel (Hightower *et al.*, 1998), et le cas de la présentation à plusieurs utilisateurs⁷⁸ dont l'auteur de la trace peut lui-même faire partie (Plaisant *et al.*, 1999).

Interactivité

La dernière dimension de la présentation des traces numériques que nous pouvons évoquer est celle de leur *interactivité*. Cette interactivité peut avoir un rôle crucial dans l'interprétation des traces numériques d'interaction, et il n'existe *pas a priori d'unique visualisation* répondant à tous les besoins (MacGavin *et al.*, 2006), d'où la création et le développement de systèmes de visualisation « *actifs* » grâce auxquels un utilisateur peut naviguer dans la trace. Au-delà de la simple navigation, ont donc également été développés des systèmes traçants « *inter-actifs* » laissant la main à l'utilisateur dans le choix d'une visualisation et de son paramétrage. Pour ce qui est des présentations « *actives* », nous pouvons y considérer celles des logs de navigations sous formes d'historiques navigable, que cela soit pour l'analyse (Cadez *et al.*, 2000 ; Chen *et al.*, 2004 ; Youssefi *et al.*, 2004) ou pour un soutien à l'activité de l'utilisateur (Hightower *et al.*, 1998, Wexelblat *et al.*, 1998).

⁷⁸ Sans rentrer dans les détails nous remarquerons que le problème d'une présentation de trace à plusieurs utilisateurs en temps réel se rapproche d'une problématique d'*awareness*.

Si l'interactivité la plus riche est appelée de tous leurs vœux par les chercheurs, il reste que les développements demandés sont conséquents, ce qui explique que beaucoup de systèmes traçants existants n'ont été sur ce point implémentés que *partiellement*, relativement aux préconisations de leurs auteurs. Parmi les exemples disponibles, nous pouvons citer l'outil de visualisation *VOLTS : Visualization of Logged Time Series* (Hawkey, 2006), lequel a été développé pour répondre de manière relativement générique à des besoins de présentation interactive de traces numériques. Cet outil permet de présenter des *logs* de navigation pour une période de temps donnée, sous la forme d'une sorte de diagramme de Gantt. Le temps est représenté horizontalement, et on distingue verticalement différents flux « d'information » (correspondant ici à des fenêtres ouvertes dans un navigateur Web), dans lesquels sont visualisés des événements sous forme de petites boîtes de couleur. Ce qui est intéressant c'est que cet outil est censé autoriser l'utilisateur de la trace (ici un analyste de l'activité) à *sélectionner dynamiquement* les attributs qu'il souhaite utiliser pour l'affichage graphique, et à ajuster le codage de couleurs en fonction des valeurs qu'il veut mettre en avant.

En termes de présentation de traces, il est possible également de s'inspirer des travaux réalisés sur la visualisation des *traces d'exécution* de programme, ou sur celle de *résumés* de séquences vidéo (Nunes et al., 2007). Confrontés à des problèmes similaires en terme d'interactivité de la présentation, les chercheurs développent des solutions adaptables. McGavin et al., (2006) par exemple propose des *plugins* de visualisation à destination des développeurs, fournissant différentes options de filtrage, synchronisation, et paramétrage gérables directement pas l'utilisateur.

2.4.2 Des traces de l'interaction à celles de l'activité

Si sur le plan technique la question qui se pose est celle de la capacité à *recupérer* des informations qu'il s'agit ensuite de *traiter* au mieux, d'un point de vue conceptuel la question est plutôt de savoir comment leur *donner du sens*. On remarquera dans les travaux présentés une évolution dans le temps du *niveau d'interprétation* des traces produites par les systèmes traçants. S'agissant des traces utilisées comme support d'analyse, on note ainsi une évolution de l'objet étudié à travers l'analyse des traces d'interactions : les *interactions* elles-mêmes au début, puis les *comportements*, et enfin les *activités*. Pour soutenir ce changement de niveau d'interprétation, on attend des traces qu'elles atteignent un niveau d'abstraction qui ne correspond pas à ce que les *logs* permettent en général de fournir, ce qui constitue une sérieuse limite au développement de systèmes traçants plus intéressants, avec des traces numériques plus riches. Bien que pointé par certains auteurs (Dillenbourg, 1999 ; Avouris et al., 2005 ; Soller et al., 2005), ce problème fondamental ne semble pas toujours être un obstacle comme en atteste le titre de la contribution de (Roussel et al., 2006) qui affirme que « *All you need is logs* ». De la manière d'aborder la problématique du sens dont les traces numériques d'interaction peuvent être porteuses, découlent deux visions, deux conceptions différentes de ces traces.

2.4.2.1 Une vision classique des traces numériques

Dans l'absolu, on pourrait dire de la première qu'elle a une vision de la trace numérique qui la rapproche de la notion de la trace ordinaire, matérielle. De ce point de vue, les traces numériques (comme les traces matérielles) existent déjà, elles sont présentes dans l'environnement numérique sous forme de *logs* qu'il suffit de conserver pour en constituer des traces complètes. Par conséquent, on cherchera à « donner du sens » à ces traces *a posteriori*, c'est-à-dire qu'on devra traiter ces traces *a posteriori* pour

augmenter leur niveau d'abstraction, pour leur attribuer une sémantique qu'elles n'ont pas au départ, et permettre ainsi une interprétation *qualitative* des comportements ou des activités, d'où les traitements statistiques divers et variés, les indicateurs, *etc.* De ce point de vue, la manière de traiter les traces numériques d'activités humaines ne diffère pas grandement de la manière dont on traite les traces d'exécution d'une application informatique pour les visualiser. On ne parle plus ici d'interactions impliquant un utilisateur mais d'évènements machines, ces traces d'exécution étant elles-mêmes constituées de *logs* : « *descriptions d'évènements qui interviennent durant l'exécution d'un programme [...] comme la création d'objet, l'appel de méthodes, l'accès ou la modification d'un champ, le chargement de classe, ou le déclenchement d'une exception* »⁷⁹ (MacGavin et al., 2006, p.154). Les techniques, notamment statistiques, que nous avons évoquées sont utilisées aussi bien pour caractériser des « comportements » d'utilisateurs que des « comportements » de systèmes (Reiss, 2003 ; Jones et al., 2004 ; Systä, 2005).

Dans ce contexte il n'est pas surprenant de constater que des auteurs ayant mis en œuvre de gros moyens de traitements statistiques pour mettre en évidence tel ou tel motif ou pattern dans une trace, essaient *a posteriori* de « plaquer un modèle théorique » sur des « découvertes empiriques ». Les modèles théoriques, qu'ils soient comportementaux ou cognitifs, se prêtent mal à l'exercice, et le passage du quantitatif au qualitatif ne peut se faire uniquement par le calcul : « *as aims and tasks of the users often stay below the surface. This makes their contextual interpretation inherently difficult and additional qualitative information is needed to support a detailed task-related evaluation of the data* » (Weinreich, 2006, p.141).

2.4.2.2 Une vision alternative des traces numériques

Une seconde conception des traces numériques consiste à « donner du sens » aux traces numériques *a priori*, c'est-à-dire enrichir *a priori* la sémantique des éléments qui seront tracés. Il s'agit globalement d'inverser la démarche et de ne plus travailler à partir de données disponibles dans les systèmes informatiques pour d'autres raisons, mais d'une analyse de l'activité pour déterminer quels sont les éléments qui doivent être tracés. Il ne s'agit pas d'ignorer la nécessité des traitements *a posteriori*, ni de rejeter les différentes techniques que nous venons d'évoquer ci-dessus. En revanche, il s'agit de changer de postulat de départ : les *logs* ne sont pas suffisants (Avouris et al., 2005) pour fournir une trace numérique à la sémantique riche qui *représente l'activité* d'un utilisateur. C'est en effet pour un tel niveau d'interprétation que l'on cherche à créer des traces. De ce point de vue les traces numériques ne sont pas présentes « par défaut » dans l'environnement, et il ne suffit pas de les récupérer. Il faut les créer *a priori*. C'est exactement la position défendue par l'approche des Systèmes à Base de Trace modélisée que nous allons décrire dans le chapitre suivant, approche dans laquelle s'inscrit ce travail de thèse.

⁷⁹ Notre traduction.

Chapitre 3

Les Systèmes à Base de Traces Modélisées

Résumé du chapitre

Ce chapitre présente l'approche des *Systèmes à Base de Traces modélisées* (SBTm) qui propose un cadre formel pour mettre en œuvre l'idée que les traces numériques doivent être explicitement définies *a priori* dans un *modèle de trace*. Les premiers travaux de l'équipe Silex (Liris) défendant cette idée ont été menés dans le cadre de l'approche MUSETTE (Modéliser les USages Et les Tâches pour Tracer l'Expérience). L'approche consistait à générer au cours de l'utilisation d'un environnement informatique une trace numérique structurée en états et transitions et composée d'éléments définis dans un *modèle d'utilisation*. La structuration obtenue devait permettre d'identifier dans la trace des *épisodes réutilisables* par des agents assistants. L'approche des SBTm qui lui succède dans les travaux de l'équipe Silex, a *généralisé* et *formalisé* les concepts qui étaient directement mis en œuvre dans MUSETTE. La *trace modélisée* qui en est le principal, est définie comme l'association d'une collection d'observés à un modèle explicite de cette collection appelé *modèle de trace*. Le modèle de trace ne surdétermine pas la trace modélisée elle-même, mais fixe son « vocabulaire ». Il détermine « en intention » toutes les traces possibles, mais ne prescrit pas la forme ou la structure finale d'une trace générée, ni ne fige son interprétation. Sur un plan technique l'approche des SBTm est sur le point de se doter d'un *framework* informatique permettant de traiter de manière générique les traces modélisées grâce à leurs modèles. Sur un plan méthodologique, l'approche des SBTm ne dispose pas jusqu'alors d'une *méthodologie* de modélisation propre, *i.e.* d'une démarche spécifique et rationalisée de conception d'un modèle de trace. Dans les applications de l'approche, la détermination du modèle de trace a jusqu'ici été improvisée en fonction des contraintes de terrain. La problématique générale de cette thèse se concrétise en un objectif spécifique : établir une *méthodologie de modélisation des traces numériques* propre aux Systèmes à Base de Traces modélisées.

3.1 Introduction

Dans le prolongement de l'état de l'art, ce court chapitre sera consacré à la présentation de l'approche des *Systèmes à Base de Traces modélisées*, proposée par l'équipe Silex du Liris⁸⁰ et dans laquelle s'inscrit notre travail de thèse. Les différents travaux de l'équipe ont abouti tout d'abord à une *généralisation* des Systèmes à Base de Trace, puis à une *formalisation* de l'approche des *Systèmes à Base de Traces modélisées* qui repose sur les concepts clefs de *trace modélisé* et de *modèle de trace*. Aujourd'hui, cette approche constitue un *cadre conceptuel* pour penser de manière *générique* les Systèmes à Base de Traces, quant à la formalisation proposée dans le cadre des Systèmes à Base de Traces modélisées, elle a rendu possible le développement (encore en cours) d'un *framework informatique* dédié à la gestion de toute trace numérique répondant aux critères de l'approche. Le développement de l'approche a été réalisé durant deux séries de travaux. Les premiers travaux, ceux de l'approche MUSETTE, ont fait un premier pas important en imposant l'idée de traces numériques structurées et réutilisables, définies dans un modèle. Puis, c'est en généralisant les concepts appliqués dans cette première approche qu'a été construite et définie l'approche des Systèmes à Base de Traces modélisées. Le chapitre débute donc par une présentation de l'approche MUSETTE (section 3.2) qui constitue la genèse de ce qu'est aujourd'hui l'approche des Systèmes à Base de Traces modélisées, qui sera présentée dans un deuxième temps (section 3.3).

3.2 L'approche MUSETTE

L'objectif initial de l'approche MUSETTE (Modéliser les USages et les Tâches pour Tracer l'Expérience) était de tracer, puis réutiliser *l'expérience d'utilisation* d'un système informatique. La notion même *d'expérience d'utilisation*, issue de travaux antérieurs (Prié 1999 ; Prié *et al.*, 1999), fût le point de départ du travail de recherche. Il s'agissait de pouvoir *décrire* l'expérience d'utilisation, de façon à la *matérialiser dans une trace* générée par les interactions de l'utilisateur avec son environnement informatique. Entendue comme une *réalisation remarquable* d'une tâche donnée, l'expérience d'utilisation matérialisée en une trace, devenait alors *réutilisable* en tant que telle. Cette réutilisation pouvait avoir diverses finalités (Prié et Mille, 2000), mais c'est l'idée de réutiliser ces traces d'utilisation pour *l'assistance à l'utilisateur* qui fût retenue et constitua le principal axe de développement de l'approche (Champin *et al.*, 2001, 2002). Rapidement après les premières ébauches, l'approche fût consolidée par une définition plus rigoureuse des concepts utilisés (Champin et Prié, 2003 ; Champin *et al.*, 2003, 2004), et par sa mise en œuvre concrète (Laflaquière, 2003 ; Laflaquière et Prié, 2003 ; Laflaquière *et al.*, 2005 ; Laflaquière et Ciaccia, 2005). L'approche MUSETTE a ouvert des pistes de réflexion intéressantes quant à l'assistance à l'utilisateur, en s'inspirant notamment du *Raisonnement à Partir de Cas* (Aamodt et Plaza, 1994; Aamodt, 2001), mais a surtout préfiguré l'approche des *Systèmes à Base de Traces modélisées*.

⁸⁰ Laboratoire d'Informatique en Images et Systèmes d'information (LIRIS UMR 5205 CNRS).

3.2.1 Les principaux concepts de l'approche MUSETTE

Le principe de base de l'approche est relativement simple. Il s'agit de mettre en place un *agent observateur* de l'utilisation d'un système informatique qui génère une trace d'utilisation. Cette trace est définie de telle sorte qu'un *analyste de trace* puisse y identifier des *épisodes significatifs* représentant des *expériences d'utilisation*, lesquels sont alors *réutilisés* par des *agents assistants* pour soutenir l'utilisateur dans sa tâche.

3.2.1.1 Agent observateur, modèle d'utilisation et trace primitive

Afin de décrire l'utilisation dans les termes appropriés à l'identification d'expériences d'utilisation particulières, l'approche MUSETTE propose de les *modéliser explicitement*, ce qui revient à déterminer *a priori* les *éléments constitutifs* d'une trace numérique générée par l'utilisation d'un système informatique. Les éléments constitutifs ainsi déterminés sont appelés les *Objets d'Intérêt* et sont décrits dans un modèle d'utilisation qui regroupe deux types d'objets – des *entités* et des *événements* – ainsi que des *relations* (typées) posées entre ces objets. Les entités sont des objets présents dans l'espace d'interaction de l'utilisateur⁸¹ alors que les événements sont des changements de l'environnement opérés par le système lui-même ou par une action de l'utilisateur. L'objectif de la trace visée étant sa *réutilisation*, ces objets d'intérêts, dotés de leurs attributs, doivent impérativement être *signifiants* pour l'utilisateur aussi bien que pour le système informatique qui les manipule.

Une fois le modèle d'utilisation établi, un *agent observateur* peut être mis en place et *générer* concrètement une *trace d'utilisation* en enregistrant les instances d'objets d'intérêt définis, de façon à ce que la trace générée respecte le modèle. L'agent observateur repose sur un ensemble de *règles d'observation* qui définissent d'une part la manière d'instancier des objets d'intérêt dans la trace à partir des divers événements machine⁸², et d'autre part la manière de *structurer* la trace elle-même. La trace ainsi générée est appelée *trace primitive*, entendue comme la « première trace signifiante », aussi bien pour le système que pour l'utilisateur lui-même (Figure 3.1).

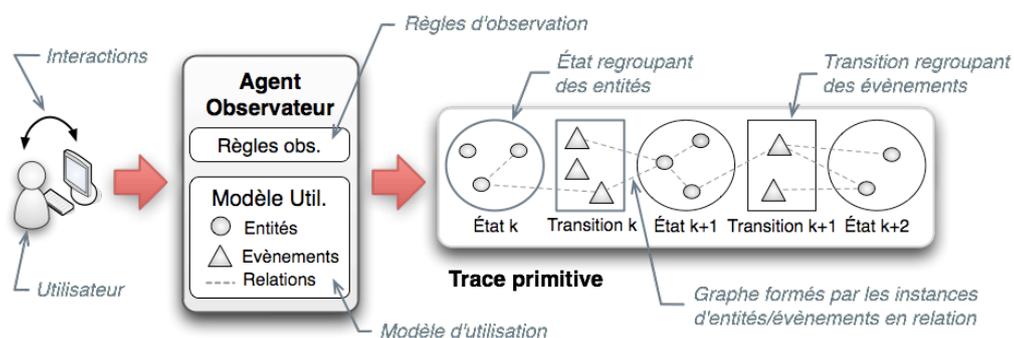


Figure 3.1 : Génération d'une trace primitives structuration en états et transitions.

La trace primitive générée se compose de deux types de *structures distinctes* : les *états* et les *transitions*. Les *états* qui regroupent des *entités*, représentent l'état du système observé (différent d'un « état-machine ») à un instant (ou une période) donné par rapport au modèle d'utilisation. Les *transi-*

⁸¹ Ne correspondant pas nécessairement à un objet du modèle de conception de l'outil manipulé (Laflaquière, 2003).

⁸² Ces événements machines peuvent être des événement directement (ou indirectement) accessibles dans le système, ou bien doivent être obtenus par la mise en place d'un code spécifique ou d'une modification de l'environnement lui-même.

tions regroupent quant à elles un ensemble d'évènements qui se produisent entre deux états. La trace primitive est donc *in fine* une *séquence alternée* d'états et de transitions, ces derniers structurant un *graphe* dont les *nœuds* sont des entités et des évènements étiquetés par l'ensemble de leurs attributs, et dont les *arcs* sont des relations.

3.2.1.2 Analyseur de trace, épisodes significatifs et signatures de tâches

La trace primitive ainsi obtenue se prête au traitement d'un *analyseur de trace*, qui a pour objectif d'en extraire des *épisodes significatifs*, *i.e.* des *segments* de la trace primitive correspondant à une expérience d'utilisation (liée à la réalisation d'une tâche spécifique), et potentiellement réutilisable en tant que telle. Ces épisodes sont identifiés grâce à des *signatures de tâche*. Une signature est un ensemble de points communs à un type de tâche particulier, impliquant les mêmes entités et/ou évènements, produits à un même moment ou dans un ordre particulier (en fonction des états et transitions) dans la trace primitive. On peut la voir comme un « motif » ou à un « pattern de tâche » (Paternò, 1994) partiellement instancié dans la trace. Au final, un segment est repéré et extrait de la trace par l'analyseur en tant qu'épisode significatif lorsque le segment contient *au moins une* signature de tâche (Figure 3.2). Ce processus alimente le système d'un certain nombre d'épisodes significatifs qui peuvent alors être exploités par des agents assistants.

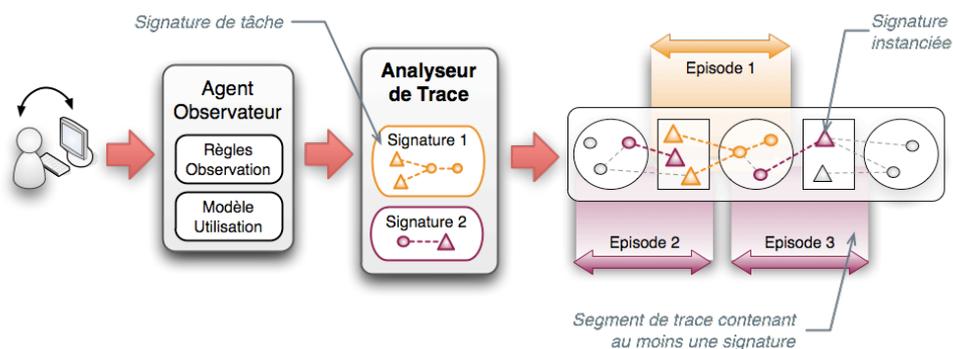


Figure 3.2 : Extraction d'épisodes significatifs, sur la base de signatures de tâches.

3.2.1.3 Agents assistants et réutilisation d'expérience d'utilisation

Plusieurs types d'agents assistants, réutilisant les épisodes significatifs, ont été imaginés. De manière classique un agent assistant « générique » prenant en charge plusieurs tâches *via* leurs *signatures*, peut explorer la trace courante et suggérer (ou accomplir), au fur et à mesure, des actions qui semblent pertinentes. On peut également proposer des agents « réactifs », mobilisés uniquement à la demande de l'utilisateur lui-même, et qui lorsqu'ils sont appelés tentent d'identifier la tâche engagée pour trouver dans la trace des épisodes similaires et s'en inspirer pour proposer automatiquement des opérations à effectuer pour prolonger l'action de l'utilisateur.

Ce type d'assistance (similaire à celle fournie par les systèmes conseillers) peut rapidement perdre de sa pertinence lorsque l'utilisateur est engagé dans une tâche complexe dont le succès ne dépend pas uniquement de la « bonne » manipulation de l'outil, mais des décisions et des choix qu'il fait relativement à l'environnement numérique dans lequel il est plongé *via* cet outil. Nous avons détaillé ce problème (Laflaquière et *al.*, 2005) en prenant comme exemple la veille informationnelle sur le *Web*, instrumentée d'une application de recherche et cartographie documentaire dédiée, *Human-links*TM.

Pour ce type de tâche, l'agent assistant peut ne pas avoir de suggestion pertinente à fournir relativement à la demande de l'utilisateur, ce qui n'empêche pas que la réutilisation des épisodes significatifs peut tout de même apporter de l'aide à l'utilisateur.

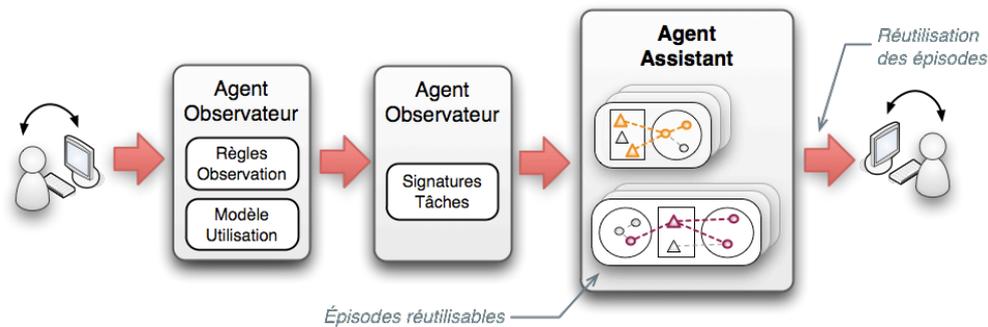


Figure 3.3 : Réutilisation(s) des épisodes significatifs par un agent assistant.

En effet, de façon plus pragmatique, l'agent peut simplement proposer à l'utilisateur de *consulter lui-même* les épisodes retrouvés. Seul l'utilisateur, en effet, est à même d'identifier clairement les épisodes qui correspondent le mieux à sa situation actuelle en tenant compte de critères qui ne sont pas accessibles au système. Dans ce cas d'assistance *médiée* ou *indirecte*, la présentation d'un épisode peut *inspirer une solution* à l'utilisateur qu'aucune déduction logique du système ne pourrait fournir sur la seule base de la trace. Notons qu'un des intérêts de l'approche tient au fait que l'interrogation de la base d'expérience est réalisée simplement en utilisant le système (sans faire appel à un langage de requête particulier), et réduisant le risque d'une aide hors de propos. De plus, il est tout à fait possible pour l'utilisateur de définir et rajouter au fur et à mesure de nouvelles signatures de tâches. L'agent peut de façon identique parcourir la trace à la recherche des épisodes contenant cette signature fraîchement définie pour les soumettre à l'utilisateur, ce qui donne une grande flexibilité au système.

3.2.2 Applications de l'approche MUNETTE

Le fonctionnement d'un système d'assistance basé sur des épisodes significatifs se rapproche de celui des systèmes de *Raisonnement à Partir de Cas* (Aamodt et Plaza, 1992 ; Aamodt, 2001), que certains travaux de recherche ont déjà tenté de transposer directement dans un contexte d'assistance à l'activité d'un utilisateur. C'est le cas par exemple des travaux de M. Jackinski et B. Trousse (1997, 1998, 1999) qui proposent d'extraire et de réutiliser de la navigation des internautes des « cas de navigation mutualisables ». Dans le cas de l'approche MUNETTE, la grande différence avec une approche de Raisonnement à Partir de Cas (RàPC) classique⁸³ repose sur le fait de ne plus considérer des *cas* dans un *cycle de RàPC*, mais d'une part des *traces*, d'autre part des *épisodes* pouvant – mais *pas* obligatoirement – être utilisés en tant que cas. On peut alors imaginer d'autres utilisations que le RàPC *stricto sensu*, *i.e.* comme nous l'avons rapidement évoqué : présentation à l'utilisateur, analyse, assistance indirecte, *etc.* Pour marquer cette distinction, l'adaptation des principes du RàPC a d'ailleurs donné lieu à la création du *Raisonnement à Partir d'Expérience Tracée* proposés dans (Philippon et al., 2005) et (Mille et al., 2006).

⁸³ Le RàPC se base sur la remémoration et la réutilisation de « cas » concrets d'utilisation de système (sous la forme Problème-Solution). La construction d'un système RàPC suppose donc de définir précisément quelle est la structure des cas, comment ceux-ci sont stockés, remémorés, utilisés, *au préalable* à la construction du système. Changer le modèle de cas revient à changer le système.

Outre ce développement théorique, divers travaux autour de MUSETTE ont exploité les principes généraux de l'approche et testé ses limites dans différents domaines applicatifs (Arana *et al.*, 2004 ; Mille et Prié, 2006 ; Laflaquière et Prié, 2009), dont celle proposée pour l'assistance à la veille sur le *Web* avec *Human-Links*TM, dont nous avons fait mention⁸⁴ (Laflaquière et Prié, 2003). De plus, la réutilisation des épisodes pour l'assistance a petit à petit été étendue à la *facilitation* de l'utilisation considérée au sens large (Mille et Prié, 2006) ainsi qu'à l'*analyse* à base de traces d'utilisation pour laquelle l'utilisateur de la trace n'est plus celui dont l'activité a été tracée. Bien que nous ne soyons pas rentrés dans tous les détails de l'approche⁸⁵ nous avons tenu à présenter les concepts importants de l'approche MUSETTE dans la première partie de ce chapitre car elle préfigurait⁸⁶ l'actuelle approche des *Systèmes à Base de Traces*.

3.3 L'approche des Systèmes à Base de Traces Modélisées

À la suite de l'approche MUSETTE, l'approche des Systèmes à Base de trace a été proposée comme une *généralisation* des systèmes exploitants des traces numériques comme *objets informatiques* à part entière. On peut considérer comme Système à Base de Trace tout dispositif qui implémente dans un système informatique, et pour tout ou partie, les trois fonctions suivantes (Figure 3.4).

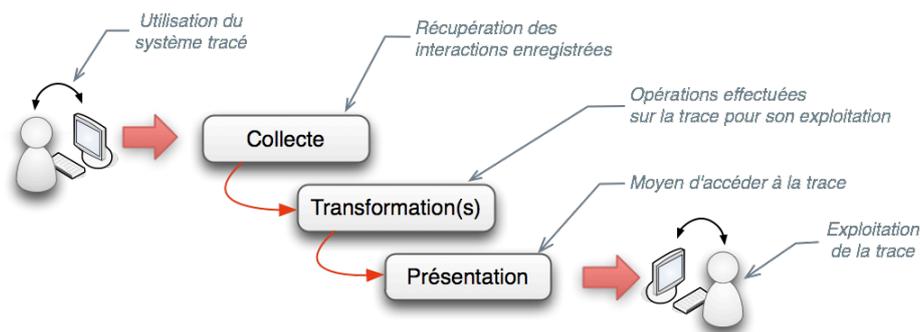


Figure 3.4 : Généralisation des systèmes traçants.

- Une fonction de collecte : qui alimente d'une manière ou d'une autre le dispositif à partir de données issues de l'interaction d'un utilisateur avec son système.
- Une fonction de transformation des traces numériques, qui globalement consiste à préparer les données à leur réutilisation en tant que trace.
- Une fonction de présentation des traces qui donne concrètement accès à la trace pour son exploitation (une visualisation, et le cas échéant fixe les moyens *d'interaction* avec la trace visualisée) ou qui fournit sous forme de *service*, des traces à une autre application (agents assistants, ou système de visualisation indépendant par exemple).

⁸⁴ Notre travail sur *Human-Links* a été réalisé à l'occasion d'un début de thèse CIFRE qui n'a duré que quelques mois entre 2003 et 2004, avant que l'entreprise partenaire (à l'origine du logiciel) ne soit contrainte à déposer son bilan.

⁸⁵ Détails qui seront prochainement disponibles sous la forme d'un chapitre de livre (Laflaquière et Prié, 2009).

⁸⁶ En fait, l'approche MUSETTE a été une ébauche théorique d'une approche plus globale (celle des Systèmes à Base de Traces), dont elle constitue *a posteriori* une *spécification*.

Nous reviendrons plus en détail sur ces différentes fonctions dans le cadre de ce qui va suivre, et pour le moment nous pouvons simplement signaler qu'un dispositif implémentant ces fonctions *concrètement* dans un système informatique, devra nécessairement mettre en place un certain nombre de fonctionnalités de stockage, de consultation, de calcul et de manipulation des traces. Il sera dans ce cas considéré comme un *Système de Gestion d'une Base de Trace* (SGBT). Au-delà de la *généralisation* que nous venons d'évoquer, l'approche que nous sommes en train de décrire propose surtout une *formalisation* que nous dénommerons l'approche des *Systèmes à Base de Traces modélisées* (SBTm) et dont l'approche MUSETTE est finalement une « spécialisation » ou une « instance ». Cette *formalisation* comporte donc un ensemble de définitions que nous allons exposer dans les sections à suivre.

3.3.1 Définition formelle du concept de trace modélisée

3.3.1.1 Trace numérique

On considère une *situation d'activité* dans laquelle un *utilisateur*, acteur de la réalisation de cette activité, exploite son environnement informatique. La réalisation en question est *réduite*, du point de vue des traces numériques qu'on peut en obtenir, aux *actions observables*. Les limites de cette réduction (de la situation considérée) sont fixées par les possibilités d'observation que nous accordons à l'instrumentation mise en place, et qui définit un *système observé*.

- Un utilisateur : sujet d'une activité qui réalise pour tout ou partie son activité dans un environnement numérique.
- Un système observé : ensemble des outils numériques utilisables pour mener à bien une activité, et instrumenté pour produire des traces numériques.

Dans ce contexte, une trace numérique sera constituée d'éléments issus de l'observation des interactions de l'utilisateur avec son environnement durant la réalisation de son activité, et que nous appelons des *observés*. Ces observés, bien qu'issus de l'observation du système utilisé, ne correspondent pas nécessairement à un événement machine, mais sont présents dans l'interaction et doivent être identifiables par le système. Du fait de la linéarité des interactions avec un environnement numérique (et donc de la nécessaire sérialisation des enregistrements), une trace numérique sera composée d'un ensemble d'observés *temporellement situés*, *i.e. organisés* et *liés* à une représentation du temps. Cette organisation peut-être directement séquentielle (chaque observé est suivi et/ou précédé par un autre) ou découler du caractère temporalisé des objets de la trace (chaque observé possède un identifiant temporel). Puisque nous nous situons dans la cadre d'une activité donnée, nous nommons *domaine temporel* la représentation du temps liée à l'activité tracée. À l'intérieur d'un domaine temporel donné, une trace décrit seulement un intervalle fini de temps. Nous appelons un tel intervalle *l'extension temporelle* de la trace. Formellement voici comment se définissent les différents concepts évoqués :

- Un Domaine Temporel : Un domaine temporel T est un ensemble dénombrable d'instants. Une fonction bijective $SuccT : T \rightarrow T$ associe à chaque instant son successeur, définissant un ordre total \leq_T dans T .
- Une Extension Temporelle : Étant donné un domaine temporel T , une extension temporelle ξ_T est un élément quelconque de $I(T)$, avec $I(T)$ l'ensemble de tous les intervalles finis sur T .

- Un observé temporellement situé : Est considéré comme observé toute information structurée issue de l'observation d'une interaction, temporellement situé dès lors qu'il est associé à une partie de l'extension temporelle de la trace à laquelle il appartient.
- Une trace numérique : On appelle trace une collection d'observés temporellement situés, structurée par leurs relations. Différents types de relations sont entretenus entre observés, la relation minimale étant la relation temporelle (liée à l'extension temporelle).

3.3.1.2 Modèle de trace et trace modélisée

Toute trace numérique étant constituée d'enregistrements, elle est nécessairement (et souvent implicitement) *prédéterminée*. L'approche des Systèmes à Base de Traces modélisées propose d'assumer cette prédétermination en définissant *a priori explicitement* les éléments constitutifs d'une trace dans un *modèle* qui en définit des types. Ce modèle de trace, associée à la trace elle-même permettra *a posteriori* l'exploitation de la trace en question, tant sur le plan des opérations informatiques que sur le plan de l'interprétation qu'on pourra donner de la trace concrètement générée.

- Modèle de trace : Un modèle de trace est un tuple $M_{TR} = (T, C, R, A, dom_R, range_R, dom_A, range_A)$ constitué :
 - o D'un domaine temporel T
 - o D'un ensemble fini C de types d'observés (ou classes) et un ordre partiel \leq_C défini sur C
 - o D'un ensemble fini R de types de relations, disjoint de C et un ordre partiel \leq_R défini sur R
 - o D'un ensemble A fini d'attributs disjoint de C et R
 - o De deux fonctions $dom_R : R \rightarrow C$ et $range_R : R \rightarrow C$ définissant le domaine et le co-domaine des types de relation
 - o De deux fonctions $dom_A : A \rightarrow C$ et $range_A : A \rightarrow D$ définissant le domaine et le co-domaine des attributs
 - o De plus, pour tout couple de relations $r1$ et $r2$ d'un modèle de trace, on doit garantir la propriété suivante : $r1 \leq_R r2 \Rightarrow dom_R(r1) \leq_C dom_R(r2) \wedge range_R(r1) \leq_C range_R(r2)$.
- Trace modélisée : l'association d'une collection d'observés temporellement situés, structurée par leurs relations, et d'un modèle explicite de cette collection d'observés. Dit autrement une trace modélisée est l'association d'une *trace numérique* et du *modèle de trace* qui permet d'en interpréter les observés. Formellement, une trace modélisée est un tuple $TR = (M_{TR}, \xi_T, O, \lambda_C, \lambda_R, \lambda_A, \lambda_T)$ constitué :
 - o D'un modèle de trace $M_{TR} = (T, C, R, A, dom_R, range_R, dom_A, range_A)$
 - o D'une extension temporelle ξ_T définie dans T
 - o D'un ensemble fini O d'observés disjoints de C , R et A
 - o D'une fonction totale⁸⁷ $\lambda_C : O \rightarrow C$ décrivant l'instanciation des observés
 - o D'une relation $\lambda_R \subseteq O \times O \times R$ décrivant l'instanciation des relations
 - o D'une fonction partielle⁸⁸ $\lambda_A : O \times A \rightarrow V$ décrivant l'instanciation des attributs
 - o D'une fonction totale $\lambda_T : O \rightarrow I(\xi_T)$ décrivant la temporalisation des observés.

⁸⁷ λ_C est défini sur chaque $o \in O$.

⁸⁸ $\lambda_A(o,a)$ peut être indéfini pour un certain $(o,a) \in O \times A$. Et V un ensemble de valeurs concrètes définies indépendamment de tout modèle.

Une Trace modélisée respectant son modèle de trace vérifie les contraintes suivantes :

- $\forall(o1, o2, r) \in \lambda_R, \lambda_C(o1) \leq_C \text{dom}_R(r) \wedge \lambda_C(o2) \leq_C \text{range}_R(r)$
- $\forall(o, a, v) \in \lambda_A, \lambda_C(o) \leq_C \text{dom}_A(a) \wedge v \in \text{range}_A(a)$

Notons que ce modèle définit des types d'observés, c'est-à-dire qu'il détermine les éléments constitutifs de la trace mais *ne détermine pas la trace directement*. Dit autrement, le *modèle de trace* définit le « vocabulaire » de la trace mais *ne la détermine pas complètement a priori*. Le modèle détermine « en intention » toutes les traces (respectant ce modèle) possibles. Quel que soit leur niveau d'abstraction (*i.e.* leur proximité aux événements informatiques), le modèle de trace en *typant* les observés, vient préciser leur sémantique et les moyens de les utiliser. Dans la définition formelle qui en est donnée ci-dessus, nous pouvons constater que le modèle de trace définit les *observés*, leurs *attributs*, ainsi que les *relations* qu'ils peuvent entretenir entre eux. On notera au passage qu'il existe une *hiérarchie de spécialisation* sur les types d'observés. Disposant des dernières définitions formelles de ces différents concepts (Settouti et *al.*, 2009), nous pouvons aborder la description plus concrète de ce qu'est un *Système à Base de Trace modélisée* (SBTm) et la manière dont il est implémenté dans un environnement informatique.

3.3.2 Les Systèmes à Base de Traces modélisées

Comme le montre le schéma ci-dessous (Figure 3.5), un tel système reprend les fonctions générales d'un Système à Base de Trace avec cette fois les moyens de gérer les *traces modélisées associées* de leurs modèles.

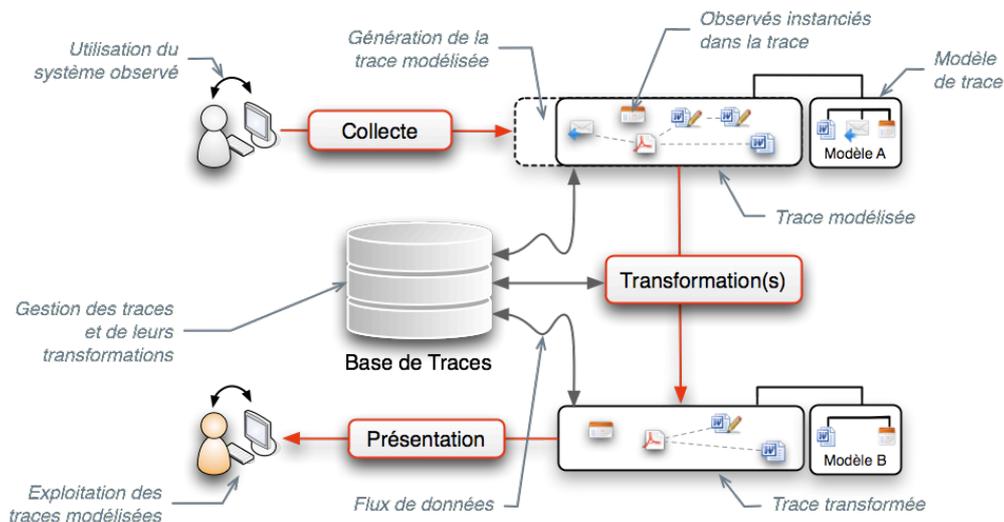


Figure 3.5 : Architecture générale d'un SBTm.

3.3.2.1 Collecte

Premier processus de l'architecture, la *collecte* alimente le SGBT avec des éléments permettant de générer, par enregistrement systématique et en respect de son modèle, une *trace modélisée*. La collecte doit disposer d'au moins une *source de traçage*, qui permet au sein du *système observé*, de capter des données générées par les interactions de l'utilisateur avec son système. Cette source de traçage peut être de natures diverses : il peut s'agir aussi bien d'un flux audio ou vidéo, que du flux formé par les *logs* du système, la condition étant qu'il doit être possible d'extraire (manuellement ou automatique-

ment) de ces flux des *observés*. La collecte peut (si nécessaire) s'appuyer sur plusieurs sources de traçage en parallèle, dont il faut alors gérer la synchronisation (Figure 3.6).

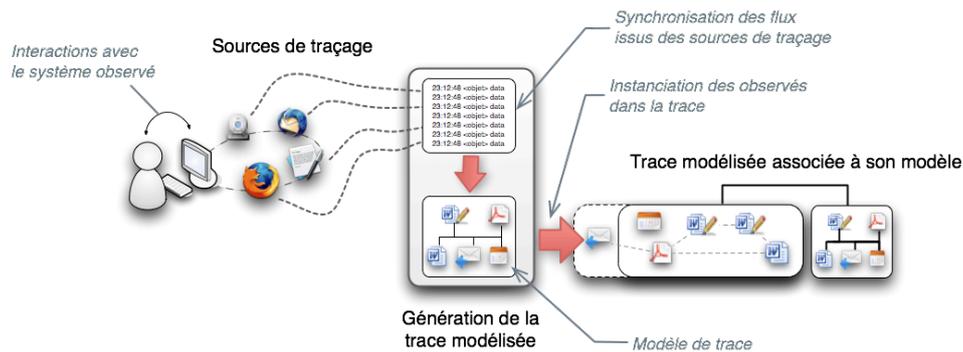


Figure 3.6 : Détail du processus de collecte des traces modélisées dans un SGBT.

Le produit de la collecte est donc une trace associée à son modèle, que l'on nomme *trace première*. Comme son nom l'indique, cette trace première, est la première dont on puisse considérer, à l'issue du processus initial de collecte, qu'elle est modélisée explicitement dans les termes du système, et donc pleinement *manipulable* et *exploitable* par celui-ci.

3.3.2.2 Transformations

Cette trace première, est stockée et manipulée (comme l'ensemble des traces modélisées gérées par le SBTm) dans une *base de traces*. Elle fera en particulier l'objet de diverses *transformations* qui auront pour but d'amener la trace modélisée à présenter un niveau d'abstraction correspondant à l'exploitation qui en est prévue : application d'autres transformations, présentation de la trace modélisée à l'utilisateur de la trace modélisée ou encore calcul d'indicateurs. On appellera *trace transformée* une trace modélisée issue de la transformation d'une trace première. Ces transformations peuvent être opérées de manière *manuelle* ou *automatique*, et ont pour condition de respecter la *cohérence* de la transformation, *i.e.* que les changements opérés sur les observés de la trace première soient tels qu'ils respectent un second modèle de trace qui sera associé à la trace transformée. Une *transformation manuelle* désigne toute modification impliquant un changement dans la composition des éléments de la trace, due à un ajout, un filtrage ou un enrichissement de ses observés par l'utilisateur du SBTm. Ce dernier doit donc garantir la cohérence entre les traces et leurs modèles lors d'une transformation manuelle. Les *transformations automatiques* sont appliquées par le SBTm lui-même selon des *modèles de transformation*, un modèle de transformation étant un ensemble de règles exprimant les opérations de sélection ou de réécriture de la trace modélisée considérée (première ou déjà transformée).

On peut distinguer *trois types* de transformations automatiques, qui peuvent bien sûr être combinés :

- Les transformations de type *sélection* : elles créent une nouvelle trace modélisée contenant tous les observés respectant un *filtre de sélection* donné. C'est ce type de transformation par exemple, qui est utilisé pour séparer les observés pertinents de la trace du bruit, pour le *nettoyage* de la trace première. Ces filtres sont des contraintes sur les observés, le domaine temporel, et les relations structurelles de la trace modélisée à transformer.

- Les transformations de type *réécriture de motifs*⁸⁹ : Elles permettent de remplacer un ou plusieurs observés par un autre observé (généralement de plus haut niveau d'abstraction). Ces transformations peuvent toucher une séquence d'observés qui se succèdent directement ou non dans la trace modélisée. L'expression de ce type de séquence d'observés représente un motif (*pattern*) que l'on peut réécrire.
- Les transformations par *fusion temporelle* : Ce dernier type de transformation ne concerne pas une, mais plusieurs traces modélisées. Les transformations en question consistent à *fusionner, assembler et/ou grouper* les observés de ces différentes traces modélisées en tenant compte de la temporalité de leurs éléments. La trace transformée par fusion suit un modèle de trace regroupant les modèles des traces fusionnées. Cette transformation est possible si les traces fusionnées possèdent des extensions temporelles homogènes⁹⁰.

Notons ici que ces transformations (comme de manière générale tout calcul effectué sur la trace modélisée) s'appuient nécessairement sur diverses fonctionnalités de manipulation élémentaires, comme la *recherche* ou de *tri* dans la trace modélisée (requête). Ces fonctionnalités sont directement dépendantes du langage utilisé pour coder le SBTm et des fonctions proposées par la base utilisée.

3.3.2.3 Présentation

Dernière partie de l'architecture, la *présentation* de la trace correspond à l'interface du SBTm pour permettre à un utilisateur d'accéder à la trace modélisée et/ou de piloter les traitements automatiques et transformations manuelles. Nous regroupons donc sous le terme de *présentation* les fonctionnalités de *visualisation* et d'*interaction* des traces modélisées à présenter (Figure 3.7).

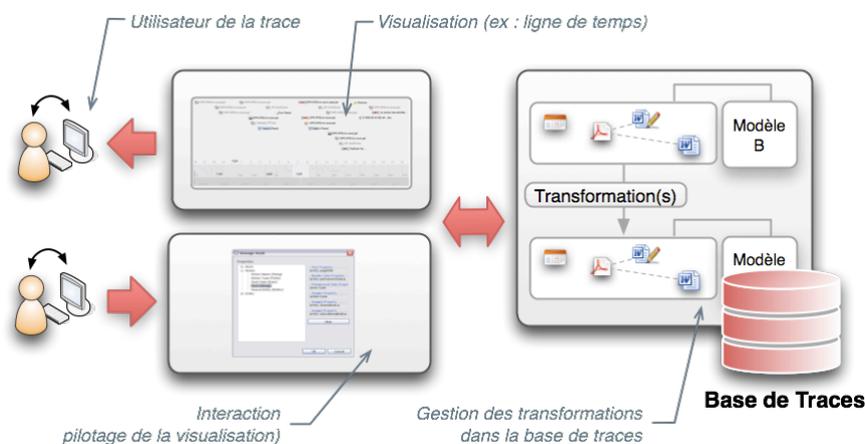


Figure 3.7 : Détail du processus de présentation des traces modélisées dans un SBTm.

La *visualisation*, tout d'abord, qui consiste à afficher de manière graphique et/ou textuelle la collection d'observés d'une trace modélisée, peut être considérée comme une transformation particulière respectant un modèle de trace dédié. Sur le même principe que les transformations décrites précédemment, la visualisation reposera sur un *modèle de visualisation*. Cette transformation n'est pas obligatoirement nécessaire, elle correspond à la mise en cohérence du système de visualisation choisi et du

89. Une représentation d'une séquence, ou d'un sous ensemble d'observés identifié, dans la trace modélisée.

90. Permettant des translations et conversions de temps (intervalles vs instants).

format de la trace modélisée à afficher. Comme nous le verrons par la suite, il suffit parfois que la trace modélisée soit exprimée sous la forme d'un fichier *XML* pour utiliser diverses interfaces de visualisation, par exemple de type « ligne de temps » (Figure 3.7).

Les outils de visualisation directement applicables, ne proposent en revanche pas tous, les moyens d'*interagir* avec la trace présentée de manière satisfaisante. De plus, il faut également prévoir la possibilité de gérer la présentation des modèles associés aux traces. Dans le strict cadre du SBTm, la phase de présentation consiste avant tout à fournir une trace transformée telle qu'elle puisse faire l'objet de visualisations et d'interactions qui peuvent être définies et appliquées en dehors du système. Des travaux sont cependant engagés pour définir des services génériques d'affichage et de manipulation de traces modélisées et de leurs modèles, en s'appuyant sur l'expérience de la création de fonctionnalités d'interaction dans des environnements particuliers (Cram et al., 2007). En revanche, il est impératif que les interactions envisagées soient déterminées en respectant les contraintes de l'environnement dans lesquelles elles prendront place.

3.3.3 Implémentations

Plusieurs applications de l'approche ont été menées et des Systèmes à Base de Traces modélisées ont été implémentés dans des situations d'activité variées. C'est le cas par exemple du travail de O. Georgeon (2006) sur la mise en place d'un SBTm destiné à l'analyse des comportements de conduite automobile (*Abstract*). Le système, mis en place à l'INRETS⁹¹, gère cinq sources de données qui sont traitées pour intégrer les traces d'interactions : vidéos des différentes caméras du véhicule, mesures prélevées sur le véhicule, informations sur l'environnement provenant d'un télémètre, les données de navigation d'un *GPS* et les événements déclenchés par l'expérimentateur.

C'est le cas aussi dans des projets liés à des EIAH, comme le projet *Ambre*, avec la mise en place d'un SBTm destiné à assister un apprenant au cours de son apprentissage de méthodes additives (Guin et al., 2002), ou le projet *Géonote*⁹² où un SBTm est mis en place pour fournir à un tuteur les traces numériques de ses élèves durant la préparation d'une étude géologique en classe de Sciences et Vie de la Terre (Settouti et al., 2007), ou encore le projet lié à *eLycée* sur lequel nous aurons l'occasion de revenir en détail par la suite. Très récemment encore, un travail a été mené pour mettre en place un SBTm visant la mise en place d'indicateurs d'activités éducatives dans l'environnement *Moodle*⁹³ (Djaouad et al., 2009) et dans un domaine d'application tout à fait différent le logiciel *Adventure* consacré à la lecture active de vidéos a été instrumenté d'un SBTm visant à soutenir l'utilisateur dans la reprise d'activité après interruption (Richard et al., 2009).

Dans tous ces exemples d'applications de l'approche, les chercheurs ont dû implémenter de manière *ad hoc* un SGBTm au sein de l'environnement concerné par les traces (système observé), le plus souvent en étant contraint de se limiter à l'instrumentation d'un outil unique. Pour faciliter la mise en place des SGBT et ainsi donner les moyens à l'approche de prendre une nouvelle ampleur, un projet de développement d'un *framework informatique générique* a vu le jour : le projet *Kernel for Trace-Based Systems (KTBS)*. Il s'agit donc de créer une implémentation générique de SBTm qui pourrait

⁹¹ Institut National de Recherche sur les Transports et leur Sécurité : <http://www.inrets.fr/> (consulté le 02/05/2009).

⁹² Voir <http://praxis.inrp.fr/praxis/projets/geomatique/geonote/> (consulté le 04/05/2009).

⁹³ Voir <http://moodle.org/> (consulté le 04/05/2009).

offrir des *services* de manipulation de traces modélisées nécessaires à l'implémentation d'un SGBT : collecter, transformer, requêter, visualiser et partager des traces modélisées. Le développement de cet atelier, encore en cours est sur le point d'être finalisé.

Pour l'heure, *KTBS*⁹⁴ supporte l'expression de modèles de trace comme des *ontologies RDF-Schéma*⁹⁵ décrivant des *concepts* et des *relations*. Les traces décrites par un modèle de trace ontologique sont constituées d'instances (individus) représentant leurs observés. Son architecture correspondant à celle décrite Figure 3.5, fournit pour commencer la gestion de la collecte. Cette collecte permet de passer (en temps réel ou différé) d'une ou plusieurs sources de traçage quelconques à une trace première stockée dans une base de traces *RDF*, et passe par deux étapes : (a) la description du modèle de la trace première contraignant et structurant la trace collectée et (b) l'instanciation du modèle pour créer la trace. Étant donné la diversité des sources de traçage et des techniques pouvant être utilisées dans cette seconde étape, le système *KTBS* définit une API standard à laquelle des collecteur *ad-hoc* peuvent se connecter. Par contre, une fois la trace première instanciée, le système permet la gestion intégrée et générique des traces : l'utilisateur peut spécifier des *transformations*, manuellement ou à l'aide de modèles de transformation. Il peut également bénéficier des services de *navigation* dans les traces modélisées.

La fin du développement de cet atelier marquera prochainement un pas important dans le développement de l'approche des Système à Base de Traces modélisées. Avec ce soutien technique, les applications basées sur la réutilisation de traces modélisées devraient voir le jour plus facilement et plus rapidement, ce qui permettra aux chercheurs de se consacrer pleinement à l'usage des traces qu'ils contribuent à créer dans les environnements informatiques.

3.4 Conclusion

Ce chapitre a présenté le cadre conceptuel des *Systèmes à Base de Traces*. Créé à la suite d'une première série de travaux sur les traces numériques comme support à l'assistance de l'utilisateur (approche MUSETTE), ce cadre propose dans un premier temps une *généralisation* des « Systèmes à Base de Trace », *i.e.* tout dispositif informatique mettant en œuvre l'enregistrement de certaines interactions utilisateur afin d'en constituer une trace numérique considérée (et manipulée) comme un objet informatique de premier ordre. Dans un deuxième temps est proposé une *formalisation* de ces systèmes avec les *Systèmes à Base de Traces modélisées* pour lesquels sont mis en place les concepts de *trace modélisée* et de *modèle de trace*. L'idée est de considérer toute trace comme une collection d'observés eux-mêmes explicités dans un modèle associé à la trace. Le modèle de trace en question est constitué d'un ensemble de types d'observés dont il définit également les attributs et les relations.

Diverses manipulations de la trace modélisée sont ainsi envisageables de manière indépendante, par le biais du modèle de trace. Du point de vue de l'implémentation concrète d'un SBTm (c'est-à-dire la mise en place d'un SGBT), une avancée prochaine de l'approche est attendue avec l'arrivée d'un

⁹⁴ Pour plus d'information voir les publications de L.S. Settouti (2006, 2007, 2008) sur la genèse de *KTBS* avec le projet ATER.

⁹⁵ Voir www.w3.org/TR/rdf-schema/ (consulté le 23/07/2009).

Framework informatique (KTBS) permettant de gérer les principales fonctions dont l'implantation d'un SBTm (SGBT) aurait besoin.

3.4.1 Un « méta-modèle »

Finalement, le cadre conceptuel que nous désignons par « SBTm » n'est ni tout à fait une approche, ni directement un modèle. Nous aimerions le voir comme un « méta-modèle » des traces modélisées, *i.e.* un modèle de la façon de mettre en place des modèles de traces dans un système informatique qui les génère. Très peu de travaux dans la littérature proposent d'adopter ce type de point de vue, même si nous avons pu relever par exemple l'approche *DGU* que ses auteurs présentent comme un « méta-modèle des approches basées sur des traces numériques dans les EIAH » (Choquet et Iskal, 2006-07), qui est composé de trois facettes : *Defining*, qui désigne la modélisation des besoins de traces (dans une situation pédagogique), *Getting* qui désigne la modélisation des moyens de traçage, et *Using*, qui désigne la modélisation des usages des traces. En ce qui concerne ce qui équivaldrait au « modèle de trace », l'approche propose de *Usage Tracking Language (UTL)* qui définit à la fois une formalisation des objets observés (*traceable-concept*)⁹⁶, et une formalisation de la trace numérique elle-même (*track*) et dont on peut voir une représentation schématique dans la figure ci-dessous (Figure 3.8).

Cette proposition reste toutefois consacrée exclusivement au domaine des EIAH et dépend des particularités de celles-ci. En l'occurrence, elle s'appuie sur le fait que dans les EIAH, les activités sont définies à l'avance et les objets pédagogiques modélisés dans le système, ce qui permet directement de réutiliser cette base formalisée pour construire des traces. De plus, l'approche ne concerne qu'une réutilisation spécifique des traces, qui vise la réingénierie de l'EIAH impliqué grâce à une évaluation réalisée sur la base d'indicateurs extraits de la trace numérique.

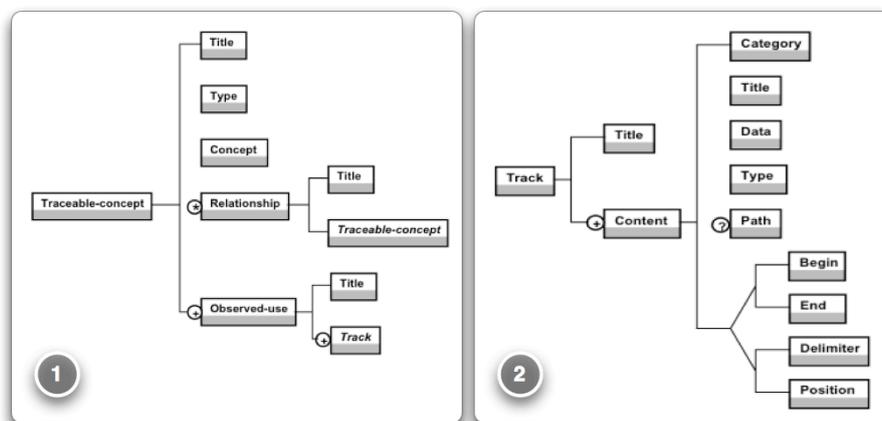


Figure 3.8 : UTL partie représentation (1) et trace (2).
Tiré de (Choquet et Iskal, 2006-07, p.10-11)

Dans ce contexte la proposition que constitue les SBTm se révèle plus ouverte. D'un point de vue conceptuel, elle permet d'envisager la gestion des traces *indépendamment* des systèmes observés, alors même que cette dépendance commence à être pointée comme un handicap pour les systèmes traçants

⁹⁶ « A Traceable-concept is a concept from which it is possible to track something, for instance, a resource and an activity are traceable concepts from which we can track the beginning, the end, the duration, etc. presents the information model of this part of UTL » (Choquet et Iskal, 2006-07, p.10).

qui « [...] *are tightly integrated with their target applications and difficult to apply to new applications* » (Nakamura et Igarashi, 2008, p.23) et sont du coup fortement sensibles au changement de l'environnement numérique et « *over time it can be difficult to continually update and refine tools to work with new versions and features of commercial software* » (Hawkey, 2006, p.4). La logique des SBTm sur ce point trouvera son aboutissement dans le système *KTBS* qui concrétisera la possibilité de concevoir rapidement des SGBT.

3.4.2 Retour à la problématique

Après avoir présenté l'approche dans laquelle nous souhaitons nous inscrire, remettons en perspective ce que nous venons de décrire relativement à ce que nous voulons atteindre. Notre objectif de départ est de se donner les moyens, dans le cadre d'une activité documentaire instrumentée, de créer une trace numérique de cette activité de façon à la rendre interprétable comme une « *image* »⁹⁷ de l'activité effective d'un utilisateur et ce quel que soit le contexte de réutilisation et le statut de l'interprétant : une représentation pour soi, pour un autrui. Il ne s'agit pas d'enregistrer quantitativement des données sans distinction, mais de constituer une « représentation qualitative » de l'activité. Le cadre conceptuel des SBTm respecte cet objectif, et apporte une réponse du point de vue de la conception d'une trace numérique comme un *objet informatique*. Il nous offre un cadre conceptuel formalisé et un support informatique indispensable, nous permet également de dissocier les problèmes strictement techniques liés à l'*implémentation* d'un SBTm (SGBT), de la question de la *modélisation des traces* et de la *conception de leur(s) usage(s)*.

En revanche, l'approche ne dispose pas pour l'heure d'une définition de *la manière* dont on met en œuvre concrètement un SBTm, *i.e.* d'une *méthodologie de modélisation propre* qui encadrerait la conduite de la modélisation d'observés. Or, la pertinence et l'effectivité d'un SBTm repose sur le ou les modèle(s) de trace qui donne à la trace modélisée son caractère *interprétable*. En termes simples, l'approche définit pour le moment *sous quelle forme* une trace doit être modélisée, mais n'indique pas *la manière* dont on peut concrètement déterminer quels éléments de l'interaction doivent, ou ne doivent pas, être enregistrés et constituer en tant qu'*observés*, des éléments constitutifs d'une trace modélisée, ni pourquoi. En d'autres termes, il manque une *méthodologie de modélisation* propre aux traces modélisées alors même que la valeur d'une trace modélisée en tant que représentation qualitative d'une activité, ne repose pas uniquement sur des *capacités techniques* de collecte ou de visualisation de données, mais relève également fondamentalement d'une *compétence d'analyse et de modélisation* de la trace. La problématique de cette thèse se précise et son objectif de départ visant la construction de traces numériques représentant l'activité se concrétise dans un objectif concret : *établir et proposer une méthodologie de modélisation de trace dans le cadre des SBTm*.

Dans les chapitres qui vont suivre nous allons donc présenter la méthodologie que nous proposons. Dans un premier temps (chapitre 4) nous exposerons les éléments théoriques auxquels elle est adossée, puis dans un second temps (chapitre 5) nous expliciterons son contenu, *i.e.* les différentes méthodes qui la constituent, et enfin dans un troisième temps (chapitre 6), nous présenterons la partie empirique de ce travail, *i.e.* l'application directe de la méthodologie sur deux terrains d'application.

⁹⁷ *eikon*, voir chapitre 1.

Chapitre 4

Cadre d'analyse et méthodes de modélisation des activités instrumentées

Résumé du chapitre

La mise au point d'une méthodologie de modélisation des traces, dédié à l'approche des Systèmes à Base de Traces modélisée, s'est appuyée à la fois sur un travail de recherche théorique et sur un travail de recherche appliquée. Ce chapitre est consacré à la présentation du premier, c'est-à-dire le travail de recherche théorique et bibliographique dont l'objectif est d'identifier des domaines de recherche dont nous pouvons d'une part utiliser *l'expertise* en termes de méthodes et techniques de *modélisation*, et d'autre par tirer une *grille d'analyse* théorique des activités que nous avons à analyser, en l'occurrence des *activités documentaires instrumentées*. Concrètement nous avons exploité trois domaines de recherche différents, tant pour leurs apports théoriques sur leur manière d'aborder et de décrire les activités instrumentées que pour leur expertise en termes de modélisation :

- (a) Le domaine des IHM pour ses approches d'analyse et de modélisation de tâche habituellement destinées à la conception de systèmes informatique et de leurs interfaces.
 - (b) Le domaine de l'Ingénierie des Connaissances pour son approche théorique des activités documentaires, mais aussi et surtout pour ses approches de modélisation des connaissances.
 - (c) Le domaine du *CSCW* pour son expertise de modélisation descriptive des activités complexes et pour ses nombreuses approches théoriques de l'activité instrumentée individuelle et collective.
-

4.1 Introduction

Comme nous venons de le rappeler à la fin du chapitre précédent la problématique générale de cette thèse est de savoir comment définir une trace numérique qui donne une « image » explicite et interprétable de l'activité effective d'un utilisateur, ou en d'autres termes une représentation « qualitative » de la réalisation de cette activité, par opposition aux données « quantitatives » habituellement enregistrées. Cette problématique se concrétise avec le choix que nous avons fait de nous inscrire dans l'approche des Systèmes à Base de Traces modélisée (SBTm). L'objectif devient alors de doter concrètement l'approche des SBTm d'une *méthodologie de modélisation des traces* numérique qui lui soit propre, contribuant ainsi à l'établissement d'une approche des traces numériques complète. Pour fonder cette méthodologie, nous avons associé un travail de *recherche théorique* à un travail de *recherche empirique*. Nous avons ainsi forgé, par son application en situation, une méthodologie de modélisation théoriquement inspirée d'autres approches de modélisation présentes dans les champs de recherche liés à l'Informatique. Ce chapitre présente le résultat du travail bibliographique qui a été nécessaire pour identifier les *éléments théoriques* dont nous avons besoin pour établir notre méthodologie de modélisation des traces. Ce travail de recherche visait des sources d'expertise à la fois en termes d'*approches de modélisation* et en termes d'*analyse des activités* humaines instrumentées, toutes deux nécessaires à une méthodologie de modélisation des traces numériques.

Nous avons cherché à capitaliser les apports de trois domaines de recherche distincts qui ne sont pas toujours directement associés et qui seront présentés dans ce chapitre : le domaine des Interaction Homme-Machine (IHM) avec les concepts d'*analyse de tâche* et de *modèle de tâche* (section 4.2) ; le domaine de l'Ingénierie des Connaissances (IC), pour son expertise en termes de modélisation de connaissances, d'activités documentaires (section 4.3) ; et enfin le domaine du *Computer Supported Cooperative Work (CSCW)* pour ses théories et ses analyses de l'activité instrumentée, individuelle et collective (section 4.4).

4.2 Apports du domaine des IHM

Une première analyse de la littérature nous a rapidement orienté vers la *modélisation de tâche* utilisée dans le domaine des IHM et qui apparaît comme la démarche la plus proche de celle que nous cherchons à mettre en place. La modélisation de tâche ne constitue pas un domaine en soi, mais fédère un ensemble de travaux spécifiques au sein du domaine des IHM où elle est présentée comme un ensemble de méthodes et outils issus de l'ergonomie (Kieras, 1996) et appliqués à l'étude des interfaces de dispositifs informatiques. La pratique de la modélisation de tâche a intégré les processus de conception et d'évaluation des interfaces depuis de nombreuses années et a contribué à placer la notion de tâche au cœur des problématiques d'IHM : « *Tasks as goal-oriented actions are one of the most important aspects of HCI* » (Rauterberg et Fjeld, 1998, p.152). En retour, les travaux d'IHM ont largement contribué au développement de la modélisation de tâche elle-même.

Grossièrement, un *modèle de tâche (task model)* est une description normalisée des actions ou opérations à mettre en oeuvre pour atteindre un but. Cette description plus ou moins abstraite et exprimée

dans un langage plus ou moins formel (parfois même opérationnalisable) peut porter sur des opérations de natures différentes : opérations mentales pour un modèle de tâche cognitive, ou actions tout à fait concrètes pour un modèle de tâche de manipulation quelconque (les différentes opérations à effectuer pour mettre en route une machine par exemple). Les modèles de tâche sont traditionnellement exploités dans les processus de conception d'interfaces de dispositifs informatiques où ils permettent de raisonner sur l'adéquation entre les opérations qu'un utilisateur désire ou doit impérativement effectuer pour réaliser une tâche donnée, et les opérations que l'interface de son dispositif propose : « *Its purpose is to provide a baseline to reason about how a computer system might fit in with the user's tasks, and which functionality, operating procedures, etc. should be provided* » (De Haan, 2000, p.99).

Dans les sections à suivre nous allons revenir rapidement sur les concepts clés de la modélisation de tâche en IHM (section 4.2.1), puis nous évoquerons plus précisément le *processus de modélisation* (section 4.2.2), et enfin nous nous intéresserons à quelques *formalismes* parmi les plus cités servant à exprimer des modèles de tâches (section 4.2.3).

4.2.1 Concepts clés de la modélisation de tâche

4.2.1.1 Le concept de tâche

Bien qu'il possède en théorie une définition très précise qui l'en distingue, la notion de tâche est parfois confondue avec celle d'*activité* que nous avons pour notre part utilisée depuis le début de ce manuscrit. Tâche et activité sont en effet deux notions très proches, qu'il est d'autant plus difficile de distinguer qu'elles sont parfois utilisées indifféremment (Raby et Borges, 2000). Ordinairement, une *tâche* est entendue comme « un travail défini et limité par autrui ou par soi-même, à exécuter dans certaines conditions »⁹⁸ et comme en attestent les définitions suivantes, en Informatique en particulier, intrinsèquement liée à la notion de but :

- « *Tasks are Activities that have to be performed to reach a goal* » (Paterno, 1999)
- « *The user's job or work activities, what the user is attempting to accomplish* » (Kieras, 1996)
- « *Un objectif à atteindre par l'utilisateur à l'aide d'un système interactif* » (Balbo, 1994).
- « *A task is defined by the goals that a system tries to achieve* » (Aamodt, 2001).

La tâche renvoie à l'idée de *prescription* des opérations qu'un acteur doit mener. Selon le cas, les opérations en question peuvent être celles d'un acteur humain ou celles d'une machine, et désigner des actions physiques, soit des opérations mentales (Sabah, 1997 ; Wright et al., 2000). Pour atteindre un *but – a priori* atteignable – cette prescription est fonction des *conditions* de sa réalisation. Dans ce cas, la réalisation de la tâche est souvent appelée *activité*. Pour bien marquer la distinction, l'ergonomie préfère même parfois opposer tâche *prescrite* et tâche *effective* (Tricot et Nanard, 1998), la seconde devant décrire les actions *effectivement* engagées par un utilisateur pour accomplir la tâche (prescrite) qui lui a été attribuée. Finalement, *tâche effective* et *activité* s'opposent de la même manière à la tâche *prescrite*. Ce qui les distingue c'est le fait que l'*activité* possède un sens plus large et inclut notamment l'engagement cognitif de l'acteur.

⁹⁸ Définition du dictionnaire en ligne : <http://www.cnrtl.fr/> (consulté le 03/06/09).

De manière générale on finit par considérer que « l'activité et la tâche représentent, respectivement, le travail réalisé et le travail prescrit » (Wisner, 1995) ou encore avec Leplat et Hoc (1983) que « la tâche indique ce qui est à faire, l'activité ce qui est fait »⁹⁹. Si les deux notions se distinguent, elles sont toutes deux rapportées au *contexte* de réalisation dans lequel la tâche s'inscrit. La tâche apparaît donc comme *définie et limitée*, associée à l'idée de *prescription* et de *but ou objectif* connu et *a priori* atteignable, en tenant compte de *son inscription* dans le contexte d'exécution.

4.2.1.2 Les modèles de tâche

Les modèles de tâche sont des descriptions formalisées de tâches de différentes natures¹⁰⁰. Ils sont exploités dans différents contextes dont les principaux, en ce qui concerne le domaine des IHM, sont la *conception* et l'*évaluation* d'interface (Tricot et Nanard, 1998).

Dans le cadre de la *conception* d'interface tout d'abord, modéliser une tâche consiste à décrire la situation de travail et l'activité présente de l'opérateur (ou du type d'opérateur) pour lesquelles on élabore un nouvel outil. Il s'agit alors de concevoir les fonctionnalités qui correspondent à cette activité. Un modèle de tâche peut aussi permettre de décrire la situation future, *i.e.* celle dans laquelle se trouvera l'opérateur une fois qu'il disposera du nouvel outil. Typiquement, on peut ainsi définir un ensemble de spécifications pour le développement d'une interface.

Dans le cadre de l'*évaluation* d'une interface, modéliser la tâche consiste à décrire la situation de travail telle qu'elle « devrait être » si l'opérateur atteignait de façon optimale le but fixé avec l'outil que l'on évalue. Le modèle de tâche sert ici de *réfèrent*, soit pour l'évaluation directe d'un outil (cet outil permet-il de réaliser la tâche de façon optimale ?), soit pour une comparaison avec le modèle de la tâche utilisé lors de la conception de l'outil (est-ce que l'utilisateur peut réaliser cette tâche avec cet outil ?). On peut par exemple construire plusieurs *scenarii* de réalisation de telle ou telle tâche et vérifier sa faisabilité en interprétant directement la représentation du modèle. Enfin, signalons que les modèles de tâches sont également utilisés pour « prédire » le comportement d'un futur utilisateur, il suffit pour cela d'avoir un modèle de tâche définissant complètement les opérations à effectuer dont on juge l'exécution par un sujet potentiel à partir de telle ou telle théorie cognitive. On peut également opérationnaliser ce modèle de tâche dans une simulation de réalisation de la tâche, de façon à mesurer chaque opération dans le contexte de différentes activités.

4.2.2 Processus de modélisation des tâches

Intéressons-nous maintenant au *processus* de création de ces modèles de tâche. La modélisation couvre globalement trois étapes que F. Paterno (1999) résume de la manière suivante. Dans un premier temps est menée une *analyse de tâche* (*task analysis*), dont l'objectif est « *to identify what the relevant tasks are* ». Dans un second temps, la *construction du modèle* lui-même qui « *describes precisely the relationships among the various tasks identified* ». Le troisième et dernier temps est celui de l'*exploitation* du modèle de tâche qui consiste à exprimer « *how activities can be performed to reach the users' goals when interacting with the application considered* ». L'application concrète du processus se révèle moins linéaire que sa présentation ne le laisse entendre. En effet, la modélisation, no-

⁹⁹ Cité par (Tricot et Nanard, 1998).

¹⁰⁰ Reflétant en fait la nature des opérations qui composent la tâche décrite.

tamment à cause des contraintes du formalisme utilisé, peut influencer voire orienter totalement la manière d'analyser la tâche. De même, l'interprétation du modèle obtenu peut remettre en cause les étapes précédentes. Ainsi un modèle de tâche pré-existant peut lui-même être servir directement de support à l'analyse (Paris et *al.*, 2003). Dans les faits, les trois étapes de ce processus sont donc non seulement indissociables mais font en réalité partie d'une seule et même démarche, ce qu'illustre la figure ci-dessous (Figure 4.1).

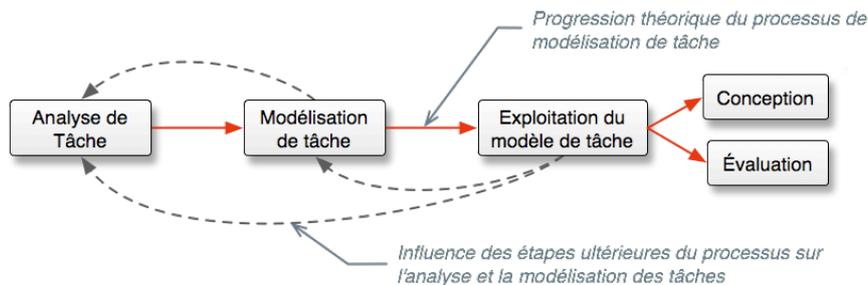


Figure 4.1 : Processus de modélisation de tâche.

4.2.2.1 Analyse de tâche

La travail *d'analyse de tâche* est relativement complexe bien qu'il soit parfois résumé un peu simplement comme un *recueil d'information* d'une part et comme une *interprétation* de ces données d'autre part : « *Task analysis requires information about the user's situation and activities, but simply collecting data about the user's task is not necessarily a task analysis* » (Kieras, 1996, p.1408). L'analyse de tâche débute souvent par un recueil informel d'informations (Paterno, 1999) et se poursuit de manière plus structurée à l'aide d'outils et techniques empruntés aux Sciences Humaines et Sociales¹⁰¹ : questionnaires, entretiens (individuels ou collectifs), observations, analyse de la documentation, étude des incidents, *etc.* Ce travail peut être centré sur des objets sensiblement différents. Soit sur *l'utilisateur*, sa composante cognitive ou bien sur le *contexte* précis dans lequel il réalise ses tâches. Soit sur le *système*, ses ressources et son « *information-flow* ». Soit encore sur *les interactions* lorsqu'il s'agit de comprendre comment celles-ci se déroulent, un peu à la manière d'une analyse dialogique : actions utilisateur et *feedback*, études des situations critiques dans l'échange d'informations ou encore des situations d'erreurs connues, des interruptions, *etc.*

4.2.2.2 Création du modèle

La création du modèle, pour mettre en relation les différents éléments dégagés par l'analyse, va reposer principalement sur un grand principe, celui de la *décomposition* des tâches. Ce principe reprend l'idée selon laquelle une tâche est réductible à une ensemble de *sous-tâches*, elles-mêmes réductibles à un groupe de sous-sous-tâches (*etc.*), jusqu'à arriver à une décomposition en *opérations simples* (interactions de bas niveau, telle une action sur un bouton ou une commande en ce qui concerne les interfaces utilisateur), également appelées *tâches élémentaires*. Le facteur principal de décomposition est hiérarchique, suivant les règles de composition logique (Paterno, 2002). Finalement la création du modèle de tâche va consister à construire progressivement une *représentation* (systématique) de la tâche analysée dans un *formalisme* permettant de réaliser cette décomposition.

¹⁰¹ À tel point que cette phase semble faire l'objet de critiques quant à son caractère « scientifique » : « *Thus a task analysis for system design must be rather more informal, and primarily heuristic in flavor compared to scientific research* » (Kieras, 1996).

4.2.3 Formalismes

Pour fournir un support à l'expression des modèles de tâches, que cela soit durant la création du modèle ou pour utiliser le résultat final, des approches structurantes sont offertes aux modélisateurs, qui proposent l'utilisation de formalismes pour représenter de manière structurée leurs modèles. On parle ici d'approche « structurante » pour signifier que l'utilisation d'un formalisme va généralement de pair avec une manière de conduire l'analyse et d'exploiter les « modèles de tâche exprimés dans un formalisme idoine » (Balbo, 1993). Il existe un nombre relativement important de formalismes dédiés aux modèles de tâche mais tous ne sont cependant pas extrêmement répandus. Nous ne donnerons ici à titre d'illustration que quelques exemples : *GOMS*, *MAD* et *UAN*.

4.2.3.1 *Goal, Operator, Method, Selection* (Card et al., 1983)

GOMS est sans doute le formalisme de ce type le plus répandu, plus parce qu'il en est un des premiers exemples que par l'ampleur de son utilisation, qui s'est amoindrie (Pirolli et al., 2002). *GOMS* est un formalisme de description du comportement attendu d'un utilisateur expert. Les modèles de tâche produits sont utilisés pour détecter les erreurs de conception les plus grossières et ce à quatre niveaux d'abstraction en fonction des objectifs recherchés par le concepteur : *tâche*, *fonction*, *argument*, *actions physiques* (Kieras et John, 1994). Une tâche étant décomposée et décrite en *GOMS*, le but est de prédire le temps mis par l'utilisateur (expert) pour atteindre un but donné en additionnant les temps d'exécution des opérations nécessaires. La valeur obtenue peut être comparée avec celle des seuils fixés dans un plan qualité. La méthode peut également servir à guider un choix de conception en comparant des performances attendues pour plusieurs solutions envisageables, sachant que *GOMS* ne traite pas en principe le cas des erreurs. Ces justifications, fondées sur une prédiction analytique et non pas sur la seule intuition heuristique, peuvent alors être documentées dans des tables ou encore dans un schéma de décision de type « Question Options Critères » dont un exemple est donné ci-dessous.

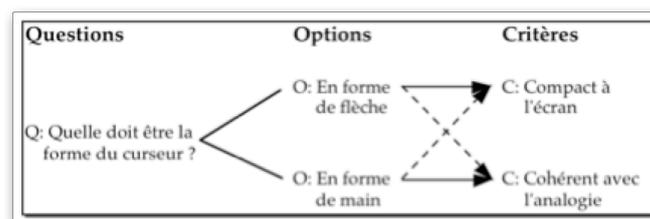


Figure 4.2 : Un exemple d'utilisation de la notation QOC. Les flèches pleines pointent les critères favorables à l'option en regard, et les flèches en pointillés pointent les critères non favorables. Tiré de (Balbo, 1994, p.20).

4.2.3.2 Méthode Analytique de Description de tâches (Scapin et Pierret-Goldreich, 1989)

La « méthode » MAD proposée par Scapin et Pierret-Goldreich (1989) est plutôt destinée à l'analyse des besoins utilisateurs. Dans l'exemple d'application proposé par (Jambon, 1996), MAD est utilisé dans le contexte du contrôle aérien pour représenter l'interaction globale d'un utilisateur avec sa station de travail. Comme la majorité des formalismes et supports de modélisation, MAD est basé sur une *décomposition hiérarchique* des tâches. Le formalisme de représentation utilisé est arborescent et exprime d'une part des relations logiques de *composition*, et d'autre part des liens *temporels*

d'enchaînements permettant d'exprimer ainsi : la séquence, le parallélisme, la boucle, l'alternative et l'option (tâche non obligatoire). MAD met l'accent sur la représentation conceptuelle, et les détails fins de l'interaction ne sont pas exprimés explicitement. Notons que MAD est une approche dotée d'outils logiciels d'édition : éditeur de tâches, d'arbres, d'objets et de schémas¹⁰².

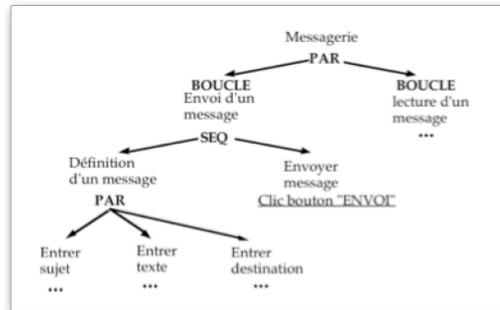


Figure 4.3 : Extrait de l'arbre logico-temporel pour une tâche de messagerie.
Tiré de (Balbo, 1994, p.57).

Dans la figure ci-dessus est présenté un exemple « d'arbre MAD » pour une application de messagerie électronique. On y constate la possibilité d'envoyer/recevoir en parallèle plusieurs messages. Envoyer un message exige que l'utilisateur définisse le contenu du message avant de signifier au système son désir de l'émettre. Spécifier le contenu demande la saisie du sujet, celle du texte et du destinataire, ces trois tâches pouvant être menées dans un ordre quelconque. Un formalisme textuel complète la représentation graphique permettant de préciser si nécessaire tâche et/ou sous tâches.

4.2.3.3 User Action Notation (Hartson et al., 1990)

UAN fait partie, comme GOMS, des approches qui sont issues de travaux bénéficiant d'apports des sciences cognitives. C'est une approche descriptive, tournée vers l'analyse prédictive ou les spécifications externes informelles (Balbo, 1994), qui s'appuie également sur une décomposition des tâches en terme de séquences de sous tâches. Le formalisme UAN de base, créé à l'origine pour faire des *évaluations pré-prototype* (Hartson et al., 1990), s'appuie sur une représentation tabulaire à trois colonnes. Comme le montre l'exemple de la figure ci-dessous, la colonne de gauche spécifie la suite temporelle des actions physiques que l'utilisateur doit effectuer pour accomplir la tâche. La colonne centrale décrit les retours d'information fournis par le système aux actions utilisateur. Enfin la colonne de droite sert à noter le changement d'état de variables pertinentes pour l'interface. Dans l'exemple ci-dessous, la tâche considérée est celle de la sélection d'un objet (une icône).

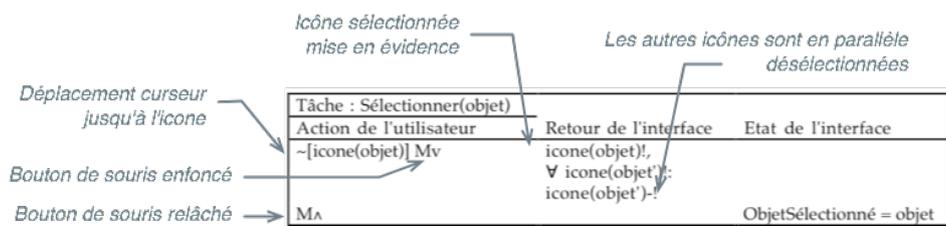


Figure 4.4 : Description UAN de la sélection d'un objet - Tiré de (Hartson et al., 1990, p.188).

¹⁰² Par exemple l'outil ALACIE (Calvary, 1998).

Dans sa forme actuelle, comme dans d'autres approches et formalismes, on retrouve des relations temporelles : *séquence*, *interruptibilité*, *substituabilité*, *alternative*, *parallélisme* (dans le sens d'ordre indépendant). De fait, les tâches *UAN* peuvent être élémentaires ou composées. Par extension, la colonne de gauche d'une description *UAN* de tâche peut comprendre des actions physiques et/ou des noms de tâche. Dans les deux cas, actions et tâches sont liés par des relations temporelles. La figure ci-dessous présente les différentes possibilités (Figure 4.5).

Relation temporelle	Symbole UAN
Séquence	A B
Attente	A (t>n) B
Disjonction répétée	(A B)*
Indépendance d'ordre	A & B
Entrelacement monodirectionnel	A → B
Entrelacement bidirectionnel	A ↔ B
Parallélisme	A B

Figure 4.5 : Relations temporelles dans UAN

UAN a été largement utilisé et bénéficie d'une bonne extensibilité. Cette extensibilité a été notamment exploitée Balbo et Coutaz (1993) pour étudier une interface multimodale en ajoutant de nouveaux éléments au formalisme de base, ou encore par *eXtended UAN* qui vise à en combler certaines « faiblesses » sémantiques¹⁰³ (Gray et al., 1994). Il existe également plusieurs outils implémentant l'approche et la notation, notamment *Quantum* (Paris et al., 2003) et *Euterpe*¹⁰⁴ (De Haan, 2000).

4.2.4 Bilan

Nous ne pouvons ici nous étendre sur la présentation des multiples approches et formalismes de modélisation de tâche. Pour trouver des informations détaillées concernant celles-ci, nous renvoyons le lecteur à des sources issues de l'ergonomie ou du domaine des IHM, notamment (Balbo, 1994), (Jambon, 1996), et plus récemment (Lucquiaud et al., 2002) qui proposent des comparaisons des différentes solutions proposées. En fait, un exposé exhaustif ne nous est pas nécessaire pour établir un bilan concernant la modélisation de trace comme source d'inspiration d'une méthodologie de modélisation de traces numériques. Il existe effectivement des points communs aux deux démarches de modélisation, à commencer par leur centre d'intérêt principal qu'est l'activité *effective* qu'il faut décrire et analyser. Nous retiendrons de la démarche de modélisation de tâche son *analyse fine*, sa prise en compte des facteurs contextuels dans la réalisation concrète d'une tâche à exécuter, ainsi que son articulation entre analyse et création du modèle.

Concernant l'utilisation des formalismes de représentation évoqués comme de potentiels outils de modélisation de trace, nous faisons preuve d'une certaine réserve. En effet, leur principe de base de décomposition de la tâche en unités élémentaires, et leur objectif de décrire de manière exhaustive les interactions du point de vue du fonctionnement du système, ne correspondent pas à notre démarche de modélisation. Nous souhaitons adopter le point de vue de l'utilisateur sur la réalisation de sa tâche, ce qui n'implique pas une description complète des opérations effectuées sur la machine. En fait un modèle de tâche ne peut être identifié ni à un modèle de trace, ni à une trace modélisée. Une trace modéli-

¹⁰³ Il est par exemple impossible d'exprimer le fait qu'une variable est modifiée en un instant quelconque entre le début et la fin d'une tâche.

¹⁰⁴ Voir <http://www.cs.vu.nl/~gerrit/gta/euterpe.html> (consulté le 10/08/2009).

sée ne décrit pas exhaustivement les opérations exécutées pour réaliser une tâche, elle n'est qu'une représentation partielle, du point de vue de l'utilisateur, de la réalisation de l'activité *dans les termes de l'activité* et non ceux de *l'interaction* homme-machine. Le modèle de trace quant à lui n'est pas un modèle de tâche, puisqu'il ne définit pas directement la trace mais *uniquement* ses constituants. Le modèle de trace se rapproche finalement plus d'un *modèle de connaissance*, ce qui nous a conduit à aborder le domaine de *l'Ingénierie des Connaissances*.

4.3 Apports de l'Ingénierie des Connaissances

Notre recherche s'élargit donc au domaine de l'Ingénierie des Connaissances (IC) qui propose un ensemble de principes méthodologiques de modélisation des connaissances dont nous souhaitons tirer partie. Après avoir resitué l'Ingénierie des Connaissances en tant que domaine de recherche (section 4.3.1), nous décrirons ses principaux concepts (section 4.3.2), puis nous nous intéresserons à ses méthodes de modélisation (section 4.3.3) en tentant d'établir un possible parallèle entre modèle de trace et modèle de connaissance.

4.3.1 L'Ingénierie des Connaissance comme domaine de recherche

L'Ingénierie des Connaissances a fait son apparition il y a une vingtaine d'année. L'objectif premier de cette ingénierie est de concevoir des *Systèmes à Base de Connaissances* (SBC) dont le fonctionnement permet d'opérationnaliser des connaissances. Ses travaux se situent dans le prolongement de ceux de l'Intelligence Artificielle (IA) sur les Systèmes Experts (Aussenac, 1989). Comme pour ses prédécesseurs, l'opérationnalisation de connaissances par un SBC vise la résolution automatique ou semi-automatique de problèmes. Dans un cadre théorique donné, une telle démarche exige la modélisation du *problème* et celle de la *méthode de résolution* du problème en question. Les connaissances nécessaires à cette modélisation sont tirées du domaine dans lequel le problème se pose et sont *formalisées* pour être *opérationnalisées* dans un système informatique. De cette démarche canonique, issue de l'IA, l'Ingénierie des Connaissance hérite d'une vision formaliste des connaissances. Elle s'en démarque cependant en visant à dépasser le formalisme mathématique historiquement lié à la résolution de problème pour opérationnaliser des connaissances exprimées en *langage naturel* :

« Là où l'informatique numérique permet classiquement d'opérationnaliser des connaissances exprimées dans le cadre d'un formalisme mathématique, l'IC aurait pour projet d'opérationnaliser des connaissances exprimées en langage naturel, et représenterait donc l'effort de recherche nécessaire à la découverte des principes de conception et de réalisation de systèmes pouvant raisonner non seulement sur des nombres, mais également sur des symboles pour résoudre automatiquement ou assister à résoudre un problème » (Bachimont, 2000, p.1).

L'IC a aujourd'hui intégré les apports de bien d'autres disciplines (Charlet, 2004), de la linguistique à la sémiotique (pour traiter de la question de l'expression des connaissances en langage naturel), en passant par la psychologie (pour la question de l'élicitation des connaissances), l'informatique (pour les opérationnaliser) ou encore l'ergonomie (pour l'appropriation des systèmes informatiques produits), les Sciences de l'Information et de la Communication et la gestion (pour intégrer l'usage de

cette instrumentation dans son environnement organisationnel). L'IC est un domaine de recherche relativement jeune qui acquiert petit à petit son indépendance relativement à l'héritage que lui lègue l'IA et elle apparaît aujourd'hui comme « un ensemble de techniques reposant sur le numérique pour aborder les connaissances » (Bachimont, 2004, p.62). Adoptant un nouveau point de vue sur les connaissances l'IC vise désormais la conception de systèmes permettant d'assister l'utilisateur à la découverte, la création et la manipulation de connaissances dont il est lui-même le seul révélateur (Cahier, 2006), autrement dit « des machines donnant à penser [plus] que des machines qui pensent » (Bachimont, 1993, p.238).

Ce changement de perspective sur la connaissance induit également un changement du point de vue sur la cognition. Abandonnant la vision cognitiviste de l'humain, considéré comme un *système de traitement de l'information*, le domaine s'est ouvert à une vision plus *artefactuelle, contextuelle et collective*, voire *sociale* de la cognition et des connaissances. Ce mouvement d'ouverture s'est récemment accéléré après avoir tardé à s'enclencher : la prise en compte du caractère collectif de la connaissance notamment, fut particulièrement absente du domaine pendant longtemps et reste encore fragile (Brown et Duguid, 2002) bien que l'importance des processus collectifs et sociaux dans ceux de la connaissance ait été démontrée à plusieurs occasions (Rot, 2005). Cette évolution est d'importance, notamment parce que, traditionnellement un travail d'ingénierie des connaissances abouti à la conception d'un SBC dont la finalité est d'intégrer une *organisation* en fonctionnement, et donc une réalité et un contexte d'usage qui dépassent de loin les seules propriétés logiques et formelles du système produit. Ainsi l'IC en tant que science de la conception de « Système à Base de Connaissance » ne peut se limiter à être une science purement formelle, une simple branche de la logique proposant des techniques de manipulation de symboles, et doit intégrer la dimension *humaine* des situations qu'elle crée par l'usage des dispositifs qu'elle déploie.

4.3.2 Concepts clefs de l'Ingénierie des Connaissances

4.3.2.1 La notion de connaissance

Plusieurs points de vue sur les connaissances cohabitent au sein même de l'IC, marquant le plus souvent l'influence d'autres domaines de recherche, notamment les Sciences Cognitives. Dans cette diversité une manière classique d'aborder la définition des connaissances est de distinguer leur caractère *explicite* ou *implicite*. Nous devons cette caractérisation au désormais fameux travail de I. Nonaka et H. Takeuchi (1995) qui popularisèrent une distinction déjà ancienne entre *connaissance explicite* et *connaissance tacite* (annexe 1). À l'origine une connaissance *explicite* se réfère à une connaissance qui peut-être exprimée sous forme de mots, dessins, ou métaphores, ce qui se traduit par le fait que cette connaissance est *exprimable* et *transmissible*, autrement dit potentiellement *formalisable*. La connaissance *tacite* quant à elle désigne une connaissance difficilement exprimable quelle que soit la forme du langage choisi (Polanyi, 1966), ce qui a pour conséquence l'impossibilité pure et simple de leur formalisation et la très grande difficulté de leur transmission¹⁰⁵.

¹⁰⁵ Dans le cas de connaissances tacites on parle également de « savoir-faire » pour désigner des connaissances parfois incarnées et procédurales (savoir faire du vélo par exemple), qui possèdent ces caractéristiques, ou encore de « connaissances informelles » par opposition à des connaissances « formelles ».

En caricaturant nous pouvons dire qu'à son origine l'IC a hérité de l'IA une vision théorique de la connaissance similaire à la conception qu'en avait le Cercle de Vienne pour qui « toute connaissance n'est connaissance qu'en tant qu'elle est formelle » (Bachimont, 2000, p.2). Cette vision a nettement évolué depuis les débuts de l'Ingénierie des Connaissances, notamment avec la construction d'une distinction entre les notions de *donnée*, d'*information* et de *connaissance*.

La notion de *donnée* et celle d'*information* furent d'abord formellement définies par la Théorie l'Information Shannon (1948), pour laquelle une information est *codée* par son émetteur, *transmise*, puis *décodée* par un récepteur. Dans ce cadre, est « donnée » toute information codée et transmise affectant un programme ou un système. Si la donnée affecte le système, l'information touche celui qui est amené à *interpréter* et de ce dernier dépendra la création d'une *connaissance* correspondant à cette information. Rapidement, les frontières entre ces différentes notions se sont effritées sous les remarques conjointes des Sciences du Langages, de celles des sciences de l'Information et de la Communication ou encore de la Systémique (Prince, 1996). Au final, on fait face à « un continuum par rapport à un processus d'action et nous plaçons des étiquettes sur des concepts manipulés par ce processus en fonction des niveaux d'entrée dans celui-ci » (Charlet, 2004, p.7). Il apparaît ainsi que c'est en fonction de l'action de celui qui les interprète que les *données* peuvent être alternativement considérées comme *informations* ou comme *connaissances*.

La pluralité des interprétations possibles à une même information conduit à renoncer à l'espoir de l'Intelligence Artificielle de modéliser complètement une situation pour en extraire le *sens* (B. Bachimont, 1994 ; Cahier, 2006). Sens et connaissance ne peuvent se concevoir en dehors d'un *cadre interprétatif* et la caractéristique majeure de la connaissance est *l'interprétation humaine* qui lui donne son existence. En un certain sens, l'Ingénierie des Connaissances a finalement dû admettre que la connaissance qu'elle cherche à modéliser *n'existe pas a priori* (LeMoigne, 1994) mais uniquement dans le *contexte* de *l'interprétation* d'une information *utilisée*¹⁰⁶. Le lien à l'action est également très fort, et on le retrouve dans des définitions récentes. Pour B. Bachimont une connaissance est « un moyen d'agir », et pour M. Zacklad elle doit être vue comme « un potentiel d'action attribué à un acteur individuel ou collectif dans le contexte d'une situation au sein de laquelle celui-ci poursuit un projet » (Zacklad, 2004, p.4).

Au sein de l'IC, la notion de connaissance continue d'évoluer notamment sous l'influence de nouvelles idées issues des Sciences Cognitives. C'est le cas des idées du « *Situated Learning* », concerné autant par *l'acquisition* que la *transmission de connaissances* et pour qui (depuis la fin des années 1980) la connaissance n'est pas une « chose », ni un « panel de descriptions », ni une « collection de règles et de faits » (Resnick, 1989 ; Brown, 1989). De même pour d'autres partisans d'une approche « située » de la cognition pour qui la connaissance se définit comme « la capacité de coordonner et de séquencer un comportement en s'adaptant dynamiquement aux circonstances toujours changeantes » (Clancey, 1995), comme « une interaction dynamique » (Barab & Duffy, 2000) ou encore une « co-production » (Brown, 1989) entre le *contenu* et le *contexte*, *l'activité* et la *situation*.

¹⁰⁶ Il s'agit selon (Charlet, 2005) d'une vision *systémique* de la notion de connaissance. Une vision épistémologique insiste plus sur le fait que la connaissance repose sur la *médiation du signe* et qu'elle n'est possible que dans un *environnement technique*, la technique donnant à l'homme la possibilité de mémoriser des connaissances de manière externe, dans des outils qu'il pourra mobiliser. Pour plus de détails sur cette question, voir (Bachimont, 1996, chapitre 7).

4.3.2.2 Une ingénierie des inscriptions de connaissances

L'évolution de la notion de connaissance n'a pas qu'un impact terminologique. L'objectif fixé aux *Systèmes à Base de Connaissances* que l'IC construit a ainsi évolué depuis ses origines : aux capacités de représenter et d'opérationnaliser des connaissances formelles, on ajoute celle de proposer des données, sources d'interprétation par l'utilisateur, de prendre en compte et d'explicitier le contexte d'utilisation de ces données, de fournir à cet utilisateur les moyens – informatiques – d'agir et donc de réécrire les résultats de son interprétation (Charlet, 2004). De son côté B. Bachimont défend l'idée que, la connaissance n'étant pas un objet matériel sur lequel une ingénierie pourrait avoir prise, la connaissance « ne peut s'appréhender techniquement, c'est à dire *via* une ingénierie, qu'à travers les inscriptions qui l'expriment » (Bachimont, 2004, p.55). Ainsi, l'IC serait plus proprement désignée comme étant une *ingénierie des inscriptions de connaissances* (Bachimont, 2004).

Sous l'impulsion de ce changement de perspective, l'IC revoit ses objectifs et ne vise plus l'édification de « systèmes intelligents » stockant, produisant ou transmettant des connaissances, mais la conception de « technologies cognitives permettant à un usager de s'approprier des connaissances pour lui anciennes (remémoration) ou nouvelles (constitution) ; des systèmes techniques rendant leur usage intelligent. À condition qu'ils explicitent des *normes d'interprétation*¹⁰⁷, les systèmes formels peuvent toutefois posséder (en partie au moins) des propriétés pour lesquelles ils avaient été pensés au départ, notamment en terme de *création* de connaissances. L'*effectivité* des modèles formels peut en effet conduire à la création de combinaisons de symboles inattendues, devenant source d'innovation potentielle pour les utilisateurs de ces systèmes, poussés à inventer des interprétations nouvelles, à réécrire les connaissances et les relations qu'elles entretiennent : « De même que l'ordinateur ne « voit » pas les images qu'il permet de construire, il ne pense pas les nouvelles inscriptions qu'il formule, mais il permet de voir du nouveau comme de penser autrement » (Charlet, 2004, p.7).

Pour désigner cette nouvelle perspective, B. Bachimont (2004a) propose le terme de « raison computationnelle », en référence à la « raison graphique » de Goody (1979), pour laquelle « l'outil informatique n'est pas tant un outil de modélisation qu'un nouveau type de support pour l'inscription de connaissances ». Ainsi, l'Ingénierie des connaissances se révèle être une « ingénierie des inscriptions de connaissances » dont l'objectif sera de concevoir des « dispositifs de manipulation de ces inscriptions en vue de leur interprétation » (Cahier, 2006), en élargissant la notion d'inscription à tout ancrage matériel de connaissance, mais en se focalisant pour l'occasion sur les inscriptions *numériques* de connaissances. Les inscriptions numériques en question font référence à l'ensemble des techniques d'inscription mobilisant le support numérique pour l'expression, la transmission, le partage et l'appropriation des connaissances (Bachimont, 2004a). Nous ne pouvons pas rentrer plus en détail dans une discussion de la notion de connaissance, et devons nous concentrer sur des aspects plus pragmatiques de la modélisation des connaissances.

4.3.2.3 Les modèles de connaissances

On peut distinguer deux types de modèles de connaissances, les *modèles de raisonnement* et les *modèles du domaine* (Bachimont, 2004a, p.125). Les modèles de raisonnement représentent et/ou opéra-

¹⁰⁷ Permettant de guider le travail intellectuel de l'utilisateur entre le sens strict des symboles et des calculs et le sens induit, issu de l'interprétation linguistique de ces symboles.

tionnalisent des processus de résolution de problèmes. L'IC se trouve là plus directement en phase avec son héritage de l'IA dont elle a revisité la manière d'opérationnaliser des concepts proposés à l'origine en psychologie, puis en sciences cognitives (Newel, 1982 ; Hoc 1987). Ne s'attachant pas à la description fine des processus cognitifs engagés, les chercheurs visent plutôt à mettre en évidence l'existence de modèles *génériques* et *réutilisables* de résolution de problème. De manière générale, la description et l'opérationnalisation de ces processus reposent sur des outils issus de l'IA, tels les langages *LISP* (Bouaud, 1989) ou *PROLOG* (Colmerauer et Roussel, 1996). Les *modèles de domaine* décrivent les connaissances relatives à un domaine d'application, par exemple des « connaissances métiers ». La représentation formalisée de cette description de ces connaissances doit permettre la tenue de processus de « raisonnement » sur ces connaissances, ce qui est le principe même d'une *ontologie* (cf. infra). Pour reprendre un exemple classique, dans le domaine médical, un modèle du domaine va décrire un ensemble de connaissances relatives aux symptômes de diverses pathologies ainsi que leurs interactions connues, et un modèle de raisonnement permet d'appliquer des raisonnements conduisant à un diagnostic ou bien à l'élaboration d'une thérapie adaptée à la pathologie identifiée.

4.3.2.4 Les Système à Base de Connaissances (SBC)

Un SBC est un système informatique manipulant des représentations symboliques selon des prescriptions formalisées lors de la modélisation des connaissances. Ces représentations symboliques renvoient à des notions du domaine, des *primitives*, qui seront manipulées en respectant la grammaire du système formel dans lequel elles s'inscrivent. Ces manipulations sont des constructions *syntaxiques*, et même si les primitives renvoient à des termes interprétables pour tout spécialiste du domaine, les règles formelles de manipulation ne sont pas les règles de *l'interprétation*. Or, l'interprétation des expressions générées par le système repose sur le contexte d'usage du SBC, dimension dont il faudra nécessairement tenir compte dans le processus de modélisation lui-même. Ainsi faut-il voir les SBC comme « des systèmes sémiotiques de manipulation d'inscriptions symboliques, dont le fonctionnement informatique doit permettre à un utilisateur d'interpréter et de comprendre le système dans le cadre de son activité et de ses usages, en utilisant les termes du domaine » (Charlet, 2004, p.8).

4.3.2.5 Les ontologies

Un aboutissement classique pour un processus de modélisation de domaine est la *représentation* d'un ensemble de connaissances fondamentales du domaine en question. Pour désigner cet ensemble de connaissances qu'elle opérationnalise, l'informatique utilise le terme d'*ontologie* qu'elle emprunte au vocabulaire de la philosophie. À l'origine l'ontologie se définit comme la « doctrine ou théorie de l'être » (Ricoeur, 1999) qui, constatant qu'il existe « quelque chose » dans le monde se demande « quoi ? ». De manière similaire l'informatique se pose le même type de question dans le cadre restreint d'un domaine de connaissance particulier. Déterminer les connaissances fondamentales d'un domaine revient à se demander « ce qui existe » dans ce domaine.

Le développement des ontologies est lié à la volonté de construire mieux et plus rapidement des SBC en réutilisant le plus possible d'éléments génériques, tant sur le plan du raisonnement que des connaissances du domaine. On peut en donner la définition suivante, fondée sur les spécifications de T. Gruber (1993) et de M. Uschold et M. Gruninger (1996) : « une ontologie implique ou comprend une certaine vue du monde par rapport à un domaine donné. Cette vue est souvent conçue comme un

ensemble de concepts (entités, attributs, processus), leurs définitions et leurs interrelations. On appelle cela une contextualisation. Une ontologie peut prendre différentes formes mais elle inclura nécessairement un vocabulaire de termes et une spécification de leur signification. » (Charlet, 2004, p.14). Ainsi une ontologie est d'abord une *conceptualisation* étant entendu qu'elle définira des *concepts* (se rapprochant en cela d'une *taxinomie*), puis un *artefact informatique* dont on veut spécifier le comportement. La manière dont *sont construites* les ontologies nous intéresse tout particulièrement.

4.3.3 Méthodes de modélisation des connaissances

4.3.3.1 Élaboration d'un Système à Base de Connaissance

Au-delà des modèles de connaissances en eux-mêmes, dont nous n'allons pas approfondir ici la présentation, ce sont les principes de base de la *création* de ces modèles qui nous intéressent ici. En d'autres termes nous nous sommes intéressés à la manière dont sont élaborés les Systèmes à Base de Connaissances. Cette élaboration inclue la création du modèle de connaissance au sens strict ainsi que la conception de l'artefact informatique qui permettra de l'exploiter, ce dernier étant à considérer avec son contexte d'usage (Bachimont, 2004a). L'objectif du modélisateur est de faire en sorte que le comportement du SBC conçu soit compréhensible pour son utilisateur. Pour cela, il doit définir son modèle au « niveau des connaissances » (Charlet, 2004), *i.e.* que le modèle « fasse sens » aux yeux de l'utilisateur du SBC.

On attribue habituellement deux aspects au travail de modélisation des connaissances : d'une part *l'élicitation* des connaissances qui vise à extraire, expliciter, faire émerger les connaissances à modéliser, et d'autre part la *représentation* de ces connaissances, qui formalisées sont éventuellement opérationnalisables. Pour être plus précis, le processus d'élaboration d'un SBC se structure classiquement en quatre étapes (Aussenac-Gilles et *al.*, 1992) :

1. Le recueil d'expertise : déterminer les données du problème et les connaissances pour le résoudre (en empruntant des méthodes à la psychologie ou à l'ergonomie).
2. La construction d'un schéma conceptuel : élaborer un cadre général abstrait de résolution (qui ne précise pas tous les détails), qui sera effectué par le système.
3. La spécialisation du schéma conceptuel : utiliser les connaissances du domaine et de l'application pour compléter et préciser le schéma conceptuel.
4. L'opérationnalisation du modèle conceptuel : choisir un langage de programmation dont les structures correspondent à celles du modèle conceptuel.

4.3.3.2 Méthodes de conception d'ontologies

Les méthodes de conception d'ontologies peuvent également constituer une source d'inspiration. De nombreuses méthodes ont été créées et proposées par les chercheurs en IC, parmi lesquelles on peut citer, à titre d'exemple, *Methontology* (Gomez-Pérez et *al.*, 2004 ; Corcho et *al.*, 2005) qui intègre la construction d'ontologies dans un processus de gestion de projet complet partiellement soutenu par

l'environnement *WebODE*¹⁰⁸. Il y a aussi *Ontospec* (Kassel, 2002) qui propose une méthode de spécification semi-informelle d'ontologies conceptuelles en langue naturelle fortement structurée et contrôlée, et qui met l'accent sur l'échange entre le modélisateur et l'expert au cours du processus de conception. Ou encore la méthode *Archonte* que nous avons choisi de détailler ici.

La méthode *Archonte* (*ARCHitecture for ONTological Elaborating*) proposée par B. Bachimont (Bachimont et al., 2002) pour construire des ontologies s'appuie sur la *sémantique différentielle*, et propose plusieurs étapes (Baneyx, 2007). Il faut commencer par *choisir les termes* pertinents du domaine et normaliser leur sens, puis justifier la place de chaque concept dans la hiérarchie ontologique en précisant les relations de similarités et de différences que chaque concept entretient avec ses concepts frères et son concept père. Il faut ensuite *formaliser* les connaissances, ce qui implique par exemple d'ajouter des propriétés à des concepts, des axiomes. Il faut enfin *l'opérationnaliser* dans un langage de représentation des connaissances. Chaque méthode est adaptée au contexte dans lequel elle est mise en œuvre. Ainsi par exemple, dans le cadre du projet *Menelas* (Bouaud et al., 1995 ; Bachimont, 2000a), voici les étapes de modélisation ontologique proposées (Charlet, 2004, p.17-18) :

1. Analyse de corpus (du corpus au signifié) : dans un domaine (ici la médecine) où les connaissances sont exprimées largement en langue naturelle, il est possible de construire un corpus textuel à partir duquel caractériser les notions utiles à la modélisation ontologique et le contenu sémantique qui leur correspond. L'analyse en question utilise largement les outils et techniques permettant de construire des RTO¹⁰⁹ pour produire des candidats termes.
2. Normalisation sémantique (du signifié au signifié normé) : ces candidats termes comportant potentiellement ambiguïtés, et définitions circulaires, il est nécessaire d'en normaliser les significations pour ne retenir, pour chacun d'eux, qu'une seule signification, qu'une seule interprétation possible par un être humain. C'est ce que propose la sémantique différentielle de B. Bachimont. On normalise ainsi un ensemble de termes en construisant un système de différences entre ces termes. Dans ce contexte, la structure construite est un arbre (Bachimont, 2000a). À la fin de cette étape, l'ontologie construite n'est pas formelle : c'est un arbre de signifiés linguistiques normés.
3. Engagement ontologique (du concept linguistique au concept formel) : à cette étape, nous avons un arbre de *primitives linguistiques* qu'il faut exprimer dans une sémantique *formelle*. Cette formalisation passe par une définition en *extension* des objets définis en *intention* dans les étapes précédentes. La structure de l'ontologie passe de l'arbre au *treillis* avec la possibilité de définir des concepts *définis*, composés de concepts *primitifs*.
4. Opérationnalisation : la dernière étape consiste à exprimer l'ontologie à l'aide d'un langage formel de représentation des connaissances qui permet de réaliser des inférences (classification, généralisation, etc.), lesdites inférences dépendant des propriétés du langage choisi.

Relativement à notre recherche nous avons là un exemple de structuration générale de l'activité de modélisation dont nous pouvons nous inspirer. Bien entendu la fin du processus, c'est-à-dire *l'opérationnalisation* des connaissances modélisées, va différer. En principe en effet, le processus

¹⁰⁸ Voir <http://webode.dia.fi.upm.es/WebODEWeb/index.html> (Consulté le 08/05/2009).

¹⁰⁹ Ressource Termino-Ontologique. Ces outils font de l'extraction terminologique et permettent d'obtenir des signifiés linguistiques avec une organisation plus ou moins structurées, souvent sous forme de réseaux.

classique de modélisation de connaissances aboutit à leur expression dans un langage plus ou moins formel qui en permettra par la suite l'opérationnalisation : « représenter des connaissances propres à un domaine consiste à décrire et à coder les éléments de ce domaine pour qu'une machine puisse les traiter, raisonner et résoudre des problèmes particuliers » (Baneyx, 2007, p.49). Plusieurs formalismes parmi lesquels les graphes conceptuels J. Sowa (1984) ou les logiques de description, permettent de représenter des ontologies, composées de concepts possédant des propriétés et des relations, et de mettre en œuvre sur cette représentation des *mécanismes d'inférence*.

Concernant notre approche, l'opérationnalisation et les propriétés inférentielles de la représentation finale d'une ontologie nous intéressent moins que l'élicitation et l'extraction des connaissances. Il y a en effet une forte similitude entre cette démarche et celle qui consiste à modéliser une trace d'utilisation. Là où l'ontologue cherche à extraire des connaissances du domaine formalisables pour permettre un raisonnement automatique, le modélisateur de trace cherche à extraire des connaissances mobilisées concrètement dans la réalisation d'une activité pour en rendre compte dans une trace numérique de cette activité.

4.3.3.3 Élicitation et/ou extraction des connaissances

Extraction de connaissances à partir de textes

Globalement, on peut distinguer deux démarches d'élicitation des connaissances dont la première consiste à extraire les connaissances d'un *corpus* constitué à cet effet. Selon le contexte on réunira un ensemble de documents issus de l'activité qui mobilise les connaissances que l'on cherche à modéliser. Le corpus regroupe selon les cas, des documents très divers, dont la nature dépend du problème à résoudre et du domaine étudié. Son volume dépasse le plus souvent les moyens de traitement manuel, d'où la création et l'utilisation de traitements automatiques et d'outils issus du *Traitement Automatique des Langues* (TAL). Ces techniques permettent d'extraire de manière plus ou moins automatique¹¹⁰ des *termes* voire des *concepts* de l'ensemble des textes du corpus que le modélisateur a ensuite la charge de valider (avec l'appui d'experts du domaine).

Un exemple de ce type d'instrumentation est la plateforme logicielle *Terminae* d'aide à l'élaboration de ressources terminologiques et ontologiques à partir de textes (Aussenac-Gilles et al., 2005; Szulman et al., 2002). Cet outil intègre un environnement d'étude terminologique, un environnement d'aide à la conceptualisation et un système de gestion d'ontologies. Parmi les fonctions proposées on trouve (Baneyx, 2007) un module de dépouillement des résultats des extracteurs de candidats termes comme *Syntex* ou *YaTeA* (Bourigault et al., 2005) et un module d'exploration de corpus grâce à des patrons lexicaux et/ou syntaxiques avec *Linguae*. La plateforme dispose d'un éditeur d'ontologie qui permet la création puis la modification de celle-ci, qui sont finalement sauvegardées au format *XML*, et peuvent être exportées en *OWL* ou *RDFS*. Quelles que soient les performances de ces traitements automatisés, la validation passe par l'expertise des acteurs du domaine (et futur utilisateurs du SBC) qui sont *in fine* les *seuls* à même de juger de la pertinence effective des concepts proposés dans le modèle de connaissance.

¹¹⁰ Avec une implication humaine variable : selon les cas on verra le traitement comme étant piloté par le modélisateur ou bien strictement sans intervention de celui-ci.

Recueil auprès d'experts du domaine

La seconde méthode consiste donc à recueillir directement les connaissances auprès d'experts du domaine. Il s'agit d'amener un acteur à *exposer et expliciter* ses propres connaissances. Dans cet exercice délicat le modélisateur est confronté à des problèmes qui relèvent directement des Sciences Humaines, et c'est en toute logique que l'IC emprunte pour la circonstance leurs outils à la Psychologie, aux Sciences Cognitives ou à l'Ergonomie. Certaines difficultés rencontrées sont d'ordre cognitif. Il est en effet de manière générale difficile à tout un chacun d'exprimer directement et hors contexte d'action ses propres connaissances. Dans le cas d'un expert la situation est encore plus compliquée car beaucoup de ses connaissances du domaine sont passées de factuelles à opératoires, *i.e.* des connaissances qui sont mobilisées inconsciemment à travers des routines. On peut y ajouter des problèmes d'ordre socio-cognitif car les experts d'un domaine ont une vision personnelle des connaissances de celui-ci, parfois difficilement partageable par tous (y compris par d'autres experts) alors même que l'ontologie doit montrer une certaine forme de consensus. Enfin il y a également de manière plus terre à terre des problèmes d'ordre socio-organisationnel, car les experts ne sont ni les plus disponibles ni forcément les plus enclin à prendre le temps d'explicitier de manière exhaustive des connaissances qui ne représentent pour eux qu'un faible intérêt.

Pour faciliter l'activité d'expression et l'explicitation des connaissances de l'expert plusieurs outils peuvent être mobilisés. Tout d'abord en dehors du temps de l'action elle-même, les questionnaires, entretiens, formels ou informels, dirigés ou non, individuels ou collectifs. Plus en prise avec l'action, on peut aussi faire appel à des mises en situation avec des séquences de verbalisation de l'action, des simulations ou encore des techniques d'auto-confrontation (Rix et Lièvre, 2005). On rencontre également, selon le domaine impliqué, des techniques plus originales comme les techniques de narration ou de « *storytelling* » (Soulier, 2003, 2007)¹¹¹.

Nous pouvons noter deux « tendances » dans les outils qui sont aujourd'hui développés pour instrumenter cette phase du processus de modélisation des connaissances. La première concerne l'utilisation de langage de *représentation graphique*. Cette forme d'expression était déjà largement utilisée par le passé mais manuellement (crayon et papier). Aujourd'hui les outils proposés prennent en charge l'édition et la manipulation de représentations graphiques pour faciliter l'explicitation des connaissances. Pour certains d'entre eux, ces outils sont de plus capable d'exporter ces représentation dans des formats de données directement compatibles avec l'opérationnalisation des connaissances représentées (Paquette, 2002 ; Héon *et al.*, 2009). Les représentations graphiques ont l'avantage d'être facilement partageable, un groupe d'acteur pouvant discuter en même temps autour d'une même représentation globale, ce qui nous amène à la deuxième « tendance », qui est celle de la « co-conception ».

Dans la « co-conception » d'une ontologie (ou d'un SBC en général), la modélisation devient le fruit d'un travail coopératif entre plusieurs acteurs participants activement au processus de création du modèle de connaissance. Ce qui est entendu comme une « co-conception », ce n'est pas seulement une implication plus prégnante des acteurs du domaine aux côtés du modélisateur, mais également les échanges constructifs entre les acteurs eux-mêmes, de sorte que dans certain cas le modélisateur de connaissances prend une posture de pilotage d'un réseau ou d'une petite communauté qui produira

¹¹¹ Les récits sont en effet une forme extrêmement riche d'expression des connaissances qu'il s'agit ici d'exploiter au mieux.

elle-même une modélisation de ses connaissances. Il est intéressant de noter que la contribution de novices du domaine est un réel apport aux démarches de modélisation. Ils apportent un point de vue critique, notamment sur l'accessibilité et l'interprétabilité d'une modélisation du domaine, ainsi que sur l'utilisabilité et la facilité de prise en main d'un système informatique mis en place. La collaboration entre expert et novice lors de la phase de recueil des connaissances est même souhaitable (Basque et al., 2008).

D'une certaine façon cette pratique de co-conception d'une représentation de connaissances rejoint les approches plus théoriques qui défendent une vision intrinsèquement collective et sociale des connaissances. De fait, l'intégration d'une dimension *sociale* au sein de l'IC a donné naissance à un courant dont on voit émerger les idées sur divers terrains d'application. C'est le cas par exemple du *Web socio-sémantique* qui, dans le prolongement du *Web sémantique*¹¹², prend en compte la nature « sociale » du *Web* (Zacklad, 2005b). Cette approche considère la co-construction d'une représentation du domaine et de l'organisation sociale de l'activité de la communauté. Reposant sur les standards du Web sémantique d'un point de vue technique, le Web socio-sémantique propose un point de vue conceptuel original en créant des outils pour une méthodologie de co-conception d'une représentation d'un domaine. Dans le cas de l'application « *DKN SeqXAM* », il s'agit de permettre la comparaison de classifications réalisées par plusieurs experts sur un ensemble de termes identiques (Zaher et al., 2006).

4.3.4 Bilan

La présentation que nous venons de donner de l'Ingénierie de Connaissances, de ses concepts et de ses outils et méthodes reste parcellaire car nous l'avons focalisée sur les seuls aspects qui pouvaient nous être directement utiles dans notre propre démarche. Il existe des différences fondamentales entre un Système à Base de Traces (SBT) et un Système à Base de Connaissances (SBC) à proprement parler. Ceci étant, la démarche de modélisation nous semble quant à elle avoir des points de similitudes importants, raison pour laquelle nous nous en inspirerons largement. Sur un plan théorique tout d'abord, nous retiendrons de l'IC sa définition des connaissances et de leurs inscriptions. Sur un plan méthodologique général nous retiendrons que le travail de modélisation de connaissances ne peut être dissocié du travail de conception d'un artefact informatique (SBC), et de manière plus détaillée nous retiendrons la décomposition en étapes d'un processus de modélisation de connaissance qui guide le modélisateur, avec un intérêt particulier pour les premières étapes d'élicitation / extraction de connaissances et de construction d'un schéma conceptuel.

Ce qui différencie nettement notre démarche de modélisation de trace et une démarche de modélisation des connaissances classique est que cette dernière vise à représenter des connaissances « stables » du domaine, c'est-à-dire « valides » pour un maximum d'acteurs sur la durée la plus longue possible, alors que nous cherchons à faire des traces numériques une représentation de connaissances qui sont liées à une expérience singulière de réalisation d'une activité qui n'a de validité que dans le périmètre de l'activité réalisée. Le seul « expert du domaine » dans ce cas est l'utilisateur qui a réalisé l'activité en question, ou bien à la rigueur, l'analyste qui pouvait l'observer à ce moment là.

¹¹² Le Web sémantique constitue une extension sémantique du Web courant dans laquelle l'information a un sens défini permettant aux ordinateurs et aux hommes de mieux travailler en coopération « The semantic web is not a separate web, but an extension of the current one, in which information is given well-defined meaning » (Berners-Lee et al., 2001). Aujourd'hui, le web sémantique s'apparente à une architecture informatique permettant entre autres l'exploitation de connaissances formalisées (Charlet et al., 2004).

Pour modéliser une trace il ne faut donc pas seulement analyser l'activité réalisée du point de vue des connaissances du domaine qu'elle mobilise de manière régulière, mais également du point de vue des connaissances du contexte qu'elle mobilise de manière singulière. En ce sens l'approche de modélisation que nous propose l'IC ne peut répondre complètement à nos besoins. Il nous fallait sur ce point compléter les outils conceptuels dont nous disposions pour aborder l'analyse de l'activité effective d'un utilisateur, des concepts nous permettant de rendre compte de la singularité de la réalisation de l'activité dans un contexte cognitif, social et organisationnel toujours renouvelé. Sachant que les terrains d'applications que nous avons à aborder sont l'apprentissage collaboratif à distance et la production collaborative de contenus, nous nous sommes tournés vers le domaine du *CSCW*.

4.4 Apports du *CSCW*

Nous avons choisi de mobiliser le domaine du *Computer Supported Cooperative Work (CSCW)* pour y puiser les *moyens théoriques* d'analyser et de décrire très précisément les activités documentaires instrumentées dont nous voulons fournir une trace modélisée. Pas uniquement en tant qu'activités abstraites, mais comme activités *réalisées effectivement*, dans une situation concrète toujours singulière. Ces apports théoriques doivent nous fournir d'une part une *grille d'analyse* de l'activité tracée, et d'autre part une *vision* de l'activité à partir de laquelle nous pourrions penser les *usages* potentiels des traces modélisées que l'activité instrumentée soit individuelle ou collective. Après avoir présenté le *CSCW* comme domaine de recherche ainsi que ses idées fondatrices (section 4.4.1), nous présenterons plusieurs approches théoriques de l'activité couramment utilisées en *CSCW*, en réservant une place centrale à la Théorie de l'Activité (section 4.4.2), que nous compléterons avec d'autres approches issues de diverses disciplines (4.4.3).

4.4.1 Présentation du *CSCW*

Le domaine de recherche du *CSCW* s'est constitué il y a une quinzaine d'années, à un moment où les innovations technologiques ouvraient la porte à l'ère de la collaboration instrumentée et au modèle naissant de l'entreprise étendue : « L'architecture des futures organisations « virtuelles » reposerait toute entière sur un réseau de machines à coopérer dotées du pouvoir d'abolir les distances et de faire naître entre ses membres une connectivité généralisée » (Cardon, 1997, p.25).

Le terme de « *CSCW* » a été utilisé pour la première en 1984 par Cashman et Greif, à l'occasion de l'organisation d'un colloque interdisciplinaire intitulé « *How to support people in their work arrangements with computers* » (Schmidt et Bannon, 1992, p.3). Le terme semble cependant avoir précédé de quelques années le champ de recherche qu'il désigne. En effet, il faut attendre décembre 1986 et la première organisation de la conférence *CSCW*, à Austin au Texas, pour que le *CSCW* se définisse lui-même comme un champ de recherche. Cette conférence biannuelle alterne depuis 1989 avec l'édition européenne, plus « universitaire », de l'*ECSCW* ou « *European Conference on Computer Supported Cooperative Work* » (Tollmar, 2001). Le *CSCW* s'est structuré rapidement dès sa création grâce à une

institutionnalisation intense, s'appuyant en quelques années sur deux conférences phares mais aussi sur des dizaines d'ouvrages, des forums internationaux, des colloques et des centaines d'articles¹¹³.

4.4.1.1 Rupture et pluridisciplinarité fondatrices

Lors de sa création le *CSCW* voulait marquer une *rupture* dans le traitement scientifique de la question des situations de travail coopératif. Bien que l'on ne puisse pas parler de rupture technologique au sens strict, le *CSCW* souhaitait se différencier de la branche de l'informatique qui se consacrait alors au développement de logiciels de communication et de partage de documents, connue sous le nom de « *Groupware* » (Schmidt et Bannon, 1992). Pour une poignée de chercheurs, il est apparu qu'une démarche de développement technologique isolée ne pouvait répondre seule aux attentes et aux questions soulevées par l'instrumentation des situations instrumentées de travail coopératif (Salembier, 2002), et qu'il était nécessaire de prendre du recul relativement à l'idée forte d'une « *enabling technology* » (Longchamp, 2003). Pointant les limites d'une approche purement technologique, ces pionniers se tournèrent vers des disciplines de *Sciences Humaines et Sociales* (SHS), pour tenter de *comprendre, analyser et expliquer* les activités coopératives qu'ils se donnaient pour objectif d'instrumenter : « *CSCW should be conceived as an endeavor to understand the nature and characteristics of cooperative work with the objective of designing adequate computer-based technologies* » (Schmidt et Bannon, 1992, p.3).

Le nombre des disciplines mobilisées s'est rapidement élargi (Psychologie, Psychologie cognitive, sociale ou des organisations, Sociologie, Ethnologie et Anthropologie, Sciences Cognitives, Ergonomie, Sciences de Gestion, Systémique, Sciences des organisations, *etc.*) de sorte que le *CSCW* est finalement devenu un « label » auquel s'identifient une grande variété de chercheurs appartenant à diverses disciplines et travaillant autour des questions de conception de systèmes informatiques et de leurs usages en situation de coopération¹¹⁴. Généralement focalisées sur l'étude de groupes restreints (Schmidt, 2002) en situation de coopération dans une activité de travail aux interactions riches¹¹⁵, on attendait des analyses menées en *CSCW* qu'elles répondent à trois types de questions (Grudin, 1994) :

- Comment les gens travaillent-ils ensemble ?
- De quoi ont-ils besoin pour travailler ensemble en tant que groupe ?
- Comment les technologies informatiques (et de communication) peuvent être développées pour supporter les personnes et leurs activités ?

Cet objectif originel, qui supposait finalement que les SHS *répondent aux questions* soulevées par la *conception* de dispositifs technologiques, n'a jamais été complètement atteint. Le lien du *CSCW* aux technologies reste fort et le *CSCW* est encore souvent considéré comme une « caisse de résonance technologique » des études portant sur les activités coopératives (Salembier, 2004). Il est également commun de caractériser le champ du *CSCW* par une classification de ses *outils*, comme c'est le cas par exemple avec la classification des « 7 familles » d'outils proposée par D. Cardon (Cardon, 1997), celle du « Trèfle de Ellis » (Gronier et Sagot, 2002), ou encore celle de la « Matrice de Johansen » (voir

¹¹³ Pour une analyse plus détaillée voir (Cardon, 1997).

¹¹⁴ A tel point que très tôt certains estimèrent que cette diversité jetait le trouble quant à l'identité et à la visibilité du domaine : « *It appeared as if the label CSCW was simply being used as a umbrella term under which a variety of people from different disciplinary perspectives could discuss aspects of computer system design and use by people* » (Schmidt et Bannon, 1992).

¹¹⁵ Le *CSCW* se positionne entre les problématiques de Systèmes d'Information d'un côté et celles des Interfaces Homme-Machine de l'autre.

annexe 1). Comme c'est le cas ici avec la Matrice de Johansen (Figure 4.6), ces classifications reposent en général sur trois types de variables (Cardon, 1997) : le *temps* (synchrone / asynchrone), l'*espace* (en présence / à distance) et les *techniques de représentation et de visualisation* (audio / photo / vidéo, écran / tableau / réalité virtuelle).

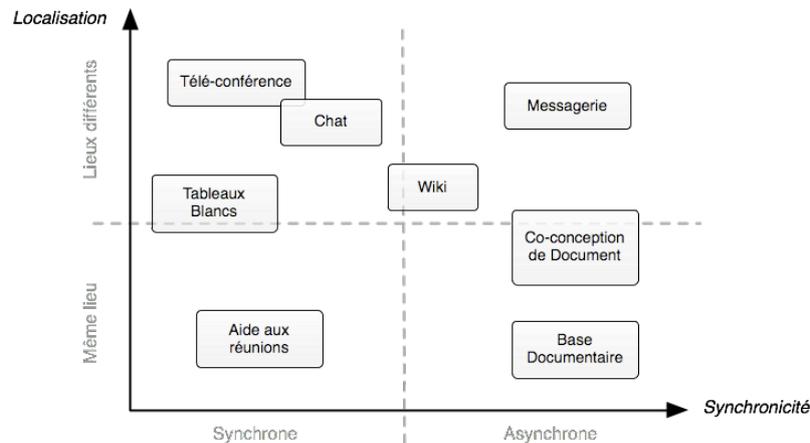


Figure 4.6 : La Matrice de Johansen permet de positionner les outils en fonctions des caractéristiques spatio-temporelles de leurs situations d'utilisation – Tiré de (Johansen, 1991).

Analyse et conception de situations instrumentées de coopération

Le *CSCW* s'efforce d'intégrer deux types de problématiques, d'*analyse* d'une part et de *conception* d'autre part (Grudin, 1994). Les premières s'appuient sur les *outils conceptuels et méthodologiques* que fournissent les SHS pour *observer, décrire et caractériser* les situations de travail étudiées¹¹⁶. Les analyses en question ne se limitent pas à des situations qu'il faut instrumenter, mais se consacrent également à l'évaluation des situations *déjà* instrumentées, notamment parce que l'impact de la mise en place d'un système de *CSCW* est tel (sur le plan social, organisationnel, humain, technique)¹¹⁷ que les résultats d'une analyse préalable perdent fortement de leur pertinence (Clases et Wehner, 2002 ; Sonnenwald et al., 2004, Cardon, 1997). Seules des approches *qualitatives* semblent à même de rendre compte de la complexité des situations de travail coopératif (Rogers et Ellis, 1994 ; Tollmar, 2001 ; Halverson, 2002) alors même que les processus de conception appellent de leurs vœux des recommandations génériques, plutôt « quantitatives ».

L'intégration, tant souhaitée, des approches de SHS à la conception puis au développement des dispositifs technologiques reste une difficulté majeure, parfois qualifiée de « *critical gap* » ou encore de « *great divide* » (Bourguin, 2000 ; Halverson, 2002). Même si des recherches sont menées sur le processus de conception lui-même pour tenter de comprendre comment réduire cette fracture (Gronier, 2002, Clases et Wehner, 2002), pour certain la *transposition directe* de méthodes de SHS à un processus de développement informatique est tout simplement impossible. Lorsque cette transposition se fait tout de même, elle se fait au prix de *modifications substantielles* qui peuvent remettre en cause l'apport de la théorie initiale dans le processus (Rosenberg et al., 1997). Du coup, comme le fait re-

¹¹⁶ Cette grille d'analyse doit permettre la mise en évidence d'aspects *critiques* de la situation de travail étudiée (de coordination, d'articulation, de gestion, etc.) afin d'adapter une réponse technologique et/ou organisationnelle.

¹¹⁷ Cet impact fait d'ailleurs lui-même l'objet d'analyses, et certains considèrent même l'introduction de systèmes coopératifs comme un véritable levier de *changement organisationnel* (Tollmar, 2001 ; Orlikowski, 2002 ; Kuuti, 1996).

marquer (Halverson, 2002), quelle que soit la théorie convoquée, le *CSCW* se tourne systématiquement vers le même ensemble d'outils et de méthodes pour la conception : « *User Centered Design* », « *Participatory design* », ou encore « *Design rationale* » (Tollmar, 2001 ; Halverson, 2002), méthodes qui soit dit en passant ne sont pas réservées aux situations de travail collectif et coopératif.

4.4.1.2 Les approches théoriques de l'activité instrumentée coopérative

Depuis ses débuts, le domaine du *CSCW* a mobilisé un nombre important d'approches théoriques qui rendent compte très diversement de l'activité humaine et dont nous ne pouvons faire une exploitation exhaustive (Figure 4.7). Nous avons donc sélectionné un sous-ensemble d'approches qui nous paraissent pertinentes relativement à nos objectifs, avec pour critère de sélection la prise en compte combinée du *caractère instrumental, cognitif et collectif ou coopératif* des activités. La première de ces approches, qui apparaît comme incontournable au sein du *CSCW* est celle de la *Théorie de l'Activité*, que nous adopterons comme *approche centrale* (section 4.4.2). Nous la renforcerons des apports de quatre *approches complémentaires* (section 4.4.3) : *l'Ergonomie Cognitive*, la *Cognition Distribuée*, *l'Action* et la *Cognition Situées*, ainsi que la *Théorie de la Coordination*.

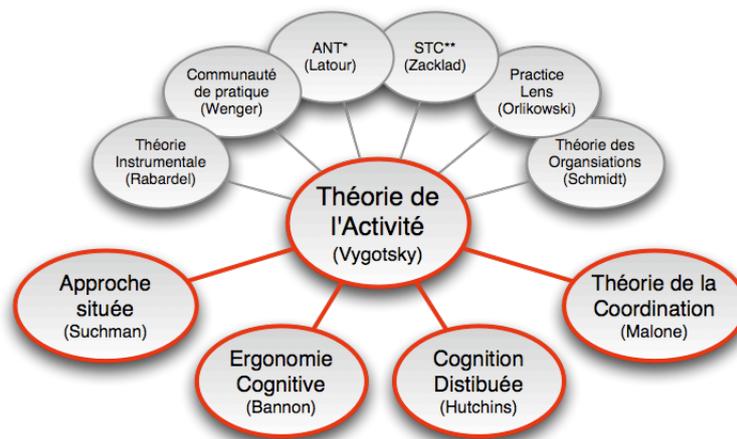


Figure 4.7 : Théories issues des Sciences Humaines mobilisées dans le *CSCW*, dont les cinq que nous avons retenues et mises en évidence.
(**Actor Network Theory*, ** *Socio-économie des Transactions Coopératives*)

4.4.2 Approche centrale : la Théorie de l'Activité

La Théorie de l'Activité prend naissance dans l'école historico-culturelle soviétique de la psychologie, fondée par L.Vygotski, dans les années 1920. Des travaux fondateurs de L.Vygotski sur sa théorie *psychologique* subsistent encore aujourd'hui des concepts forts, comme l'idée que les fonctions mentales humaines apparaissent sur le plan *inter-individuel avant* d'apparaître sur le plan *intra-individuel*, (la pensée étant alors considérée comme un dialogue intériorisé), l'idée que les processus mentaux sont *médiés* par des *outils techniques* ou *psychologiques* dont la genèse est sociale, ou encore l'idée d'une *zone proximale de développement* définissant la zone entre ce qu'un apprenant est capable d'accomplir seul et ce qu'il saurait faire avec de l'aide extérieure.

La Théorie de l'Activité doit beaucoup de sa notoriété actuelle à un prolongement des travaux de L. Vygotski par un de ses étudiants puis collègue, A. Leontiev, qui a fortement contribué à façonner la

Théorie de l'Activité que l'on connaît aujourd'hui sous la forme d'un « cadre conceptuel » plus que d'une théorie au sens strict (Hemmecke et Stary, 2004) et dont nous allons détailler les concepts de *structuration de l'activité*, de *rôle de médiation des outils*, et de *dynamique de développement de l'activité*, nous évoquerons également les apports plus récents de Y. Engeström et de K. Kuuti.

La structuration de l'activité et médiation de l'outil

La Théorie de l'Activité était la première approche à prendre *l'activité* comme *unité d'analyse* de l'étude psychologique¹¹⁸. L'activité y est définie comme un « *système cohérent de processus mentaux internes, d'un comportement externe et de processus motivationnels qui sont combinés et dirigés pour réaliser des buts conscients* » (Bourguin, 2000, p.42), et possède trois *niveaux hiérarchiques* (Figure 4.8) : les *activités*, au niveau le plus haut, naissent d'un *besoin* (ou motivation) qui se concrétise en un *objet*, physique ou mental (livre à écrire ou campagne humanitaire à mener). Ces activités sont réalisées à travers des chaînes *d'actions*, individuelles ou collectives, partageant cet objet. Les actions sont elles-mêmes réalisées à travers des chaînes *d'opérations*, au niveau le plus bas. Ces opérations sont en principe inconscientes et correspondent à des actions dont le modèle s'est montré suffisamment fiable pour qu'une orientation consciente soit inutile.

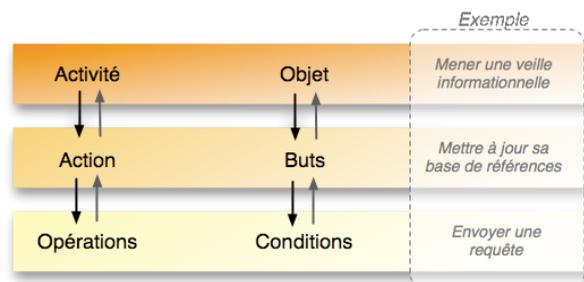


Figure 4.8 : Niveaux hiérarchiques de l'activité humaine dans la Théorie de l'Activité.

Pour la Théorie de l'Activité, il n'y a pas isomorphisme entre les buts à atteindre et les façons d'y arriver. Les opérations sont autant guidées par l'environnement que par la *structure* de l'activité dans laquelle elles sont incorporées. Cette *structure* met en scène un *sujet actif* qui réfère à l'individu (ou au groupe d'individus) dont le travail est choisi comme point de vue pour l'analyse. L'*objet* de l'activité y est quant à lui modelé et transformé à l'aide d'outils pour devenir le *produit* de système ainsi constitué. Les relations entre *sujet* et *objet* sont ainsi *médiatisées* et *culturellement construites* par l'*outil*. C'est cette médiation que met en avant le célèbre triangle « sujet - outil - objet » représenté ci-dessous (Figure 4.9).

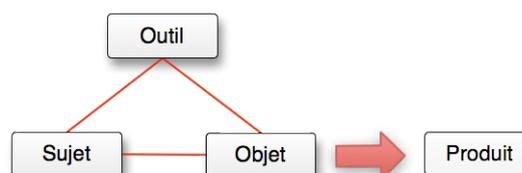


Figure 4.9 : Structure d'une activité dans la Théorie de l'Activité.

¹¹⁸ Les travaux de L. Vygotski, puis de Leontiev, forgent une théorie psychologique qui permet d'articuler les sciences sociales et les sciences du comportement, réconciliant deux unités d'analyse (le système social d'un côté, l'individu de l'autre) en une seule : l'activité.

L'outil constitue la clef du « passage d'un mode d'action direct sur le monde à l'action à travers des médiateurs » (Lonchamp, 2003). Cette idée a été utilisée par Vygotski, qui l'a appliquée aux outils psychologiques, pointant notamment la complexité de l'articulation des processus internes (mentaux) et processus externes (comportementaux), processus qui se déploient à travers l'outil et doivent impérativement être étudiés conjointement. Ainsi, pour la Théorie de l'Activité, les *outils* sont aussi bien *techniques* (un marteau) que *psychologiques*¹¹⁹ (un calendrier), et quelle que soit leur nature, ils peuvent être créés pendant le *développement* de l'activité et sont porteurs d'une certaine *culture liée* à ce développement¹²⁰.

Dynamique du développement des activités et concept de contradiction

Les éléments constitutifs d'une structure d'activité (Figure 4.9) sont en perpétuelle évolution, et peuvent *changer de statut*, un « produit » peut par exemple devenir un « outil » à un moment donné au cours de l'activité. Cette propriété permet de rendre compte du caractère *dynamique* de l'ensemble de la structure d'activité. La situation de travail et ses conditions évoluent conjointement, conférant un caractère réflexif à toute activité. Les actions ont tendance à transformer les situations, et donc les conditions qui entrent en jeu dans la réalisation des opérations : « ainsi, l'activité apparaît comme continuellement influencée par une situation qu'elle ne cesse de modifier » (Bourguin, 2000, p.45)¹²¹. La *contradiction* est le moteur de cette dynamique : « *in activity theory contradictions are seen as immanent tendencies that drive change and development* » (Guy, 2005). À l'origine, le concept de contradiction était utilisé (par Vygotski lui-même) dans une acception « forte », *i.e.* l'impossible co-existence de deux notions, idées, ou choix (Clot, 2002). Il est aujourd'hui considéré sous une forme « faible », c'est-à-dire toutes les formes d'inadéquations observables dans le fonctionnement d'un système d'activité: contradictions ou tensions structurelles entre activités, ou entre outils et pratiques (Engeström et al., 1997 ; Collins et al., 2002)¹²².

Apports à la Théorie de l'Activité

La Théorie de l'Activité s'est largement enrichie à l'occasion de son utilisation au sein du *CSCW*. Les travaux de Y. Engeström sur les aspects *supra-individuels* de l'activité sont un premier exemple de cette contribution du *CSCW* à l'évolution de la théorie (Engeström, 1997 ; Engeström, 2005). Afin de décrire des réseaux d'activités inter-reliées (*i.e.* des systèmes d'activités), Y. Engeström propose d'introduire trois nouveaux éléments : les *règles*, la *communauté* et la *division du travail* (Figure 4.10). La *communauté* est composée de multiples individus ou groupes d'individus qui partagent un même objet général (but/motivation). La *division du travail* quant à elle, désigne à la fois la division horizontale des tâches entre les membres d'une communauté et les divisions verticales, de pouvoir ou de statut, et médiatise la relation communauté-objet. Enfin, les *règles* médiatisant la relation sujet-communauté font référence aux normes explicites et implicites, conventions et relations sociales au sein de la communauté qui contraignent les actions et interactions dans le système d'activité.

¹¹⁹ Outils dans ce dernier cas symboliques, internes ou externes (Lonchamp, 2003 ; Hemmeke et Stary, 2004 ; Decortis et al., 2000).

¹²⁰ Les outils incorporent notamment l'expérience « accumulée à la fois dans les propriétés structurelles de l'outil (forme, matériaux, etc.) et dans la connaissance de l'utilisation de l'outil » (Bourguin et Derycke, 2005).

¹²¹ Cet intérêt pour l'idée de *développement* sous-jacente à la Théorie de l'Activité, s'accordent particulièrement bien avec les approches ethnographiques, qui s'intéressent tout particulièrement à l'étude de l'histoire et du développement des pratiques.

¹²² P. Collins propose de façon détaillée de voir trois types de « tensions » : *dans* les éléments d'un système d'activité (dans les Règles par exemple), *entre* ces mêmes éléments et enfin *entre* différents systèmes d'activités.

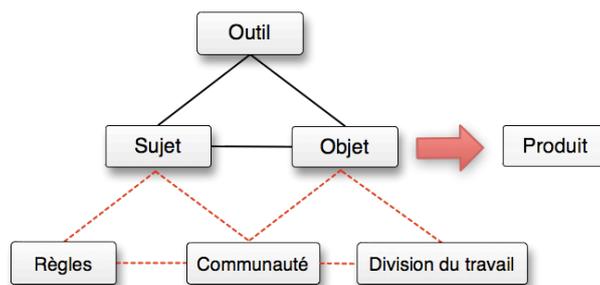


Figure 4.10 : Structuration d'un système d'activités selon Engeström (1997).

Dans cette structure étendue, les *contradictions* sont toujours à l'origine de la dynamique de *développement* de l'activité, aspect sur lequel a beaucoup insisté K. Kuuti (1991, 1993, 1996), autre grand contributeur à la Théorie de l'Activité. Ce qui est intéressant dans cette dynamique, dans le fait que le sujet puisse lui-même modifier son objet, c'est qu'elle n'est pas continue mais passe par des *périodes de stabilité* qui permettent de maintenir une certaine direction. Cette proposition permet de structurer une approche d'analyse en fonction du repérage de ces différentes phases, qui sont particulièrement pertinentes dans le cadre d'activités instrumentées par des *outils informatiques*.

Le travail de K. Kuuti s'est en effet principalement concentré sur la discussion des apports potentiels des *technologies de l'information* en relation avec le modèle à trois niveaux (opération/action/activité) et des différents composants de la structuration de l'activité. On retrouve dans le tableau ci-dessous le résultat de son analyse, qui reprend les apports potentiels des technologies de l'information en fonction de leur mode : passif (au niveau des opérations), actif (au niveau des actions) et expansif (au niveau des activités). La première catégorie concerne les systèmes classiques, la seconde les systèmes coopératifs actuels, la troisième permet d'envisager les futurs systèmes (Bourguin, 2000).

	Outil	Règles	Division du travail	Sujet «pensant»	Objet	Communauté
Passif	Automatisation de routine	Contrôle implicite et imposé	Fixée et imposée	Déclenche des actions prédéterminées	Données	Implicite et invisible
Actif	Support aux actions de transformation et de manipulation. Rendre les outils et procédures visibles	Pensées partagées et visibles	Coordination, Organisation visible du travail	Recherche l'information	Matériel partagé	Réseau visible Support à la communication
Expansif	Automatisation De nouvelles routines ou construction d'outil	Construction, négociation des règles	Réorganisation	Apprend, comprend	Construction de l'objet	Construction de la, ou d'une nouvelle communauté

Figure 4.11 : Classification des supports de travail de K. Kuuti - Tiré de (Bourguin, 2000, p.53)

La Théorie de l'Activité a également bénéficié de nombreuses mises en œuvres. On peut citer, entre autres, le travail de J. Bardram (Bardram, 1998), qui reprend dans sa thèse les niveaux hiérarchiques de l'activité et propose le concept d'*anticipation* comme clef de l'activité, en tant qu'elle guide les trois niveaux de celle-ci : l'anticipation (de la réalisation de l'objet) motive l'activité, elle est le but de l'action et la base d'orientation des opérations. On peut également mentionner le travail de (Bourguin, 2000) sur la conception d'un système d'apprentissage coopératif assisté par ordinateur, mais également diverses analyses de situations de travail fondées sur la Théorie de l'Activité : logistique de transport routier (Decortis et al., 2000), support technique clientèle chez HP (Collins et al., 2002) et réseau de santé pour (Kuutti, 1992).

4.4.3 Approches complémentaires

Nous venons de présenter en détails la Théorie de l'Activité que nous considérons comme centrale et que nous utiliserons comme un socle théorique sur lequel nous souhaitons capitaliser les apports d'autres approches que nous allons maintenant évoquer.

4.4.3.1 L'Ergonomie Cognitive

Contrairement à ce que pourrait laisser croire son nom, l'Ergonomie Cognitive (EC) n'est pas tant un sous-domaine de l'ergonomie qu'une branche du domaine des IHM¹²³, qui affirme d'ailleurs en être à l'origine (Hoc, 1998). Son objectif général est d'appliquer une devise ergonomique de la meilleure adaptation possible à l'homme de ses moyens technologiques de production, milieux de travail et de vie (Lonchamp, 2003), à des situations d'activités mobilisant des artefacts informatiques dits cognitifs. Les travaux en EC concernent avant tout les problématiques d'interfaces homme machine, mais les situations d'activité coopérative ont peu à peu gagné du terrain et ces travaux ont fini par converger nettement avec les problématiques du *CSCW*. L'activité coopérative a ainsi été abordée dès les années 1980, à travers la notion de *représentation partagée* construite par des acteurs, ainsi que les mécanismes de *régulation de la charge (cognitive) de travail* (Leplat et Hoc, 1983 ; Leplat 1991 ; Hoc, 1998). Sur le plan méthodologique l'EC s'est notamment distinguée en défendant l'idée que ce sont les observations de terrain qui priment, et qu'à ce titre ce sont aux *observations* d'être à l'origine de toute *conceptualisation* et *modélisation* d'une activité donnée : « *The approach is essentially pragmatic, experience-driven and conducted through field studies* » (Decortis *et al.*, 2000).

Cognition et tâches cognitives

Empruntés à l'Ergonomie générale, les concepts de *tâche prescrite* et de *tâche effective*, sont des éléments forts de l'Ergonomie Cognitive. Dans une terminologie simplifiée, et de façon générale en ergonomie, on parle de *tâche* pour indiquer la *tâche prescrite* (*i.e.* l'aspect explicite du travail) et *d'activité* pour indiquer la *tâche effective* (*i.e.* l'aspect implicite du travail)¹²⁴. Dans le premier cas, il s'agit de ce qu'il est demandé à l'opérateur de faire, avec éventuellement des précisions sur comment le faire. Il s'agit donc d'objectifs ou de buts fixés, ainsi que des conditions organisationnelles qui contraignent la réalisation de ces objectifs (moyens, ressources, obstacles). Dans le second cas, celui de la tâche effective, il s'agit de la réponse individuelle à la tâche prescrite (Lonchamp, 2003). Les tâches ne peuvent se définir en dehors de leur *contexte* réel d'exécution, et reposeront sur les *compétences*, les capacités *d'adaptation*, et les *stratégies cognitives* mobilisées par un opérateur les réalisant.

Fidèle au paradigme cognitiviste, l'EC considère la *cognition* comme un traitement d'information symbolique¹²⁵ (Hoc, 1998) : chaque agent *reçoit, traite, puis transmet* une information à un autre agent dans la situation considérée. Logiquement, l'EC adopte une *unité d'analyse* centrée sur l'*individu* et une méthodologie reposant fortement sur la *modélisation* des *systèmes* autant que des *utilisateurs* eux-mêmes (Cardon, 1997). De façon générale, l'EC utilise une modélisation qui permet *a minima* de dé-

¹²³ Si cette démarche a été fortement soutenue par l'école française d'ergonomie, l'EC plonge bel et bien ses racines d'abord dans le domaine des IHM. C'est en 1982 à Amsterdam qu'a eu lieu la première conférence européenne d'EC bien que ce nom n'apparaisse pas encore dans son intitulé (on lui a préféré celui de *Mind and Computer*).

¹²⁴ Nous avons évoqué cette question en évoquant la modélisation de tâches (section 5.2).

¹²⁵ Selon le modèle du Système de Traitement d'Information (STI) .

gager des *buts*, des *rôles* et des *tâches*. L'idée centrale est de mettre en place une sorte de « système d'épreuves » permettant d'effectuer des mesures « objectives » dont les résultats sont censés donner des éléments de jugement quant au fonctionnement d'un système homme-machine, en tenant compte par exemple du *temps d'exécution* d'une tâche¹²⁶.

Coordination des buts et régulation de la charge de travail

Dans le cas d'une situation de travail coopératif, l'étude de l'activité est considérée comme « réductible » à l'étude de multiples situations individuelles articulées. L'EC aborde en effet le travail coopératif en tentant de répondre à deux questions (Leplat, 1991) : comment la dimension collective intègre-t-elle l'activité individuelle, et comment l'activité individuelle est déterminée et s'adapte à l'activité collective ? Et quels sont les mécanismes de régulation interne et le rôle des artefacts dans ces mécanismes ? L'intégration de l'activité collective dans l'activité individuelle se traduit par plusieurs phénomènes dont la *coordination des buts* et la *régulation de la charge de travail*.

La vision centrée sur l'individu de l'EC met en avant le caractère idiosyncrasique du passage de la tâche à l'activité : chaque membre d'un collectif a sa *propre interprétation* des *buts* de son activité (et de celle du groupe). Au-delà des divergences engendrées par une interprétation biaisée de l'un des membres (par manque de connaissances spécifiques par exemple), l'EC tente de mettre en lumière des *buts conflictuels* qu'il est difficile de coordonner. C'est le cas par exemple dans le contrôle aérien où les critères de « réussite » peuvent être opposés « rapidité vs confort » par exemple. La *régulation* de la charge de travail est un phénomène observable qui concerne directement *l'intégration* de l'activité collective à l'activité individuelle.

Il se fonde sur la *conscience mutuelle* que les opérateurs ont de la répartition de la charge de travail. L'augmentation de la charge de travail d'un opérateur, induisant une baisse de la communication et d'éventuelles erreurs de synchronisation par exemple, est ainsi tout à fait comprise par les autres. Ces mêmes opérateurs peuvent, en communiquant, tenter de se coordonner pour mieux répartir les charges de travail (Giboin, 2004). Les *mécanismes de régulation* nécessaires¹²⁷ reposent en grande partie sur les moyens de communication et de diffusion de l'information dans une situation donnée.

Artefacts et ressources de l'environnement dans un référentiel commun

Dans son analyse d'une situation d'activité instrumentée, l'EC se préoccupe également du *contexte* de réalisation. En l'occurrence, il s'agit de déterminer d'une part les *facteurs de complexité* de l'environnement, des acteurs, des tâches et de l'activité, et d'autre part de s'intéresser aux *ressources* présentes dans l'environnement (Decortis et al., 2000). Parmi ces ressources, les *artefacts* jouent un rôle clef dans la *coordination* des opérateurs : leur utilisation varie en fonction de la situation et d'un opérateur à l'autre, mais ils restent à la fois un moyen de traiter un problème et un moyen de communication avec les autres opérateurs.

Cette coordination, pour artefactuelle qu'elle soit, ne prendra effet qu'à condition qu'il existe un *référentiel commun* aux différents opérateurs impliqués. Défini par De Terssac et Chabaud (1990), la no-

¹²⁶ Nous pouvons faire ici un parallèle avec ce que nous avons dit sur la modélisation de tâche (section 4.2).

¹²⁷ On peut distinguer les mécanismes de *régulation fonctionnelle* des mécanismes de *régulation structurelle*. Les premiers concernent les corrections dynamiques des effets de perturbations *externes*, les seconds des adaptations du système en présence de modifications *internes*.

tion de *référentiel (opératif) commun* désigne une « représentation fonctionnelle commune aux opérateurs, qui oriente et contrôle l'activité que ceux-ci exécutent collectivement » (Giboin, 2004). Il s'agit en fait de « l'intersection » des représentations fonctionnelles des différents opérateurs. Ces représentations évoluent à partir de l'interprétation des prescriptions, en fonction des compétences et de la pratique des opérateurs en question. Plusieurs définitions de la notion de référentiel commun ont vu le jour, et de nombreuses notions similaires sont à signaler : « contexte partagé », « terrain commun », « représentation fonctionnelle partagée », *etc.* (Giboin, 2004 ; Lonchamp, 2003).

Méthodologie d'analyse

À partir des réflexions de ses débuts, dites de « génie cognitif » et portant avant tout sur les systèmes eux-mêmes, l'EC est peu à peu arrivée à considérer que son objectif n'est plus de développer uniquement des outils, mais de *situations de travail* toutes entières en tant qu'elles intègrent les dimensions humaines et technologiques du problème. La conception de ces situations de travail repose sur l'analyse de terrain réalisée à différents *niveaux de précision* (Decortis *et al.*, 2000). Une analyse à « gros grains » permet tout d'abord de donner une description générale de l'environnement de travail. Par la suite une analyse à « grains fins » permet de détecter des micro-mécanismes cognitifs qui guideront la suite de l'analyse (Bressole *et al.*, 1996) dont l'objectif final consistera à *remonter* vers une conceptualisation de l'activité.

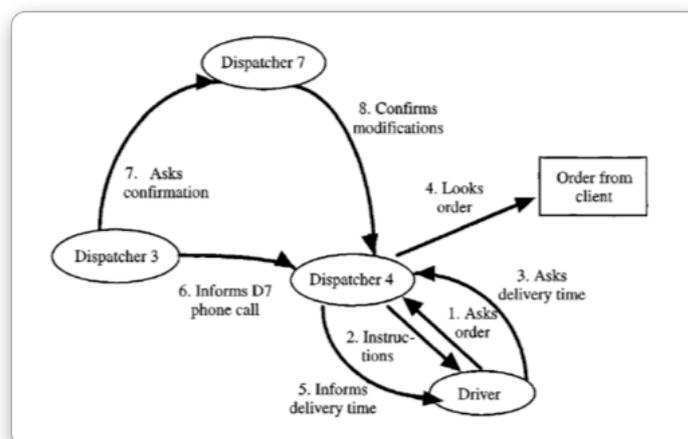


Figure 4.12 : Exemple de schéma utilisé dans l'analyse micro (échange entre deux agents, entendu et réutilisé par d'autres agents dans le même espace) - Tiré de (Decortis *et al.*, 2000, p.22).

L'analyse « micro » peut consister par exemple à étudier par le détail les échanges des agents impliqués dans une activité donnée. F. Decortis par exemple s'intéresse à ce niveau d'analyse au « *mutual awareness* » en s'appuyant pour cela sur une décomposition et une représentation de certains échanges, tels que présentés dans la figure ci-dessus (Figure 4.12). Ce type d'analyse a été mené dans diverses situations, comme ici dans un bureau de gestion logistique d'un transporteur routier par exemple (Decortis *et al.*, 2000) et souvent dans des situations à risque comme dans les salles de contrôle de centrale nucléaire (Rognin, 1999) ou de trafic aérien (Dumazeau et Karsenty, 2000).

4.4.3.2 La Cognition Distribuée

La théorie de la *Cognition Distribuée* (CD) est le fruit du travail de E. Hutchins et de ses collègues de l'université de Californie à San Diego dans les années 1990. Elle fût mise au point dans un effort de recherche visant à décrire et analyser des situations de travail, dotées d'*artefacts technologiques*, et considérées comme des *systèmes computationnels complexes* (Decortis et al., 1997). Les premiers exemple de situations étudiées furent le pilotage d'un navire et celui d'un avion (Hutchins, 1994, 1995 ; Hutchins et Klausen, 1996). L'originalité (et l'apport principal) de la Cognition Distribuée réside dans le changement de *perspective* qu'elle propose sur l'activité cognitive, la considérant la *cognition* comme le fait d'un *système sociotechnique* dans lequel agents humains et artefacts font circuler et transforment de l'information. Cette approche qui associe *contextualisation sociotechnique* et *perspective computationnelle* de la cognition adopte une méthodologie d'analyse empruntée à l'anthropologie cognitive et à l'ethnométhodologie.

Le système sociotechnique

Le changement de perspective que propose E. Hutchins consiste à passer (au sens premier) d'une perspective de « *knowledge in the head* » à une perspective de « *knowledge in the world* » en considérant que « *a process is not cognitive simply because it happens in a brain, nor is a process non-cognitive simply because it happens in the interactions among many brains* » (Halverson, 2002, p.248). Relativement aux courant dominants au début des années 1990¹²⁸, la CD ne traite pas d'une « nouvelle forme » de cognition mais d'un changement « d'agent cognitif ». Sans remettre en cause les principes d'une approche computationnelle et représentationnelle de la cognition, elle propose simplement de les appliquer au-delà de l'individu.

L'activité cognitive est dès lors considérée comme le produit d'un système *sociotechnique* entendu comme un *système fonctionnel* englobant un ensemble d'*individus*, d'*artefacts* et de leurs *interrelations* dans le contexte environnemental dans lequel ils sont situés (Decortis et al., 1997). La cognition n'est donc plus un phénomène localisé¹²⁹ (Rogers, 1997) de traitement d'information centré sur l'individu, mais distribué au sein d'un système plus large, dans lequel les outils deviennent des agents de l'activité cognitive globale : « *rather than trying to map the findings of cognitive psychological studies of the individual, we should map the conceptualisation of the cognitive system on a new unit of analysis: the sociotechnical system* » (Hutchins, 1995a, p.3).

Trajectoires d'information, changement d'état représentationnels et outils

C'est au sein de ce système *sociotechnique* (devenu système cognitif à part entière) que la Cognition Distribuée étudie la *trajectoire* des *informations* et des *connaissances* ainsi que leurs *transformations* au cours d'une activité : « *How is the information introduced into the system? How is it maintained? How is it transformed and where is it propagated?* » (Decortis et al., 2000, p.26). Lorsque l'information se propage dans le système sous forme de représentations, elle va *changer d'état* et de *medium*.

¹²⁸ Au début des années 1990, c'est encore le paradigme cognitiviste et représentationliste, centré sur l'individu qui domine

¹²⁹ Dans un espace et un temps.

Une information peut, par exemple, entrer dans le système étudié sous une forme écrite (un e-mail, un fax), pour être ensuite lue, puis passer par un état représentationnel différent dans la mémoire d'un participant avant de ressortir du système, sous forme orale lors d'une communication téléphonique. C'est donc en s'appuyant sur les concepts de *medium* et d'état représentationnel que la Cognition Distribuée parvient à caractériser la circulation et l'organisation de l'information au sein d'un système fonctionnel (Rogers et Ellis, 1994). On comprend à quel point le rôle des *artefacts* y est crucial.

De façon générale dans la CD, une grande attention est apportée aux artefacts du système en tant que *media* permettant la propagation de représentations, sans distinction spécifique. L'introduction de la notion d'*outil* permet cependant de raffiner l'analyse puisqu'ils sont en effet distingués selon leur degré de technologie (du papier à l'ordinateur) faisant ainsi varier sensiblement, par exemple, le fait qu'ils soient partageables et accroissent ainsi le nombre de trajectoires possibles de l'information dans le système sociotechnique.

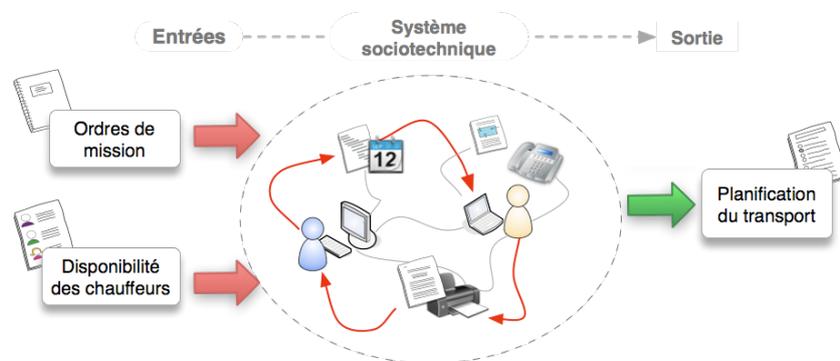


Figure 4.13 : Exemple d'un système sociotechnique avec ses entrées/sorties.

Ce changement de perspective modifie bien entendu profondément le statut des *représentations* qui sont au cœur des approches cognitives¹³⁰. De même, pour les classiques *entrées et sorties*, qui deviennent les entrées et sorties du système fonctionnel dans sa globalité. Si nous prenons l'exemple de l'étude de (Decortis *et al.*, 2000) concernant une entreprise de transport routier, le système sociotechnique comprendra la salle où se trouvent les contrôleurs logistiques, les artefacts leur permettant de travailler, et les contrôleurs eux-mêmes. Les *entrées* du système sont les ordres de mission des clients et les disponibilités des chauffeurs et des véhicules. La *sortie* sera la planification du transport à réaliser effectivement (Figure 4.13).

Articulation de l'activité collective dans la cognition distribuée

Sur le plan de l'activité coopérative la CD associe la coopération au partage de *connaissances spécifiques*, et affirme que la capacité d'organiser la coopération dépend de l'organisation sociale et physique du système sociotechnique. La coordination des actions des membres quant à elle n'est pas atteinte par un plan prédéfini. Selon la CD, celle-ci se construit dynamiquement et s'appuie sur une *compréhension intersubjective* qui s'appuie elle-même sur des *connaissances générales* des participants sur la situation de travail. En écho à cette dynamique et à la dépendance à l'environnement, la structure glo-

¹³⁰ Pour simplifier nous parlerons de représentation « au sens cognitiviste » pour désigner le type particulier de représentation interne sous-tendant toute l'approche cognitiviste, et les représentations dans un sens plus large ici, correspondant à une représentation informationnelle, i.e. l'état d'une information à un moment donné dans le système.

bale de réalisation de l'activité *émerge* des interactions locales entre les membres du collectif (Longchamp, 2003). Pour être plus précis, en ce qui concerne les activités collectives et coopératives, la CD s'intéresse aux *interactions* qui ont lieu entre agents humains du système¹³¹. Elle propose pour leur étude trois concepts clefs : *l'horizon d'observation*, la *connaissance partagée*, et la *compréhension intersubjective*.

Dans un système sociotechnique, *l'horizon d'observation* se définit comme « *the surroundings a person can perceive and monitor in addition to his/her own task* » (Decortis *et al.*, 2000, p.27). De manière générale, cet horizon d'observation désigne, dans un espace matériel du système fonctionnel, *l'agencement* physique des composants, une cloison empêchant par exemple de voir ce que fait son voisin dans le bureau d'à côté. Cette possibilité, cette « fenêtre » (Salembier, 2002), est importante car elle a un rôle fondamental dans la *correction des erreurs* (nécessaire pour réagir à l'erreur d'un autre agent du système), dans *l'apprentissage* (par observation, et sous le contrôle d'un pair dans cet horizon) ou encore la *gestion de l'aide* (détection du besoin d'aide). Tous les agents ont leur horizon d'observation propre. Les horizons en question ne sont pas nécessairement équivalents, n'autorisant pas aux acteurs la même facilité d'accès aux informations.

Il en va de même pour les connaissances. La répartition des connaissances dans le système n'étant pas homogène, on cherchera à éviter que le recouvrement des connaissances des différents participants ne soit trop faible (c'est un facteur de risque d'erreur). Au contraire, il faut soutenir la *construction* d'une *connaissance partagée spécifique* entre les composants du système, au point que dans certains cas extrêmes, un composant puisse en remplacer un autre : « *shared task knowledge is a distributed cognition feature which enables the monitoring of the ongoing activity by other knowledgeable participants who will be able to perform the task when one of the members can't* » (Decortis *et al.*, 1997).

Bien entendu, la création de connaissances partagées spécifiques doit pouvoir s'appuyer sur des connaissances générales partagées entre les participants, des connaissances relatives à la « manière dont les choses sont supposées aller et comment elles se déroulent habituellement ». Cette connaissance est utilisée pour construire et négocier la *compréhension de la situation en cours*, dite *compréhension intersubjective*. Cette *intersubjectivité* permet une communication qui va au-delà de la signification des énoncés verbaux. Ce concept pourrait être rapproché de celui de « *mutual knowledge* » et qui fait référence à ce qu'un participant sait du savoir d'un autre participant avec lequel il interagit.

Méthodologie d'analyse

L'analyse d'une activité dans le cadre de la Cognition Distribuée comporte, dans les grandes lignes, trois étapes : (a) *définition fonctionnelle* du système cognitif (sociotechnique), (b) *détermination* des représentations et des processus en jeu dans le système, (c) *énumération* des instanciations physiques de ces représentations et des algorithmes contrôlant le ou les processus. Leur mise en œuvre s'appuie sur un certain nombre de méthodes classiques pour le recueil de données : observations des pratiques dans le cadre du travail, simulations, entretiens, enregistrements vidéos, *etc.* Cette démarche semble toutefois comporter quelques difficultés pointées dans la littérature : « *It is not a methodology that one can readily pick off the shelf and apply to a design problem* » (Rogers, 1997, p5).

¹³¹ Avec un soin particulier réservé aux interactions qui constituent un point d'entrée ou de sortie du système fonctionnel.

La manière dont doivent être déterminées les *limites* du système fonctionnel, en particulier, n'est pas suffisamment claire¹³², et le *coût* élevé d'une analyse portant sur un système « à géométrie variable » est également pointé du doigt (Longchamp, 2003). Il en est de même pour la tendance des analyses à se concentrer sur des laps de temps très courts, privilégiant en général des *situations critiques*, des cas *d'erreur* ou de *défaillance*. En se focalisant sur les *trajectoires* et le *statut représentationnel* des informations circulant dans le système sociotechnique, l'analyse semble également ignorer les *structures statiques* comme les *organisations*, les *rôles* (Decortis et al., 2000), les *pouvoirs* (Longchamp, 2003), alors même que la cognition est sensée être *socialement* située. Malgré cette faiblesse méthodologique, et quelques critiques appuyées (Button, 1997, 2008), la Cognition Distribuée a souvent été utilisée pour l'analyse de diverses activités, pour la navigation maritime ou aérienne (nous l'avons signalé en introduction), mais également pour analyser des activités de logistique (Decortis et al., 2000), des pratiques d'ingénierie (Rogers, 1992, 1993), des activités de conception dans l'industrie de la construction (Perry, 1997), ou encore des activités de contrôle aérien (Halverson, 1992, 1994, 1995).

4.4.3.3 Les approches situées : Action, Cognition et Apprentissage Situés

Cognition située

L'idée, déjà ancienne¹³³, que la cognition et l'action ne peuvent être considérées que par rapport à une *situation*, toujours particulière, dans laquelle elle se déroule, a poussé certains chercheurs à mettre l'accent sur la prise en compte du *contexte*. Les tenants de l'approche située ont amplifié cette idée en incluant le contexte (classiquement vu comme un ensemble de valeurs prises par des paramètres physiques à un moment donné) dans une situation plus large, incluant notamment d'emblée les aspects *sociaux* et *historiques* de l'acteur (Salembier, 1996). L'aspect social est un des facteurs importants à prendre en compte dans le cadre d'une analyse du fonctionnement d'un acteur en tant qu'élément de fonctionnement d'un groupe ou d'une société : « *Toute action est improvisée à l'intérieur d'un champ de significations organisées socialement* » (Agre, 1987).

La dimension sociale est pour Clancey (1997), l'une des trois perspectives principales de la cognition située. Plus qu'une simple influence les partisans de la cognition située proposent « *les interactions sociales comme étant une source potentielle de structure pouvant contribuer à la cognition* » (Worham, 2001). Il s'agit par exemple du « *savoir social* » que l'on a d'un interlocuteur, et que l'on utilise inconsciemment pour mener une conversation en choisissant spécifiquement un vocabulaire adapté. Une autre dimension du contexte « *étendu* » proposé par l'approche située est *l'histoire* de l'acteur. L'histoire de l'acteur n'est pas limitée au sens premier du terme, et inclue « *non seulement le passé, mais aussi le présent et le futur de l'acteur* » (Salembier, 1996). On entend par « *futur* » de l'acteur, les éléments anticipés et/ou attendus par celui-ci. Si les fonctions et les rôles sociaux sont importants, il faut se rappeler qu'ils se sont développés à travers une *histoire* dont ils sont inséparables. La manière dont l'acteur peut tirer partie de son passé est donc bien sûr à prendre en compte mais tout aspect de la cognition influencé par la perception du temps et du déroulement de l'action l'est aussi.

¹³² « *Mentioning the system immediately raises a number of questions. How do we define its boundaries ?* » (Decortis et al., 2000).

¹³³ L'origine du terme « *situatedness* » est attribuée à Mead G.H. (1940).

Cette redéfinition est un élément clef de l'approche située, sur lequel de nombreux chercheurs de ce mouvement s'accordent. Au lieu de considérer un agent « dans » un environnement, comme « *a cherry in a bowl* » pour reprendre l'expression de Dewey, l'approche située préfère parler de l'agent et de l'environnement comme deux parties d'un *tout*. Cette approche change naturellement la nature des relations étudiées. Au lieu de regarder l'agent et son environnement comme deux entités distinctes et séparément descriptibles, dont il faudrait caractériser les relations, ce nouveau point de vue nous invite à considérer les interactions de deux parties d'un même système comme le problème central. Il s'agit alors pour étudier ce système de traiter l'agent et son environnement en terme de *participation* à une activité, de suivre leur « *trajectoire* », en tenant compte de leurs interactions.

La première conséquence est que la nature des relations agent / environnement, passant de l'état de relation interne /externe, deviennent des relations internes d'un même système¹³⁴. Les interactions avec l'environnement *ne sont pas médiées* par un encodage passant de l'extérieur vers l'intérieur ; l'environnement est plutôt perçu de façon « *directe* », selon le terme de Gibson (1977). Du même coup, seconde conséquence, il faut redéfinir le *cadre d'étude* d'un acteur dans une situation donnée : les échanges acteur/contexte ne sont pas tous observables au niveau de ce qui se passe « *dans la tête de l'acteur* », car tous les éléments contextuels interagissent en même temps avec l'acteur, et le seul cadre possible pouvant prendre en compte tous les éléments participants à une situation donnée, est celui de *l'activité*. Les éléments clefs ne sont plus alors des symboles, mais les interactions entre tous les éléments du contexte (y compris l'individu) pouvant jouer un rôle significatif dans l'action, c'est à dire utilisable et *utilisés* par l'acteur. Les acteurs utilisent des éléments de cet environnement pour supporter leur action, de telle manière qu'il ne semble pas nécessaire que les acteurs se représentent tous les éléments pertinents de la situation. Cette dernière remarque nous permet de faire la transition qui nous amène à présenter succinctement l'approche de l'Action Située.

Action Située

L'Action Située est une proposition de l'anthropologue L. Suchman, formée à l'école de l'ethnométhodologie¹³⁵, et qui travaillait alors à *Palo-alto*. Elle s'est fait connaître grâce à son ouvrage « *Plan and Situated Actions* » en 1987, puis à quelques échanges fameux, par articles interposés, avec Vera et Simon¹³⁶. La mise en avant du caractère situé de toute action s'accompagne chez Suchman de la remise en question de la « conception classique de la planification et du rôle fonctionnel des plans, héritée de la tradition cognitiviste » (Salembier, 1996). Selon la tradition cognitiviste le plan constitue non seulement une description mais aussi une prescription intégrale de l'action (Quéré, 1999). L'exécution de l'action correspond alors à la réalisation effective d'un programme pré-déterminé. Il ne s'agit pas pour L. Suchman de nier l'existence des plans mais de redéfinir leur rôle et leur statut, « le rôle principal [est attribué] non plus aux plans mais aux actions situées ; les plans étant une chose produites parmi d'autres par ces actions situées » (Suchman, 1993).

¹³⁴ L'idée de traiter le système comme un « tout » correspond aux théories de Maturana et Varela (1987), pour qui le système perceptif est « *informationnellement clôt* ».

¹³⁵ Le thème de recherche clef des ethnométhodologues est de comprendre comment les processus sociaux s'auto-organisent au moyen d'un certain nombre d'opérations d'acteurs dans le cours même de leur effectuation ; tout l'intérêt étant de mettre à jour les « *ethno-méthodes* » mises en œuvre dans ces processus d'auto-organisation

¹³⁶ Il s'agissait notamment d'un numéro spécial de *Cognitive Science* (Volume 17), 1993 (Suchman, 1993 ; Vera et Simon, 1993).

Dans le modèle du plan, le premier rôle est tenu par la planification et la délibération, avant même l'engagement de l'action, comme si paradoxalement la clef de l'action se trouvait avant même son déclenchement quand l'acteur n'agit pas encore. L'idée d'un processus dans l'effectuation d'une action, se développant durant celle-ci, orienté vers une finalité et ayant une temporalité est nouvelle (Quéré, 1999), et redonne à l'action la dimension dynamique qu'elle perd avec le modèle du plan : « *Rather than attempting to abstract action away from circumstances and represent it as a rational plan, the approach is to study how people use their circumstances to achieve intelligent actions* » (Suchman, 1987, p. 50). La volonté de replacer la cognition dans l'effectuation de l'action est un élément important. On rejoint dans cette démarche les idées de Maturana et Varela (1987) pour qui « la cognition est une action effective, et toute action est cognition ». Au-delà de la remise en cause du statut de la planification de l'action, c'est également le caractère *symbolique* des représentations internes qui est questionné. L'accent est mis sur les processus *d'interaction*, entre acteurs et environnement. On peut alors admettre l'existence de représentations symboliques mais « les représentations ne sont plus au centre de l'esprit, [...] elles émergent des interactions des processus mentaux avec l'environnement » (Clancey, 1997).

Apprentissage situé

À la fin des années 1980, sous l'impulsion de quelques chercheurs (Resnick, 1989 ; Brown, 1989) s'est créé le courant du « *situated learning* » qui reprend l'argument de la nécessaire *contextualisation* de l'enseignement¹³⁷. Nous avons déjà indiqué que pour les tenants de « l'apprentissage situé », la connaissance n'était pas une « chose » (section 4.3.2), et nous pouvons ici ajouter qu'ils la considèrent comme une abstraction analytique, comme l'énergie, et non comme une entité que l'on peut prendre en main. On ne peut pas par exemple, inventorier toutes les connaissances de quelqu'un.

La connaissance définie comme *interaction dynamique* ne peut être séparée de l'activité. Brown et al. parlent de la séparation entre « savoir » et « faire » en ces termes : « l'activité dans laquelle la connaissance est développée et déployée [...] n'est ni séparable, ni auxiliaire de l'apprentissage et de la cognition. Au contraire c'est une part intégrante de ce qui est appris » (Brown, 1989). À chaque expérience la connaissance dont on a besoin est reconstruite en fonction de la situation, et son utilisation ne peut consister en une recherche puis une application des éléments retrouvés en mémoire. Selon l'approche classique ces arguments posent problèmes quant au *transfert de connaissance*¹³⁸. Le courant de l'apprentissage situé répond qu'un « *savoir-faire* » peut très bien servir à plusieurs tâches contrairement aux *concepts*. Chaque situation est re-conceptualisée et est en ce sens unique.

Selon l'approche située, il ne faut pas à confondre *abstraction* et *généralisation*. En effet, *l'abstraction* conceptuelle de l'enseignement des mathématiques n'implique pas forcément la *généralisation* des techniques apprises à tous les problèmes fondamentalement similaires (Greeno & Moore, 1993). C'est le cas dans les systèmes experts : le contexte est limité aux descriptions que l'on peut en faire, et la généralisation se fait sur une forte conceptualisation. Les modèles courants de résolution de problèmes semblent appauvris. Le problème à résoudre et les connaissances nécessaires sont limités à des descriptions, ce qui dans la théorie fonctionne mais reste mal adapté à la pratique. L'expertise par exem-

¹³⁷ Idée que l'on peut faire remonter à Dewey.

¹³⁸ Anderson et al. (1996), prennent l'exemple d'un éditeur de texte. Si une personne est habituée à un éditeur de texte, elle sera plus rapide à savoir se servir d'un nouvel éditeur qu'une personne ne s'étant jamais servi d'un éditeur de texte. Il y aurait donc transfert de connaissance.

ple n'est « *pas simplement le fait de savoir beaucoup de règles mais aussi de savoir comment faire une bonne interprétation* » (Clancey, 1995). Ainsi la pratique se trouve en fait très éloignée de la théorie.

4.4.3.4 La Théorie de la Coordination

Nous avons jusqu'ici évoqué des théories et approches plutôt *psychologiques* et *cognitives*, mais le *CSCW* propose également des approches plus *systémiques*, issues de domaines différents, comme la *gestion* ou la *sociologie*. C'est le cas par exemple avec la *Théorie de la Coordination* que nous avons choisi de présenter.

La Théorie de la Coordination (TC) a été développée et proposée par T. Malone et ses collègues du « centre de la science de la coordination » du *MIT* au début des années 1990. Il s'agit d'une approche interdisciplinaire qui se propose d'intégrer plusieurs interprétations de la notion de *coordination* (Malone et Crowston, 1990) issues de diverses disciplines : la psychologie, l'économie, la théorie de l'organisation, l'informatique. Toutes ces disciplines ont en effet une approche particulière des phénomènes de coordination, et la TC souhaite en faire émerger des principes directeurs lui permettant de développer une « nouvelle technologie de la coordination » (Longchamp, 2003). Le postulat de base est donc que les formes d'organisation dépendent avant tout des *mécanismes de coordination* utilisés pour gérer les dépendances entre tâches et ressources, lesdits mécanismes étant des mécanismes de traitement de l'information. Un rôle central est donc réservé à la question de la diffusion de technologies de l'information : « *Coordination theory provides an approach to the study of processes. In this view, the form a process takes depends on the coordination mechanisms chosen to manage dependencies among tasks and resources involved in the process. These mechanisms are primarily information-processing and so the use of new media will particularly affect their cost, perhaps changing which are preferred* » (Malone et Crowston, 1994).

Coordination et dépendances

Selon Malone et Crowston (1994) le travail de groupe doit être analysé au niveau de son *organisation*, c'est-à-dire en termes d'*acteurs*, qui effectuent des *activités interdépendantes* qui elles-mêmes requièrent des ressources pour atteindre des *buts*. La *coordination* se définit alors comme la *gestion des dépendances* entre ces activités. Sur un plan terminologique, notons que la TC, la coopération (qui n'exclut pas les buts conflictuels), la compétition (qui suppose un gain pour l'un et une perte équivalente pour l'autre), et la collaboration (qui suppose le projet commun de deux parties égales) sont considérées comme *des formes particulières de coordination*. Dans une organisation donnée, les activités peuvent être en relation de différentes manières : relation de partage de ressources, relation producteur/consommateur, les relations temporelles et les relations tâches/sous-tâches, *etc.* Les dépendances créées par ses relations sont classées en trois catégories (Figure 4.14). De ce noyau de trois catégories minimales, il est possible de développer d'autres types de dépendances, par *agrégation* ou *spécialisation*. Par exemple, la dépendance de flux peut-être spécialisée en une dépendance *temporelle* (une activité ne peut commencer que lorsqu'un autre a terminé de produire la ressource nécessaire) ou *spatiale* (lorsque les activités sont localisées dans des lieux différents).

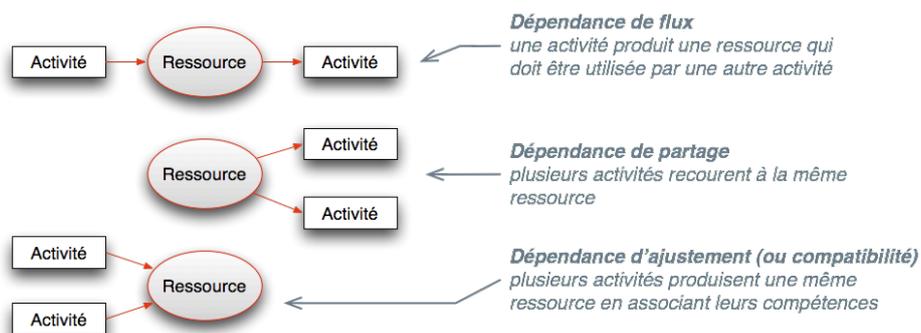


Figure 4.14 : Catégories de dépendances dans la Théorie de la Coordination.

Mécanismes et coûts de la coordination

Toute organisation en fonctionnement va mettre en œuvre des *mécanismes de coordination* pour gérer les problèmes de dépendances entre les activités. Plusieurs mécanismes peuvent constituer une réponse à un même type de dépendance. On aboutit ainsi à des *formes* différentes d'organisation, et il est toujours possible de constituer un mécanisme *alternatif* (Lonchamp, 2003). L'originalité de l'approche pluridisciplinaire de la TC est de chercher à s'inspirer de mécanismes de coordination pris dans des situations les plus diverses, pour en tirer profit lors de l'analyse d'une situation de coordination humaine. Prenons l'exemple des dépendances de partage : quels mécanismes de coordination mis en évidence dans des domaines ou disciplines différents répondent à ce type de problème ? L'économie propose le *mécanisme de marché*, celui des *enchères*, la gestion parle de *budget*, de *décisions managériales*, l'informatique propose de son côté *l'allocation des ressources*, les *systèmes de priorités*, etc.

Il ne s'agit pas pour la TC de tenter d'appliquer ou de reconnaître directement un de ces mécanismes dans le cadre d'une situation de coordination humaine. Il s'agit de tirer de leur comparaison, de leur mise en commun, une nouvelle façon d'aborder les mécanismes de coordination de façon générale. Pour ce qui est des mécanismes de coordination humains en particulier, l'approche souligne évidemment leur particularité et insiste sur les questions de *motivation*, *d'incitation* et *d'émotion*. De plus, dans le cadre d'une organisation humaine en situation, la question de la coordination ne saurait être réduite à celle de la seule *gestion systémique* des dépendances « les processus de prise de décision et de communication doivent être considérés également » (Lonchamp, 2003). La modification de l'un ou l'autre de ces processus peut générer de nouvelles formes organisationnelles tout comme de nouveaux mécanismes de coordination.

Selon la TC, la coordination quelle que soit sa forme ou son type, engendre des *coûts*, dont dépendent évidemment des *mécanismes* de coordinations considérés, qui eux-mêmes dépendent des *moyens de leur mise en œuvre*. C'est essentiellement là qu'interviennent les *artefacts* et en particulier les outils informatiques et les technologies de l'information et de la communication disponibles. Ces technologies sont particulièrement étudiées pour leur capacité à réduire les coûts de communication (économie de temps, gestion de la communication asynchrone, instantanéité des échanges, etc.). Cette réduction des coûts peut autoriser la mise en place de mécanismes qui auraient été trop coûteux sinon pour l'organisation. La TC se donne ainsi pour objectif applicatif d'analyser des organisations en situation de travail et de déterminer les coûts de coordination dans une configuration organisationnelle donnée, pour être à même ensuite d'en proposer une évolution qui les minimise : soit on modifie les technolo-

gies pour mieux répondre aux exigences d'un mécanisme en place, soit on change de mécanisme de coordination, avec l'aide éventuelle d'une refonte des technologies en place.

Méthodologie

Dans le contexte de l'analyse d'une organisation, la TC suggère de suivre une méthode constituée d'un certain nombre d'étapes. La première étape consiste à *délimiter le système* étudié, c'est-à-dire donner le périmètre de l'organisation considérée (en l'occurrence, on se restreint en général à une équipe, ou un service, rarement plus). La seconde étape est consacrée à la collecte et au dépouillement de données, qui reposent classiquement sur une série d'entretiens et sur le dépouillement d'un certain nombre de documents (documents de procédure, documents échangés entre les acteurs identifiés, *etc.*). Ensuite sont menées trois identifications successives. L'*identification des activités*, tout d'abord, à la sortie de laquelle on pourra produire une sorte de modèle de processus de l'activité, en identifiant les *buts*, les *acteurs*, les *étapes* et les *informations échangées*. Le modèle obtenu (qui peut se présenter simplement sous la forme d'un tableau) est ensuite complété par l'*identification des ressources* qui sont effectivement mobilisées par le processus, puis celle des *dépendances*.

L'analyse de ces dépendances peut se faire de deux façons. La première consiste à identifier les tâches qui semblent faire partie d'un mécanisme de coordination et à déterminer la dépendance concernée, qui est une méthode plutôt « *top-down* » (Malone et Crowston, 1994). La seconde méthode consiste à exploiter la liste des tâches et ressources et d'en extraire systématiquement les dépendances existantes, plutôt « *bottom-up* » (Malone et Crowston, 1994). Ainsi les dépendances de partage, par exemple, peuvent être facilement repérées en analysant les ressources utilisées par plus d'une tâche, de même pour les dépendances de tâche / sous-tâche. En dehors de ces recommandations générales, les auteurs eux-mêmes affirment qu'il n'y a pas *a priori* une unique bonne façon de mener cette analyse ou de réaliser ces différentes identifications. Pour ce qui est de la mobilisation de disciplines parfois très éloignées les unes des autres, Malone et Crowston proposent de considérer qu'il existe deux grands types d'analyse (Malone et Crowston, 1994, p.100) : les analyses paramétriques (*parametric analysis*), et les analyses comparatives (*baseline analysis*).

Pour les premières, il s'agit de mobiliser une théorie abstraite dont les paramètres varient selon les situations (et de déterminer lesquels varient : motivation, capacité cognitive, coût de communication, *etc.*). Pour les secondes, il s'agit de comparer plusieurs situations relativement à un modèle de coordination de *référence*. Il faut ensuite expliquer les différentes *déviations observées* à l'aide d'autres théories (comportementales, cognitives, sociales, *etc.*). La TC compte moins d'applications que les précédentes théories évoquées, nous pouvons toutefois citer le travail présenté dans (Crowston et Kammerer, 1998), dont l'objectif était de se servir d'une analyse fondée sur la Théorie de la Coordination pour évaluer et reconcevoir un outil support à l'activité d'une équipe technique chez IBM, en charge de traiter différents types d'erreurs et autres problèmes d'exécution de leurs programmes chez des clients. Sans rentrer dans les détails, nous signalerons tout de même que la solution proposée au problème de dépendance identifié consiste à créer un autre type de dépendance dont on ne sait pas s'il est plus gérable que le premier.

4.5 Conclusion

Dans ce chapitre nous avons présenté un résumé du travail théorique et bibliographique que nous avons effectué à la recherche d'éléments théoriques et méthodologiques pouvant servir de fondement à une méthodologie de modélisation des traces dédiées au Systèmes à Base de Traces modélisées. Notre démarche nous a conduit à mobiliser trois domaines de recherches dont nous avons exploité les ressources bibliographiques : IHM, IC et *CSCW*. Nous retenons de ce travail à la fois des éléments purement méthodologiques mais également des éléments théoriques qui nous aident à établir un cadre théorique pour l'analyse de l'activité que nécessite une modélisation de trace. Le bilan de ce travail est le suivant :

- Dans le domaine des IHM, nous avons étudié particulièrement les approches de *modélisation de tâches*. Nous avons expliqué que les modèles de tâches ne pouvaient être assimilables ni à un modèle de trace, ni à une trace modélisée elle-même. Toutefois, nous retenons de la démarche de modélisation de tâches la forme générale de son processus, avec une analyse fine de l'activité qui précède la création d'un modèle de tâche à proprement parlé.

- Dans le domaine de l'Ingénierie des Connaissances, nous nous sommes intéressé au processus de modélisation des connaissances qui se rapprochait plus d'une démarche de modélisation de trace. Nous retenons d'une démarche de modélisation des connaissances sont architecture générale avec d'une part un recueil d'expertise et les outils et techniques qui permettent de le mettre en œuvre, d'autre part la formalisation d'une représentation de connaissances. Les modèles de trace partagent avec les modèles de connaissances le fait d'être constitué de primitives qu'il s'agit de déterminer. Nous avons également appris de nos lectures que les Systèmes à Base de Traces modélisées, à l'instar des Systèmes à Base de Connaissances, ne doivent pas être uniquement considérés comme une représentation formelle mais aussi comme un artefact informatique plongé dans une situation d'usage donnée. Sur un plan théorique nous retenons également de notre lecture du domaine les notions de connaissance et d'inscription de connaissances que nous réutiliserons par la suite.

- Dans le domaine du *CSCW*, nous avons principalement cherché des éléments concernant l'analyse et la modélisation d'une activité instrumentée complexe et en situation, que celle-ci soit entendu au niveau individuel ou au niveau collectif. Nous retenons en particulier la Théorie de l'Activité comme une approche structurante de l'analyse de l'activité dont nous avons besoin. Elle permet en effet de dégager des dimensions dans lesquelles nous pouvons exploiter les apports d'autres théories, parmi lesquelles la Cognition Distribuée, l'Ergonomie Cognitive ou la Théorie de la coordination. Nous souhaitons combiner ces différents apports pour construire une grille d'analyse de l'activité instrumentée, individuelle et collective. Nous allons voir dans le chapitre suivant, comment ce travail théorique et bibliographique a été mis à profit pour proposer une première méthodologie complète de modélisation de trace numérique.

Chapitre 5

Vers une méthodologie de modélisation des traces numériques

Résumé du chapitre

Ce chapitre présente la méthodologie de modélisation de trace numérique que nous souhaitons intégrer à l'approche des Systèmes à Base de Traces modélisées. Sur un plan théorique nous proposons d'y considérer une activité documentaire instrumentée (ADI) comme une activité cognitive dont on peut caractériser plusieurs dimensions (son caractère documentaire, instrumental et situé). Nous affirmons également qu'une *trace modélisée* doit être conçue et gérée comme une véritable *inscription de connaissances* qui doit être *interprétable* comme une représentation de l'activité. Elle doit être conçue et considérée *dans* et *avec* son contexte d'interprétation, et exprimée à un *niveau de description* correspondant à celui des connaissances actualisées dans la réalisation de l'activité tracée. Sur un plan pratique, la méthodologie de modélisation se compose pour le reste de trois autres éléments : (a) un *processus générique* de modélisation de trace, (b) une *spécialisation* de ce processus en fonction d'un contexte de modélisation et d'usage, (c) une *dynamique d'intégration* de la modélisation à la mise en place concrète d'un SBTm dans une situation d'activité donnée. Le processus de modélisation qui est au cœur de la proposition est lui même composé de six étapes regroupées en deux phases : une phase d'analyse descriptive de l'activité, et une phase de construction du modèle de trace. Les différentes étapes proposées visent à encadrer et à guider au mieux le travail du modélisateur de trace qui doit garder une approche de modélisation guidée par l'analyse de l'activité plus que par les contraintes techniques liées au terrain. La *spécialisation* d'un processus générique de modélisation est liée à la nécessaire *conception des usages* du SBTm qui doit accompagner le travail de modélisation de trace. Quelle que soit le contexte d'usage le modélisateur de trace doit viser un premier modèle de trace capable de rendre le SBTm implémenté opérationnel – c'est la phase d'amorçage du SBTm en question. Par la suite il pourra ajuster le modèle de trace et faire évoluer le SBTm mis en place dans une phase d'évolution du système qui tiendra compte des usages réels des traces dans la situation d'activité considérée.

5.1 Introduction

Ce chapitre présente notre *proposition de méthodologie de modélisation des traces numériques* dans le cadre de l'approche des Systèmes à Base de Traces modélisées (SBTm), qui définit une trace modélisée comme l'association d'une collection d'observés et d'un modèle de cette collection. Nous aspirons à ce que cette trace modélisée puisse être utilisée comme une représentation qualitative de l'activité effective dans des situations d'activités documentaires instrumentées. Pour ce faire nous proposons de définir une méthodologie de modélisation de trace dédiée, *i.e.* de donner les méthodes nécessaires au choix des observés et relations constitutifs d'une trace pour une situation d'activité donnée. Depuis que le cadre conceptuel des SBTm a été défini, plusieurs travaux de modélisation de trace ont été concrètement réalisés sur divers terrains d'application (voir chapitre 3). Ces modélisations n'étaient toutefois guidées que par les contraintes liées à la situation et au savoir faire empirique du concepteur du SBTm implanté. Notre travail constitue donc une *première tentative* de rationalisation et de formalisation à ce niveau de généralité d'une *démarche de modélisation des traces propre aux SBTm*.

Compte tenu du caractère exploratoire de ce travail de recherche, il serait peut-être plus juste de qualifier ce qui va suivre « d'approche » méthodologique. Toutefois, par facilité de langage plus que par prétention, nous utiliserons le terme de « méthodologie ». Nous souhaitons souligner aussi par ce terme le fait que la démarche proposée prend en compte une situation d'implémentation d'un SBTm dans sa globalité. La méthodologie proposée se fonde sur l'association d'une recherche théorique et bibliographique, évoquée au chapitre précédent, et sur notre expérience en tant que modélisateur, dont nous reparlerons au chapitre suivant. Elle comporte quatre éléments, qui seront décrits dans les principales sections de ce chapitre. Tout d'abord un *cadre théorique* pour aborder les Activités Documentaires instrumentées et redéfinir les traces modélisées (section 5.2). Puis un *processus générique* de modélisation (section 5.3) qui structure le travail du modélisateur de trace en une progression par étapes, suivi d'une *spécialisation* du processus de modélisation (section 5.4) qui consiste à *spécifier* le processus générique en fonction du contexte d'usage des traces modélisées. Enfin, une *dynamique* d'implantation d'un SBTm intégrant le travail de modélisation de trace dans une situation d'activité donnée (section 5.5).

5.2 Cadre théorique

Ce que nous appelons ici « cadre théorique » est un ensemble d'idées, concepts et principes avec lesquels nous comptons comprendre et décrire les Activités Documentaires Instrumentées (ADI) que nous voulons tracer. Ce cadre qui est directement inspiré du travail bibliographique présenté au chapitre précédent, représente pour nous un élément important de notre démarche. En effet, il doit nous permettre de rendre compte des activités que nous abordons à deux titres : tout d'abord en nous donnant une *grille d'analyse de l'activité* que nous voulons tracer, ensuite en nous permettant de définir le contexte dans lequel les traces modélisées sont susceptibles d'être réutilisées à terme (*i.e.* dans l'activité elle-même).

5.2.1 L'activité caractérisée par ses dimensions

Pour commencer, nous empruntons à la Théorie de l'Activité les composants d'une activité (sujet, outil, objet) pour en faire des dimensions de l'activité (section 4.4.2). Chaque dimension donne une clef de lecture distincte de l'activité et constitue un espace dans lequel nous souhaitons exprimer le *caractère situé, instrumental et documentaire* des activités qui nous intéressent. De manière générale, nous considérons une ADI comme une activité cognitive, un travail intellectuel reposant sur la manipulation de documents dans un environnement numérique. Il est à noter que l'environnement numérique en question constitue à la fois un outil de manipulation des documents auxquels il donne accès et un espace de travail dans lequel prennent place les manipulations. De plus ce dernier devient un espace communicationnel partagé par plusieurs acteurs impliqués dans une activité collective.

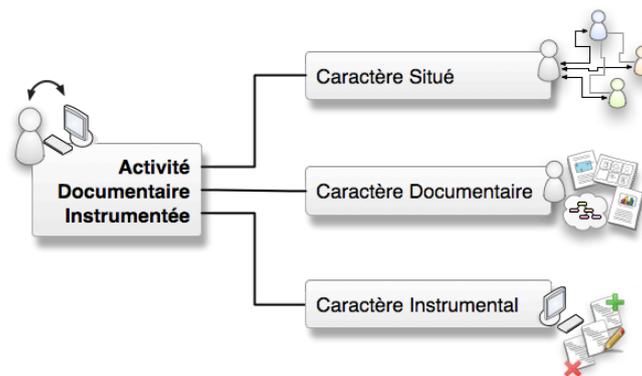


Figure 5.1 : Dimensions de description d'une Activité Documentaire Instrumentée.

5.2.1.1 Caractère situé

La première dimension que nous souhaitons aborder est celle du « sujet » de l'activité et de son caractère « situé ». Nous abordons les activités documentaires instrumentées comme des activités situées dans un espace socio-organisationnel et une histoire. Rarement une activité documentaire peut être, de manière générale considérée comme une activité purement et strictement individuelle. Il en est de même pour les activités qui nous intéressent ici. L'apprentissage à distance ou la production de contenus de formation prennent place dans des contextes socio-organisationnels riches qui influencent directement le déroulement de l'activité des utilisateurs. Toute compréhension et toute description d'une activité documentaire instrumentée passe donc nécessairement dans un premier temps, par la prise en compte du contexte social, organisationnel et historique de la réalisation de l'activité qui a lieu à un moment donné.

5.2.1.2 Caractère documentaire

La deuxième dimension que nous pouvons aborder est celle de « l'objet » de l'activité et de son caractère principalement documentaire. La grande variété des documents numériques, ou de ce qui peut être considéré comme tel, invite à s'arrêter quelques instants sur la notion de « document numérique ». L'arrivée du numérique a semble-t-il donné un second souffle aux discussions qui animent depuis longtemps la question d'une définition de la notion de document. Cette définition est communément empruntée à S. Briet pour qui un document est « tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène physique ou intellectuel »

(Briet, 1951, p.7). On retient de cette définition l'*intentionnalité* (« aux fins de ») du document qui doit servir de preuve ou qui est en tout cas présenté ou perçu comme tel¹³⁹, et son *indexicalité* car tout objet ayant la propriété d'être classé parmi d'autres dans un ensemble de relations organisées acquiert le statut de document. Les caractéristiques du numérique bouleversent certains des repères qui permettaient de stabiliser la notion de document : son accès, son indexation, sa lecture, sa conservation, sa validité, *etc.* sont autant de fonctions que le numérique permet de réaliser différemment. Le lecteur intéressé trouvera dans les travaux de R. Pédaouque une synthèse de ces réflexions¹⁴⁰.

Document numérique

On retiendra modestement des réflexions R. Pédaouque que l'arrivée du numérique change profondément le rapport au document. Il établit un nouveau « contrat de lecture » (Pédaouque, 2003, p.3) découplant de l'équation *Document numérique = structure + données*. Un document n'est pas formé d'un seul bloc mais de plusieurs segments distincts. Chaque segment peut être considéré comme une production sémiotique contenant plusieurs briques informationnelles (données et informations) rassemblées dans une certaine unité informationnelle. Leur mise en relation, autant que leur contenu informationnel, sont porteurs de sens : « *Humans combine [characteristic structures] into complex ones (letters into words, icon into plant maps, etc.) and associate the complex [characteristic structures] with a meaning, so defining complex [characteristic pattern]* » (Fogli et al., 2005). Le lecteur doit posséder les connaissances nécessaires et un référentiel commun avec l'auteur pour s'approprier la structuration du document. On ne peut donc pas réduire un document à un ensemble de segments signifiants car leur organisation, leur agencement (un segment comme un titre introduisant des sous-titres et des paragraphes) et leur présentation (mises en forme différentes entre les titres et les paragraphes), font du document un tout indivisible (Bringay, 2006). Et dans le cas d'un document numérique, le lecteur et son dispositif informatique de lecture vont participer concrètement, voire activement à la construction du sens du document lui-même.

Système d'organisation des connaissances

À une autre échelle, c'est également les nouvelles possibilités de structuration qui caractérisent les environnements numériques constitués de multiples documents. Les nouvelles possibilités de manipulation, notamment les hyperliens qui transforment les documents en « hyperdocuments » (Bachimont, 2000) modifient les espaces documentaires numériques. On y établit de véritables *Systèmes d'Organisation des Connaissances* (Zacklad, 2007) : classifications multiples, thésaurus, ontologies formelles, ontologies sémiotiques, folksonomies, *etc.* Les nouvelles possibilités d'organiser des contenus ouvrent de nouvelles façons de faire supporter ou soutenir aux artefacts une partie de l'activité cognitive, ainsi que de nouvelles possibilités d'y inscrire des connaissances. L'environnement est organisé par un acteur pour penser, et le numérique lui offre de nouvelles modalités d'organisation.

L'extériorisation de la cognition, notion présente sous diverses formes en Ingénierie des Connaissances (section 4.3), dans la Théorie de l'Activité, la Cognition Distribuée ou encore l'Ergonomie Cogni-

¹³⁹ Un exemple anecdotique, proposé par Briet est celui d'une antilope d'espèce encore inconnue. L'animal gambadant dans la brousse n'a rien d'un document, cependant ce même animal, capturé, importée par un zoo et placée dans un parc pour y être étudiée le devient alors. Elle devient alors la preuve physique utilisée par ceux qui l'étudient.

¹⁴⁰ Un collectif de chercheurs auto-baptisé Roger Pédaouque a travaillé durant plusieurs années sur la question du document numérique, et a produits plusieurs ouvrages de réflexion sur le sujet (Pédaouque, 2005 ; Pédaouque, 2007).

tive (section 4.4), est également mobilisée concrètement concernant les environnement documentaires, en particulier par le domaine du *Personal Information Management*. On y étudie en effet la façon dont un utilisateur mobilise son environnement numérique pour faciliter ou soutenir sa propre activité documentaire. Par exemple la manière dont il organise personnellement et classe ses documents, dont il crée des piles et des files de documents pour gérer son activité documentaire (Whittaker et Hirschberg, 2001 ; Stumpf et al., 2008). Dans le contexte d'activité documentaire numérique qui est le notre, nous retiendrons donc qu'une partie non négligeable de l'activité cognitive dont nous parlons se matérialise dans l'organisation concrète de l'ensemble de l'environnement documentaire de travail.

Documents pour l'action

Une autre particularité importante du document numérique est le fait qu'il incorpore la possibilité de partage et de manipulation par plusieurs acteurs : les possibilités de co-écriture, d'annotation (Lortal, 2006 ; Zacklad, 2005a). En considérant le document numérique non seulement comme une production sémiotique mais comme un projet transactionnel, M. Zacklad décrit parfaitement cette nouvelle dimension. Il définit ainsi le « Document Pour l'Action » (DoPA) qui remet en perspective chaque articulation des fragments d'un document relativement à l'activité dans laquelle il est plongé.

5.2.1.3 Caractère instrumental

Dernier point à aborder, le caractère instrumental des ADI. En fait nous avons en grande partie et implicitement évoqué les principales questions en décrivant les manipulations des documents dans la section précédente. Nous tenons donc simplement à souligner que la médiation de l'outil informatique dans l'interprétation de document est un élément majeur de ces activités. En l'occurrence la manipulation de certains outils exige des connaissances avancées qui seront mises à l'épreuve à chaque utilisation. Pour prendre l'exemple de la recherche d'information, un outil comme le moteur de recherche peut, selon le point de vue, être considéré à la fois comme document et comme application informatique. Il faut aussi rappeler que cette manipulation, même si elle est globalement maîtrisée doit être entendue et considérée dans un mouvement de « genèse instrumentale » (Rabardel, 1995) permanent, chaque nouvelle utilisation venant renforcer l'expérience de l'utilisateur et ses connaissances pratiques liées à la manipulation de l'outil numérique. Concernant l'analyse de l'activité qui devra être menée, il faut être attentif aux possibilités et aux contraintes qu'apporte un outil particulier à une situation d'activité.

5.2.2 L'activité effective caractérisée par sa structure singulière

La caractérisation générique des ADI n'est pas suffisante et nous devons du point de vue des traces modélisées nous interroger sur leur *réalisation concrète*. De manière générale, nous venons de caractériser une ADI comme une activité cognitive située, documentaire et instrumentale d'interprétation, d'exploitation, d'échange et de manipulation d'un ensemble de documents numériques. Nous considérons sa réalisation comme une expérience à chaque fois renouvelée *d'actualisation* de connaissances, idée que nous allons tenter d'expliquer par ce qui suit.

Tout d'abord et de manière générale, nous considérons que la réalisation d'une ADI mobilise nécessairement des connaissances « du domaine ». Il s'agit là des connaissances que l'acteur (re)construit à partir des documents qu'il manipule, des connaissances directement liées à l'interprétation de ces do-

cuments. Mais pour mobiliser concrètement ces connaissances du domaine l'acteur doit également mobiliser *d'autres connaissances* nécessaires à la réalisation de l'activité, des connaissances liées à la *manipulation* de l'environnement numérique (savoir faire dans un environnement numérique, processus métacognitifs, *etc.*) et des connaissances liées au *contexte social et historique* de l'activité (connaissances sociales, émotionnelles, *etc.*), des connaissances procédurales et contextuelles des connaissances « incarnées », pour les opposer aux connaissances du domaine que l'Ingénierie des Connaissances se donnent pour but de « désincarner » justement.

Ce qui fait la singularité de la réalisation d'une activité, ce qui en fait une « expérience vécue » par l'acteur c'est d'abord que la situation de réalisation n'est jamais strictement la même. Le contexte de l'activité évolue en permanence, c'est un fait constatable directement, mais l'acteur lui-même évolue. Pour être plus précis chaque réalisation concrète constitue un enrichissement de ses propres connaissances, susceptibles alors d'être mobilisées à une autre occasion. C'est ainsi que se constitue ce qu'on a l'habitude d'appeler « l'expérience » d'un acteur. Ceci nous conduit à envisager la réalisation d'une ADI non seulement comme la mobilisation de connaissances mais également comme l'enrichissement de ces mêmes connaissances. Ainsi de manière dynamique et non réversible chaque réalisation de l'acteur est une *actualisation* de ses connaissances. Les connaissances qu'il mobilise, qu'il s'agisse de connaissances du domaine ou de connaissances de la situation, s'enrichissent de leur mobilisation.

Nous considérons que dans la réalisation de son activité, un acteur agit de manière *cohérente et structurée*. Cela ne veut pas forcément dire que l'acteur agit de manière *rationnelle* du point de vue d'un observateur, mais que de son point de vue l'acteur suit une démarche cognitive structurée, construite à partir de la mobilisation de l'ensemble de ses connaissances dans une situation donnée. Cette démarche cognitive est inaccessible directement à l'observation. On peut même dire qu'elle échappe en partie à l'acteur lui-même étant donné que certaines de ses actions relèvent d'automatismes dont il n'est pas nécessairement conscient¹⁴¹. Cette démarche se manifeste toutefois sous la forme d'une organisation des actions de l'acteur, de ses interactions si on parle d'un utilisateur d'environnement numérique. Ainsi, du point de vue de l'utilisateur au moins, chaque réalisation concrète d'une activité est une expérience singulière et un ensemble structuré de ses actions pour atteindre son but.

5.2.3 Les traces modélisées comme inscription de connaissance

Dans le cadre des activités documentaires instrumentées qui nous intéresse, la réalisation d'une activité mobilise et actualise les connaissances d'un acteur, ses connaissances du domaine bien sûr mais aussi des connaissances procédurales et contextuelles. La mobilisation de ces connaissances permet à l'acteur d'agir de *manière structurée* de façon à atteindre un but qu'il se fixe lui-même. La manifestation de cette structuration de l'activité est *l'organisation des interactions* menée par l'acteur¹⁴². Chaque réalisation d'activité est singulière, chaque organisation des interactions l'est aussi. Une trace modélisée est une collection d'observés structurée par leur relation, et nous défendons l'idée que si une trace modélisée reflète cette organisation singulière, elle peut être interprétable comme une *inscription de connaissances* qui ont été mobilisées pour réaliser l'activité.

¹⁴¹ Voir à ce sujet la Théorie de l'Activité notamment (section 4.4.2).

¹⁴² Kaptelinin (2003) rappelle que cette organisation n'est qu'une *manifestation* de la structure de l'activité effective.

Or, pour *une activité donnée* l'organisation des interactions n'est interprétable qu'à un *certain niveau de description*, un niveau de description qui permet d'exprimer une trace au « *niveau des connaissances* » mobilisées et actualisées dans la réalisation d'une activité. En modélisant la trace, *i.e.* en déterminant explicitement des observés et leurs relations dans un modèle de trace, le modélisateur *introduit une sémantique* attachée aux éléments constitutifs de celle-ci, de façon à ce que leur mise en relation soit *a posteriori* interprétable *non* comme une *simple succession temporalisée d'interactions*, mais comme la *manifestation d'une activité structurée*.

Prenons l'exemple d'une activité de recherche d'information sur le *Web* dans le cadre d'une veille ponctuelle. En plus de ses connaissances du domaine, le veilleur va mobiliser d'autres connaissances, notamment relatives au fonctionnement des outils de recherche pour trouver des informations intéressantes. La mobilisation de ces connaissances va permettre au veilleur d'appliquer une *stratégie*, si basique soit-elle. Pour simplifier, disons qu'il n'y en a que deux : la première stratégie consisterait à utiliser un annuaire *Web* et de naviguer en profondeur à partir de chaque résultat fourni par l'annuaire, la seconde consisterait à partir d'un moteur de recherche et d'explorer les résultats avec une faible profondeur dans la navigation, une exploration « en largeur » des résultats du moteur si on peut dire.

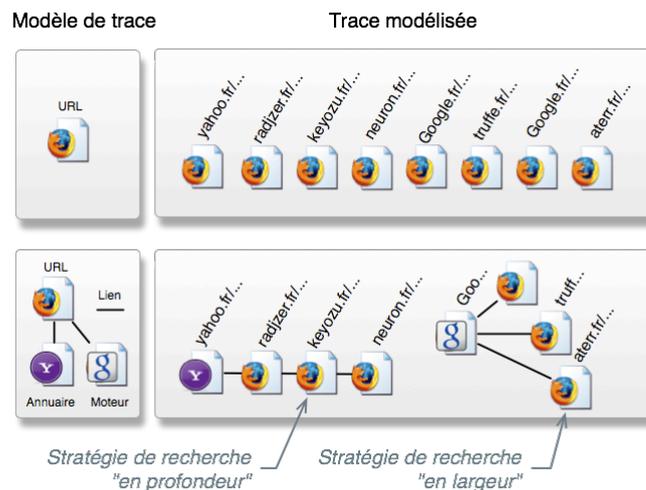


Figure 5.2 : L'organisation des interactions pour une activité donnée n'apparaît qu'à un niveau de description particulier que le modélisateur cherchera à atteindre avec une trace modélisée.

Concevoir une trace modélisée de cette activité de recherche comme une *inscription des connaissances* mobilisées par le veilleur pour la réaliser signifie que le *niveau de description* de la trace modélisée générée par son activité doit permettre d'y *interpréter directement* les stratégies qui ont été utilisées. Nous avons schématiser cette idée dans la figure ci-dessous (Figure 5.2) : l'organisation des interactions n'est interprétable qu'à un niveau de description particulier, que le modèle de trace doit permettre d'atteindre. Il faudra donc au modélisateur déterminer et chercher à exprimer une trace à ce niveau de description, tout comme une ontologie est conçue au « *niveau des connaissances du domaine* » (Charlet, 2004).

La mise en forme finale de la trace modélisée, c'est-à-dire sa visualisation au sens strict, aura bien entendu un rôle important dans la mise en évidence d'une organisation des différentes interactions, mais en amont le point crucial pour le modélisateur reste de déterminer quels sont les éléments de cette organisation et le niveau de description requis. Il est à noter également que le modèle de trace ne

détermine pas complètement une trace modélisée, et ne fixe pas son interprétation finale. La trace modélisée n'est modélisée « qu'indirectement », au sens où seuls ses éléments constitutifs (observés et relations) sont déterminés par le modèle. Chaque trace modélisée concrètement générée par un SGBT représentera une organisation d'observés singulière : la manière dont les observés seront instanciés, mis en relation et temporalisés dépend tout simplement de l'activité de manipulation elle-même. Le modèle de trace détermine un ensemble de « possibles » – l'espace de toutes les organisations possibles d'observés dans une trace, pas sa forme définitive.

Conçue comme une inscription de connaissances, une trace modélisée donne accès à ce qui d'ordinaire échappe à l'objectivation et à l'interprétation directe et repose uniquement sur les capacités mnésiques des acteurs (un tiers ou soi-même). Par le biais d'une trace modélisée, exprimant la structure de l'activité au niveau de description adéquat, l'activité effective devient *objet de connaissances* : on peut interpréter directement sa propre activité comme celle des autres, on peut comparer deux réalisations, revenir *a posteriori* sur les moyens et les choix qui ont permis d'atteindre un but et qui ne se sont pas nécessairement inscrits dans l'environnement. De plus, en tant qu'instanciation d'un modèle de trace qui en explicite les éléments, une trace modélisée se prête à *l'interprétation critique*. Non seulement l'explicitation des observés dans le modèle de trace éclaire l'interprétation d'une trace modélisée concrète, mais encore, la visualisation de la trace modélisée peut être *modulée* en exploitant le modèle en question par diverses *transformations* systématiques, offrant *plusieurs lectures possibles* d'une même trace modélisée. Dans le cas où cette modulation de la visualisation est pilotée directement par le lecteur, ce dernier pourra concrètement *construire* son interprétation à travers plusieurs lectures. Nous concevons donc les traces modélisées comme des inscriptions de connaissances à part entière, qui donnent accès à l'activité effective et se prêtent à l'interprétation critique.

C'est dans cet esprit que nous avons conçu notre méthodologie de modélisation. Celle-ci doit guider le modélisateur qui aura pour objectif de concevoir une trace modélisée comme l'inscriptions des connaissances mobilisées dans la réalisation singulière d'une activité donnée par un utilisateur en situation. L'enjeu du travail de modélisation sera donc de déterminer les observés de manière à ce qu'une trace modélisée – constituée de ces observés instanciés – soit exprimée au niveau de description correspondant aux connaissances mobilisées, et que l'on souhaite exploiter dans une réutilisation de la trace. Le travail de modélisation dépend donc directement de l'activité en question, et de manière plus large du contexte de modélisation.

5.3 Processus de modélisation générique

Le processus de modélisation générique présenté dans cette section a pour objectif *d'encadrer* et de *guider* le travail de création d'un *modèle de trace* lié à la mise en place d'un SBT dans une situation d'activité donnée. Avec comme objectif de principe l'obtention d'une trace modélisée entendue comme une inscription de connaissance, le travail de modélisation repose fondamentalement sur *l'analyse* de l'activité tracée. Les contraintes techniques liées à la collecte des observés, qui est souvent prise comme point de départ, seront intégrées au cours de processus lui-même. Le processus de modélisation se divise en deux grandes phases : une phase *d'analyse descriptive de l'activité* visant à faire res-

sortir les *observés* constitutifs d'une trace modélisée interprétable comme une image de l'activité (section 5.3.1) et une phase de *création du modèle de trace* proprement dit, *i.e.* un typage et une formalisation des observés et de leur relation dans un modèle (section 5.3.2).

5.3.1 Phase d'analyse descriptive de l'activité

La première grande phase est celle de *l'analyse descriptive* de l'activité tracée¹⁴³. Cette analyse menée par le modélisateur doit le conduire à *comprendre* comment l'activité menée par un utilisateur s'inscrit dans son environnement numérique, pour être à même d'en produire une *description* marquant la progression et l'articulation de l'activité engagée. Concrètement, cette phase comporte trois étapes : il faut tout d'abord *déterminer le périmètre de modélisation*, puis *décrire l'activité effective* selon ses différentes dimensions, et enfin en établir une *synthèse* de ces descriptions.

5.3.1.1 Détermination du périmètre de la modélisation

La première étape consiste à définir ce que nous avons appelé le *périmètre de modélisation*. Il s'agit de déterminer concrètement *l'activité tracée* et *le système observé* mais également de délimiter le domaine de *pertinence* du modèle de trace défini et des traces modélisées qui y seront utilisées.

Quelles que soient les modalités d'observation, rappelons qu'aucune activité humaine (*a fortiori* documentaire) n'est jamais complètement accessible à un observateur et que seule une partie de l'activité réalisée est observable. Dans notre cas, cette restriction est encore plus forte, puisque seule l'utilisation de l'environnement numérique (un ensemble d'interactions) est observable. L'activité que l'on souhaite tracer déborde en principe de l'utilisation des outils numériques qu'elle mobilise : tout d'abord parce que l'action commence et se prolonge *au-delà* de l'environnement numérique et ensuite parce que la part intellectuelle, cognitive de cette activité *ne s'inscrit pas nécessairement* dans l'environnement, et encore moins nécessairement dans l'environnement numérique *observable* depuis la machine. *L'activité tracée* sera donc finalement entendue comme l'activité qui mobilise une partie déterminée de l'environnement numérique observable, qui réciproquement sera défini comme un *système observé*, *i.e.* un ensemble d'outils et de ressources numériques nécessaires à la réalisation de l'activité considérée et instrumenté pour la collecte.

Cette première étape, qui peut sembler simple, peut se révéler épineuse et nécessiter une étude préliminaire sur le terrain. Les difficultés sont similaires à celles que rencontre un modélisateur face à la détermination du système *sociotechnique* dans le cadre de Cognition Distribuée : on doit déterminer de manière circulaire un système observé et une activité qui le mobilise. Il existe deux risques dans cet exercice : soit le modélisateur se focalise sur l'utilisation de l'environnement numérique et risque de ne pas prendre en compte des éléments majeurs de l'activité ; soit, il se focalise d'abord sur une activité donnée, et il risque de ne pas être capable de déterminer un système observé pertinent, soit que l'activité ne s'inscrive pas suffisamment dans l'environnement numérique, soit que le nombre d'outils à considérer est difficilement gérable.

¹⁴³ Si nous devions faire le parallèle avec la modélisation de connaissances cette phase correspondrait à celle « d'élucitation ».

Par exemple, pour une activité de Recherche d'Information sur le *Web*¹⁴⁴, plusieurs situations peuvent se présenter. Si le veilleur dispose d'une plateforme dédiée, qui prend en charge les principales opérations de la veille, le système observé pourra être focalisé sur ladite plateforme. Dans le cas contraire (Figure 5.3), on ne peut se limiter à un seul outil fusse-t-il le navigateur *Web*, l'activité allant bien au-delà de la stricte navigation sur Internet. Il faudrait ainsi déterminer un système observé composé de plusieurs outils : outils de communication (indispensables à l'activité de diffusion mais également de recherche elle-même), outils de gestion des flux d'informations (agrégateur de flux RSS par exemple), ou encore outils documentaires (éditeurs de texte ou HTML) nécessaires à la construction des rapports de veille, *etc.*

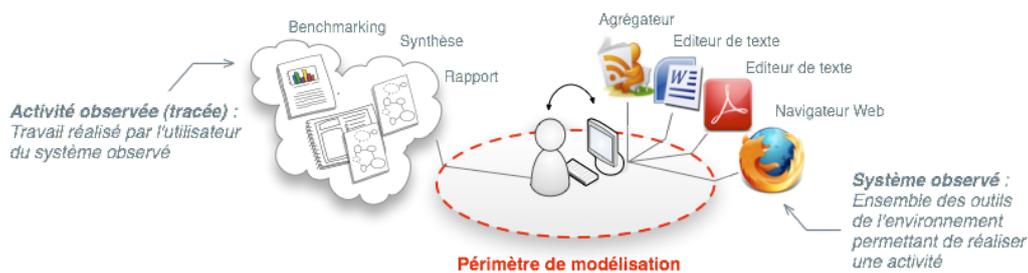


Figure 5.3 : Illustration d'un périmètre de modélisation.

Sur la base de notre propre expérience de terrain, nous pouvons formuler quelques recommandations. Il ne s'agit pas à proprement parler de prescriptions mais d'indications sur la conduite possible du travail pratique du modélisateur. Le modélisateur pourra commencer par déterminer *l'activité prescrite*, en s'appuyant par exemple sur la documentation disponible¹⁴⁵, puis observer la manière dont l'environnement est mobilisé pour mener *effectivement* cette activité : quels outils sont concrètement utilisés et comment ? Dans cette première observation, il faut être attentif aux *alternatives possibles* en termes de *supports* et *ressources* pour la réalisation : certaines opérations sont-elles réalisables de plusieurs manières, avec des outils différents, et si oui lesquels ? Quelles sont les modalités d'utilisation de l'une ou l'autre solution ? Sont-elles également observables ? Ces quelques questions sont d'un point de vue pratique un bon point de départ.

5.3.1.2 Description analytique de l'activité selon ses dimensions

La seconde étape consiste à *analyser* de manière approfondie l'activité effective. Le modélisateur s'appuiera pour cela sur la constitution de plusieurs *descriptions* de l'activité en question, en empruntant des points de vues différents. L'objectif de ces descriptions est de mettre en évidence un maximum d'éléments d'interactions tels qu'ils sont perçus par l'utilisateur dans le cours de son action, et parmi lesquels seront sélectionnés des observés. Cette analyse exploite les différentes dimensions de l'activité définies précédemment (section 5.2) pour en faire autant de points de vue sur l'activité, et forme une sorte de *grille* de lecture et d'analyse descriptive de l'activité. Pour résumer, cette analyse descriptive aborde l'activité selon trois dimensions principales, celle du *sujet*, de l'*objet* et enfin de l'*outil* (Figure 5.4).

¹⁴⁴ Nous prenons toujours l'exemple de la RI dans le cadre d'un travail de veille informationnelle.

¹⁴⁵ Descriptifs de projet, de missions, cahiers des charges, *etc.*

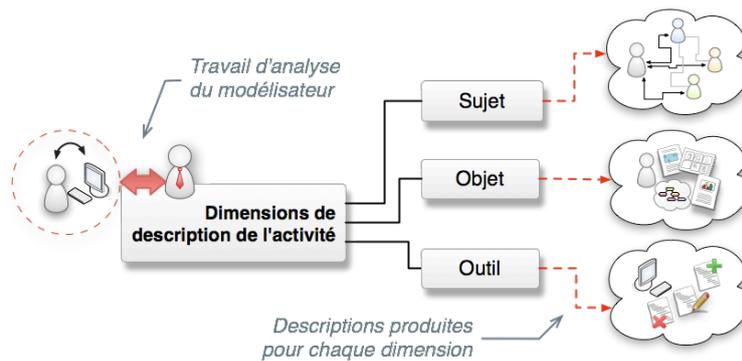


Figure 5.4 : Dimensions de l'activité utilisée comme grille d'analyse.

(1) Dimension du sujet

L'analyse portant sur le sujet de l'activité consiste principalement à *décrire le contexte socio-organisationnel du sujet* de l'activité, décrire comment ce contexte se matérialise dans l'activité interactionnelle du sujet et comment celle-ci s'inscrit dans l'activité collective si celle-ci existe. Le modélisateur cherchera à définir son rôle, ou sa position dans l'organisation, dans le suivi d'un *workflow*, sa participation à une communauté, à un groupe de travail, *etc.* Selon l'activité en question, le modélisateur pourra faire appel à diverses techniques d'analyse et de description, notamment *graphiques*, que propose le *CSCW* (section 4.4).

(2) Dimension de l'objet :

Concernant la dimension de l'objet, dont nous en avons souligné le caractère principalement documentaire, le modélisateur s'attachera à caractériser et à décrire les différents documents manipulés durant l'activité. Il faudra déterminer leur statut, leur présence dans l'activité, leur organisation, la manière dont ils sont perçus et exploités : certains documents ne sont utilisés qu'en tant que partie d'un corpus plus large, d'autres ne sont utilisés que très partiellement (sections, paragraphes)¹⁴⁶ d'autres encore sont produits tout au long de l'activité ou au contraire ont une durée de vie éphémère, certains sont partagés, d'autres non, certains documents sont des *ressources*, d'autres des *supports* de l'activité, d'autres enfin en sont le *produit*. Pour déterminer clairement les éléments manipulés du point de vue de l'utilisateur lui-même, le modélisateur pourra faire appel à des techniques d'explicitation et de représentation des connaissances empruntées au domaine de l'IC en s'inspirant directement de *commonKads* par exemple (Schreiber et al., 2000).

(3) Dimension de l'outil :

Concernant la *dimension de l'outil*, le modélisateur aura pour but de décrire *l'utilisation* des outils numériques en termes *d'opérations*. Dans le cadre des activités documentaires numériques qui nous intéressent ici, cette description portera principalement sur les possibilités de *manipulation* des documents, et de *communication* de l'utilisateur. L'objectif n'est pas ici de définir complètement des *modèles de tâche* (section 6.2.2), bien que le résultat puisse éventuellement s'en rapprocher. La description de l'utilisation qui en ressort s'appuiera en grande partie sur le modèle de conception de l'outil pour

¹⁴⁶ Il est important pour le modélisateur de déterminer quelles « unités documentaires » sont réellement perçues et utilisées par les acteurs, au-delà du découpage en fichiers présents dans l'environnement. On fait ici référence à notre cadre théorique (au début de ce chapitre).

des actions comme ouvrir, fermer, supprimer, *etc.* mais également sur des opérations qui n'en relèvent qu'indirectement comme modifier, créer une nouvelle version, déplacer, *etc.*

5.3.1.3 Recoupement des descriptions

Pour clore cette première phase du processus de modélisation de trace, la dernière étape aura pour objectif de faire émerger une liste « d'observés potentiels » ou « observés candidats ». Jusqu'ici le processus de modélisation restait dans une phase d'analyse descriptive dans laquelle l'engagement des choix du modélisateur n'a été sensible que dans la détermination des limites du périmètre de modélisation, le reste du travail étant principalement descriptif. Ici les choix du modélisateur conditionneront la pertinence du résultat final. Pour éviter un choix arbitraire précipité, nous avons organisé la sélection définitive des observés en deux temps, l'un étant la fin de la première phase, l'autre le début de la seconde. Dans un premier temps (qui nous concerne ici) l'objectif est de fournir un *maximum* d'observés potentiels, raison pour laquelle nous avons utilisé plusieurs dimensions d'analyse de l'activité (qui multiplient les chances de mettre à jour des éléments d'interaction présents à l'utilisateur durant son activité). Mais l'objectif est aussi de retenir des observés potentiels, c'est-à-dire des éléments dont qui sont suffisamment *pertinents*, raison pour laquelle nous proposons de croiser les descriptions, l'idée étant que si des éléments d'interactions sont présents de manière saillante dans plusieurs dimensions (ou si on peut les reconstituer de cette façon) alors ils ont un intérêt particulier et peuvent être considérés comme observés potentiels.

Concrètement, chacune des descriptions dans sa dimension met au jour des éléments de l'interaction présents à l'utilisateur durant la réalisation de son activité. Ces éléments, qui apparaissent comme caractéristiques de l'activité réalisée sont de natures très différentes car il peut s'agir par exemple d'un document particulier, d'une action de l'utilisateur comme une recherche, ou un changement de valeur dans un champ de base de donnée, ou encore un échange avec un autre acteur. Tous ces éléments peuvent prétendre au statut « d'observé potentiel ». Ils sont donc de natures différentes mais également de niveaux d'abstractions différents. Quoi qu'il en soit, le modélisateur doit déterminer un ensemble de ces observés potentiels. Dans l'absolu, le choix de ces observés potentiels dépend totalement du modélisateur et de sa perception de l'activité. Il pourrait décider d'une liste de manière arbitraire avec pour seul argument que les éléments d'interaction sont « manifestement importants » à la description de l'activité telle qu'elle est perçue par l'utilisateur. Pour compléter le recoupement des descriptions, nous proposons au modélisateur d'ajouter une dimension d'analyse supplémentaire.

Nous proposons une démarche un peu plus systématique, en nous appuyant sur une analyse de la *dynamique temporelle* de l'activité. L'idée est de donner une base au modélisateur pour « faire émerger » indirectement et non décider arbitrairement d'un ensemble d'observés potentiels. Pour cela nous proposons de mettre en évidence certains éléments d'interactions grâce à leur caractéristiques temporelles, *i.e.* parce qu'ils apparaissent soit de façon régulière dans la réalisation de l'activité, soit au contraire de manière exceptionnelle. Dans les deux cas, on peut faire de ce constat un argument pour justifier le choix d'un « observé potentiel ». L'idée d'utiliser la dynamique temporelle de l'activité effective a un double intérêt. D'abord elle va aider à faire émerger des observés potentiels comme nous venons de le dire, mais donne également une idée de l'échelle temporelle qu'il faudra attribuer à la trace modélisée finale. Le *temps* peut être ainsi vu comme une quatrième dimension de l'analyse descriptive (Figure 5.5).

Concrètement nous proposons donc de caractériser *l'amplitude*, la *densité* et la *fréquence* de la réalisation d'une activité ou d'une partie de celle-ci. L'amplitude est le laps de temps nécessaire à cette réalisation particulière, la densité est le nombre d'interactions nécessaires à cette réalisation rapporté au temps, et enfin la fréquence est le nombre d'apparitions de cette réalisation rapporté au temps de l'observation. Nous englobons ici dans « réalisation de l'activité » tout élément d'interaction qui constitue la réalisation de l'activité. Selon la nature des éléments choisis on sera amené à constater qu'un document particulier apparaît très souvent dans l'activité d'un utilisateur A, et qu'un utilisateur B effectue une recherche systématiquement chaque semaine, *etc.*

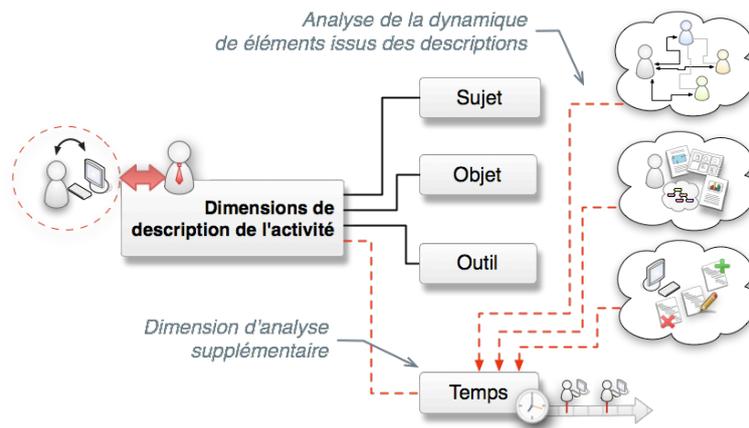


Figure 5.5 : Les dimensions formant la grille de l'analyse descriptive.

5.3.1.4 Recommandations

Le résultat attendu au terme cette phase du processus générique est un ensemble d'observés potentiels qui auront retenu l'attention du modélisateur notamment par le biais de leur caractéristiques temporelles. Nous ne prescrivons pas un format particulier pour les différentes descriptions de l'activité impliquée dans cette phase. Nous conseillons néanmoins, concernant la description de la dimension « sujet » d'utiliser des schémas permettant de mettre en évidence les relations entretenues entre les acteurs au cours de l'activité effective ; concernant la dimension « objet » une description sous forme de liste/tableau ou un schéma structuré ; enfin concernant la dimension « outil » une description sous forme de liste(s). Pour réunir l'ensemble des informations nécessaires et construire différentes descriptions de l'activité, le modélisateur devra faire appel, selon la situation, à toute technique lui permettant de comprendre les tenants et aboutissants de l'activité effective : l'observation *in situ*, l'observation participative, les entretiens (guidés ou non, simples ou en groupe), l'utilisation de questionnaires, l'exploitation de simulations ou de situations provoquées, l'utilisation du « *think-aloud* » voire des techniques d'explicitation ou de narration.

5.3.2 Phase de création du modèle de trace

La seconde phase du processus¹⁴⁷ est la création du modèle proprement dit, et consiste à faire émerger des descriptions de l'activité obtenues dans la phase précédente, des *observés potentiels* qui seront *sélectionnés* puis *typés* afin de constituer un *modèle de trace* qui sera *complété* et *consolidé*.

¹⁴⁷ Dans le parallèle avec le processus de modélisation des connaissances, il s'agirait ici de la phase de « représentation ».

5.3.2.1 Sélection des observés

La première étape consiste à *sélectionner des observés* parmi les observés potentiels que la première phase de ce processus aura mis en évidence. À nouveau, cette sélection pourrait être faite de manière arbitraire par le modélisateur, mais nous proposons de la réaliser par le biais d'une *schématisation* de l'activité, par exemple sous forme de graphe. Le modélisateur procèdera de la façon suivante :

- Choisir plusieurs situations d'activité effective – des « extraits représentatifs » de l'activité. Les situations en question seront enregistrées pour être schématisées en détail par la suite.
- Schématiser la réalisation effective de ces « extraits » sous forme de *graphes* à partir des objets mis en évidence dans la synthèse de la phase précédente (Figure 5.6). Le graphe dont il est question ici ne doit pas comporter de contraintes particulières, il s'agit simplement d'un moyen de représentation. Les objets qui y sont représentés, qui s'articulent pour décrire l'activité, peuvent être de toute nature (objets de l'interface, opérations, actions, évènements ponctuels ou duratifs), tout comme les relations qui sont posées entre eux.
- Comparer les différents graphes descriptifs et y sélectionner des observés. La comparaison entre les graphes amène à repérer des éléments qui sont *communs* ou au contraire ceux qui apparaissent comme *caractéristiques* d'une opération. Dans les deux cas le modélisateur devra juger ceux qui doivent constituer des *observés potentiels*, *i.e.* des éléments du niveau de description de l'activité engagée. L'objectif est de déterminer quel « vocabulaire » serait nécessaire et suffisant à la description de *tous* les extraits choisis.

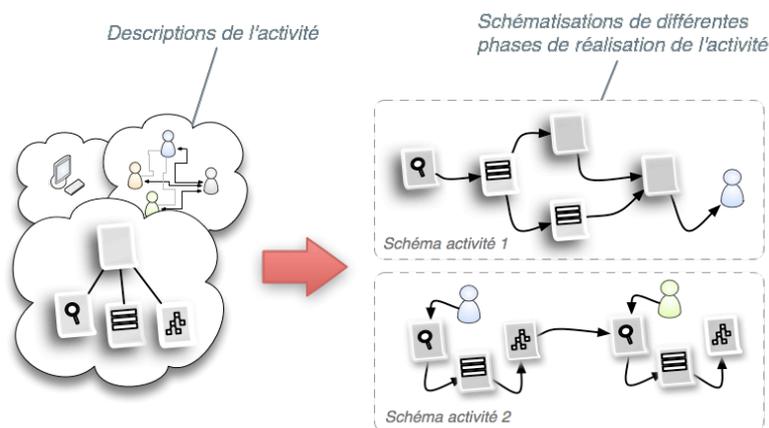


Figure 5.6 : Sélection des observés par schématisation de l'activité.

Finalement, cette schématisation revient à créer, à la main, une trace modélisée « sans son modèle », une trace telle qu'il serait pertinent qu'elle soit générée par le SBT mis en place (et selon le contexte de modélisation bien sûr). Le choix des éléments qui seront explicitement représentés dans ces schémas, consiste en fait à *sélectionner des observés potentiels* parmi tous les éléments composants les descriptions de l'activité données par l'analyse.

Il est difficile de donner à ce stade un exemple simplifié, l'intérêt de cette étape étant de donner *en détail* une description schématisée de l'activité. Cette étape sera toutefois illustrée dans le chapitre suivant, qui présentera notre travail de terrain. Concernant les recommandations que nous pouvons formuler sur cette étape, soulignons qu'il faut éviter de choisir de schématiser uniquement des situa-

tions particulièrement problématiques ou difficiles pour l'utilisateur. En revanche la comparaison entre une réalisation « normale » et une réalisation « extrême » peut être intéressante. De plus, pour entamer ce type de schématisation, il est parfois plus simple de partir d'un *événement* local, une suite d'opérations courtes, caractéristique de l'activité à schématiser, que l'on étend ensuite à l'ensemble de l'extrait choisi.

Soulignons le fait qu'il existe plusieurs manières d'exprimer les observés modélisés, leurs attributs et leurs relations. Le modélisateur peut choisir par exemple de ne modéliser que des actions (ou événements), de sorte que la trace modélisée sera constituée d'une suite d'actions (les objets sur lesquels elles portent étant des attributs de l'action), ou au contraire des observés qui sont uniquement des objets (entités), les actions portant sur eux étant alors attributs de l'objet. De même pour les relations, qui peuvent ne pas être explicites dans le formalisme utilisé : une relation peut-être un observé en elle-même, avec pour attributs les observés entre lesquels elle fait le lien, ou bien être implicite et n'être qu'un attribut commun aux deux observés mis en relation (avec une référence croisée). Il est inutile de détailler toutes les combinaisons possibles, mais attirons l'attention sur le fait que des expressions équivalentes avec une syntaxe différente peuvent avoir un impact sur l'interprétation de la trace modélisée visualisée au final.

5.3.2.2 Typage et formalisation des observés et de leurs relations

Une fois les observés potentiels et leurs relations sélectionnés dans la schématisation précédente, l'étape suivante consiste à en établir des *types* qui constitueront le modèle de trace. Le principe se rapproche de ce qui se fait en modélisation de connaissances dans une approche *ascendante*, *i.e.* lorsqu'il s'agit de créer des *types* d'objets formalisés à partir de leurs *instances* (Bachimont, 2004). Quelle que soit la nature des éléments considérés (événement, action, entité, objet, ou relation), le modélisateur devra donc tenter de les *regrouper*, les *comparer*, afin de les *généraliser*, de les *abstraire* en *types d'observés*. Pour chaque observé typé, le modélisateur se basera sur les caractéristiques des instances pour en déduire les *attributs* utiles et nécessaires¹⁴⁸. Ce typage s'accompagne d'une *hiérarchisation* des types d'observés créés (entités / événements), dans la mesure où cette hiérarchisation est possible et cohérente. Selon la nature des observés retenus il se peut qu'une hiérarchisation complète (une hiérarchisation unique pour l'ensemble des types) ne soit en effet pas possible.

5.3.2.3 Consolidation du modèle

La dernière étape, du processus est ce que nous avons appelé la « consolidation du modèle » et qui comprend plusieurs opérations qui achèvent la mise en place du *modèle de trace*. Tout d'abord, la mise en place d'une ébauche de modèle peut faire apparaître des observés qui n'avaient pas été envisagés, ou ne s'étaient pas imposés dans la schématisation de l'activité. Cette étape marque également l'entrée en ligne de compte de manière plus concrète des *contraintes* liées à la collecte : il faut s'assurer que l'instanciation des types d'observés constituant le modèle est possible, même si celle-ci implique la mise en place de *règles d'instanciation* complexes qu'il faut alors définir. L'ensemble des *règles d'instanciation* des observés doivent être déterminées par ailleurs avec l'informaticien responsable de l'implémentation du SBT. Ces règles peuvent être plus ou moins complexe en fonction des

¹⁴⁸ Selon la manière d'exprimer les relations entre deux types d'observés, il est possible que celles-ci fassent partie des attributs à définir.

observés à obtenir. Même un observé simple en apparence peut exiger la mise en place de règles de collecte plus complexes. Par exemple définir que le type « document consulté » ne sera instancié sur un fichier que si celui-ci est visualisé plus de 30 secondes en premier plan ; règles d’instanciation des relations : par exemple deux observés de type « page web visitée » ne sont mis en relation par un lien « lien http » que si l’utilisateur a effectivement cliqué sur un lien de la page (et non le bouton retour).

On peut marquer la fin du processus générique de modélisation au point où le premier modèle de trace complet est produit. Avant d’être implémenté et intégré au SBT, il doit encore être *validé* en passant par une « simulation de trace modélisée » réalisée à la main. Pour s’assurer de la cohérence et de la faisabilité du produit de sa modélisation, le modélisateur aura donc une dernière opération à effectuer. L’opération en question consiste à enregistrer¹⁴⁹ une activité sur un court laps de temps, puis à « générer à la main » la trace modélisée en instanciant les observés du modèle ainsi que leur relation en suivant les règles qui y sont définies. Le résultat pourra être soumis et discuté avec les utilisateurs potentiels de la trace modélisée, ce qui permet également d’identifier leurs besoins en terme de présentations de la trace, besoins qui dépendent fortement du contexte d’usage de la trace que nous allons évoquer dans les sections à suivre avec la *spécialisation* du processus générique de modélisation. En attendant, nous allons clore la présentation du processus générique par deux dernières remarques.

5.3.3 Commentaires

Résumé

Le processus de modélisation générique que nous venons de détailler constitue le cœur de notre proposition de méthodologie de modélisation. On peut le résumer de la manière suivante (Figure 5.7). Tout d’abord, il est composé de deux grandes phases, dont la première (phase d’analyse descriptive) vise à analyser l’activité de l’utilisateur pour en tirer un ensemble d’observés potentiels et la seconde (phase de création du modèle) à sélectionner et créer un modèle des observés constitutifs d’une trace modélisée. Les trois étapes de la première phase orientent le modélisateur pour produire l’ensemble le plus riche possible d’observés « potentiels » après avoir délimité son périmètre de modélisation. Les trois étapes de la seconde phase, de manière plus classique, conduisent l’utilisateur à formaliser un modèle à partir des observés qu’il aura finalement retenus comme pertinents pour constituer une trace modélisée relativement à l’activité qu’il veut tracer.

Application du processus

La première remarque concerne la présentation linéaire que nous venons de faire des étapes du processus. Nous avons effectivement conçu de manière très détaillée des étapes de modélisation avec pour but d’encadrer le plus étroitement possible le travail du modélisateur, et d’éviter à celui-ci de faire face à des choix de modélisation rendus trop complexes par la gestion simultanée d’un excès de contraintes. Cela ne signifie pas pour autant que dans l’application du processus de modélisation ces étapes soient strictement indépendantes : des retours en arrières, des boucles de modifications ne manqueront pas d’intervenir et seront même nécessaires. L’ordre des étapes tel qu’il est proposé nous permet cependant de marquer une progression de la modélisation qui pourrait être décrite comme une combinaison des approches de modélisation spécifiques aux domaines dont nous nous sommes inspirés.

¹⁴⁹ Enregistrement (vidéo ou de capture d’écran) permettant de visualiser toutes les opérations sensées être observables par le SBT.

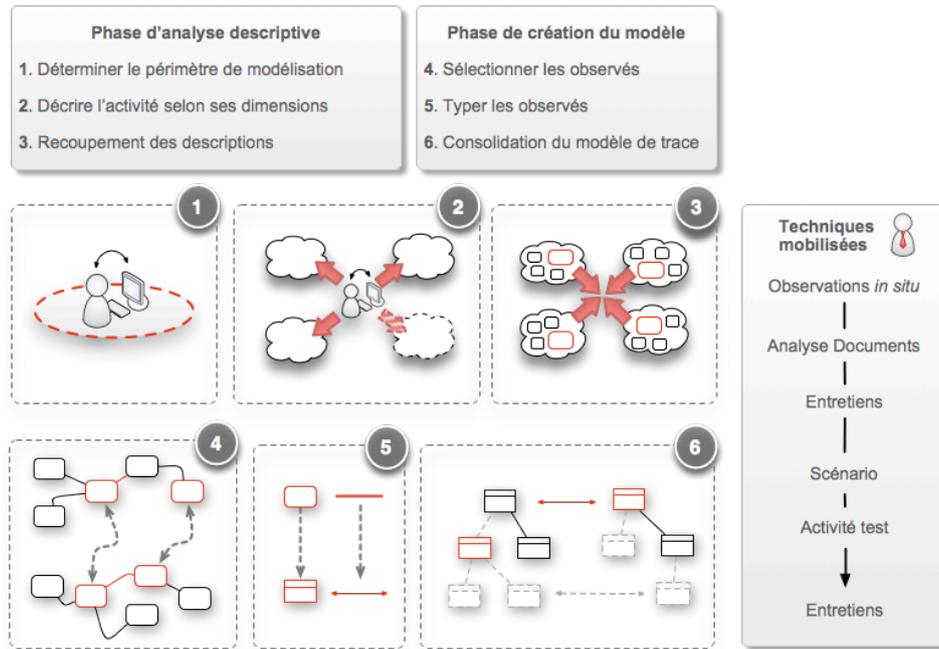


Figure 5.7 : Résumé du processus générique de modélisation de trace avec ses étapes et les principaux outils méthodologiques d'analyse mobilisés.

Combinaison de plusieurs approches de modélisation

Le travail de modélisation que nous proposons à travers notre processus, notamment en ce qui concerne l'analyse descriptive de l'activité s'apparente en effet à la combinaison des différentes approches de modélisation. En caricaturant la démarche d'analyse en question on retrouve en effet, avec l'analyse du sujet une modélisation de type *CSCW*, celle de l'objet documentaire une modélisation de type *IC* et avec celle de l'outil une modélisation de type *IHM*, à laquelle il faut ajouter l'analyse de la temporalité de l'activité, indispensable à la modélisation d'une trace, mais non spécifique à l'une de ces approches. L'objectif du modélisateur de trace *n'est pas de réaliser complètement ces modélisations* (au sens où les modèles produits seraient opérationnalisables), mais le fait d'engager la démarche de modélisation et l'analyse liée à la représentation de ces différents types de modèle conduit à un ensemble de descriptions de l'activité qui mettra concrètement en avant tous les types d'éléments descriptifs d'une activité donnée, ce qui est l'objectif de cette association.

Un double mouvement de modélisation

La dernière remarque concerne l'engagement du modélisateur dans son travail de modélisation. En prenant un peu de recul par rapport au processus qui structure le travail de modélisation de trace, on peut considérer que ce dernier se compose de *deux mouvements de modélisations opposés* (Laflaquière et al., 2008). Dans un premier mouvement, le modélisateur caractérise de manière « descendante » un objectif d'expression de la trace modélisée et dans un second mouvement il détermine les éléments constitutants (les observés) en intégrant, de manière « ascendante » cette fois, les contraintes liées au Système à Base de Trace implanté. Le modélisateur cherchera le *point d'équilibre* entre ces deux mouvements en les faisant *converger* vers un modèle de trace qui permette une trace modélisée qui atteigne l'objectif *d'interprétabilité* tout en respectant les *contraintes techniques* liées au système concret lui-même.

Comme nous l'avons expliqué au début de ce chapitre, notre méthodologie part du principe que l'on cherche à créer une trace modélisée à un *niveau de description* qui soit en cohérence avec l'activité observée, un niveau qui correspond à son *interprétabilité* en tant que représentation de l'activité effective. Le travail de modélisation, au-delà de l'application stricte du processus générique que nous venons de proposer, va dépendre du *contexte de modélisation* d'une part et surtout du *contexte d'usage prévu* du SBT implanté d'autre part. Le travail de modélisation est ainsi indissociable de la conception du SBT plongé dans une situation d'usage, raison pour laquelle une spécialisation du processus de modélisation est nécessaire.

5.4 Spécialisation du processus de modélisation

De manière classique nous envisageons deux grands types de situation d'usage des traces modélisées : un support à l'*analyse* de l'activité ou un *support* de l'activité elle-même. Deux types de situations d'usage(s) qui engendrent des *contextes de modélisation* très différents, et des *engagements de modélisation* différents pour reprendre un terme de l'Ingénierie des Connaissances.

5.4.1 Traces modélisées pour l'analyse de l'activité

5.4.1.1 Contexte de modélisation

Dans ce premier type de situation, le modélisateur peut avoir à mener concrètement un travail de modélisation *de terrain* relativement différent selon le contexte auquel il fait face. Plusieurs cas de figures peuvent se présenter. Dans le cas le plus simple, le modélisateur est à la fois le *concepteur du modèle* de trace, le *codeur* qui implémente concrètement le SBT et aussi *l'analyste* qui exploite les traces modélisées (Figure 5.8-1). De notre expérience, ce n'est pas le cas le plus courant, et le modélisateur aura parfois à travailler avec des *analystes extérieurs* au processus de modélisation, des experts de l'activité tracée qui seront finalement les *utilisateurs* du SBT. Plus souvent encore, le modélisateur aura à travailler de concert avec les *codeurs* qui réaliseront l'instrumentation de l'environnement et le déploiement du SBT (Figure 5.8-2).

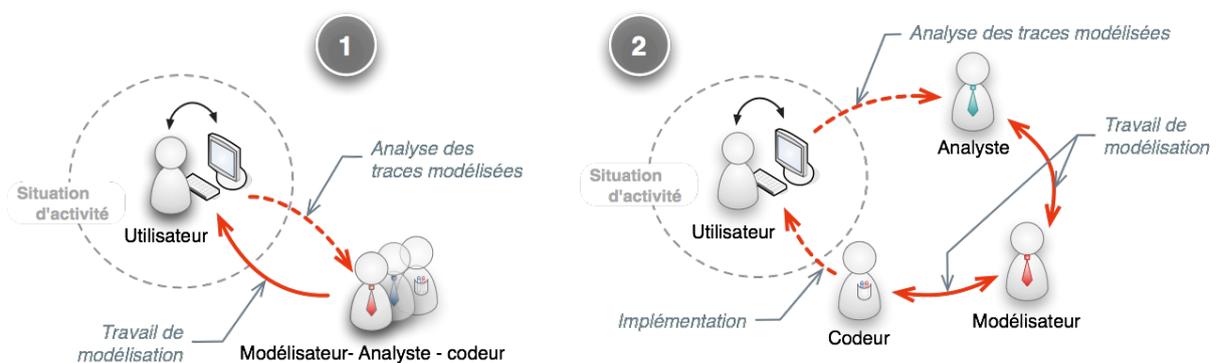


Figure 5.8 : Contextes de modélisation possibles dans l'optique d'une traces modélisées exploitée comme support d'analyse : (1) le modélisateur travaille seul ou (2) collabore avec d'autres acteurs.

Quel que soit le cas de figure, l'objectif du modélisateur de trace reste de créer une trace modélisée capable de *représenter*, de « tenir lieu de » l'activité *pour l'analyste* qui l'exploitera. Dans ce cas le *niveau de description* pertinent que doit atteindre la trace modélisée est à *déterminer directement en fonction des objectifs d'analyse*. De manière générale en effet, un analyste cherche à étudier un phénomène ou un type de phénomène particulier (par exemple les stratégies de recherche d'information dans l'exemple que nous avons pris précédemment). Le modélisateur cherchera donc à s'accorder à ces objectifs d'analyse pour fixer le niveau de description et guider sa sélection d'observés devant être décrits dans le modèle de trace.

5.4.1.2 Conception d'usage(s)

Dans le cas d'une exploitation des traces modélisées pour l'analyse de l'activité, notre objectif est de créer une trace modélisée qui représente de manière qualitative l'activité réalisée. Par conséquent la mise en place d'un Système à Base de Traces modélisées n'a pas tant pour objectif de constituer un « corpus de traces » à exploiter, que d'être un moyen pour l'analyste de « dialoguer à distance » (dans le temps et dans l'espace) avec l'activité qu'il observe. L'architecture informatique et la formalisation des traces modélisées d'un SBTm offre un outil d'analyse qui permet à l'analyste de *construire* - littéralement - son *interprétation*. Grâce à une gestion des traces modélisées par le biais de leur modèle, de nombreuses fonctionnalités de présentation des traces modélisées sont imaginables et permettent d'offrir plusieurs lectures d'une même trace modélisée.

Il est possible qu'avec le premier modèle de trace implémenté, le travail de l'analyste se révèle fructueux, et même qu'il puisse constater dans la trace des phénomènes qu'il n'avait pas *a priori* cherchée à observer. Toutefois, il est peu probable qu'un seul modèle réponde complètement aux besoins de l'analyste pour *construire son interprétation*. Car c'est bien de cela qu'il s'agit, la construction d'une interprétation analytique, combinant plusieurs plans, se basant sur les différents niveaux de description disponibles. Le processus de modélisation doit donc être (en partie) réitéré, de manière à produire plusieurs modèles de traces ou au minimum prévoir des *transformations* possibles des traces modélisées obtenues, pour offrir plusieurs niveaux de lecture des traces en question.

Du point de vue de l'analyse, les caractéristiques d'un SBT permettent d'articuler analyse *quantitative* et *qualitative*. Non seulement le modélisateur pourra prévoir, lors du choix des observés, plusieurs niveaux de lecture de la trace modélisée, mais il est possible de laisser à l'analyste lui-même, lors de l'exploitation de la trace, la possibilité de créer de nouveaux observés à partir des observés présents dans la trace. Dit autrement, l'analyste peut créer ses propres transformations de trace. Dans ce cas, le travail du modélisateur sera clairement la conception d'un outil interactif : on dépasse les simples fonctionnalités d'une *navigation active* de la trace pour passer à une *exploitation interactive*. C'est ce qui constitue dans le fond le réel *projet de modélisation* de trace, et de *conception* d'un SBT destiné à soutenir une analyse de l'activité, et c'est également ces éléments qui orienteront les choix du modélisateur dans son travail de modélisation.

Une illustration de ce type de projet nous est donnée dans la thèse de O. Georgeon (2006). Le travail d'analyse de l'activité repose sur une ensemble de traces modélisées qui seront transformées (ré-écrites) afin de changer de niveau d'abstraction. L'activité dont il est question ici est celle de la conduite automobile. La collecte se fait via une multitude de capteurs qui donnent des informations sur

l'état de la voiture (vitesse, position du volant, *etc.*) et les actions de l'utilisateur (regard, position, *etc.*). Un échange entre analyste et modélisateur permet la mise en place de transformations nécessaires pour *construire une analyse*.

5.4.2 Traces modélisées pour la réflexivité de l'activité

5.4.2.1 Contexte de modélisation

Le second type de situation d'exploitation considéré consiste à *présenter* les traces modélisées de son activité à *l'utilisateur du système lui-même* (*i.e.* à celui qui en est à l'origine), pour lui donner une *image* de sa *propre activité* dans le temps. À nouveau le modélisateur de trace peut avoir affaire à différents contextes de modélisation, similaires à ceux que nous avons présentés précédemment : soit le modélisateur cumule les missions de modélisation et de codage mais également d'expert de l'activité tracée, soit il est amené à collaborer avec d'autres acteurs. Cette fois cependant, c'est l'utilisateur observé lui-même qui sera mis en quelque sorte en *position d'analyste* de sa propre trace, et la trace modélisée devra *intégrer* l'activité en cours, et n'en sera pas indépendante (Figure 5.9).

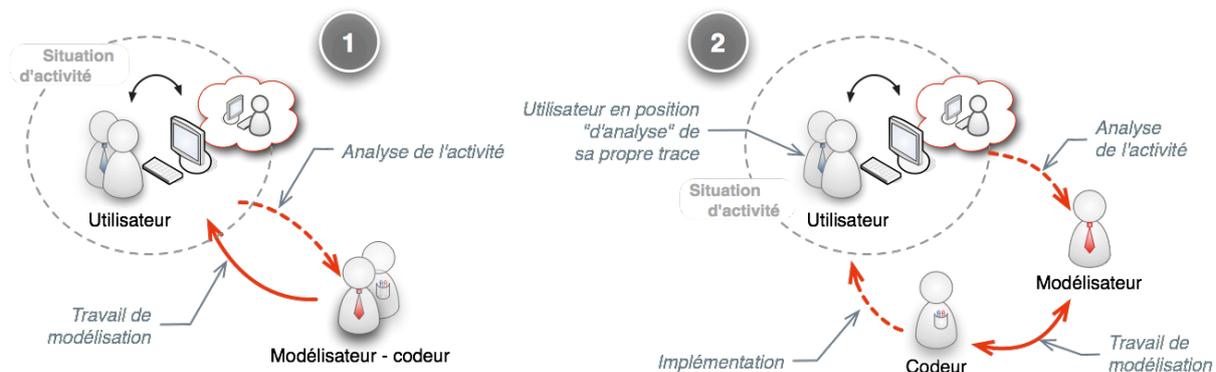


Figure 5.9 : Contextes de modélisation possibles dans l'optique d'une trace modélisée pour la réflexivité de l'activité : (1) le modélisateur implémente seul ou (2) collabore avec d'autres acteurs.

Quel que soit le cas de figure, la trace modélisée dans cette situation devra cette fois « *re-présenter* » l'activité, ou en d'autres termes « rendre présent à nouveau » à l'utilisateur sa propre activité passée. En fait la confrontation d'un utilisateur à sa propre trace conduit à deux types d'interprétation possibles de sa part. En termes d'usages de la trace modélisée, il s'agit pour le modélisateur de fournir une trace dont nous disons qu'elle soutient la *réflexivité* de l'activité qui se compose de deux dimensions, l'une de *remémoration* l'autre de *métacognition*.

5.4.2.2 Conception d'usage(s)

Présenter à un utilisateur sa propre trace le conduit en effet soit à se *remémorer ce qu'il a fait*, et la trace est alors interprétée comme une (re)contextualisation des éléments manipulés par le passé, soit à *objectiver* sa propre activité, c'est-à-dire interpréter non seulement ce qu'il a fait mais *comment il l'a fait* ce qui implique des *processus métacognitifs*. Concernant la trace modélisée à obtenir, il n'est pas question de viser soit l'une, soit l'autre de ces interprétations possibles de la trace : ces deux dimensions de l'interprétation sont *indissociables*. On peut toutefois remarquer que dans une trace modélisée composée d'observés mis en relations, ces observés constituent leur propre contexte, c'est-à-dire que

la présentation d'une trace modélisée induit en principe directement une contextualisation des observés instanciés dans la trace.

Concernant le travail de modélisation, il est impossible dans cette situation de s'appuyer sur un objectif particulier d'analyse pour déterminer le *niveau de description pertinent* pour un premier modèle de trace. Le modélisateur doit concevoir lui-même, *en fonction* de l'activité et de l'utilisateur impliqué, des situations d'usage de la trace modélisée *au cœur même* de l'activité pour en déduire le niveau de description à atteindre. La phase d'analyse de l'activité du processus générique de modélisation que nous avons proposé revêt dans ce cas une complexité et une importance toute particulière. La trace modélisée visée devant soutenir à la fois des processus mnésiques et métacognitifs, le choix des observés devra passer par la détermination de ce qui pour l'utilisateur lui-même, constituent les « termes de son activité ». Dans cette démarche on se rapproche de ce que Kiss et Quinqueton (2004) désigne comme des « *stances* » qu'ils définissent comme « *atomic units of subjective reality* ».

Très concrètement, un moyen d'obtenir ce type d'explicitation nous semble être la combinaison d'entretiens et de réalisations à haute voix. Par exemple : dans un premier temps, le modélisateur discute un scénario plausible d'activité. Dans un entretien *a priori*, il demandera à l'utilisateur de décrire ce qui va constituer la réalisation de l'activité à venir. Dans un troisième temps, ce dernier réalise l'activité (enregistrée) en question en commentant à haute voix ses actions. Enfin, dans un entretien *a posteriori*, le modélisateur demandera à l'utilisateur de comparer et d'expliquer la comparaison entre ce qu'il avait prévu de faire et ce qu'il a réellement fait. Les résultats de ce genre d'investigations peuvent conduire à mettre en évidence des éléments saillants pour l'utilisateur dans son activité mais qui ne sont pas présents dans l'environnement numérique, et pas non plus nécessairement observables, ni même facilement explicitables pour l'utilisateur.

Il faut donc prévoir, dans le travail de modélisation une association étroite de l'utilisateur, et une analyse fine de son activité. Pour soutenir ce travail d'explicitation le modélisateur aura tout intérêt à prévoir également un certain nombre de *boucles de validation* dans le processus de modélisation. Comme dans la situation précédente l'utilisateur confronté à sa trace en réalisera une *interprétation* qui dépendra du contexte d'interprétation et des moyens qui lui seront donnés de construire cette interprétation. Il est bien entendu inutile d'espérer créer un modèle de trace parfait qui permettrait de produire pour l'utilisateur, quelles que soient les conditions courantes, une vision adéquate de ce qu'il a fait, et de comment il l'a fait. Il faut dans ce cas, comme dans le précédent envisager l'existence de plusieurs présentations possibles de la trace avec des *transformations* adéquates. Il faut également envisager l'évolution du modèle de trace lui-même, voire la mobilisation de plusieurs modèles de trace. À nouveau, le travail de modélisation d'une trace ne se limite pas ici à déterminer les termes du modèle de trace. Il s'agit de concevoir les usages du SBT, de concevoir le SBT lui-même comme un instrument, devant permettre à l'utilisateur d'interpréter sa propre activité, qu'il s'agisse d'une simple contextualisation ou d'une analyse critique de celle-ci.

5.5 Dynamique de mise en place d'un SBTm

Quelle que soit la situation d'usage, le travail de modélisation ne s'achève pas totalement avec l'implémentation du premier (ni même d'un unique) modèle de trace. Au contraire, l'effort de modélisation continue dans un certain nombre d'ajustements, de mise en place de transformations liées à de nouvelles visualisations, voire des adaptations profondes du modèle de trace en place. En fait, on peut considérer que le travail de modélisation n'est jamais totalement abouti (Laflaquière et *al.*, 2008). À la suite de la mise en place du premier modèle, il est possible (voire probable) qu'il faille proposer des transformations de la trace modélisée, ce qui revient à définir de nouveaux modèles de traces, notamment pour répondre aux besoins d'une visualisation interactive.

Comme dans les démarches d'Ingénierie des Connaissances le travail de modélisation est indissociable du travail de conception du système informatique – ici le SBTm implémenté (ou SGBT) - qui sera utilisé *in fine*. Pour que ce dispositif soit fonctionnel il faudra donc plusieurs itérations du processus de modélisation, l'établissement de compromis, notamment lorsque ce travail implique d'autres acteurs : analystes, et/ou codeurs. Il faudra ajouter, modifier, compléter le modèle de trace et faire évoluer le SGBT pour l'adapter aux premiers usages réels des traces modélisées, et pour garantir de façon pérenne *l'interprétabilité* et *l'accessibilité* des traces modélisées, dans le cadre d'une activité donnée.

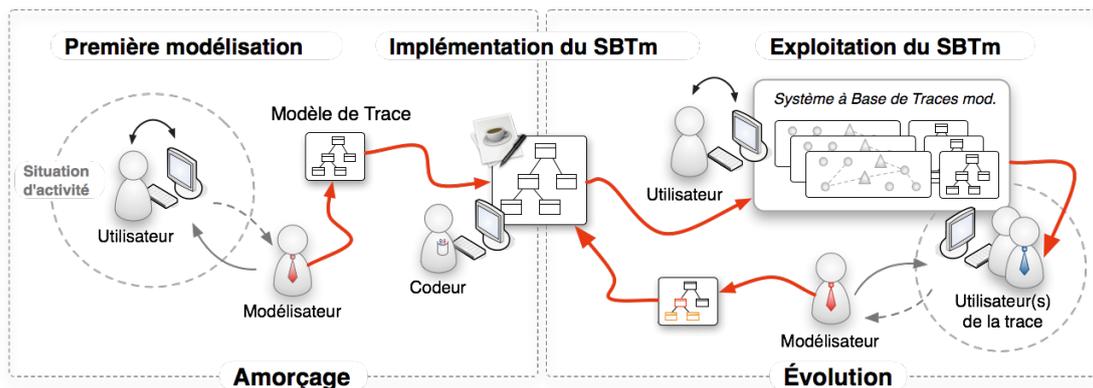


Figure 5.10 : Les deux temps de conception et de déploiement d'un SBTm.

Par la suite, l'évolution du dispositif et du modèle de trace, devra accompagner celle de l'activité elle-même. En effet, d'une part l'activité elle-même évolue et ensuite, comme pour tout autre outil, des usages du SGBT vont se développer¹⁵⁰, se mettre en place et faire naître du même coup de nouvelles attentes. Pour tenir compte de cette situation, nous suggérons que le déploiement d'un SBTm et le travail de modélisation allant avec soient exécutés en deux temps (Figure 5.10). Dans un premier temps l'application du processus de modélisation aura pour but de converger, par ajustements successifs de la modélisation, vers un premier modèle de trace permettant de générer une trace modélisée interprétable. Il s'agira de la *phase d'amorçage* d'un SBTm. Cette phase aboutit à une première implémentation du système et à l'intégration effective du modèle de trace défini.

Dans un second temps, et une fois que les premières traces modélisées sont exploitables et exploitées, l'évolutivité des SBTm sera mise à profit pour faire évoluer les traces modélisées en fonction de leurs

¹⁵⁰ Notamment du fait de son appropriation progressive ou « genèse instrumentale » (Cerrato, 2005).

usages. Durant cette seconde phase, le modélisateur travaillera à l'adaptation de la trace modélisée et de sa présentation à la situation d'usage effective : définitions de transformations des traces modélisées, développement de fonctionnalités de visualisation en rapport avec ces transformations, développement de fonctionnalités d'interaction avec la trace (*etc.*) mais également à nouveau analyse de l'activité, identification de nouveaux besoins et éventuellement remise en cause du modèle de trace initial. Ces ajustements instaurent une sorte de *cycle* permanent d'amélioration des traces modélisées dans leur situation d'usage, principe classique dans le domaine de la conception de systèmes informatiques (en particulier d'interfaces) et également appliqués dans certains travaux sur des systèmes traçants comme dans (Ritter, 1994) ou (Pirolli *et al.*, 2002).

5.6 Conclusion

Nous avons présenté dans ce chapitre l'ensemble des éléments qui constituent notre proposition de méthodologie, incluant un cadre théorique, un processus de modélisation de trace générique et ses spécialisations, ainsi qu'une démarche dynamique à deux temps d'implantation d'un SBTm. Pour conclure ce chapitre nous allons revenir sur la manière dont la méthodologie proposée articule les apports théoriques présentés au chapitre précédent. La démarche globale proposée par la méthodologie de modélisation présentée dans ce chapitre, composée d'un processus générique de modélisation, d'une spécialisation de ce processus et d'une dynamique intégrant dans une même dynamique modélisation de la trace et conception du SGBT ne dépareille pas des approches globales rencontrées dans les domaines que nous avons mobilisés, même si sur le dernier point elle s'approprie une posture plus spécifique à l'IC et à la conception de SBC.

Le processus de modélisation générique de trace lui-même s'inspire plus spécifiquement des processus de modélisation d'IHM et surtout de ceux d'IC. On retrouve ainsi dans les deux phases de notre méthodologie que sont l'analyse descriptive d'une part et la création du modèle d'autre part, les deux phases qui structurent un processus de modélisation des tâches avec d'une part une mise en évidence des tâches à décrire, et d'autre part une description formalisée des tâches en question. Dans le détail, les six étapes qui entrent dans ces deux phases sont plutôt inspirées des processus de modélisation des connaissances de domaine et de construction d'ontologie, notamment en ce qui concerne les étapes de la phase de création du modèle.

Au sein de ce processus de modélisation générique, la phase d'*analyse descriptive* remobilise quant à elle les trois domaines de l'IC, des IHM et du *CSCW*. Grossièrement on peut dire qu'elle leur emprunte leur démarche d'analyse propre pour les intégrer en une seule et même démarche d'analyse de l'activité nécessaire à la modélisation de trace. Ainsi par exemple la première étape du processus, la délimitation du périmètre de modélisation s'inspire de la définition du système sociotechnique présenté par la Cognition Distribuée que nous avons empruntée au *CSCW*.

De manière plus forte encore, l'étape de description analytique de l'activité, qui repose sur trois dimensions principales (empruntées à la Théorie de l'Activité)¹⁵¹, associe en fait des démarches d'analyse relevant des trois domaines précités : l'analyse et la description de la dimension « sujet » se rapproche ainsi d'une démarche d'analyse typique du *CSCW*, celle de la dimension « objet » d'une

¹⁵¹ Dimensions sujet, objet et outils auxquelles nous avons ajouté celle de la temporalité.

démarche d'IC et celle de la dimension « outil » d'une démarche d'analyse de tâche en IHM. La méthodologie que nous proposons s'inspire donc largement des différents domaines IC, IHM et *CSCW*, dont nous avons articulé les apports méthodologiques (Figure 5.11). Toutefois, la méthodologie proposée telle qu'elle est présentée ici n'est pas seulement le fruit d'une combinaison d'approches théoriques. Sa création repose également et pour une bonne part sur un travail de recherche appliquée. En l'occurrence, ce travail de « recherche-action » a été réalisé sur deux terrains d'application distincts, que nous allons présenter dans le chapitre suivant.

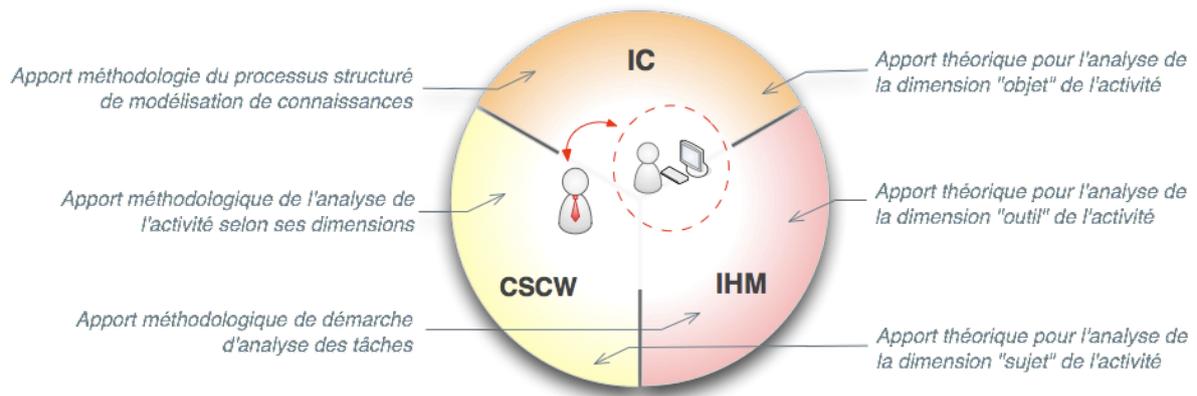


Figure 5.11 : La méthodologie de modélisation des traces proposée combine les apports méthodologiques et théoriques des domaines de l'IHM, de l'IC et du *CSCW*.

Chapitre 6

Recherche action pour une méthodologie de modélisation des traces numériques

Résumé du chapitre

La méthodologie de modélisation de traces proposée doit autant aux autres approches de modélisation dont nous nous sommes inspirés pour en donner une base, qu'à son application directe sur le terrain qui lui a donné sa forme définitive. En l'occurrence, nous avons pu travailler sur deux terrains d'applications au sein desquels nous avons mené un travail de modélisation de trace. Le premier travail de modélisation a été réalisé pour une activité *d'apprentissage collaboratif à distance* avec la plateforme *eLycée*. Nous avons tiré de cette première expérience plusieurs enseignements concernant la méthodologie de modélisation à adopter : la nécessité d'une structuration stricte, la nécessité d'une séparation claire entre analyse de l'activité et problèmes d'implémentation, enfin la nécessité d'une démarche de conception des usages. Le second travail de modélisation a été mené pour une activité de *production de contenus pédagogiques* réalisée par une équipe de développeur de formation professionnel utilisant un atelier de production collaboratif (*Emulsion*). Contrairement au projet précédent, nous avons pu mener à bien toutes les étapes de notre méthodologie de modélisation pour accomplir une phase d'amorce du SBTm devant être *in fine* mis en place. L'usage envisagé des traces que nous avons conçues dans le contexte de ce terrain était une forme de capitalisation des connaissances mobilisées dans la création des formations. Pour vérifier qu'il était possible à un développeur d'interpréter une trace modélisée de son activité pour y retrouver sa stratégie de développement pédagogique pour une formation donnée, nous avons mis en place une expérimentation dans laquelle nous avons confronté un développeur à sa propre trace modélisée, en développant pour cela une solution technique *ad hoc* autour de l'outil de visualisation *Timeline*. Le protocole expérimental utilisé était celui de l'interruption / reprise de l'activité. Les résultats de cette première expérimentation suggèrent une situation plus complexe que ne prenait pas complètement en compte notre hypothèse de départ d'une facilitation de la reprise d'activité grâce à la présentation d'une trace modélisée. Celle-ci a effectivement permis une objectivation de l'activité tracée, sans en permettre une totale réappropriation pour en reprendre la réalisation. Il a fallu pour cela le concours d'un autre acteur dans un échange collaboratif. De ce constat découlent de nouvelles questions, qui invitent à dépasser le cadre de notre problématique de départ en discutant la notion de trace modélisée d'activités conjointes.

6.1 Introduction

Ce chapitre présente la partie appliquée de notre travail de recherche sur la conception d'une méthodologie de modélisation de trace. Rappelons que pour mettre au point la méthodologie proposée au chapitre précédent, nous nous sommes appuyés d'une part sur un cadre théorique issu de la littérature (chapitre 4) et d'autre part sur une mise en pratique directe du travail de modélisation de trace, une « recherche-action » durant laquelle nous avons éprouvé et ajusté les apports théoriques pour en faire une méthodologie de modélisation effective.

Comme en Ingénierie des Connaissances, le travail de recherche sur les traces numériques ne peut s'abstraire des situations concrètes dans lesquelles une activité est tracée, ce qui nous a incité en partie à parler « d'ingénierie des traces numériques » (Laflaquière et al., 2008). En l'occurrence les traces modélisées ne peuvent être considérées comme une inscription de connaissances que dans la mesure où ces connaissances sont *concrètement mobilisées* dans l'activité qui est tracée. Ce qui implique qu'une modélisation ne peut réellement être mise à l'épreuve *que dans sa confrontation à une situation d'activité réelle*.

Suivant cette logique, nous avons tenu pour cette thèse à travailler sur des terrains d'expérimentation en situation réelle et exclu une approche fondée uniquement sur une expérimentation en milieu contrôlé avec des activités fictives et provoquées. Ce choix comporte des contraintes spécifiques : socio-organisationnelles tout d'abord, car il faut avoir complètement accès à la situation d'activité tracée¹⁵², techniques ensuite car il faut pouvoir instrumenter l'environnement numérique¹⁵³. Dans les deux exemples de terrain d'application que nous allons présenter dans ce chapitre, nous avons bénéficié de circonstances favorables qui ont rendu possible d'une part l'accès à l'activité réelle des utilisateurs et d'autre part à l'instrumentation, au moins partielle, de l'environnement numérique. Le premier terrain est une situation d'apprentissage collaboratif à distance avec la plateforme *eLycée*TM (section 6.2), le second une situation de production de contenus de formation avec la plateforme *Emulsion*TM dans une grande entreprise (section 6.3). Pour terminer, nous dresserons un bilan du travail de modélisation effectué et des conclusions qui en sont issues (section 6.4).

6.2 Traces modélisées d'une activité pédagogique dans un EIAH

La première activité réelle en situation que nous avons eu à aborder est une activité d'apprentissage se déroulant au sein d'un EIAH collaboratif : *eLycée campus*TM. En tant que « modélisateur de traces », nous prenons place parmi d'autres acteurs engagés sur le terrain dans la mise en place d'un Système à Base de Trace *ad-hoc*, intégré à l'EIAH lui-même. Nous avons donc endossé notre rôle avec un double objectif en vue : d'une part *fournir concrètement un modèle de trace* destiné à intégrer le Système à Base de Traces alors en cours d'implémentation, et d'autre part *définir les contours d'une méthodologie* permettant d'encadrer et soutenir la création dudit modèle de trace.

¹⁵² Environnement de travail des acteurs, documents, échanges entre collaborateurs, plans, brouillons, organisation personnelle et collective, mais également connaissances, stratégies, pratiques métier, autant d'informations sensibles pour d'évidentes raisons.

¹⁵³ Ce qui dans le cadre d'une activité professionnelle n'est pas évident : logiciels propriétaires, protection des données, etc.

La présentation qui va suivre ne doit donc pas être vue comme l'application d'une méthodologie de modélisation qui n'en était alors qu'à ses prémises, mais comme la conception d'une méthodologie tenant compte des problèmes rencontrés par le modélisateur de trace au cours de son travail. Nous présenterons dans un premier temps le *contexte de modélisation* qui fût le notre (section 6.2.1), autour de l'activité en situation, des acteurs impliqués et du projet de modélisation relatif à l'implantation du SBT. Dans un second temps nous évoquerons du travail d'analyse et de modélisation concrètement effectué (section 6.2.2 et section 6.2.3) et nous nous ferons le bilan (section 6.2.4).

6.2.1 Contexte de modélisation

Pour commencer, rappelons que *eLycée* est une *start-up* fondée en 2006 et qu'elle propose un service d'enseignement du français sur une plateforme d'enseignement à distance, une « classe virtuelle ». L'offre vise en particulier des enfants (de 8 à 14 ans) francophones scolarisés aux États-Unis et ne pouvant pratiquer le français dans le cursus scolaire américain classique¹⁵⁴. L'objectif n'étant pas une *initiation* à la langue, mais le *maintien* des connaissances et le *développement* de la pratique du français, la démarche pédagogique proposée s'éloigne du classique schéma « leçons / exercices », et repose sur une vision constructiviste de l'apprentissage. Pour résumer, il s'agit de considérer que l'apprentissage doit se faire *dans et par* la pratique et *en interaction* avec ses pairs.

Ainsi, l'équipe pédagogique propose à ses apprenants différentes activités pédagogiques dont l'objet principal n'est pas l'apprentissage de règles de la langue, mais la mise en situation d'expression, conduisant les apprenants à prendre la parole et à réaliser des exercices de manière collaborative en français. Les échanges nécessaires à la réalisation de l'exercice sont parfois, de ce point de vue, plus « riches d'enseignements » que le résultats de l'activité en lui-même. En plus de la pratique de la langue, c'est également un *contact culturel* qui est recherché et les activités pédagogiques qui ne portent pas directement sur les règles et exceptions du français donnent l'occasion d'aborder divers traits de la culture francophone. Bien entendu, si cette démarche pédagogique est possible c'est que l'environnement numérique mobilisé permet un travail collaboratif synchrone à distance, d'un véritable « groupe classe » virtuel.

6.2.1.1 Équipe et activités pédagogiques



Figure 6.1 : Acteurs impliqués dans les activités de *eLycée*.

Le fonctionnement de la « classe virtuelle »¹⁵⁵ fait intervenir plusieurs acteurs (Figure 6.1) dont les membres de la petite équipe pédagogique de *eLycée*. Parmi eux, trois à quatre pédagogues / enseignants préparent en amont les séances et les supports pédagogiques nécessaires. Un *Webmestre* a en charge le site *Web*, qui est à la fois un outil de communication et un espace où stocker les ressources et supports pédagogiques utilisés lors des séances de classe, mais également la gestion des

¹⁵⁴ En général il s'agit d'enfants dont le français est la (une des deux) langue(s) maternelle(s).

¹⁵⁵ Nous détaillerons plus tard ce qui constitue techniquement cette « classe virtuelle ».

comptes utilisateurs et des problèmes de connexion. Un informaticien (codeur) est quant à lui en charge du développement des outils de la plateforme en ligne. Enfin bien entendu, les séances de classe réunissent en ligne des *apprenants*, jusqu'à une dizaine en simultané qui sont *clients*¹⁵⁶ du service proposé par *eLycée*, accompagnés d'un *tuteur* issu de l'équipe pédagogique.

Les activités pédagogiques proposées sont *documentaires* et relativement *ouvertes*, visant avant tout à provoquer l'échange et la collaboration des apprenants. Il s'agit généralement de les confronter au traitement de *ressources* documentaires, consultables en ligne au sein de l'environnement et exploitant les possibilités des supports multimédia. Les échanges écrits et oraux provoqués par le traitement de ces ressources débouchent sur une *production* (le plus souvent écrite) exécutée (généralement) en binômes par les apprenants. Durant quelques minutes les apprenants en binômes travaillent dans un espace virtuel privé, avec l'aide du tuteur qui se connecte de groupe en groupe. De manière générale, une activité pédagogique type se déroule donc comme suit :

- (1) Une séance débute par une activité d'introduction à laquelle tous les apprenants participent collectivement. Il s'agit pour le tuteur d'orienter la séance en fonction de la thématique choisie, et de présenter le reste de l'activité.
- (2) Une activité plus approfondie est ensuite proposée. Effectuée en binôme, il s'agit en général de traiter un document et d'en tirer une production écrite, également réalisée en binôme.
- (3) Une fois terminée cette production, le collectif est à nouveau réuni et chacun fait part au groupe classe de sa production. La séance, qui aura duré environ deux heures, se termine par un bilan de la session animé par le tuteur.
- (4) En dehors de ces séances collectives, les apprenants sont invités à réaliser d'autres productions et à suivre des séances individuelles en tête à tête avec le tuteur (toujours à distance). Ces séances individuelles, plus courtes, permettent au tuteur d'apporter un soutien individualisé, en revenant sur les productions réalisées durant les séances collectives ou en répondant aux questions que l'apprenant n'aurait pas eu l'opportunité de poser durant les séances collectives.

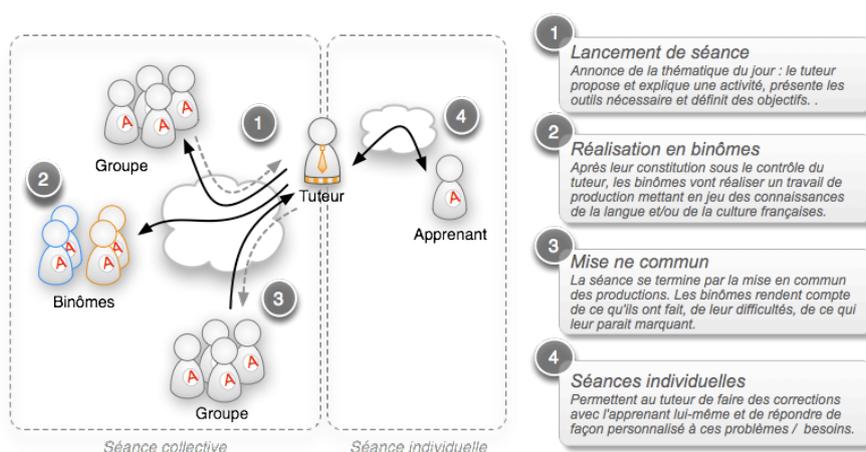


Figure 6.2 : Structure générale de l'activité des participants à la classe virtuelle.

¹⁵⁶ Point important car ne permet pas une liberté totale concernant les observations de l'activité notamment.

6.2.1.2 Projet de modélisation initial

Dans le cadre de l'implantation d'un SBTm au sein d'un EIAH comme *eLycée*, le projet de modélisation initial était de créer des traces modélisées destinées à la *réflexivité de l'activité des apprenants*, qui comme nous l'avons signalé, est un facteur déterminant dans une activité d'apprentissage et un challenge pour les traces numériques en général. Dans le projet initial la mise en place d'une trace numérique prenant place dans l'environnement des apprenants ne comportait pas d'objectif précis quant à la réutilisation concrète de ces traces par les apprenants, autrement dit la conception d'un SBTm passait également par la constitution de proposition d'usages pertinents pouvant intégrer concrètement les activités pédagogiques. Sur un plan technique, la plateforme *eMédiathèque* étant encore un développement, le choix du concepteur a été de développer directement les fonctionnalités de collecte, de gestion des traces modélisées et de leur visualisation au sein même de l'application.

6.2.2 Travail d'analyse et de modélisation

Cette première expérience de modélisation nous a permis de poser les premiers éléments de structure de notre démarche pour modéliser des traces, en commençant par la détermination du système observé et l'analyse des activités pédagogiques réalisées.

6.2.2.1 Système observé

Techniquement, la « classe virtuelle » que propose *eLycée* est en réalité une plateforme en ligne constituée de trois éléments (Figure 6.3).



Figure 6.3 : Les outils formant l'environnement proposé par *eLycée*.

(1) Un outil de *visioconférence*, en l'occurrence *Marratech*TM, qui offre un espace de travail collaboratif à toute la classe : vidéo, chat (public ou privé), tableaux blancs partagés (permettant d'écrire, de dessiner, de manipuler des images, etc.).

(2) Un outil de *navigation Web partagée*, *eMédiathèque*, développé tout spécialement pour permettre aux apprenants d'explorer pages Web et ressources multimédia de façon synchrone et coordonnée.

(3) Un *site Web*, qui regroupe les informations relatives à l'organisation des séances mais également un ensemble de liens utiles (pointant vers un dictionnaire en ligne, un glossaire, *etc.*), un ensemble de ressources exploitables, et un espace personnel permettant de gérer ses propres documents de travail.

Il faut préciser que le développement d'*eMediathèque* avait pour objectif final d'intégrer directement les fonctionnalités de visioconférence (chat, tableau blanc, puis éventuellement vidéo). Nous avons donc engagé notre travail de modélisation en considérant que *eMediathèque* serait en mesure de prendre rapidement en charge toutes les fonctionnalités¹⁵⁷ nécessaires à la classe virtuelle. Au final, le *système observé* retenu est constitué théoriquement d'*eMediathèque* seul et en pratique d'*eMediathèque* et *Marratech*TM.

6.2.2.2 Observations et analyse de l'activité

L'analyse des activités pédagogiques se déroulant dans cet environnement a demandé plusieurs séries d'observations. Ces observations avaient lieu *in situ*, si l'on peut dire, le seul moyen observer l'activité en train de se réaliser étant d'être *connecté en même temps* que les autres participants à la plateforme, et de suivre les opérations sans intervenir. Sur le plan technique ces observations ont l'inconvénient d'exiger la connexion d'une personne supplémentaire ce qui alourdit la charge sur la bande passante.

Le but de l'observation n'étant pas de ralentir la progression générale, certaines observations ont dû être interrompues. L'alternative qui aurait consisté à enregistrer les séances avait le même inconvénient, l'enregistrement devant être réalisé sur le serveur *Marratech*TM¹⁵⁸. Dans ces conditions, peu d'enregistrements complets ont pu être effectués. De plus, la position de « participant muet » de la classe virtuelle est également délicate d'un point de vue pratique. Les travaux en binôme, par exemple, s'effectuent grâce à des tableaux blancs et des conversations « privés » dans l'outil de visioconférence, ce qui ne permet pas de savoir ce qui se passe concrètement tout au long de l'activité.

La première série d'observations a été réalisée lors d'une semaine de test de la plateforme, juste avant le lancement officiel de l'offre d'*eLycée*, avec une dizaine de participants¹⁵⁹, pour qui des activités avaient été conçues spécialement. Ces premières séances étant avant tout consacrées aux réglages du dispositif technique, elles ne nous ont permis qu'une analyse superficielle du déroulement des activités pédagogiques. Une seconde série d'observations a été réalisée plus tard, en conditions réelles d'utilisation, c'est-à-dire avec des clients de la plateforme. Nous avons ainsi pu suivre, toujours en tant que participant « muet » une quinzaine de séances avec des participants se trouvant à différents endroits des Etats-Unis¹⁶⁰. Finalement, le *contexte de modélisation* était le suivant (Figure 6.4) : les observations nécessaires à *l'analyse de l'activité* se trouvaient centrées sur l'utilisation de *Marratech*TM (qui devait disparaître à terme au profit des nouvelles fonctionnalités de *eMediathèque*), avec les contraintes que nous venons d'évoquer, et le *système observé* quant à lui était centré sur *eMediathèque*, dont les fonctionnalités devaient soutenir l'ensemble des activités pédagogiques. Dans les deux cas il s'agissait de tenir compte des interactions des *apprenants* comme de celles du *tuteur*.

¹⁵⁷ De plus, une version *beta* d'*eMediathèque* prenant en charge le tableau blanc et le chat a pu être utilisée durant notre intervention mais uniquement en condition expérimentale, *i.e.* pas avec les utilisateurs réels.

¹⁵⁸ Les données sont stockées sur un serveur : il s'agit d'un enregistrement vidéo qui est rejouable dans *Marratech*TM uniquement. Les Tableaux blancs sont également enregistrés, mais à part, dans un format qui permet de les rééditer *a posteriori*.

¹⁵⁹ Des élèves du secondaire en France.

¹⁶⁰ Ce qui implique des sessions de travail à suivre entre 00h00 et 02h00 du matin, heure de Paris, souvent sans possibilité d'enregistrer. Nous avons suivi ces séances sur une période de 2 mois, début 2007. La première série avait eu lieu fin novembre 2006.

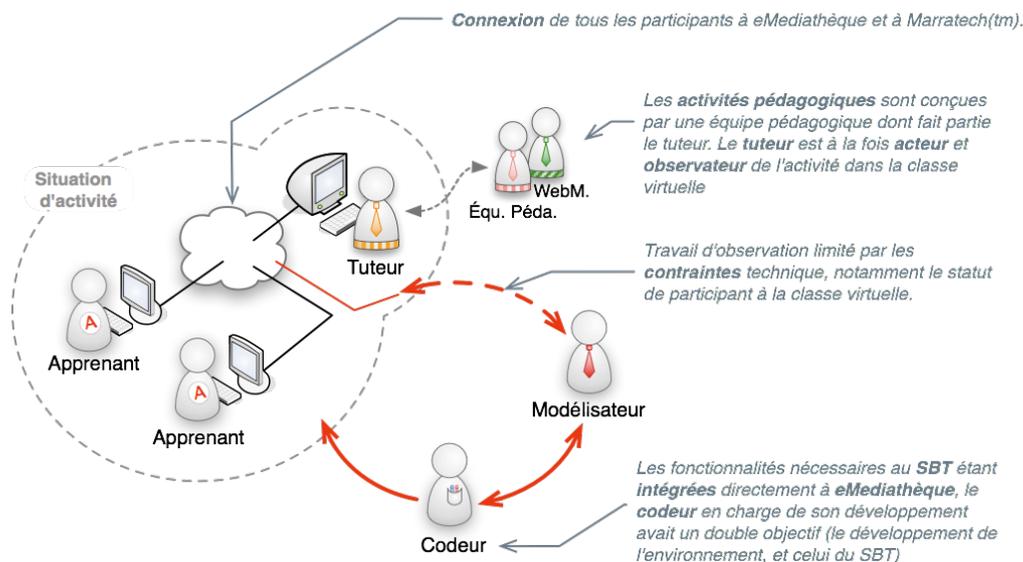


Figure 6.4 : Contexte de modélisation autour de l'activité de la classe virtuelle.

6.2.2.3 Exemple d'activité pédagogique proposée

Pour illustrer le type d'activité pédagogique proposé nous avons choisi de présenter brièvement l'exemple d'une activité de traduction d'une planche de bande dessinée (Figure 6.5). Cette activité de traduction (anglais > français), est proposée à des participants en binôme. L'intérêt est ici la confrontation à un langage particulier, dans lequel des expressions spécifiques à une langue sont employées. Cela permet d'une part de rencontrer du *vocabulaire*, mais également de *discuter* la meilleure traduction possible, étant entendu que la traduction de certaines expressions ne peuvent se faire mot à mot. Il est d'ailleurs intéressant de voir à quel point le volume d'échanges (en français) entre deux participants peut-être grand relativement à la taille de l'expression à traduire.



Figure 6.5 : Exemple d'activité pédagogique : de traduction de bande dessinée.

Chaque binôme se connecte donc à la page originale à traduire (une page *Web* contenant l'image de la page) sur *eMediathèque*. En parallèle, les deux élèves ouvrent un tableau blanc qu'ils vont partager et sur lequel ils collent l'image de la planche avec des bulles vides. En créant une zone de texte pour chaque bulle, ils complètent leur traduction. Pour négocier la traduction, les membres du binôme s'appuient sur un *chat privé* (doublé d'une conversation audio-vidéo, privée également). Les ressource

ces dont les participants disposent, outre la planche originale elle-même, sont un dictionnaire en ligne, un glossaire présent des les médias disponibles directement dans *eMediathèque*, ainsi qu'un accès au *Web* et aux sources qu'il peut fournir (en évitant toutefois les traducteurs automatiques).

6.2.3 Création d'un modèle de trace et visualisation

Rappelons que notre objectif en termes de modélisation était à la fois de fournir un modèle de trace *et* de poser les jalons d'une méthodologie de modélisation. Le contexte et les contraintes liées au développement de la classe virtuelle ont favorisé le premier aspect. En effet, pour les besoins du développement nous avons travaillé à l'élaboration rapide et directe d'un modèle de trace destiné à mettre à l'épreuve rapidement l'outil de *visualisation* de trace intégré à l'interface d'*eMediathèque*. Ce n'est donc qu'*a posteriori* que nous avons tiré les conclusions quant à la méthodologie de modélisation. En caricaturant, nous avons été contraint de passer directement d'une *analyse préliminaire* de l'activité à un *modèle de trace* opérationnel pour la visualisation.

6.2.3.1 Modèle de trace simplifié

Afin de créer rapidement un premier modèle de trace implémentable pour le module de visualisation de trace d'*eLycée*, nous avons effectué une analyse rapide et effectué le constat suivant : Premièrement, dans la majorité les activités pédagogiques proposées par *eLycée*, l'apprentissage est contextualisé par l'exploitation de *documents ressources* d'une part et la création de *documents produits* d'autre part. Dans l'exemple précédent, il s'agit de la planche originale et des différentes ressources en ligne d'une part, et de la planche à remplir d'autre part ; deuxièmement, dans ce type d'activité pédagogique l'acquisition de *vocabulaire* est un objectif majeur. Cette acquisition se fait dans le contexte de la production demandée : soit elle est le fruit d'une recherche d'un apprenant par lui-même, soit le fruit de l'échange entre deux apprenants.

Une contrainte technique venait s'ajouter aux facteurs liés à l'activité elle-même. En effet, la trace modélisée à produire devait être visible depuis un module de l'interface qui est *a priori* présent en permanence dans l'interface de l'apprenant. Par conséquent le modèle de trace devait être adapté à la fois pour une visualisation en *temps réel* et éventuellement une visualisation *a posteriori*. Nous avons donc du envisager deux usages correspondant à cette configuration du système à base de trace modélisée mis en place. La visualisation temps réel permettant une présence permanente de la trace dans l'interface de l'apprenant devait ainsi lui faire prendre conscience de l'activité en cours, alors qu'une visualisation *a posteriori*, devait permettre à l'apprenant de recontextualiser son apprentissage. Pour satisfaire cet objectif en tenant compte des différentes contraintes, nous avons produit un modèle de trace simple (Figure 6.6) reprenant grossièrement les idées principales concernant les documents manipulés et les interactions de base des apprenants dans *eLycée*.

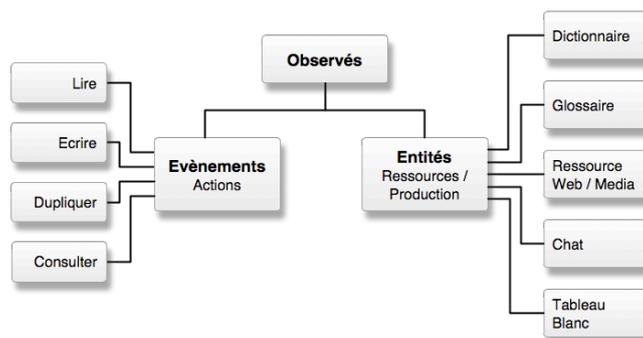


Figure 6.6 : Modèle de trace simple pour une activité pédagogique de traduction d'*eLycée*.

En parallèle, nous avons proposé deux visualisations pour la trace modélisée. En termes de mise en œuvre dans le SBTm implémenté cela signifiait un modèle de trace de base et une transformation de la trace modélisée générée afin de passer (à la demande de l'utilisateur) d'une trace modélisée « temps réel » à une trace modélisée « synthétique ».

6.2.3.2 Visualisation proposées

En complément du modèle de trace, nous avons donc proposé l'idée d'une visualisation différenciée des traces modélisées adaptée aux usages envisagés. Pour la trace modélisée « temps réel » la solution qui s'imposait était celle d'un affichage sous forme d'une liste des observés se succédant dans le temps. Pour la trace modélisée « synthétique » une élaboration plus sophistiquée de la visualisation devait être menée. C'est ce que nous avons fait avec une réflexion sur la finalité que pourrait avoir la présentation d'une telle trace à un apprenant dans le contexte d'une activité pédagogique comme celle présentée plus haut de traduction d'un planche de bande dessinée. Dans ce genre d'activité pédagogique, il est bénéfique d'aménager en fin de séance un moment de « bilan » pour revenir sur l'activité réalisée durant la séance : récapituler les notions (ici les termes) apprises, faire part des difficultés rencontrées ou au contraire des aspects positifs de l'activité réalisée. On a en effet montré depuis longtemps que les apprenants se représentent relativement mal l'activité d'apprentissage qu'ils ont menée (Gay et Mazur, 1993), d'où l'intérêt de ce type de bilan.

Dans cet esprit nous avons proposé que la trace puisse être visualisée sous une forme synthétique pour mettre en évidence les termes susceptibles d'avoir été appris dans leur contexte d'occurrence. On peut attendre d'une telle présentation de la trace modélisée premièrement qu'elle soutienne l'exercice de rappel demandé lors du bilan de la séance, et deuxièmement qu'elle renforce également la mémorisation des termes en question. L'idée de cette visualisation « synthétique » est donc de filtrer la trace pour articuler des éléments particuliers : des termes qui ont été (re)découverts par les apprenants, en l'occurrence ici des termes qui auront été cherchés dans le dictionnaire ou le glossaire en ligne (en accès direct par la plateforme).

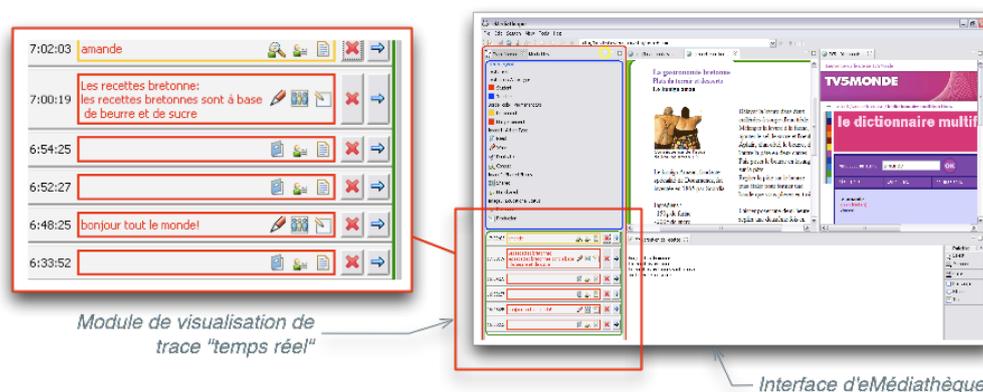
Pour simplifier, nous avons proposé que ces mots ou expression soient présentés avec un maximum d'élément de contexte, afin d'offrir le plus d'indices de rappel possible. En l'occurrence, un mot cherché dans le dictionnaire en ligne à un instant donné dans le cadre de l'activité de traduction de bande dessinée, est lié à la bulle dans laquelle il apparaît (observé « tableau blanc »), à la page du dictionnaire qui en donne la définition (observé dictionnaire) et à la conversation qui a lieu dans le chat lors de la traduction de la bulle en question (observé chat). Au final, la trace modélisée présentée sous sa forme

« synthétique » devait se présenter comme une liste temporalisée de mots rattachés à leur contexte d'utilisation et/ou de découverte pour un apprenant.

6.2.3.3 Validation du dispositif de visualisation

Le modèle de trace simple que nous avons proposé a été implémenté dans une version *beta* de l'application *eMediathèque* intégrant un *chat* et un tableau blanc, ainsi qu'un module de visualisation. L'objectif de l'équipe à ce moment là étant avant tout de valider techniquement (et rapidement) le module de visualisation en question, seul la première des deux visualisations que nous avons suggérées a été concrètement implémentée. Pour valider concrètement le module un test en condition réelle n'étant pas envisageable (la plateforme étant en exploitation commerciale) nous avons organisé une expérimentation en situation contrôlée.

Nous avons dans ce but repris l'activité pédagogique de traduction de bande dessinée que nous avons adaptée pour l'occasion. Une dizaine de membres du LIRIS a accepté de réaliser cette expérimentation qui reproduisait au mieux les conditions de déroulement¹⁶¹ d'une séance normale d'*eLycée*. Nous avons en effet préparé l'exercice comme une véritable séance de classe virtuelle. Nous avons prévu une petite activité d'une dizaine de minutes pour prendre en main les fonctionnalités nécessaires à la réalisation de l'activité, préparer des consignes claires pour les binômes « d'apprenants » qui étaient laissés en autonomie durant la réalisation de la traduction. Pour les besoins d'une analyse ultérieure, chaque poste de travail était instrumenté avec une petite application de capture dynamique d'écran.



Le travail effectué à l'occasion de cette expérimentation a effectivement permis de valider techniquement le module de visualisation de trace « temps réel ». Celui-ci a par la suite été amélioré dans sa partie technique et toujours pour une présentation « temps réel » (Figure 6.7). Contrairement à ce que nous avons espéré, nous n'avons pas pu confronter les sujets à une trace « synthétique » de leur activité ce qui aurait éventuellement permis de valider l'approche non seulement sur ces aspects techniques mais également sur le potentiel des traces modélisées à soutenir une activité de mémorisation¹⁶².

¹⁶¹ A noter l'absence de conversation audio-vidéo habituellement disponible dans *Marratech*TM mais dont le module dans *eMediathèque* n'était pas encore disponible.

¹⁶² L'implémentation de la transformation de trace modélisée nécessaire à la présentation d'une trace synthétique n'a pas été réalisée à temps.

6.2.4 Bilan

Nous dressons un bilan contrasté de ce premier travail de modélisation sur le terrain. Nous avons bien fourni un modèle de trace simple qui a été implémenté pour valider *sur le plan technique* le module de visualisation de trace, mais ni la trace modélisée elle-même en tant que support à la réflexivité de l'activité de l'apprenant, ni notre démarche de modélisation n'ont pu être évaluées au sens strict. Dans les termes de la méthodologie présentée au chapitre précédent, nous n'avons pas eu la possibilité de compléter une phase d'amorçage du SBTm qui devait être mis en place.

Primo et concernant la trace modélisée, il n'est pas possible d'évaluer l'impact d'une visualisation d'une trace « temps réel » avec une expérimentation unique, courte, en laboratoire, avec des utilisateurs pour qui c'était le premier contact avec l'application *eMediathèque*. La « prise de conscience » de son activité par un apprenant durant la réalisation de son activité est un phénomène subtil difficile à observer et mesurer, et la mise au point d'une trace modélisée (et la modélisation qui va avec) avec pour objectif ce type de réflexivité de l'activité demande un travail de recherche approfondi en amont de la réalisation technique. Deuxio, la visualisation « synthétique » de la trace modélisée était une piste intéressante car il était imaginable de mesurer son impact (sur la mémorisation des mots à termes et plus simplement encore comme support à une activité de « bilan » en fin de séance), malheureusement cette seconde visualisation n'a pas pu être implémentée. Le modèle de trace proposé l'avait été en intégrant d'entrée ces deux visualisations. Il était difficile d'analyser la pertinence de ce modèle en dehors du cadre complet dans lequel il avait été imaginé. Du même coup la validation de la visualisation elle-même sur le plan de son utilisabilité, ne pouvait être conduite. Les résultats obtenus nous paraissent donc bien en deçà de ce que nous aurions pu espérer au regard du temps et du travail fourni autour de ce projet. Néanmoins, nous avons su tirer des leçons de cette première expérience de modélisation sur le terrain en termes de méthodologie de modélisation.

6.2.4.1 Leçons pour une méthodologie de modélisation de traces

Les difficultés rencontrées nous ont offert l'occasion de réfléchir plus concrètement au processus de modélisation. De manière générale tout d'abord, il est clair que pour une première modélisation on ne doit pas faire l'amalgame de plusieurs problématiques, comme c'était le cas ici avec une visualisation « temps réel » et une visualisation « *a posteriori* ». Ensuite, définir une finalité générale comme ici « la réflexivité de l'activité des apprenants dans *eLycée* » ne donne pas au modélisateur une direction claire pour conduire son travail d'analyse et de modélisation. Il est nécessaire de *concevoir explicitement les usages* de la trace modélisée, qui n'aura de pertinence mesurable *que* par rapport à ces usages prévus. Par conséquent, un travail de modélisation de trace ne peut être uniquement guidé par les contraintes techniques de l'implémentation d'un SGBT, ce qui fût finalement le cas ici.

Ces leçons ont été mises à profit pour structurer la méthodologie de modélisation que nous souhaitons proposer. Il apparaît ainsi clairement que le travail d'analyse de l'activité conduisant à identifier des observés pertinents doit être dissocié des problèmes techniques d'implémentation (ici de visualisation). Nous avons donc renforcé la séparation entre le travail de modélisation et celui de l'implémentation du modèle. Nous avons rigidifié les étapes du processus de modélisation pour encadrer le travail de modélisation et aider le modélisateur de trace à rester au plus près de ses objectifs de départ qui doivent inclure des usages des traces modélisées, ce que nous avons également appelé un

« projet de modélisation ». De notre travail de modélisation en collaboration avec d'autres acteurs du projet nous avons retenu également la technique qui consiste à réaliser des traces modélisée « à la main », *i.e.* des traces fictives ou simulées qui permettent de sélectionner des observés.

6.2.4.2 Réflexion sur les traces modélisées dans le EIAH

Pour terminer, notons que ce premier terrain d'application nous a également apporté matière à réflexion concernant la place et le rôle des traces modélisées dans des EIAH. En effet, comme c'était le cas ici, on attend généralement de la présentation de traces à un apprenant qu'elles apportent une *réflexivité de l'activité d'apprentissage*. Cela implique que par simple présentation, les traces soutiennent, voire suscitent, la tenue de processus *métacognitifs* chez l'apprenant. Dans ce cas, créer un dispositif technique qui permette de présenter directement à un apprenant une trace de son activité numérique est-elle la meilleure façon de procéder ?

L'existence du dispositif technique en lui-même n'est pas à remettre en cause bien entendu. En revanche la manière d'envisager l'usage de ce dispositif peut l'être, en effet donner directement une trace représentant son activité à un apprenant n'est pas nécessairement l'idée la plus pertinente. Théoriquement tout d'abord, un processus métacognitif *aboutit* à une représentation de sa propre activité. Ce qui signifie que c'est *dans la construction* de cette représentation que se situe l'intérêt de ce processus et non uniquement dans son résultat. On pourrait donc considérer que présenter directement une représentation de son activité à un apprenant ne soutient pas forcément le processus sensé *construire* cette représentation.

Cette remarque est particulièrement importante dans le cas où l'on considère un apprenant qui ne maîtrise pas encore pleinement l'activité métacognitive. La tenue d'une activité *métacognitive* est en soi un objectif pédagogique : les apprenants « apprennent à apprendre ». Dans ce cas l'exploitation des traces modélisées pour une réflexivité des activités d'un apprenant *ne peut être envisagée en dehors d'un processus temporel d'appropriation, d'un apprentissage, d'un usage qu'il faut concevoir et accompagner*. En deux mots, la trace modélisée doit devenir un objet pédagogique à part entière et doit faire l'objet d'un apprentissage, du développement d'une pratique.

Concernant spécifiquement *eLycée* ensuite, nous pensons que la trace modélisée ne doit pas être directement présentée aux apprenants que progressivement. Nous suggérons même que la trace modélisée doit d'abord s'adresser au tuteur, qui sera en position d'analyste de la trace de l'apprenant. Dans un deuxième temps la trace modélisée doit être partagée par le tuteur et l'apprenant lui-même pour en construire ensemble une *interprétation*, et introduire ainsi petit à petit une capacité à analyser sa propre trace chez l'apprenant. Dans cette situation intermédiaire la trace modélisée sert de support à l'échange entre apprenant et tuteur. Finalement la trace modélisée peut être introduite, comme tout autre objet d'apprentissage dans l'environnement du travail de l'apprenant, en proposant des activités spécifiques impliquant son exploitation.

6.3 Traces modélisées d'une activité de création de contenus

Les leçons d'une première modélisation de trace sur le terrain nous ayant permis de définir de manière plus rigoureuse l'organisation de notre démarche méthodologique, notre objectif en abordant un second terrain d'application était de prendre en charge de manière plus indépendante le travail de modélisation, et de réaliser le plus complètement possible une *phase amorce* de déploiement d'un Système à Base de Traces modélisées. Le terrain d'application en question concerne l'activité d'une équipe *d'ingénierie pédagogique* d'un service spécialisé au sein d'un grand groupe industriel français, qui a pour objectif de produire les contenus et supports pédagogiques de formations professionnelles à destination de ses collaborateurs. Après avoir présenté plus clairement dans la section à suivre (section 6.3.1) cette activité de production de contenus, nous présenterons notre travail de modélisation en suivant l'articulation imposée par notre méthodologie. Nous commencerons donc par une section présentant le travail réalisé pour la phase d'analyse descriptive de l'activité à tracer (section 6.3.2), puis nous reviendrons sur la phase de création d'un modèle (section 6.3.3), que nous avons couplée à une expérimentation (section 6.3.4 à 6.3.6) liée au projet de modélisation que nous avons spécifiquement développé sur ce terrain.

6.3.1 Contexte de modélisation

Ce second terrain d'application est relativement plus compliqué que le précédent et prend place dans le cadre socio-organisationnel complexe d'un grand groupe industriel. Le nombre de personnes (directement et surtout indirectement) impliquées par l'activité de production de ces contenus et de ces formations est potentiellement élevé. Cette activité subit de fait l'influence d'une multitude de facteurs, allant de simples problèmes techniques au niveau local jusqu'aux choix stratégiques à l'échelle du groupe en terme de qualité ou de modernisation de la gestion des formations de ses agents.

6.3.1.1 Création de formations et production de contenus pédagogiques

La formation professionnelle des agents de ce groupe, qu'il s'agisse de leur formation initiale ou continue, est principalement pilotée par leurs sites de rattachement où sont dispensées en grande partie ces formations¹⁶³. Pour répondre à de nouvelles exigences normatives¹⁶⁴, la création de ces formations est depuis quelques années *numérisée* et *centralisée* au plan national. Les supports de formations papiers sont ainsi petit à petit remplacés par des supports de présentation vidéo-projetés. Si les formations restent dispensées localement, la centralisation de leur conception au Service d'Ingénierie (SI) garantit une *homogénéité* des compétences des agents formés sur tout le territoire. La numérisation des supports pédagogiques, en s'appuyant sur les standard du *elearning*, permet quant à elle de faciliter la gestion centralisée des formations¹⁶⁵ et ouvre également la voie à d'autres types de formations que les traditionnelles séances en présentiel (environnement d'auto-formation ou d'auto-évaluation, apprentissage à distance, simulations, *etc.*). Dans ce contexte, la mission du (SI) est de répondre à des demandes de formations, formulées par les sites de production en fonction de leurs besoins. Ces deman-

¹⁶³ Chaque site de production disposant de ses propres formateurs, aptes à former les agents aussi bien sur des notions génériques, que sur des particularités locales liées au site. Les formateurs en question ne le deviennent qu'après avoir plusieurs années d'expérience de terrain.

¹⁶⁴ Notamment en termes de qualité certifiée.

¹⁶⁵ Notamment leur *validité* : les contenus de formation, qui sont sensés être disponibles et accessibles, doivent être à jour des dernières (et régulières) modifications techniques, si possible sans avoir à remettre en cause toutes les formations qui font mentions de ces éléments.

des peuvent concerner tout type de fonction, quels que soient le domaine d'expertise, l'expérience ou le cursus de l'agent. En réponse à ces demandes, le service crée des formations sous la forme de stages, d'importance variable (avec un public plus ou moins large et une durée allant de quelques heures à plusieurs jours), c'est-à-dire conçoit les *activités pédagogiques*, leur *encadrement* (emploi du temps, consignes, exercices, etc.), et développe les *contenus* et *supports pédagogiques* nécessaires.

6.3.1.2 Membres de l'équipe et structure de leur activité



Figure 6.8 : Les membres de l'équipe impliqués dans la production de formation.

La création des formations réunit les compétences de différents acteurs dans une équipe d'ingénierie pédagogique (Figure 6.8) composée d'une quinzaine de membres¹⁶⁶ et placée sous l'autorité du *responsable* de service qui attribue les missions et les moyens, et valide sa production. Les contenus et supports pédagogiques sont produits par des « *développeurs* », qui ont la particularité d'être des agents expérimentés. D'abord experts de terrain reconnus, ils ont été formateurs sur le terrain avant d'intégrer cette équipe. Les formations sont quant à elles conçues en mode projet, chaque projet étant confié à un développeur qui prendra pour l'occasion le rôle de « *chef-produit* ». Les formations conçues impliquent la création ou la réutilisation de contenus pédagogiques sur lesquels interviennent également des « *graphistes documentalistes* » qui ont en charge la mise en forme des supports¹⁶⁷, leur indexation dans les serveurs du service, et fournissent également un appui pour la conception d'animations ou de schéma complexes, ou encore la recherche de certains types de ressources.

Autour de cette équipe qui travaille dans un même lieu interviennent plus ou moins directement dans la production plusieurs acteurs extérieurs : les demandeurs de formation, les formateurs sur le terrain, les experts métier et les constructeurs qui peuvent être mobilisés sur des questions techniques, etc. On pourrait également ajouter à la liste les membres d'une équipe prestataire¹⁶⁸ développant un nouvel outil de travail pour l'ensemble de l'équipe d'ingénierie pédagogique. Au cours de notre intervention sur place, l'ensemble de l'équipe nous a offert sa collaboration, et nous avons pu compter en particulier sur plusieurs développeurs et chefs-produit qui ont accepté de se rendre disponibles pour nous permettre de faire nos observations, de même pour l'équipe prestataire. Comme l'illustre le schéma ci-dessous (Figure 6.9), l'activité de l'équipe s'articule en trois temps.

(1) Une demande de formation est émise depuis un site de production¹⁶⁹. Cette demande exprimée sous la forme d'un cahier des charges de demande qui définit des besoins en termes de *compétences*. En réponse à cette demande un cahier des charges de l'offre est proposé (et négocié si besoin) par le SI, fixant cette fois des *objectifs pédagogiques généraux* (OPG) qui correspondent au développement des compétences demandées. Le cahier des charges en question servira par la suite de document de référence pour le lancement du projet de développement.

¹⁶⁶ Au moment de notre intervention fin 2007.

¹⁶⁷ Dans le respect d'une charte graphique propre au groupe.

¹⁶⁸ Il s'agit d'une équipe de la société *Solunea* (www.solunea.fr) dont nous aurons l'occasion de reparler par la suite plus en détail.

¹⁶⁹ Le processus en externe du SF est ici simplifié.

(2) Le développement proprement dit est piloté par un développeur, choisi pour être chef-produit en fonction des caractéristiques techniques et pédagogiques de la demande. Il *conçoit* la formation dans son ensemble et *réalise* ou coordonne la réalisation des contenus et supports pédagogiques nécessaires. Pour ce faire il traduit les objectifs pédagogiques généraux en *objectifs pédagogiques spécifiés* (OPS), qui sont du point de vue pédagogique des unités (ou modules) indépendantes. Les contenus et supports de ces OPS doivent être développés de façon à être *réutilisables* dans différents contextes, de sorte qu'une formation puisse être construite comme l'assemblage de plusieurs d'entre elles¹⁷⁰. Chaque contenu développé est *validé* en interne avant d'intégrer définitivement la formation, qui sera à son tour *validée* dans son ensemble¹⁷¹.

(3) Une fois la formation validée, elle est livrée puis exploitée sur le terrain. Cette exploitation, comme d'autres types d'activités dans le groupe, fait l'objet d'un *retour d'expérience* qui fera remonter, si besoin, des remarques ou des demandes de modification. Celles-ci seront prises en compte, soit pour une correction immédiate (cela peut-être le cas pour la mise à jour d'un contenu technique par exemple), soit à travers une nouvelle demande et un développement ultérieur.

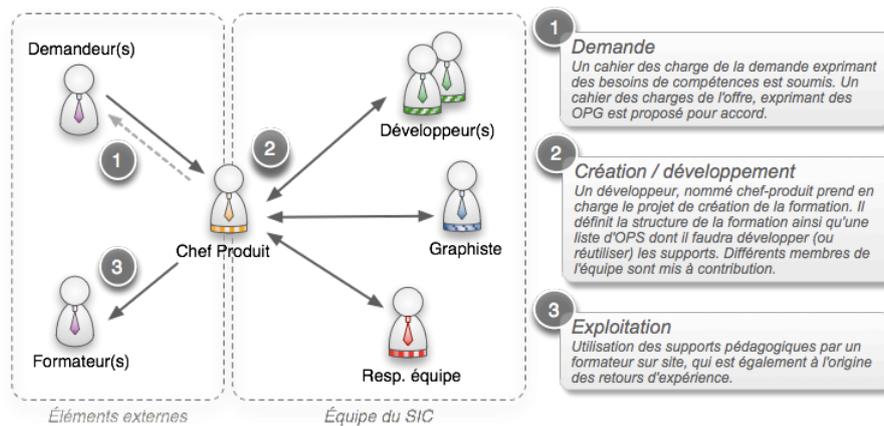


Figure 6.9 : Structure générale de l'activité de l'équipe d'ingénierie pédagogique.

6.3.1.3 Projet de modélisation initial

Dans l'optique de l'implantation d'un Système à Base de Trace, c'est principalement le *second point*, celui de la réalisation d'un projet de développement qui nous intéresse, et tout particulièrement, l'activité documentaire instrumentée du *développeur endossant le rôle de chef-produit*. De manière tout à fait intéressante en effet, le développeur en question doit gérer, dans un *même* espace numérique, partiellement partagé avec le reste de l'équipe, la conception d'une formation dans son ensemble et le développement concret des contenus et supports pédagogiques qui vont avec, le tout reposant quasi exclusivement sur la manipulation et la création de documents numériques.

Dans ce contexte notre projet de modélisation repose sur le déploiement, à terme, d'un Système à Base de Traces modélisées implémenté (ou SGBT) délivrant des *traces modélisées* réutilisées directement dans l'activité pour soutenir une *réflexivité* de l'activité d'un développeur. Nous avons donc engagé un travail de modélisation en ce sens, appliquant notre méthodologie nouvellement redéfinie pour com-

¹⁷⁰ Objectif de *capitalisation* du SF, qui souhaite diminuer autant que possible les coûts de développement de nouvelles formations.

¹⁷¹ Une formation « pilote » est menée sur site en condition réelle avant validation complète.

pléter au mieux une *phase d'amorce* du Système à Base de Trace modélisée devant être implanté¹⁷². Les sections à suivre présentent ce travail de modélisation, en commençant par la phase *d'analyse descriptive de l'activité* devant être tracée.

6.3.2 Phase d'analyse descriptive de l'activité

Première phase de notre processus de modélisation, l'analyse située vise à *déterminer le périmètre de modélisation*, puis à *décrire l'activité* qui sera tracée de façon à faire émerger des *observés potentiels* (voir section 5.3.1). L'activité en question étant d'une grande complexité, signalons que serons amenés à résumer certaines informations et omettre volontairement certains détails, parfois pour des raisons de confidentialité, mais souvent pour en simplifier la lecture et la compréhension.

6.3.2.1 Détermination du périmètre de modélisation

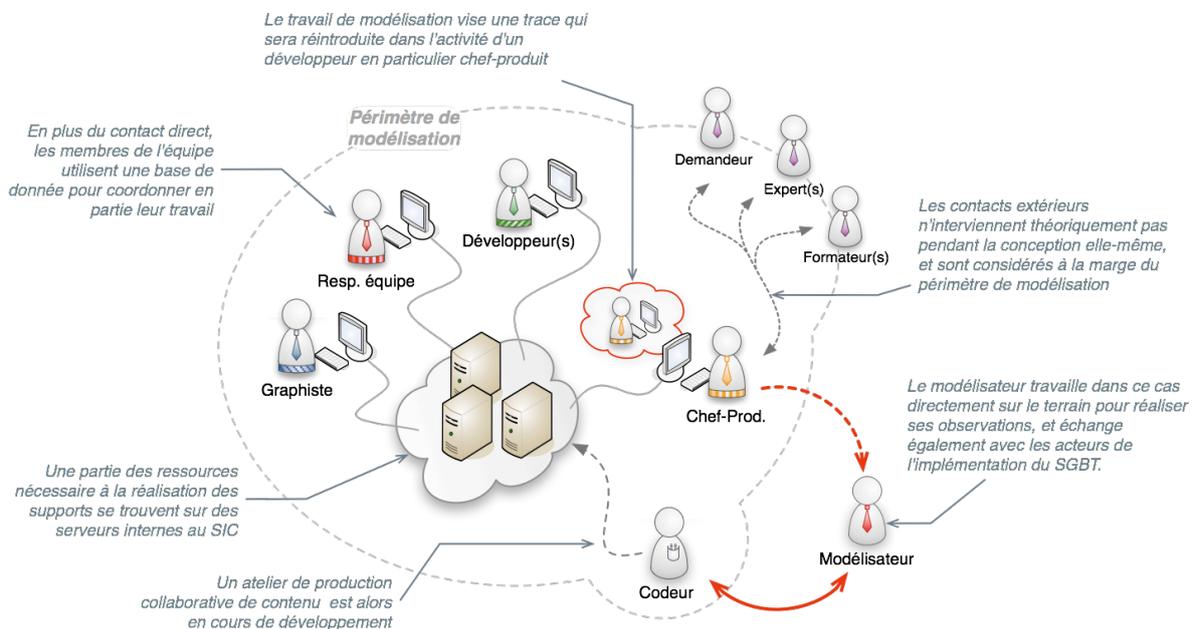


Figure 6.10 : Périmètre de modélisation de trace pour l'activité des développeurs chef-produit.

Déterminer le périmètre de modélisation revient à *co-déterminer* le *système observé* et l'*activité* qui sera tracée. Cette première étape s'est révélée plus compliquée que prévue. Plusieurs entretiens informels en face à face (tableau blanc à l'appui), complétés de plusieurs entretiens téléphoniques, ainsi qu'une étude des documents disponibles n'ont en effet pas été suffisants pour établir la liste de tous les acteurs et outils pouvant être impliqués et mobilisés dans l'activité de développement. Pour déterminer ce périmètre de modélisation (Figure 6.10) nous avons finalement eu recours à une « simulation improvisée » : nous avons demandé à un développeur de réaliser rapidement, dans son environnement de travail habituel, les opérations nécessaires à la conception d'une formation et à la production d'un support correspondant à un module (OPS).

¹⁷² Précisons que l'instrumentation de l'environnement était envisagée dans le cadre du développement d'un nouvel outil de travail pour l'équipe, qui sera décrit ci-après.

Activité tracée

Comme nous avons déjà largement évoqué l'activité en question dans l'introduction de cette section, il nous est inutile d'y revenir en détail. Remarquons simplement qu'il s'agit d'une activité documentaire, dont la réalisation passe par la manipulation et la création de nombreux documents numériques, partagés ou non. Nous pouvons aussi en profiter pour pointer quelques différences notables avec la situation précédente. Outre l'intervention importante d'éléments extérieurs à l'équipe, toute l'activité d'un développeur, en particulier dans sa dimension collective, ne s'inscrit pas systématiquement dans l'environnement numérique comme ça pouvait être le cas dans la classe virtuelle : beaucoup d'échanges ont lieu par téléphone ou en face à face. De plus il existe certaines ressources techniques uniquement disponibles au format papier par exemple. En revanche, il est possible de mettre en place une observation *in situ*, de réaliser des enregistrements et enfin de mener un travail de modélisation indépendant du développement de l'environnement.

Système observé

Le système observé est par définition l'ensemble d'outils permettant la réalisation de l'activité tracée. En l'occurrence ici, au moment de notre intervention, l'équipe est en phase d'intégration d'un *nouvel environnement de travail* qui cohabite pour un temps avec le système « initial ». Pour faire simple, nous allons d'abord décrire l'ensemble initial d'outils, puis le nouvel environnement, et enfin nous en tirerons les conclusions.

- Système initial

Le système initial est une association de plusieurs outils, qui ne sont pas nativement dédiés à la production de contenus pédagogiques. Tout d'abord, *plusieurs serveurs de contenus* sont disponibles sur un réseau interne et sont consultables directement par *l'explorateur de fichier* du système d'exploitation. Les principaux documents liés au processus de production d'une formation (cahier des charges par exemple) sont rédigés avec *Microsoft Word*TM (éventuellement *Excel*TM). S'agissant de la production de support pédagogique (OPS), elle se fait sous *Microsoft PowerPoint*TM. Ce dernier est complété le cas échéant, par un outil de dessin technique ou un logiciel de création de schéma. Le logiciel *Breeze*TM est utilisé pour transcoder les supports au format *flashpaper*¹⁷³ (qui garantit leur intégrité lors de leur exploitation). Pour gérer les fichiers des supports pédagogiques durant leur développement, une base de donnée sous *Microsoft Access*TM a été créée en interne. Elle est utilisée pour déposer des documents (ou des versions de documents) et comme elle permet un accès partagé, elle sert de plateforme commune, ainsi que de support à l'organisation d'un cycle de validation pour les différents documents concernés par cette procédure (cf. infra). Enfin, la gestion des communications électroniques (*email*, poste restante et notifications) mobilise le logiciel *Lotus Notes 5*¹⁷⁴.

- Nouveau système

L'outil qui constituera l'essentiel du nouvel environnement de travail est un atelier de production multimédia dédié au *elearning*, dénommé *Emulsion*TM et développé par la société *Solunea*¹⁷⁵. L'objectif de cette application client/serveur est de remplacer à terme l'ensemble des outils constituant le système

¹⁷³ Format équivalent à *pdf*, qui constitue une impression dans un fichier (comme *pdf*, c'est *Adobe*TM qui en est propriétaire).

¹⁷⁴ Lotus Notes (IBM Software Group) permet un accès partagé à des applications (emails, calendriers) sur un serveur Lotus Domino.

¹⁷⁵ Voir <http://www.solunea.fr/produits.html> (consulté le 20/06/2009).

précédent en gérant l'accès aux différents serveurs de contenus, la gestion de tous les contenus internes, la fabrication de supports pédagogiques (OPS) de type *PowerPoint*TM mais avec des possibilités d'animation bien plus avancées, leur publication dans un format permettant de maîtriser les modifications *a posteriori*, édition de schéma, gestion des OPS dans les différents stades de leur création, *versionning*, etc. En un mot, *Emulsion* est sensé offrir aux développeurs du SI un environnement de travail unifié, gérant de manière centralisée, les contenus, les ressources, la production, les droits ainsi que les étapes de production. En revanche, *Lotus Notes 5*TM, reste l'outil privilégié pour la gestion des communications électroniques (Figure 6.11).

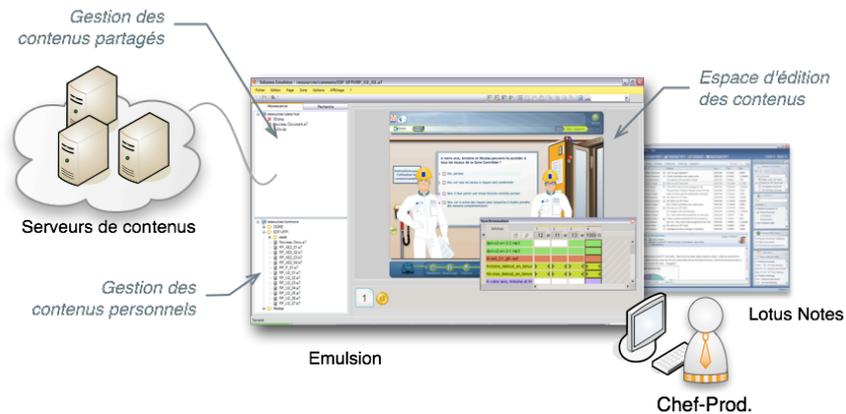


Figure 6.11 : Système observé centré sur *Emulsion* (atelier de production de contenus multimédia).

Cette situation comporte avantages et inconvénients : d'un côté le fait que le développement d'*Emulsion* soit encore ouvert permet d'envisager une instrumentation pour collecter des données, de l'autre un double travail d'analyse s'impose, l'un concernant le système exploité jusqu'ici, l'autre concernant le système qui sera réellement instrumenté pour les traces. Cette situation nous a finalement obligé à modéliser l'activité de manière générique, de façon à être en phase avec les deux environnements, nous détachant ainsi des particularités de chaque instrumentation, ce qui correspond finalement aux objectifs fixés par notre approche. Autre inconvénient notable, il a fallu fournir un effort conséquent, tant sur le plan technique qu'organisationnel pour que l'atelier soit déployé et fonctionnel sur le terrain, afin qu'il soit concrètement utilisé et donc soutenir une activité observable et traçable¹⁷⁶.

6.3.2.2 Description analytique de l'activité

Deuxième étape, la *description analytique* de l'activité consiste à aborder l'activité selon quatre dimensions principales, pour en tirer des descriptions complémentaires (section 5.3.1.2). Pour la première réelle application de notre méthodologie sur ce point, nous avons eu à traiter d'une activité relativement complexe. Pour ne pas en faire ici un exposé trop long, nous insisterons sur les points principaux et résumerons les résultats des analyses que nous avons conduites sur le terrain. Le trait principal de l'activité des développeurs chef-produit est la combinaison de la *conception* des formations dont ils ont la charge, et du *développement* de contenus et supports pédagogiques (OPS).

¹⁷⁶ Nous disposions pour ce faire de très peu de temps, environ 6 mois pour l'ensemble du projet de modélisation, durant lesquels nous avons rencontré plusieurs problèmes techniques liés au déploiement d'*Emulsion* au SF auquel nous avons participé activement.

Plusieurs entretiens ont été nécessaires, sur le terrain, pour analyser l'activité. Ces entretiens ont été menés principalement auprès de deux développeurs de l'équipe¹⁷⁷. Nous avons débuté par des *entretiens informels*, puis nous avons utilisés des *entretiens guidés*. Dans tous les cas, c'est un point important, ces entretiens étaient réalisés dans l'environnement de travail *habituel*, et cet environnement était *mobilisable* par l'acteur pour illustrer son propos. Concernant nos observations *in situ*, nous avons demandé à ces développeurs de réaliser *leur activité habituelle* en commentant les étapes de leurs travail, soit sur exemples fictifs, soit (le plus souvent) sur des exemples réels tirés de leur activités de développement alors en cours. Nous avons donc réalisé différentes descriptions de l'activité selon les dimensions définies dans notre processus de modélisation.

Dimension du sujet

Aborder l'activité par la dimension du *sujet* revient à (re)situer ce dernier dans son environnement socio-organisationnel. La création de ces formations est une activité collaborative qui implique la coordination et l'articulation du travail de production de contenus. Il s'agit donc de déterminer quelle place le développeur (en particulier chef-produit) occupe dans les collectifs que l'activité fait émerger, et inversement comment la dimension collective s'inscrit dans le travail individuel. À la base, il existe bien sûr une *organisation hiérarchique* liée au fonctionnement du service comme toutes les autres entités de l'entreprise (Figure 6.12).

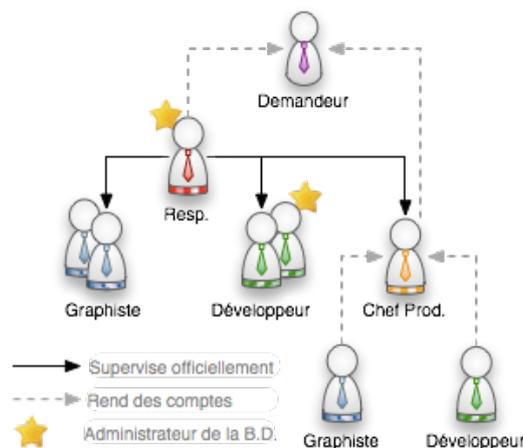


Figure 6.12 : Organigramme simplifié de l'équipe.

Cette structure hiérarchique est cependant modifiée *dans et par* l'action : le collectif se structure en fonction des contraintes de l'environnement numérique qu'il contribue lui-même à définir¹⁷⁸. Ainsi les droits de la base de donnée, par exemple, sont gérés par les agents qui l'ont mise en place, sans rapport direct avec leur position hiérarchique. Cette structure hiérarchique est également altérée dans le cadre des responsabilités liées au *cycle de validation* des modules (OPS) produits (Figure 6.13). Les développeurs changent ainsi de positions les uns par rapport aux autres, en fonction des projets, voire même occupent plusieurs rôles différents en même temps, sur des projets de développement parallèles.

¹⁷⁷ Des entretiens complémentaires ont été réalisés avec les autres membres de l'équipe.

¹⁷⁸ Idée maîtresse des approches de l'équipe Tech-cico notamment autour du Web socio-sémantique (Zacklad, 2005b ; Cahier et al., 2004).

Comme le montre le schéma ci-dessous le cycle de validation d'un module est relativement complexe. Un développeur crée son module en état *brouillon*. Une fois terminé, ce document vérifié par un autre développeur et un graphiste qui (si aucune modification n'est à signaler) font passer le produit dans son état *validé*. De même pour le chef produit en premier lieu, qui est à l'origine de la demande de développement du module (OPS), et qui le fera passer en état *approuvé*, puis en deuxième lieu le responsable de l'équipe qui autorise sa *publication* définitive au sein d'une formation¹⁷⁹.

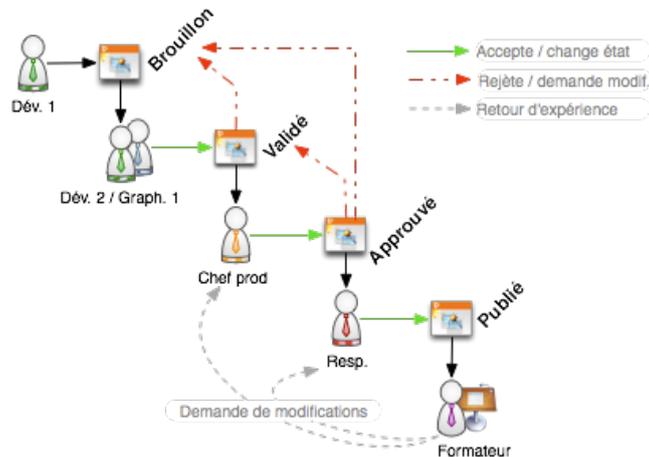


Figure 6.13 : Cycle de vie des contenus de formation produits (OPS).

Toujours du point de vue organisationnel et collectif, il est à noter plusieurs *dépendances* de l'activité d'un développeur, et notamment pour un chef-produit, pour lequel l'avancement de la conception d'une formation dépend du travail de réalisation de modules (OPS) qu'il doit *déléguer* à d'autres développeurs. Le travail d'*articulation* nécessaire au bon déroulement de l'activité globale du service est conséquent et le système en place n'instrumente pas explicitement cette part de l'activité. D'où le détournement d'usage d'une Base de Données *Access*TM mise en place en interne sur un serveur du service. Un champ est réservé au statut d'un contenu en production, marquant ainsi son état d'avancement dans le cycle de validation (sans notification automatique). Autrement dit, dans les faits, que le système porte ou non cette information les acteurs s'assureront de sa transmission à la personne concernée par un moyen de communication plus direct, que cela passe par le téléphone, l'email ou encore le face à face au détour d'un bureau. De nombreuses interactions de ce type ont lieu durant l'activité, et seule une partie passe par l'environnement informatique et est potentiellement traçable (Figure 6.13).

Dimension de l'objet

Aborder l'activité d'un développeur par la dimension de son *objet* consiste à caractériser (dans le cas d'une activité documentaire comme ici) les *documents manipulés*, et *l'espace de travail organisé* qu'ils constituent dans leur ensemble. Au-delà d'une stricte caractérisation des *documents* eux-mêmes il s'agit surtout de caractériser *ce qu'ils représentent* du point de vue de l'utilisateur qui les manipule, et les *connaissances* que ce dernier met en œuvre pour les manipuler.

¹⁷⁹ Une fois publié, le module (OPS) sort du cycle de validation et ne peut y entrer à nouveau qu'en faisant l'objet d'une nouvelle demande.

Étant donné le grand nombre de documents manipulés, une description exhaustive des documents utilisés dans la création d'une formation et la production de ses supports serait trop lourde. Nous avons choisi dans la figure ci-dessous (Figure 6.14) de mettre en valeur les « unités » documentaires directement manipulées lors de la création d'une formation. Dans l'environnement de travail des développeurs, une formation est matérialisée par un *Dossier Pédagogique* (DP).

Ce dossier comprend un ensemble de documents de référence notamment le *Cahier des Charges* (CdC) et un *Emploi du Temps* (qui servira de planning à l'ensemble de la formation), ainsi qu'un ensemble de Modules (OPS) de formation. Chaque module contient une ou plusieurs présentation(s)¹⁸⁰ ainsi qu'un *document de référence* (Doc. Réf.) qui commente ou justifie la démarche pédagogique à appliquer en réponse à la demande, donne les références utilisées, ainsi que diverses informations sur l'évaluation du module par exemple.

Les présentations quant à elles sont classiquement composées d'une série de « diapositives » que nous avons ici qualifiées de « scènes » car elles peuvent être constituées d'un simple texte mais peut tout aussi bien comporter une animation complexe impliquant des personnages agissant à l'écran. Pour simplifier nous avons limité les composants d'une scène à trois unités documentaires : du texte directement saisi par le développeur, différents médias (images, sons, vidéos)¹⁸¹ et enfin des schémas. Ces derniers ne sont pas considérés comme de simples images, ni même comme une animation : ce sont des dessins techniques réalisés avec des outils spécifiques dans un format dédié qui font l'objet d'un traitement particulier dans l'espace de travail partagé du SI. Ils sont en effet stockés et gérés dans une classification séparée.

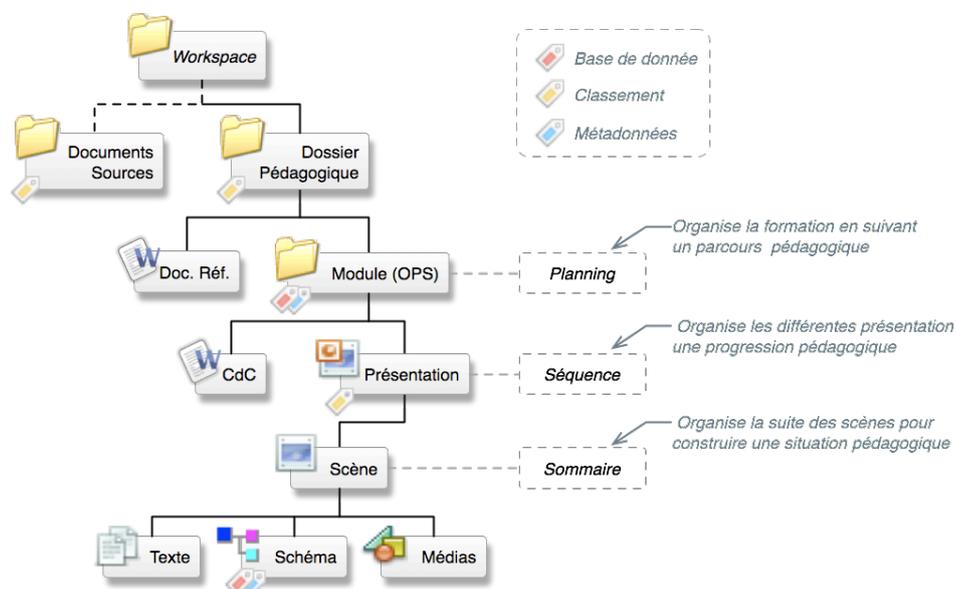


Figure 6.14 : Les principales « unités documentaires » manipulées par les développeurs lors de la création d'une formation et leur type d'indexation.

Toutes les « unités documentaires » n'ont pas le même statut au sein de l'environnement de travail (les diapositives ou les blocs de texte ne sont même pas des documents du point de vue du système de fi-

¹⁸⁰ Soit *Powerpoint*TM, soit *Emulsion*TM.

¹⁸¹ Il s'agit principalement de désigner ici la bibliothèque des objets, animés ou non, disponible dans *Emulsion*TM pour réaliser des contenus.

chier) et ne sont en particulier pas accessibles de la même manière. Dans le schéma ci-dessus (Figure 6.14) nous avons discrètement mentionné la manière dont ces unités étaient indexées : soit elles sont simplement classées dans un système de dossiers, parfois avec des métadonnées spécifiques, soit elles sont stockées dans une base de données spécifique (comme c'est le cas pour les modules).

L'ensemble des informations données ici est très simplifié. Il faut en effet compter sur le fait qu'il existe dans le système plusieurs copies et/ou plusieurs versions d'un même document dans l'environnement de travail partagé du service, et plus encore sur les machines de chacun de ses membres. Il faut en effet compter sur une double organisation des éléments l'une individuelle et l'autre partagée. Malgré les efforts d'amélioration de l'indexation des documents partagés (ajouts de métadonnées spécifiques notamment) l'environnement partagé et son organisation ne semble pas correspondre aux activités qui le sollicitent, et il arrive qu'une organisation « locale » soit produite par le développeur sur sa propre machine.

Il est intéressant de regarder de plus près la *structuration* de ces unités documentaires, pas tant sur le plan d'une relation de « composition » que sur le plan de l'organisation des « composantes » entre elles lors de leur création. En fait, l'activité de développement d'une formation va requérir un travail d'organisation des contenus qui va porter sur trois niveaux (au moins). Durant la création d'une formation le développeur va ainsi chercher à maintenir une pertinence pédagogique de son produit tant au niveau de la formation, que des modules ou des situations pédagogiques. Chaque niveau a un moyen d'exprimer cette organisation (Figure 6.14). Le *planning* de la formation est un produit de la *démarche pédagogique* du développeur : le temps consacré à chaque module selon le niveau, les pré-requis ou le moment de la journée sont autant d'éléments pris en compte et utilisés pour correspondre au style pédagogique choisi. De même au niveau d'un module (qui rappelons-le est censé être utilisable dans divers contextes) qui est pensé comme une *séquence* de situations permettant d'acquérir des connaissances et au niveau d'une scène qui doit amener une situation pédagogique particulière. Il peut s'agir par exemple de provoquer une situation de contradiction pour inviter les apprenants à réfléchir à une solution possible, ou une scène purement descriptive.

Nous n'allons pas nous étendre sur la description car ce qui est intéressant de notre point de vue c'est le fait que l'expression de la démarche pédagogique du développeur s'exprime à tous les niveaux et que lors de la construction d'une formation l'organisation de chaque niveau va avoir un impact sur les autres. Le choix d'une démarche pédagogique globale pour la totalité de la formation (par exemple de courtes séances axées sur la participation active des apprenants) va impacter l'emploi du temps bien sûr mais également l'organisation des séquences de présentation ainsi que le contenu desdites présentations. Inversement, la réutilisation d'un module existant peut influencer sur les choix pédagogiques généraux, voire orienter la création des modules manquants. De notre point de vue c'est dans cette organisation à plusieurs niveaux que seront mobilisées et actualisées les connaissances « situées » des développeurs.

Au-delà de leurs compétences techniques sur le sujet abordé, cette organisation demande un savoir-faire pédagogique et un savoir-faire spécifique à la réalisation d'une formation dans le contexte particulier qui est le leur. Qu'il s'agisse de la progression, du parcours ou des séquences pédagogiques, l'organisation est le fruit des connaissances *métier* (en tant qu'experts de terrain), des connaissances des contextuelles des *situations de formation* (en tant qu'anciens animateurs de formation), et des

connaissances strictement *pédagogiques* des développeurs pour adapter un contenu ou une situation, un rythme d'activité à un public particulier, son niveau, son expérience, sa propension à participer. Dans notre analyse de l'activité, cette organisation des documents à plusieurs niveaux nous a guidé pour déterminer ceux d'entre eux qui étaient utiles à une description de la réalisation de l'activité des développeurs. Nous avons également mobilisé d'autres moyens. Nous avons aussi par exemple cherché à identifier des « documents pour l'action » (Zacklad, 2005a) au rang desquels nous comptons le cahier des charges, la demande de développement des modules ainsi que les OPS eux-mêmes ainsi que le *dossier pédagogique* sur lequel vont intervenir (directement ou non) différents acteurs impliqués.

Dimension de l'outil

Aborder l'activité du point de vue de la *dimension de l'outil* consiste à décrire l'utilisation des outils numériques au sein du système observé. Dans le cas présent, il s'agit simplement de dresser la liste des *opérations* qui y sont réalisées du point de vue de l'utilisateur qui manipule des documents (il se peut donc qu'il n'y ait pas une stricte équivalence avec le modèle de conception des différents outils impliqués). Nous avons fondé notre analyse plutôt sur l'atelier *Emulsion* pour lequel nous avons dressé cette liste¹⁸². Les opérations vont donc concerner l'ensemble de l'interface de ce logiciel, à savoir la partie de l'interface consacrée à la *navigation* dans les différents types de contenus, ainsi que la partie réservée à l'*édition* des supports pédagogiques (les présentations correspondant à des OPS). Dans la figure ci-dessous (Figure 6.15), sont présentées les principales opérations en question.

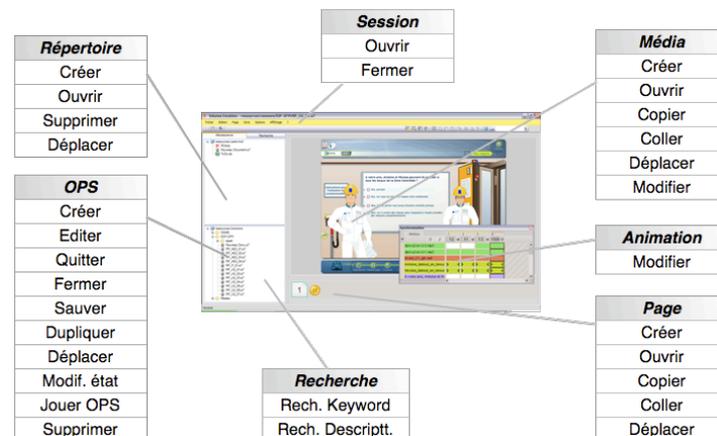


Figure 6.15 : Principales opérations possible depuis l'interface d'*Emulsion*TM.

6.3.2.3 Recoupement des descriptions

La dernière étape de cette première phase du processus de modélisation a pour objectif de faire émerger une liste « d'observés potentiels ». Un des moyens proposé au modélisateur pour y arriver est de recouper et synthétiser les descriptions de l'activité en s'aidant si besoin de la structure temporelle des interactions. Étant donné les natures des représentations en présence, il est difficile de proposer un formalisme de représentation permettant de les unifier. Il s'agit donc plus pragmatiquement de réunir, en une seule représentation le maximum des informations recueillies précédemment. Cette représentation doit également être abordable par les utilisateurs observés, ici les développeurs, afin de pouvoir discuter de la validité de l'ensemble de la description. Pour notre part nous avons réalisé quelques

¹⁸² Les opérations retenues sont également valables dans l'environnement initial.

représentations graphiques synthétiques, dont un aperçu est présenté ci-dessous (Figure 6.16). Grâce à cette schématisation, nous avons grossièrement représenté la trajectoire des différents documents manipulés par différents acteurs, sur différents supports, au cours de la fabrication d'un contenu et la conception d'une formation. Très concrètement, nous avons imprimé ce type de schéma sur grand format papier pour en soumettre le contenu aux membres de l'équipe d'ingénierie pédagogique. Nous avons ainsi pu valider et compléter une description précise de l'activité à tracer¹⁸³.

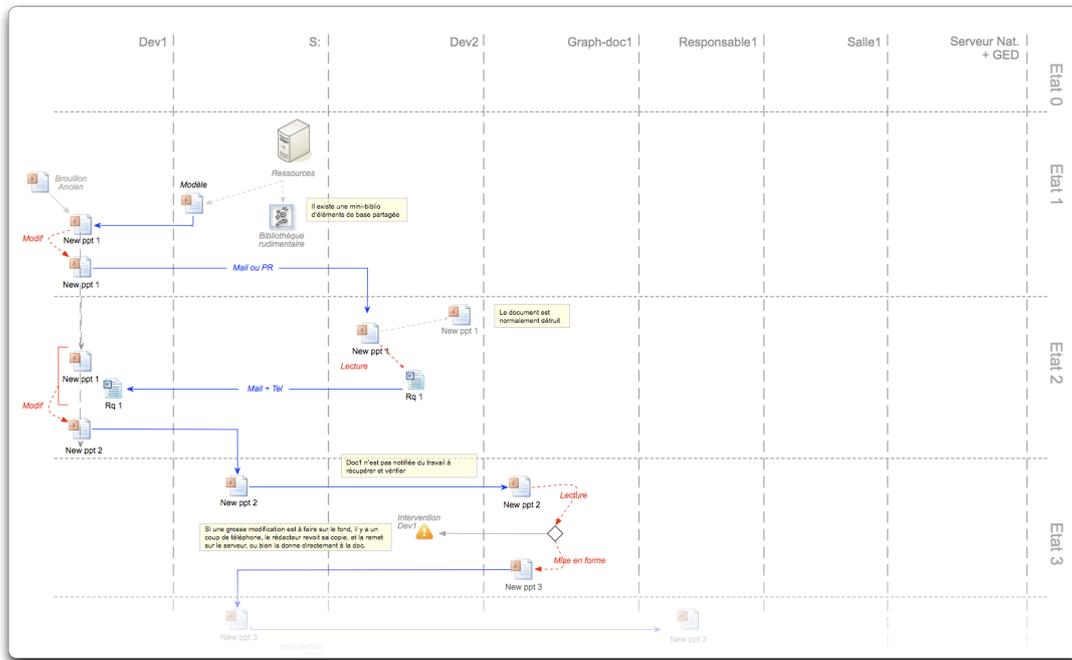


Figure 6.16 : Aperçu d'une description globale de la réalisation d'une activité de développement d'un module de formation (OPS).

Le second point concerne donc l'analyse de la *dimension temporelle* des éléments d'interaction mis en évidence dans les étapes précédentes. Rappelons qu'il s'agit d'analyser la « *structure temporelle* » des interactions des dimensions précédentes (principalement ici les unités documentaires de la dimension objet de l'activité). Étant donnée la variabilité des phénomènes observés, il n'est évidemment pas question de faire des mesures précises. Il s'agit plutôt de donner une règle générale, une tendance, une *approximation* de ces *caractéristiques* temporelles. La première caractéristique – sans doute triviale – est que l'ensemble des interactions est structuré autour des sessions de travail qui correspondent *grosso modo* aux journées de travail des développeurs. La seconde, pour ce qui concerne l'activité d'un développeur, est qu'elle est composée d'une double activité documentaire : la *conception de formations* et la *réalisation de supports pour les modules* (OPS) qui ne sont pas indépendantes nous l'avons vu mais ont des caractéristiques temporelles *différentes*.

Prenons par exemple la *conception de formation* (projet de développement global d'une formation). L'*amplitude* temporelle est relativement grande, bien que très variable (un projet peut-être mené en quelques jours au minimum, et peut courir sur plusieurs mois dans certains cas). La réalisation de

¹⁸³ Il est intéressant de noter qu'en marge du travail de modélisation, la discussion autour de ce support a permis une prise de conscience de la part des membres de l'équipe de leur propre activité selon un angle inhabituel. Ils y ont eux-mêmes pointé certaines faiblesses, notamment autour du processus de validation et ont décidé de sa simplification pour une implantation dans le nouvel environnement (*Emulsion*).

l'activité est en fait *morcelée* et *répartie* dans le temps¹⁸⁴, de sorte qu'il est quasi-impossible de réaliser cette activité d'un seul bloc, de manière *continue*, et les interruptions, plus ou moins longues sont fréquentes. Pour compenser cette discontinuité de la réalisation, plusieurs développements sont menés en parallèle. Cette réalisation est peu dense, c'est-à-dire comporte peu d'opérations effectuées en un espace de temps réduit. Si des modifications sont apportées à la formation, elles porteront sur l'organisation des modules (OPS), ou de l'emploi du temps, à un grain de précision relativement grossier, par rapport à la réalisation de l'autre activité de développement des contenus et supports. Au contraire, la réalisation concrète d'un module (OPS) implique plutôt de multiples opérations de modifications sur une courte période de temps de quelques minutes, donnant des interactions plus *denses*, une *amplitude* plus restreinte¹⁸⁵, et de fait une plus grande continuité.

Nos observations nous ont également conduit à remarquer la *fréquence*¹⁸⁶ relativement élevée de certaines interactions, notamment de longues navigations dans des dossiers organisant les contenus (ce qui correspond en fait à une *recherche* exploratoire), ou encore de « copier-coller ». Ce dernier point s'explique en partie par le fait que bon nombre d'objets (textuels notamment), intégrés aux présentations sont issus de documents techniques, normés, avec un vocabulaire précis. Pour s'assurer de l'intégrité du contenu, les développeurs usent donc largement de la possibilité de copier-coller des segments de texte, des images, des schémas, *etc.* Nous remarquerons au passage que ce type d'interaction a l'avantage d'identifier clairement un document source : si tout ou partie de son contenu a été « copié-collé » dans un autre document, il est techniquement facile de l'identifier comme tel.

6.3.3 Phase de création du modèle de trace

À ce point le processus de modélisation rentre dans sa seconde phase, celle qui doit aboutir à un *modèle de trace* prêt à être implémenté, en respectant les étapes suivantes : *sélectionner* les observés, *typer* les observés dans un modèle de trace et *consolider* ce modèle (voir section 5.3.2). Rappelons que dans le cadre de ce second terrain d'application nous souhaitons réaliser un *phase amorce* complète (voir section 5.5), c'est-à-dire créer un modèle de trace, l'implémenter et obtenir les premières traces modélisées générées par une activité réelle des utilisateurs du système observé.

Or, l'implémentation complète d'un SGBT n'était matériellement pas envisageable dans les délais imposés dans le cadre de notre thèse, la phase amorce que nous avons engagée ne pouvait atteindre son ultime étape, l'intégration du modèle de trace dans un SGBT fonctionnel produisant une trace modélisée de l'activité des développeurs. Nous souhaitons pourtant faire aboutir notre démarche en amenant notre travail de modélisation à son terme et avoir l'occasion de confronter un développeur à sa propre trace d'activité. Rappelons qu'en ce qui concerne ce terrain, le projet de modélisation était de créer des traces modélisées soutenant une réflexivité de l'activité d'un développeur, *i.e.* des traces donnant à l'utilisateur un *image interprétable* de sa propre activité.

Cette confrontation revêtait une importance cruciale à nos yeux, car seule la confrontation d'un développeur à une trace modélisée *présentée* de son activité réelle, permet d'évaluer *le potentiel* de ces

¹⁸⁴ En effet, la création d'une formation sous la responsabilité d'un chef-produit est dépendante de l'activité d'autres acteurs (demande de développements de modules), et de procédure (demande, validation)

¹⁸⁵ Peut ne prendre que quelques heures, mais aller jusqu'à plusieurs jours, plus rarement semaines.

¹⁸⁶ Entendue comme : occurrences relativement au temps et occurrences relativement au nombre total des interactions.

traces à soutenir une réflexivité de l'activité – évaluation que nous comptons bien mener, même s'il ne s'agissait que d'une première estimation. L'objectif que nous poursuivons en provoquant cette confrontation est de vérifier l'*interprétabilité* d'une trace *présentée en situation d'usage*, l'interprétation de la trace en question devant permettre à l'utilisateur lui-même de *recontextualiser* l'objet de son activité (les documents) et d'*objectiver* sa propre activité en y portant un regard critique, ce qui constitue pour nous, deux dimensions du caractère *réflexif* d'une activité.

6.3.3.1 Association du travail de modélisation à une expérimentation

Nous avons donc choisi de coupler la réalisation de la seconde phase du processus de modélisation à la mise en place d'une *expérimentation* visant à confronter des utilisateurs *directement* à une trace modélisée présentée dans leur environnement numérique de travail, en créant cette trace modélisée de manière totalement *ad-hoc*. Pour mener à bien ce projet nous avons donc du intégrer la réalisation des trois dernières étapes du processus de modélisation à la mise en place puis à l'exécution de l'*expérimentation* mise en place. C'est la raison pour laquelle, dans les sections qui vont suivre, la présentation des dernières étapes du processus de modélisation seront insérées dans la description de l'expérimentation que nous avons menée : l'étape de *sélection des observés* est ainsi associée au travail de la première phase de l'expérimentation (section 6.3.4), et les étapes de *typage* des observés et de *consolidation* du modèle sont liés au travail d'implémentation d'un outil de visualisation de trace utilisé dans la seconde phase de l'expérimentation (section 6.3.5). Avant cela nous allons immédiatement présenter l'expérimentation dont il est question.

6.3.3.2 Présentation de l'expérimentation

Compte tenu des caractéristiques de l'activité effective des développeurs et de notre objectif de leur procurer par les traces modélisées une image interprétable de leur propre activité nous avons choisi de mettre en place un protocole expérimental classique « d'interruption / reprise de l'activité. Il s'agit de provoquer une situation d'interruption relativement prolongée de l'activité en faisant l'hypothèse que lors de la reprise de celle-ci, une trace modélisée présentée constituera pour l'utilisateur un soutien, non seulement mnésique (re-contextualisation) mais métacognitif (objectivation). L'idée que les traces numériques peuvent jouer un rôle de facilitation à la reprise d'activité est un thème récurrent. Au sein même de l'équipe du LIRIS de récents travaux se sont penchés sur le sujet dans d'autres domaines d'application tout à fait différent (Richard et *al.*, 2009).

Une des raisons à avoir guidé notre choix pour ce protocole est que les interruptions plus ou moins prolongées de l'activité de développement des formations (quelle que soit sa durée propre) est un phénomène courant. Ces interruptions sont parfois gênantes car de l'aveu même des développeurs, la reprise d'une activité de développement d'une formation, qui nécessite une forte implication et dépend de multiples facteurs contextuels, est un exercice difficile. La difficulté réside en particulier dans le fait de devoir se réapproprier la « stratégie », les « directions pédagogiques » qui avaient été définies à l'origine, ou encore la « logique de l'organisation » du développement. En effet, la « stratégie pédagogique » choisie pour le développement d'une formation ne peut-être totalement fixée à l'avance. C'est au cours de la réalisation des contenus que le développeur détermine si oui, ou non, la stratégie présentée au départ, et éventuellement mentionnée dans le cahier des charges du projet, est adaptée. Or, ni la stratégie, ni l'organisation de l'activité ne sont réifiées dans l'environnement de travail des dévelop-

peurs. Une trace modélisée correctement définie pourrait donc être pertinente en devenant leur support d'inscription. Concrètement, l'expérimentation est composée de deux parties : l'une durant laquelle intervient l'interruption de la réalisation de l'activité en cours et l'autre correspondant à sa reprise. Nous allons détailler chacune de ces parties, ainsi que le travail consacré à la mise au point d'une trace modélisée visualisable dans les sections à suivre.

6.3.4 Expérimentation : première Partie

6.3.4.1 Préparation

L'interruption d'activité ne pouvait être opérée sur n'importe quelle activité. Celle-ci doit répondre à un certain nombre de caractéristiques ce qui nous a demandé un travail conséquent de préparation. Elle devait être *réaliste* quant à sa difficulté et sa thématique, d'une *durée* suffisante (ici deux jours), et se déroulant dans un environnement numérique de travail intégrant *Emulsion*TM. Nous avons opté pour une demande de formation réelle devant être traitée par l'équipe (plusieurs mois plus tard), que nous avons transformée de façon à obtenir *deux* demandes, chacune d'elle devant être confiée à un chef-produit parmi les développeurs du SI, et nécessitant l'assemblage d'au moins quatre modules (OPS). Ce choix tenait compte également du processus de modélisation en cours. En effet il s'agissait *aussi* de mettre en place la réalisation d'une activité « représentative » qu'il nous faudrait ensuite *schématiser* à la main sous forme de graphes pour y sélectionner les *observés* (voir section 5.3.2).

Pour couvrir au maximum les situations rencontrées habituellement par les chef-produits, nous avons fait en sorte que certains OPS pré-existent et soient *réutilisables*, alors que d'autres restaient à développer (ou faire développer)¹⁸⁷. Nous avons pour l'occasion rédigé l'ensemble des documents de travail nécessaires (cahier des charges, formulaire de demande de développement d'OPS, modèle de dossier pédagogique, *etc.*) de façon à ce qu'ils correspondent aux documents habituels tout en tenant compte des particularités de l'utilisation de l'atelier *Emulsion*. A ce sujet, nous avons pour l'occasion également dû déployer sur place avec l'assistance du concepteur une version d'*Emulsion* modifiée pour permettre la conservation de l'ensemble de ses *logs* serveur, ce qui a nécessité un certain nombre de tests¹⁸⁸ ainsi que la réindexation de tous les médias et contenus existants dans la base de l'application précédemment installée sur place.

6.3.4.2 Déroulement

L'activité préparée a été réalisée¹⁸⁹ sur *deux jours* par *deux développeurs* de l'équipe, que nous appellerons PL et MC, chacun d'eux ayant reçu pour mission la réalisation d'une formation. L'ensemble de la réalisation s'est déroulée dans l'environnement de travail habituel : chacun a travaillé sur sa propre machine, dans son propre bureau, et pouvait si besoin utiliser toutes les ressources à sa disposition, y compris non numériques. Ainsi les développeurs pouvaient comme à leur habitude organiser des discussions en face à face ou bien se téléphoner. Chaque développeur a donc joué le rôle de « chef-produit » pour une formation donnée, tout en restant à la disposition d'un autre projet de développement en tant que développeur au sens strict. PL et MC se sont rencontrés trois fois durant ces deux

¹⁸⁷ Le développement de certains de ces modules devait se faire avec *PowerPoint*TM, d'autres avec *Emulsion*TM.

¹⁸⁸ La version en question a d'abord été déployée au laboratoire et testé pendant plusieurs jours pour s'assurer de son bon fonctionnement.

¹⁸⁹ En partie seulement puisque le but était de les interrompre avant qu'ils n'aient pu finir.

jours pour se mettre d'accord sur une *stratégie pédagogique* à adopter, sur les deux formations (l'un étant chef-produit, l'autre développeur pour chacune d'elles). Le problème principal relatif à la stratégie était le *découpage* des objectifs généraux de la formation (OPG) en modules (OPS) de formations indépendants.

Le premier découpage, qui n'avait tenu compte *que* de la *demande*, a été remis en cause par le début du développement concret des supports. MC en tant que chef-produit a par exemple ainsi revu la conception globale de sa formation pour intégrer un contenu proposé par PL, ce qui donnait une toute autre orientation pédagogique que celle prévue au départ. Pour résumer, il s'agissait d'une formation destinée à des experts de terrain. Logiquement, la formation avait initialement été prévue pour éviter la présentation d'informations de base concernant le métier. Or c'est précisément ce que décida de faire PL, la démarche pédagogique n'étant plus alors de soumettre des informations aux apprenants mais de les faire réagir face à des informations simplistes pour qu'ils dégagent d'eux-mêmes les informations importantes. Nous ne pouvons pas, pour des raisons de confidentialité rentrer dans le contenu des documents en jeu. On notera au passage combien il peut être difficile de capitaliser, simplement à travers la conservation de contenus, de telles stratégies pédagogiques. Sur un plan plus documentaire, ce changement s'est traduit dans l'espace numérique de MC par une série de modifications importantes touchant à tous les niveaux d'organisation : dossier pédagogique, OPS, présentations.

6.3.4.3 Enregistrements

Les activités de PL et MC, en plus de notre observation sur place, ont bien entendu été enregistrées. En l'occurrence, le dispositif mis en place était le suivant (Figure 6.17) :

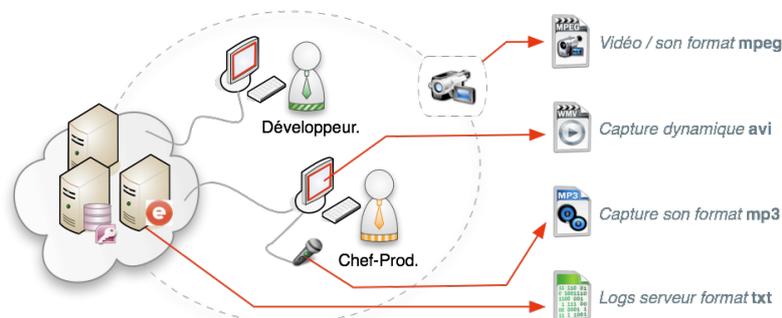


Figure 6.17 : Données enregistrées lors de la réalisation de l'activité et leurs formats.

- Audio-vidéo : nous avons filmé de manière continue la réalisation de l'activité à l'aide d'une caméra numérique. L'enregistrement a pris pour sujet l'un des deux développeurs lorsqu'ils travaillaient séparément, les deux lorsqu'ils travaillaient dans le même bureau.
- Capture d'écran et son : leurs machines de travail étaient équipées d'un petit logiciel de capture dynamique d'écran, qui bien que peu stable a permis d'enregistrer toutes les opérations effectuées sur les machines, ainsi que le son enregistré à partir d'un micro dont nous avons équipé

chaque machine¹⁹⁰. Cet enregistrement sonore permettait de faciliter la synchronisation entre la vidéo et la capture d'écran.

- *Logs* : la version d'*Emulsion* déployée spécialement nous a permis de récupérer également un ensemble de *logs* serveur.

6.3.4.4 Interruption et organisation d'un exercice de relève

À la fin des deux jours, nous avons comme prévu *interrompu* le travail de développement, *sans* bien sûr mentionner le fait que celui-ci devrait être repris prochainement. Avant de clore complètement la session de travail, nous avons organisé un exercice supplémentaire. Nous souhaitions avoir une idée des éléments qui étaient, du point de vue d'un développeur, essentiels à la reprise d'un projet de développement et nécessaires à sa description. Nous avons donc organisé une simulation de relève, *i.e.* une *transmission de projet*. Il s'agissait d'annoncer à PL que la responsabilité du projet en cours lui été retirée, et qu'il fallait en transmettre les éléments à un troisième développeur (MM). Une rencontre en face à face a donc été provoquée, dans le bureau de PL, devant sa machine. Cette séance, d'une durée d'environ une heure a été enregistrée (audio-vidéo et capture d'écran), et retranscrite par la suite.

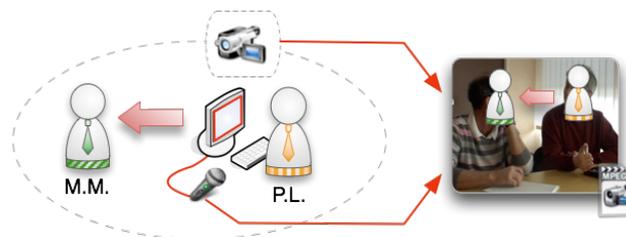


Figure 6.18 : Organisation et enregistrement d'un exercice de relève.

6.3.5 Création de la trace modélisée

L'intervalle de temps correspondant à l'interruption de l'activité a été mis à profit pour créer une trace modélisée, ce qui consistait à réaliser les opérations du processus de modélisation tout en préparant la visualisation de la trace modélisée. Nous avons donc suivi les étapes du processus, en commençant par celle de *sélection des observés*. Pour cela nous avons repris des extraits de la réalisation de l'activité ayant eu lieu pendant les deux jours de la première partie de l'expérimentation, pour les schématiser sous forme de graphes. Puis nous avons comparé ces différents extraits pour rassembler les observés qui étaient nécessaires à une description minimale commune. Rappelons que cette étape consiste en fait à produire à la main une trace modélisée « simulée », qui doit préfigurer ce que sera la trace modélisée *in fine*.

Concernant les deux dernières étapes du processus, à savoir le *typage des observés* et la *consolidation du modèle de trace* constitué, elles ont été directement influencées par l'outil de visualisation que nous avons dû utiliser (cf. *infra*). Pour le typage par exemple, nous avons dû passer d'une schématisation qui portait sur des observés (c'est-à-dire les nœuds du graphe en question) qui étaient indifféremment des entités et des évènements, à des observés uniquement exprimés en évènements, ceci pour respecter

¹⁹⁰ Le petit logiciel en question est *Camstudio 2.0* (<http://camstudio.org/>), libre et gratuit il possède à peu près les mêmes fonctionnalités beaucoup du plus connu *Camtasia* (<http://www.techsmith.fr/camtasia.asp>). Le manque de stabilité évoqué désigne des plantages qui nous ont coûté la perte des données concernant plusieurs minutes du travail d'un des développeurs.

le format imposé par l’outil de visualisation. De même, une partie des attributs des types d’observés a été prévue pour correspondre à la structure imposée par l’outil de visualisation. Au final, nous avons créé un *modèle de trace* qui tient compte des contraintes imposées par l’outil de visualisation, lui-même nécessaire pour la seconde partie de l’expérimentation. Les choix que nous avons dû faire concernant ces contraintes apparaissent dans le traitement que nous avons du mettre en place pour obtenir une trace modélisée présentée, et que nous allons maintenant détailler.

6.3.5.1 Chaîne de transformation mise en place

Pour créer directement une trace modélisée présentable à l’utilisateur, nous avons fait appel à un petit outil appelé *Timeline*¹⁹¹ : il s’agit d’une application *javascript* qui permet de visualiser un ensemble d’événements temporalisés sur une frise dans laquelle on peut naviguer interactivement de manière simple (description détaillée ci-après). L’application affiche des *événements* qui lui sont fournis sous la forme d’un fichier de données au format *XML*¹⁹² avec une structure spécifique : un événement est composé de onze champs tous au même niveau, correspondant à ses attributs. Pour créer un tel fichier à partir des enregistrements effectués sur place, nous avons exploité des outils existants et *créé nous-mêmes les outils nécessaires* pour former une sorte de « chaîne de production manuelle assistée » d’une trace modélisée de l’activité de nos développeurs du SF. Nous pouvons la résumer de la manière suivante (Figure 6.19).

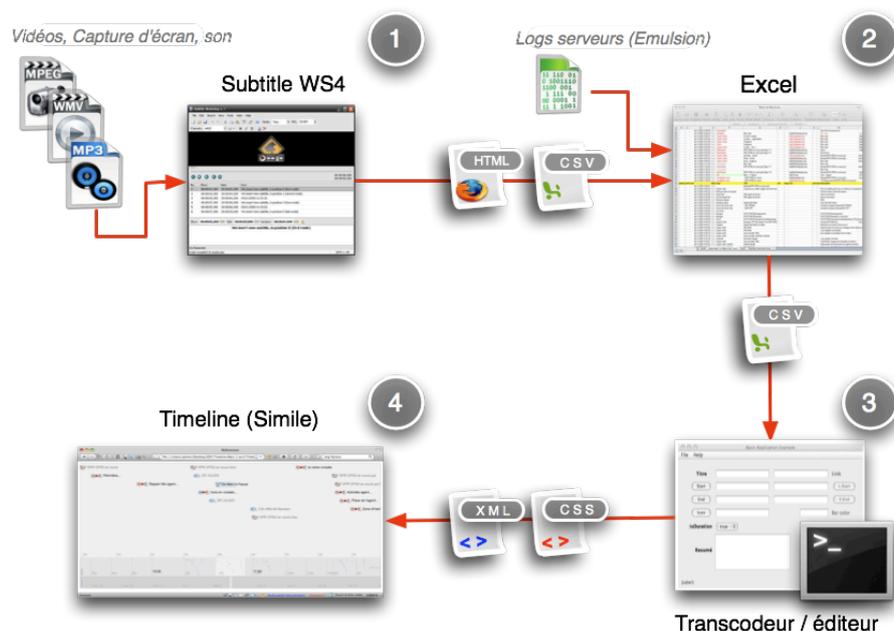


Figure 6.19 : Étapes et outils nécessaires à la présentation d’une trace modélisée en *Timeline*.

(1) Nous avons visualisé de façon synchronisée les enregistrements audio-vidéo et les captures dynamiques d’écran et nous avons utilisé un outil de sous-titrage vidéo (*Subtitle Workshop 4*)¹⁹³ pour « instancier » à la main un certain nombre d’observés¹⁹⁴.

¹⁹¹ Application proposée par *SIMILE*, pour plus d’informations voir <http://www.simile-widgets.org/timeline/> (consulté le 02/08/2009).

¹⁹² *Extensible Markup Language*, langage à balise permettant de constituer des fichiers de données structurées (arbre) et donnant la possibilité d’extension, c’est-à-dire de définir de nouvelles balises.

¹⁹³ Voir le site officiel sur <http://www.urusoft.net/> (consulté le 20/07/2009).

(2) L'outil de sous-titrage permet d'exporter l'ensemble du fichier de sous-titre dans divers formats, dont le format *html*, que nous avons utilisé et que nous avons nettoyé pour en obtenir un fichier au format *csv*¹⁹⁵, pris en charge par *Microsoft Excel*TM. Grâce au tableur nous avons pu intégrer les données issues des *logs* du serveur *Emulsion*, après les avoir eux-mêmes convertis au format *csv*. Nous avons dû, au passage, nous assurer de la fiabilité des *logs* en croisant, sur certaines portions des enregistrements, les données des *logs* et celles obtenues grâce au « sous-titrage manuel ». Les données issues des *logs* ont donc été ajoutées, puis l'ensemble des données a été *trié* temporellement. Pour correspondre au modèle de données imposé par la *Timeline*, nous n'avons conservé que des événements (chaque ligne du fichier correspond à un événement, portant sur un ou plusieurs objets, *i.e.* document ou partie de document)¹⁹⁶.

(3) Nous avons du créer une petite application, un *transcodeur*, permettant de passer du format *csv* nettoyé et trié dans le tableur à un format *XML* qui respecte la structure de donnée exigée pour l'affichage dans la *Timeline*. Nous avons également codé une autre petite application, un *éditeur d'événement*, qui prend en entrée le fichier de données de la *Timeline* et propose une interface (Figure 6.20) pour ajouter facilement un événement, directement dans le fichier *XML*. Ce petit outil nous a été utile pour compléter la trace avec certains observés manquants ou qui n'étaient pas apparus à cause d'un problème technique¹⁹⁷.

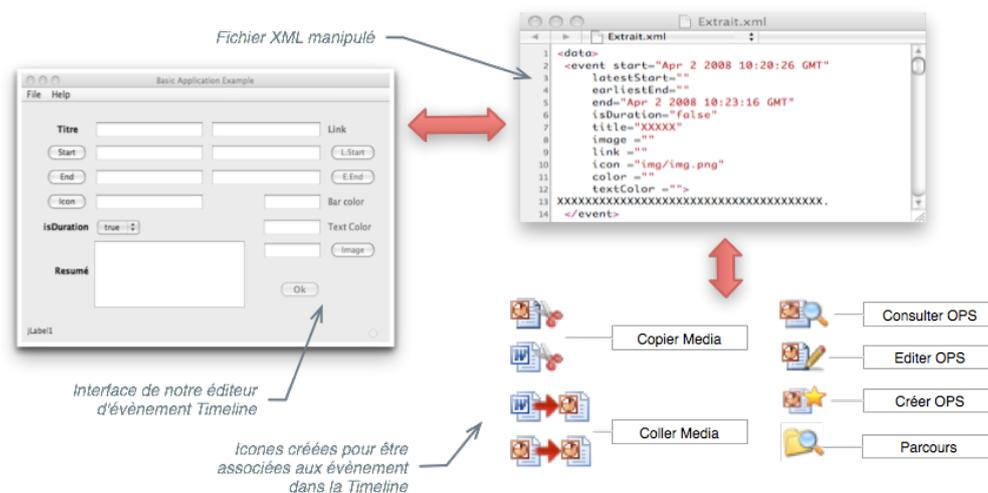


Figure 6.20 : Interface de l'éditeur d'événements et exemples d'icônes créées spécialement.

(4) Nous avons également paramétré la *Timeline* en modifiant le fichier de configuration afin « d'optimiser » l'affichage selon nos besoins. Il s'agissait entre autres choses, de définir le nombre et la taille des frises (trois en l'occurrence) ainsi que les options d'affichage, de déterminer un niveau de zoom sur différentes périodes de temps, de créer un *jeu d'icônes* symbolisant les observés (des événements ici) pour remplacer les icônes utilisées par défaut, *etc.* (Figure 6.20). Une fois les fichiers de données et de paramétrage en place, la *Timeline* a été mise en accès sur le réseau (incorporée dans une page *HTML*) pour les développeurs.

¹⁹⁴ Pour faciliter la suite du traitement, chaque observé était noté en respectant une mise en forme particulière : date0/date1/titre/contenu/.

¹⁹⁵ *Comma-separated values* (CSV) est un format ouvert, dédié à des données tabulaires sous forme de valeurs séparées par des virgules (ou point-virgule). Ce schéma est pris en charge par les tableurs tels que *Microsoft Excel*TM.

¹⁹⁶ Nous n'avons conservé que des éléments d'interaction dont nous avons vérifié la possible collecte automatique.

¹⁹⁷ Notamment pour les *logs* du serveur d'Emulsion.

6.3.6 Expérimentation : seconde partie

Après quelques semaines d'interruption, une autre séance de travail a été organisée au sein du service par nos soins. Contrairement à ce qui leur avait été annoncé, il a été demandé à PL et MC (les développeurs concernés) de reprendre le projet entamé lors de la première partie de l'expérimentation. Aucun d'eux n'avait repris ce projet entre temps. Les consignes étaient simples, il fallait reprendre le dossier, relancer le développement des modules (OPS) et terminer la conception de sa formation pour chacun. Comme nous l'avons déjà mentionné, la difficulté de l'exercice consiste non seulement à reprendre connaissance des divers documents alors produits et/ou utilisés, mais il faut surtout se *réapproprier la démarche pédagogique qui avait alors été appliquée*. Le même exercice a été demandé à l'un et l'autre des développeurs, indépendamment.

6.3.6.1 Déroulement

Chacun disposait de son environnement de travail habituel, et était équipé du même dispositif d'enregistrement que lors de la première partie de l'expérimentation (capture dynamique d'écran, caméra numérique, et enregistrement audio). De plus, il était demandé aux participants de produire à haute voix une explication de ce qu'ils faisaient, et de justifier leurs actions le cas échéant. Seul l'un des deux développeurs (MC) avait accès à une trace de son activité passée sous la forme d'une *Timeline* intégrée à son environnement de travail (en ligne, réseau local) et que nous pouvons présenter un peu plus en détail.

6.3.6.2 Détails sur la visualisation

La trace modélisée finalement présentée au développeur était telle que le montre la figure ci-dessous (Figure 6.21)¹⁹⁸. La *Timeline* finale se compose de trois frises, représentant trois échelles de temps (minutes, heures et jours). Cette combinaison a été choisie pour correspondre aux mieux à l'amplitude temporelle de l'activité tracée (voir section 6.3.2.2). Pour améliorer la lisibilité, certaines zones ont été « temporellement dilatées » : la *Timeline* permet en effet de modifier l'échelle temporelle entre deux points (fonctionnalité « *Hotzone* »), la répartition des éléments sur la bande restant automatique. Comme le montre l'extrait ci-dessous, chaque élément est présenté par un *titre* et une *icône*. Un clic sur un élément fait apparaître une bulle contenant les informations détaillées sur l'événement et les objets impliqués. Le contenu qui y figure correspond au champ *description* du format d'un événement dans une *Timeline*. Ce champ est rendu sous forme de *html* interprété, ce qui autorise la présence de liens hypertextes. Nous avons utilisé cette propriété pour donner la possibilité d'accéder aux documents directement depuis la *Timeline*, et avons structuré ce contenu pour y faire figurer certains attributs. Dans l'exemple donné dans l'extrait ci-dessous, on peut apercevoir ainsi que pour un événement « coller média », la description donne le lien vers le document source (ici un document *Word*) et un autre vers le document cible (ici une présentation)¹⁹⁹.

¹⁹⁸ Pour des raisons de confidentialité certaines parties du texte ont été masquées par des « XXX ».

¹⁹⁹ La réalisation entièrement *ad-hoc* impose que les documents doivent ne pas avoir été déplacés depuis l'enregistrement effectué durant la première partie de l'expérimentation, ce qui par chance n'était pas le cas sur la machine de MC. Il s'agit bien évidemment de la dernière version enregistrée du document, mais à terme grâce au système de *versionning* d'Emulsion on pourrait revenir à des versions à date.



Figure 6.21 : Extrait de la *Timeline* proposée à MC comme aide à la reprise d'activité.

Nous avons bien entendu, en amont, réservé un moment pour présenter à MC la manipulation de la *Timeline* : fonctionnement du défilement, principe des bulles d'informations, *etc.* Il est à noter que la manipulation elle-même n'a pas engendré de difficultés et l'aide du modélisateur durant la session de reprise n'a été demandée à aucun moment.

6.3.6.3 Constatations générales et rapide bilan

Voici en quelques mots les constatations que nous avons pu faire relativement aux observations et enregistrements réalisés²⁰⁰. La plus évidente d'entre elles est qu'*aucun des deux développeurs n'a été en mesure de reprendre seul son activité* et il a fallu une rencontre en face à face de PL et MC pour que tous les éléments soient remémorés, les documents et acteurs identifiés, et que le développement puisse reprendre son cours. Pour être un peu plus précis, chaque développeur a passé environ une heure avec son environnement de travail pour tenter de reprendre l'activité. L'un et l'autre n'ont pas rencontré les mêmes difficultés et nous pouvons résumer comme suit.

Premier développeur / chef-produit

Pour PL : de mémoire, le développeur retrouve à peu près le nom et le rôle des acteurs impliqués par le projet de développement et même le nom de code de ce dernier. Il se tourne ensuite vers les documents concernés par le développement en question qu'il tente de retrouver dans son espace de travail. Dans l'impossibilité de les retrouver en y navigant directement, il cherche à retrouver les *emails* (seuls éléments temporalisés) qui avaient été échangés alors en espérant y trouver en pièce jointe lesdits documents. Cette tentative ne donne pas de résultat immédiat : en effet les emails retrouvés sont antérieurs à la première partie de l'expérimentation et portent sur des documents qui ne sont pas liés avec le développement en question, ce qui laisse croire à PL qu'il a tout simplement perdu les documents

²⁰⁰ Nous attirons l'attention du lecteur sur le fait que le traitement de ces enregistrements pourrait être très largement approfondi par un analyse plus minutieuse et avec l'aide de spécialistes de l'analyse interactionnelle, ce que nous n'avons pas eu moyen de faire dans le cadre temporelle de cette thèse mais qui devrait être réalisé sous peu.

recherchés. Après plusieurs tentatives, il découvre l'erreur en trouvant une fiche de demande de développement d'OPS que MC lui avait soumis, mais n'est pas capable de reconstituer l'ensemble des éléments nécessaires à une reprise effective, et en particulier les choix pédagogiques qu'il avait lui-même proposés.

Second développeur / chef-produit

Pour MC, les difficultés rencontrées ont été sensiblement différentes mais le résultat pas plus positif du point de vue de l'exercice demandé. Tout d'abord, MC rencontre des difficultés pour se souvenir des noms des acteurs impliqués, et du rôle qu'il tenait lui-même dans le développement. Puisqu'il en dispose, il commence à parcourir la trace modélisée. Il la commente rapidement en y repérant très vite des éléments clefs, tel que le cahier des charges, l'ouverture d'un modèle de présentation (qu'il utilise à chaque fois), suivi d'une demande de développement d'OPS, *etc.*, séquence qu'il décrit, après avoir cité les grandes étapes à travers ces documents, de la manière suivante « ça c'est classique ». Il continue à parcourir la trace modélisée pour aboutir à une situation qu'il n'est pas capable d'expliquer : il constate qu'il a envoyé deux fois un même module à quelques heures d'intervalle alors qu'il a entre temps entamé le développement d'un autre module. Malgré l'ouverture de différents fichiers il ne trouvera pas l'explication jusqu'à en discuter directement avec PL (en fait cette particularité venait du changement de stratégie pédagogique qui avait eu lieu en cours de développement).

Justement, durant cette courte entrevue, il est manifeste que l'échange a permis aux deux développeurs de recontextualiser très vite le développement entamé, y compris la stratégie pédagogique adoptée, ce qu'ils n'avaient réussi à faire seuls, y compris avec la trace modélisée. Nous ne chercherons pas ici à trouver une élégante explication théorique à cette situation (même si une argumentation socio-constructiviste paraît tout indiquée). Signalons simplement ce que les développeurs ont eux-mêmes identifié comme l'élément clef qui a servi « d'amorce mnésique » dans la conversation. Et la réponse est très intéressante même si ce type d'observation reste anecdotique : il s'agissait du mot « disséquer ». Ce mot a été utilisé la première fois comme nouveau titre à une partie du développement qui fut modifiée par le changement de stratégie pédagogique, il a donc été largement évoqué lors de la première phase de l'expérimentation entre le développeur et son chef-produit. Il semble *a priori* logique que ce mot ait revêtu une importance particulière. Le mot lui-même était parfaitement visible sur la trace présentée, il faisait partie du texte d'un événement « copier-coller » dont l'extrait accompagnait l'icône dédiée, et nous pouvons affirmer que MC a vu ce mot dans la *Timeline*, d'abord parce qu'il a lu le mot à haute voix, et ensuite parce qu'il a ouvert l'élément de la *Timeline* correspondant au copier-coller en question. Pourtant dans ce cas là, le terme « disséquer » ne semble pas avoir eu le même effet déclencheur.

Plusieurs lectures pour des résultats contrastés

Une lecture superficielle de ces constatations pourrait conclure à un échec pur et simple de la trace modélisée à soutenir la reprise d'activité, invalidant l'hypothèse faite au départ de cette petite expérimentation. Il est cependant risqué de tirer une conclusion définitive sur cette unique observation. De plus, les raisons invocables pour expliquer la « contre-performance » de la trace modélisée sous forme de *Timeline* à soutenir la reprise de l'activité seraient aussi diverses que nombreuses. Sur la trace présentée elle-même tout d'abord : la présentation simplifiée de la *Timeline* qui ne propose qu'une mise

en relation temporelle d'actions (sans autres relations visibles) et/ou s'éloigne trop du type de représentation riche que nous voulons mettre en place avec les SBT. Sur les conditions ou le protocole expérimental ensuite : la première partie de l'expérimentation était trop courte, les différentes pièces à produire trop simplifiées, le temps d'interruption trop long, l'utilisation d'*Emulsion* pas assez maîtrisée.

Nous pouvons cependant faire une lecture plus fine des résultats observés. Nous remarquons qu'effectivement la trace présentée n'a pas permis de faire mieux qu'un environnement classique, mais nous notons également que l'exercice de reprise était suffisamment difficile pour qu'aucun des deux développeurs ne retrouve *seul* la démarche pédagogique globale et l'organisation du développement qui donnaient une cohérence aux documents retrouvés. Mais un autre niveau de lecture est possible, moins général mais non moins intéressant. Il s'agit de regarder plus en détail ce qui s'est passé autour de l'utilisation de la *Timeline*, ou plutôt de ce qui ne s'est *pas* passé.

Bien que cela soit sa première confrontation à ce type d'interface, MC n'a pas rencontré de problème majeur quant à la *manipulation* de la *Timeline*. Il n'a pas demandé d'aide pour l'interpréter, il n'a pas remis en cause le fait que cela soit bien son activité (il fait même confiance à la trace et se demande pourquoi il a envoyé deux fois le même fichier). Il ne fait appel au reste de l'environnement que pour aller consulter des documents. Nous constatons que le modèle de trace sous-jacent est relativement pertinent : les observés sélectionnés ont permis à MC de retrouver rapidement les éléments documentaires du projet de développement ; la mise en avant des « copier-coller » semble pertinente notamment parce qu'elle a permis de faire apparaître dans la trace des éléments importants, qui auraient pu être le soutien mnésique que l'on attendait, c'est le cas du terme « disséquer » qui a servi de « déclic » à MC et qui était bien présent dans sa trace.

Mais plus important encore, lors de la confrontation de MC à sa trace, celle-ci lui a permis d'*objectiver* son activité. Ses commentaires montrent en effet qu'il y a reconnu une *façon de faire* « classique » qui consiste d'une part à partir de cahier des charges, et de copier-coller tous les éléments dans une première présentation, de les structurer, puis de les répartir dans autant de présentations (type *Powerpoint*TM). Dans le prolongement, la trace lui a permis d'identifier ce qui apparaissait comme une anomalie dans cette façon de faire, à savoir la production à quelques minutes d'intervalle du même module (OPS).

Certes, il ne parvient pas à se remémorer pourquoi *mais* remarquons que l'explication qui aurait justifié cette « anomalie » aurait permis de retrouver la stratégie pédagogique alors mise en place, parce que précisément, l'explication en question *est le changement de stratégie décidée* d'un commun accord entre les développeurs lors de la conception de la formation. Ce qui montre que même avec une présentation interactive très simple (la *Timeline*), la trace modélisée que nous avons créé *avait le potentiel* d'être non seulement un support *mnésique*, un « index temporalisé des documents utilisés » si on peut dire, mais aussi et surtout le support de processus *métacognitif*, de prise de conscience de sa propre activité, d'interprétation critique de celle-ci, un support se prêtant à un niveau d'interprétation qui va au-delà de la simple manipulation des documents dans l'environnement. En ce sens, les constatations que nous avons pu faire lors de cette expérimentation n'invitent pas forcément à conclure à un échec de l'exercice.

Pour revenir à une remarque plus générale sur ce qui s'est passé lors de cette expérimentation nous ne manquons évidemment pas de remarquer que les deux développeurs ont pu *reconstruire ensembles* une situation de reprise, ce qu'ils n'avaient pas réussi à faire seuls. Plus frappant encore est le fait que durant sa tentative de reprise dans son environnement de travail PL avait quasiment annoncé que ce serait le cas : « je peux retrouver avec MC... à deux on peut retrouver ». Cette dernière remarque nous invite à revenir sur la *dimension du collectif* dans les traces modélisées, et tout particulièrement en ce qui concerne *l'usage collectif* de ces traces : que serait-il advenu si PL et MC avaient eu accès non seulement à leur trace, mais à celle de l'autre ? Cette présentation aurait-elle permis ce que l'échange verbal a produit ?

En fait, un nombre important de questions émergent du simple fait de considérer les traces modélisées dans des situations d'activités – et donc d'usages potentiels – collectives. Il était bien entendu évident, dès le départ de notre travail sur le terrain qu'il était indispensable de tenir compte de la dimension collective voire collaborative des activités observées. Et nous en avons tenu compte tout au long de la construction de notre processus de modélisation. En revanche, la confrontation de la trace modélisée à une situation d'usage concrètement collective ouvre une nouvelle perspective et un nouveau regard sur les traces modélisées que nous discuterons dans le prochain et dernier chapitre de ce manuscrit. Il nous faut avant cela conclure ce chapitre consacré à notre travail de recherche appliquée d'une *méthodologie de modélisation* des traces numériques dédiée au SBTm.

6.4 Conclusion

Nous avons présenté dans ce chapitre la partie empirique et pratique de notre travail de recherche sur l'établissement d'une méthodologie de modélisation de trace numérique. Nous avons décrit les deux terrains d'application pour lesquels nous avons concrètement conduit une modélisation de trace tout en façonnant une méthodologie pour cette démarche.

Le *premier terrain* d'application a concerné une classe virtuelle, un environnement numérique soutenant une activité d'apprentissage collaboratif du français aux Etats-Unis : *eLycée*. Nous y avons engagé un travail de modélisation de trace en intégrant un projet existant d'implémentation d'un SBTm lui-même développé comme partie intégrante de la plateforme proposée (*eMediathèque*). Le projet de modélisation initial étant tourné vers la productions de traces modélisées soutenant une activité réflexive des apprenants, les observations et analyse menées ont conduit à la création d'un modèle de trace qui a surtout été orienté par la mise en œuvre concrète d'un module de visualisation de ces traces. Concernant la conception d'une méthodologie, les difficultés rencontrées durant ce premier travail de modélisation, qu'ils relèvent du contexte de modélisation ou des faiblesses de notre démarche, nous ont conduit à faire progresser l'ébauche de méthodologie que nous mettions en place vers une démarche *rigoureuse et structurante*.

Le *second terrain* d'application présenté a impliqué une activité de création de contenus de formations par une équipe d'ingénierie pédagogique travaillant avec un atelier de production multimédia. Cette équipe crée des formations pour les collaborateurs du groupe et fabrique les contenus et supports pédagogiques nécessaires pour que la formation soit ensuite dispensée sur site. Notre projet de modélisa-

tion sur ce terrain visait l'implantation d'un SBTm, lié à l'atelier de production de contenus multimédia, délivrant des traces modélisées susceptibles d'être un support à l'activité des *développeurs* de formation et de contenus pédagogiques, en conférant une dimension réflexive à leur activité. Concernant la méthodologie de modélisation que nous mettons en place, nous avons pu mener à bien la réalisation complète d'une *phase d'amorçage* de l'implantation d'un SGBT (SBTm concret) ce qui nous a permis d'affiner les étapes de notre processus de modélisation de trace.

De plus, nous avons associé au travail de modélisation sur ce second terrain d'application la réalisation d'une *expérimentation* visant à confronter concrètement des utilisateurs à une trace modélisée de leur activité au sein de leur environnement de travail habituel. Nous avons choisi un protocole d'interruption / reprise d'activité, l'idée étant d'évaluer le potentiel d'une trace modélisée à soutenir concrètement une certaine réflexivité de l'activité. Techniquement, nous avons adapté un outil existant (*Timeline*) pour construire une trace modélisée utilisable dans le cadre de l'expérimentation. Les résultats de cette expérimentation tendent à montrer que les traces modélisées ont bien le *potentiel* pour être un véritable support de réflexivité de l'activité mais que ce potentiel ne pourra s'exprimer que le cadre d'usages qui restent encore à développer.

Bien que ce travail comporte un certain nombre de limites et mérite certainement d'être approfondi sur plusieurs points, nous avons atteint l'objectif que nous nous étions fixé dans la mesure où nous avons créé, proposé et appliqué une *première méthodologie complète* de modélisation de trace. Nous complétons ainsi l'approche des Systèmes à Base de Traces modélisées qui dispose maintenant d'une méthodologie de modélisation propre.

Le travail empirique, effectué sur les deux terrains d'application que nous venons de décrire, nous invite pourtant déjà à dépasser le cadre de la réflexion menée jusqu'ici. En effet, les activités documentaires instrumentées ont toujours, à des degrés divers, une dimension *collective*, et comme nous l'avons déjà souligné elles impliquent de manière plus ou moins forte, plusieurs acteurs qui échangent, partagent, créent des documents qui leur permettent d'agir ensemble. C'est le cas des activités pédagogiques d'*eLycée* (où l'apprentissage passe par la production conjointe de documents) comme de l'activité de production de contenus (où la création concrète des modules et des formations engage plusieurs membres de l'équipe d'ingénierie pédagogique).

Or dans les deux cas, nous avons fini par envisager un usage collectif des traces : dans le premier cas la possibilité que le tuteur et l'apprenant reviennent ensemble sur l'interprétation de la trace de ce dernier, dans le second la possibilité que les deux développeurs aient pu être confrontés à leurs traces ensemble, puisque c'est ensemble qu'ils ont su reconstruire les connaissances nécessaires pour agir, *i.e.* reprendre leur travail de développement. Cela nous conduit à poser le problème, ou pour le moment la question, du concept de « trace collective ». Nous avons en effet jusqu'ici pris en compte la dimension collective des activités tracées dans notre modélisation en l'intégrant à un point de vue d'un utilisateur unique, en respectant ainsi le cadre conceptuel de départ imposé par l'approche des Systèmes à Base de Traces modélisées. La complexité des usages potentiels des traces modélisées en situation d'activité collective nous invite toutefois à réinterroger ce cadre conceptuel initial, voire à le dépasser.

Chapitre 7

Traces d'activité conjointes

Résumé du chapitre

Dans ce chapitre nous discutons l'approche que nous avons adoptée en abordant la question des traces modélisées plongées dans le contexte d'activités collectives, coopératives ou collaboratives que nous appelons de manière générique *activités conjointes*. Le cadre conceptuel des Systèmes à Base de Traces modélisées, que nous avons adopté dans cette thèse, permet de prendre en compte le caractère conjoint d'une activité que l'on souhaite tracer. Dans la méthodologie de modélisation que nous venons nous même de proposer, ce caractère fait même spécifiquement l'objet d'une dimension de l'analyse descriptive. La problématique des traces modélisées se trouve réinterrogée par le fait qu'un grand nombre d'usages apparaissent lorsqu'une de ces traces est réutilisée collectivement. Or, le cadre conceptuel d'origine ne permet de les instrumenter facilement. C'est la raison pour laquelle nous proposons une extension de ce cadre en introduisant la notion de *sujet* d'une trace. Formellement il ne s'agit que d'un identifiant, une métadonnée attachée à chaque observé d'une trace et par voie de conséquence à la trace modélisée elle-même. Dans la pratique il permet de distinguer des *traces modélisées individuelles* et des *traces modélisées conjointes*. Ces dernières sont des traces modélisées qui sont constituées à partir de plusieurs traces individuelles. Dans le cas d'une trace conjointe croisée il s'agit de faire l'union des observés de chacun des sujets de sorte que les interactions de chacun y soient présentes. Dans le cas d'une trace conjointe enrichie, il s'agit plutôt d'une fusion dont le résultat est une trace dont le sujet est « un groupe ». Cette extension permet d'envisager l'instrumentation de multiples usages potentiels des traces conjointes. Sur le plan de l'analyse de l'activité d'un groupe tout d'abord, bien que la démarche ne soit pas très différente de celle décrite précédemment, mais aussi et surtout sur le plan de l'usage des activités conjointes au sein même de l'activité. Nous proposons de lister de manière systématique l'ensemble de ses usages et d'en tirer de nouvelles pistes de recherche pour les travaux à venir.

7.1 Introduction

Nous avons conçu et proposé une méthodologie de modélisation des traces propre au Systèmes à Base de Trace en nous inscrivant dans le cadre conceptuel des Systèmes à Base de Traces modélisées. Dans ce cadre comme dans la méthodologie proposée, toute activité documentaire instrumentée peut être tracée, y compris si celle-ci est collective. Ce fût le cas notamment pour les deux terrains d'application sur lesquels il nous a été donné de travailler. La modélisation de trace tient compte du caractère collectif de l'activité d'un utilisateur au sens où des observés impliquant d'autres acteurs de la dite activité peuvent tout à fait être modélisés puis apparaître dans une trace modélisée donnée.

Pour simplifier, dans le cadre conceptuel des Systèmes à Base de Traces modélisées les *aspects* d'une activité *collective* sont bien pris en compte pour apparaître – si c'est utile – dans la trace modélisée d'un utilisateur, mais ils le sont *du point de vue de cet utilisateur en particulier*, ou plutôt du point de vue de sa machine (depuis laquelle l'activité est observée). Comme l'illustre le schéma ci-dessous (Figure 7.1), on pourrait résumer la situation en différenciant un « périmètre de modélisation » qui tient compte de l'ensemble de la situation tracée et un « périmètre de réutilisation de la trace modélisée » centré sur le point de vue d'un seul acteur. Par la suite, soit la trace modélisée est exploitée par un analyste extérieur à l'activité, soit elle est réintroduite dans l'environnement de l'acteur dont l'activité effective a été tracée.

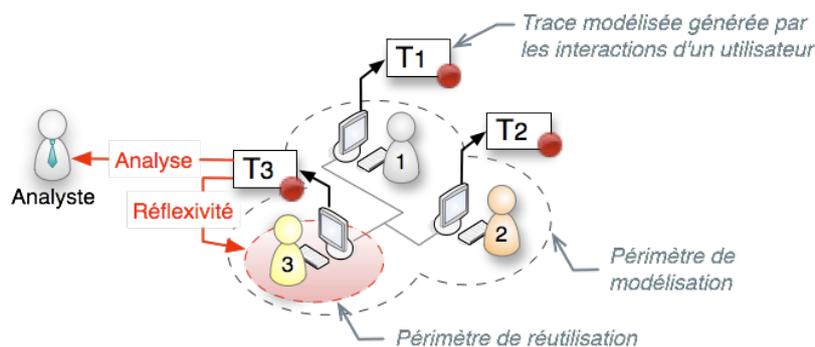


Figure 7.1 : Schématisation de la modélisation et de la réutilisation d'une trace modélisée dans le cadre d'une activité collective selon l'approche des Systèmes à Base de Traces modélisées.

Dans l'absolu, cela ne nous empêche pas d'envisager une trace « collective » au sens où nous pouvons fabriquer une collection des actions de tous les acteurs en juxtaposant les différents points de vue impliqués. Cependant le cadre conceptuel des Systèmes à Base de Traces modélisées nous semble limité relativement aux *multiples usages collectifs* qui se profilent lorsqu'une trace modélisée est réutilisable par un groupe d'acteurs. Il ne permet pas d'instrumenter facilement une activité avec une trace « collective » entendue autant comme la trace de plusieurs acteurs que comme l'usage de cette trace par l'ensemble de ces mêmes acteurs. C'est la raison pour laquelle nous proposons une *extension* de ce cadre conceptuel pour y considérer des *traces modélisées conjointes*, pensées comme les traces modélisées d'une activité collective, coopérative ou collaborative, supportant un usage lui-même collectif, coopératif ou collaboratif. Dans les sections à suivre, nous présenterons dans un premier temps l'extension du cadre conceptuel des Systèmes à Base de Traces modélisées que nous proposons et son contexte (section 7.2). Puis, dans un second temps, nous reviendrons plus en détail sur les multiples usages que ce cadre conceptuel étendu permet d'envisager (section 7.3).

7.2 Systèmes à Base de Traces modélisées conjointes

Nous avons souligné, dans l'introduction qui précède, le double sens possible de la notion de « trace modélisée collective ». Elle peut en effet être collective en tant que trace de l'activité de plusieurs acteurs, mais également en tant qu'objet d'un usage collectif par ces différents acteurs. En fait cette notion à double sens pose un problème de définition lorsque les deux sens sont cumulés. Il semble que la distinction fondamentale que nous avons faite entre trace modélisée comme support d'analyse et trace comme support de l'activité par réflexivité, se dissolvait dans le contexte d'activités conjointes où les usages possibles sont démultipliés. Pour être plus précis le problème se pose à partir du moment où l'on essaie de rendre compte d'une réflexivité de l'activité lorsque le sujet de l'activité est un groupe d'acteur.

Considérer l'analyse de traces modélisées « collectives » par un analyste en dehors de l'activité ne pose pas de problème particulier. En revanche, considérer une réflexivité des activités de plusieurs acteurs, en tant que groupe, à travers leurs traces respectives a beaucoup moins de sens (Figure 7.2). Un acteur exploitant la trace modélisée générée par son voisin est plutôt dans une posture d'analyse, mais en posture de réflexivité s'il exploite une trace modélisée de l'ensemble de son groupe par exemple. En fait, on se rend rapidement compte qu'il n'est pas possible de considérer une situation d'usage singulière. L'usage « collectif » d'une trace « collective » renferme en fait une myriade de situations différentes, qu'il s'agit de pouvoir instrumenter avec une approche des traces modélisée adaptée.

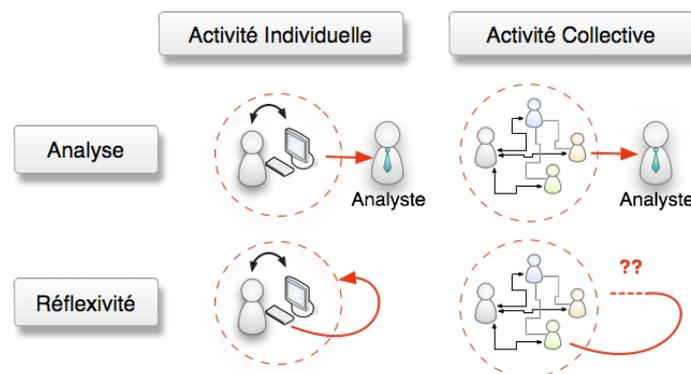


Figure 7.2 : Distinction des finalités des traces selon leur nature individuelle ou collective.

De plus, lorsque nous désignons par « collectif » l'implication de plusieurs acteurs dans une activité nous simplifions grandement le problème. En effet, comme nous l'avons suggéré également en introduction, il existe plusieurs types d'activités « collectives » : les activités collectives, coopératives et collaboratives. Puisque ces activités ne sont plus seulement susceptibles d'être un phénomène étudié par le biais d'une analyse de traces, mais bien le contexte d'usages de traces modélisées, commençons par préciser ce que désignent ces qualificatifs.

7.2.1 Les activités conjointes

Lorsqu'il est question d'activités impliquant plusieurs acteurs, l'utilisation des termes « collectif », « coopératif » et « collaboratif » peut prêter à confusion. Au sein du *CSCW*, domaine qui se consacre pleinement à l'étude de ce type d'activités, les chercheurs proposent différentes distinctions entre ces termes, en fonction du type d'engagement des acteurs, des outils utilisés, ou des contraintes de la si-

tuation d'activité elle-même. Pour notre part nous souhaitons désigner l'ensemble de ces activités sous le terme générique « d'activité conjointe ».

Ainsi, une activité sera dite conjointe à partir du moment où l'on peut établir un lien à travers l'activité entre plusieurs acteurs qui « co-agissent ». Ce lien est établi, selon le cas, sur le fait que ces acteurs utilisent les mêmes outils ou bien exploitent les mêmes ressources ou encore partagent les mêmes buts. Nous distinguons ensuite trois formes particulières d'activités conjointes que sont les activités collectives, coopératives et collaboratives, que nous distinguons en utilisant les critères de distinction classiques en *CSCW* et que nous pouvons résumer comme suit (Tableau 1).

Formes d'activité conjointe	Collective	Coopérative	Collaborative
			
Mêmes outils / ressources	Oui	Oui	Oui
Conscience réciproque	Non / implicite	Oui	Oui
But commun	Non / implicite	Oui	Oui
Interdépendance	Non / implicite	Non	Oui
Articulation	Non	Implicite	Explicite

Tableau 7-1 : Caractérisation de trois formes d'activités conjointes.

Précisons que nous entendons par *conscience réciproque* (ou de groupe) le fait que chaque acteur du groupe est informé de l'existence (ou connaît) les personnes qui sont impliquées avec lui dans une activité donnée et peut accéder à leur travail. Par *articulation* de l'activité nous désignons les mécanismes, explicites ou non, permettant à plusieurs acteurs de coordonner leurs actions, ce qui est parfois nécessaire lorsqu'il y a interdépendance des tâches c'est à dire que les conditions devant être réunies pour la réalisation d'une tâche demandée à un acteur peuvent inclure qu'une ou plusieurs autres tâches soient achevées par d'autres acteurs. Nous avons ici emprunté des notions rencontrées dans notre travail de recherche sur le *CSCW* notamment en ce qui concerne l'articulation du travail collaboratif dans la Théorie de la Coordination par exemple (section 4.4).

La confusion entre les différents types d'activités conjointe est répercutée dans la littérature sur les traces « collectives » ou « collaboratives » qu'il arrive de rencontrer. La notion d'usage conjoint des traces d'interaction est toujours plus ou moins présent dans les travaux de recherche depuis les premiers travaux sur les systèmes traçants (Chapitre 2). Les systèmes de recommandations, ou de partage de tags (Millen, 2007) peuvent ainsi être décrits comme un usage collectif de traces d'interactions. De même pour les systèmes donnant à des utilisateurs leur état d'avancement, ce qui permet une coordination implicite et peut donc être vu comme un usage coopératif des traces d'interaction. C'était également le cas pour le système de « *Edit Wear and Read Wear* » de W. Hill et J. Hollan en 1992, que l'on peut voir comme un usage collaboratif des traces d'interactions. En l'occurrence l'enrichissement du code sur lequel travaillent plusieurs informaticiens, leur permet de faciliter leur coordination.

Concernant les Systèmes à Base de Traces, c'est-à-dire des systèmes dans lesquels les traces sont des objets informatiques à part entière créés pour représenter l'activité effective, les travaux sur les possibilités d'usages collectifs sont bien moins nombreux. Les seules tentatives semblent se restreindre à la possibilité d'une « analyse de la collaboration » par des indicateurs, ou à un « retour d'information » aux utilisateurs par ces même indicateurs. Il est donc tout à fait intéressant de creuser cette question

dans le cadre conceptuel particulier que nous offre les Systèmes à base de Traces modélisées. La question est de savoir comment, dans ce cadre particulier, instrumenter des situations telles que nous en avons rencontré sur nos terrains d'expérimentation : un utilisateur qui voudrait partager sa propre trace modélisée avec un autre lors d'un échange en face à face ou encore un tuteur qui pourrait à distance partager la trace modélisée d'un apprenant pour l'accompagner dans son interprétation (avec *eLycée*).

7.2.2 Les traces modélisées d'activités conjointes

Avant de rentrer dans le détail de notre proposition d'extension des Systèmes à Base de Traces modélisées, nous souhaitons revenir sur la justification de notre démarche. En effet, rien n'indique *a priori* qu'un utilisateur quelconque soit susceptible d'exploiter les traces modélisées d'un autre utilisateur, ni même qu'il aurait intérêt à le faire dans le cadre d'une activité conjointe. Or, dans le cadre de notre expérimentation (terrain d'application sur la production de formations), nous avons pu constater qu'un développeur était capable d'objectiver son activité par l'interprétation d'une trace modélisée même sans réussir à se la réapproprier totalement, ce qui indiquerait qu'il pourrait faire de même avec l'activité d'un *autre* développeur – d'autant plus si l'activité en question les impliquait tous les deux. De plus, l'intérêt du partage de la trace modélisée de l'activité passée est ici évident puisque les deux acteurs en question n'ont été capables de se remémorer leur activité conjointes que lors d'un échange en face à face permettant de confronter leurs souvenirs et documents respectifs. Il nous paraît donc tout à fait pertinent de tenter d'instrumenter l'ensemble des situations d'usages conjoints qui pourrait naître de la mise à disposition d'un Système à Base de Trace conjointe.

7.2.2.1 Notion de sujet d'une trace modélisée

Nous ne proposons ici qu'une *extension* du cadre conceptuel des Systèmes à Base de Traces modélisées (présenté au Chapitre 3) ce qui signifie que nous en conservons les principaux concepts et que nous les adaptons au contexte des activités conjointes. Sur le plan conceptuel, l'apport essentiel réside dans l'introduction de la notion de *sujet* de la trace. On peut décrire ce sujet comme un *méta-donnée* ajoutée à une trace modélisée caractérisant la situation d'observation et par extension chaque observé de la trace. De manière générique, nous considérons que le sujet sera un simple identifiant, ou ensemble d'identifiants, et ne prenons pas position sur le fait que cet identifiant désigne une personne à un moment donnée ou bien cette personne dans une situation donnée ou encore la même personne sur plusieurs situations différentes.

7.2.2.2 Formalisation et nouvelle définition des observés

Nous avons choisi de formaliser la notion de sujet au niveau de chaque observé, ce qui simplifie la description des traces modélisées constituées d'observés provenant de différentes situations d'observation. Formellement, on définit *un sujet* comme un identifiant ou un ensemble de sujets. La récursivité de la définition du *sujet* permet à une trace ou un observé de porter sur un groupe (explicite) d'acteurs. L'identifiant peut ainsi représenter soit un *acteur individuel* : une personne, ou un acteur artificiel ; soit un *acteur groupe* : un groupe de personnes considéré comme une entité opaque (réification ou personnification du groupe), voire même un ensemble de groupes (exemple : l'ensemble composé de deux équipes, elles mêmes composées de plusieurs personnes).

Dans le cadre conceptuel des SBTm « classique » nous avons défini un observé comme toute information structurée issue de l'observation d'une interaction (section 3.3.1.1). Désormais, dans le cadre des Système à Base de Traces Conjointes, tout observé est donc *associé à un sujet*. Du coup, le sujet de la trace est lui-même déduit en « agrégeant » les sujets des observés qui la composent. Il existe plusieurs manières de faire cette agrégation. Il est possible de faire notamment l'un des deux calculs suivant : soit le sujet d'une trace est simplement « constitué » de l'ensemble des sujets de ses observés ; soit le sujet est le résultat de l'union ensembliste des sujets des observés. Considérons une trace dont les observés ont respectivement pour sujets {a,b}, {c,d} et a. Le premier calcul donnera pour sujet de la trace l'ensemble {{a,b}, {c,d}, a} ; le second donnera pour sujet {a, b, c, d}. De cette formalisation découle un dédoublement du concept de *trace modélisée*, en trace modélisée *individuelle* et trace modélisée *conjointe*

7.2.2.3 Traces modélisées « individuelles » et traces modélisées « conjointes »

Trace modélisée individuelle

Nous définissons une *trace modélisée individuelle* comme une *collection structurée et pourvue d'une extension temporelle, d'observés temporellement situés issus de l'observation de l'utilisation d'un système observé par un acteur dans le cadre d'une activité donnée*. Une trace modélisée individuelle est générée par l'instanciation d'un modèle de trace individuelle explicite et contient l'identification de l'acteur concerné. Cette définition est identique à celle d'une trace modélisée « classique » à ceci près qu'il est obligatoire que ses observés (et donc elle-même) permettent explicitement l'identification de l'acteur qui interagit avec le système observé depuis sa machine.

Trace modélisée conjointe

Sur la base de ce qui précède nous pouvons désormais proposer la définition suivante de ce que nous appellerons une *trace modélisée conjointe* : dans le cadre de l'utilisation par au moins deux acteurs distincts d'un système observé au cours d'une activité conjointe donnée, *une trace modélisée conjointe est définie comme une collection, pourvue d'une extension temporelle, d'observés temporellement situés, associée à un modèle de trace conjointe explicite, telle que deux ensembles d'observés liés aux deux acteurs distincts puissent y être distingués*. La notion de trace modélisée conjointe implique un lien constitutif²⁰¹ à au moins deux traces modélisées individuelles, correspondant aux deux acteurs distincts considérés, dont les observés constituent les traces modélisées individuelles. Deux types de *traces modélisées conjointes* peuvent être distingués (Figure 7.3) :

- Une *trace modélisée conjointe croisée* est constituée de l'ensemble des observés attribués à l'un ou l'autre des acteurs impliqués (observés), elle permet de décrire une utilisation « croisée » de l'environnement, c'est-à-dire par au moins deux *utilisateurs distincts*.

Dans le cas d'une trace conjointe « croisée » le modèle d'une trace modélisée conjointe M_{TC} sera soit le modèle commun aux traces individuelles M_{Ti} des acteurs, soit l'union de plusieurs modèles de trace individuelle différents $M_{TC} = \cup M_{Ti}$.

²⁰¹ Soit par construction de la trace conjointe à partir des traces individuelles, soit parce qu'on peut retrouver des traces individuelles dans la trace conjointe.

- Une *trace modélisée conjointe enrichie* est construite à partir d'une trace conjointe croisée complétée d'observés non liés aux acteurs de l'activité conjointe, mais liée au groupe d'acteurs lui-même considéré comme acteur.

Dans le cas d'une trace conjointe « enrichie » cette fois, le modèle M_{TC} sera le modèle de la trace conjointe croisée complété d'un modèle d'enrichissement, correspondant à l'acteur groupe

$$M_{TC} = \cup M_{Tii} + M_{TG}.$$



Figure 7.3 : (a) Deux traces modélisées individuelles, (b) une trace conjointe croisée, (c) une trace conjointe enrichie.

7.2.2.4 Remarque sur l'implémentation d'un Système à Base de Traces conjointes

Les différentes opérations que nous venons d'évoquer et qui sont intrinsèquement liées à la définition des traces conjointes²⁰² devront être réalisées selon la situation dans des configurations relativement différentes. La gestion des différents modèles de traces et des transformations afférentes au croisement ou à l'enrichissement d'une trace conjointe doit être réalisée notamment en fonction de l'architecture concrète du système observé implémenté. L'extension conceptuelle qui introduit simplement la notion de sujet engendre par ailleurs un besoin d'extension du *framework* informatique qui est encore en cours de développement. La première difficulté sera rencontrée dans le processus de collecte. Il s'agit par exemple de vérifier, en fonction d'architectures diverses, qu'il sera toujours possible de générer une trace modélisée conjointe croisée. Nous décrivons donc, ci-dessous, trois configurations de collecte en fonction d'architectures typiques rencontrées dans des situations d'activités conjointes instrumentées.

Collecte serveur (modèle de trace conjointe « unique »)

Une première situation possible est celle d'une activité conjointe pour laquelle tous les acteurs exploitent une même application serveur. En principe, toute action réalisée est enregistrable directement depuis ce dernier, ce qui permet d'envisager une collecte directement sur le serveur en question. Dans ce cas l'obtention d'une trace conjointe croisée²⁰³ peut s'appuyer sur un unique modèle de trace permettant de générer directement une trace conjointe *croisée*, une trace conjointe *enrichie* étant alors obtenue par *transformation* de cette dernière.

²⁰² Pour alléger le texte nous parlerons de « trace conjointe » ou de « trace individuelle » pour désigner respectivement une trace « modélisée » conjointe et une trace « modélisée » individuelle.

²⁰³ Pour alléger le texte nous parlerons de « trace conjointe croisée » pour désigner une « trace modélisée conjointe croisée ».

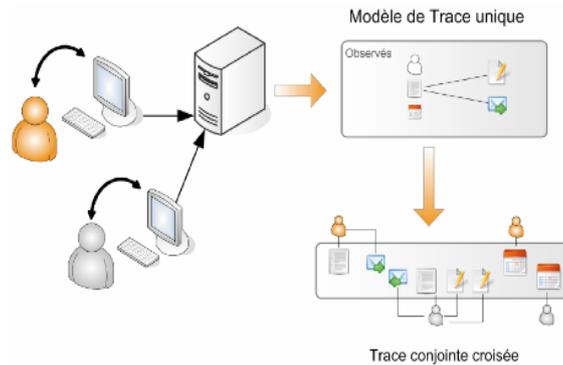


Figure 7.4 : Collecte serveur.

Collecte client (modèles de trace individuelle identiques ou différents)

La seconde situation, à l'opposée de la première, est celle où l'activité passe par l'exploitation d'un environnement client pour chaque acteur. L'environnement local de chaque acteur constituant l'unique source de traçage, la collecte s'effectue donc du côté client, en respect d'un modèle de trace individuelle. Deux cas sont alors à envisager : soit les modèles de trace conjointe sont identiques, soit ils ne le sont pas. Pour obtenir une trace conjointe croisée, il est dans ce cas nécessaire de centraliser les traces modélisées individuelles et de passer par un modèle pivot permettant de la générer : dans le cas d'un modèle de trace individuelle identique pour tous les acteurs, il s'agira simplement de fusionner les différentes traces (sur la base d'un alignement de l'extension temporelle) ; dans le cas de modèles de trace individuelle différents, le modèle pivot devra en faire l'union afin d'obtenir la trace conjointe croisée recherchée.

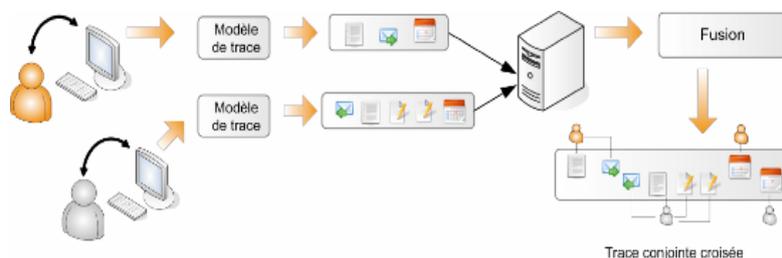


Figure 7.5 : Collecte client.

Collecte mixte (modèles de trace individuelle identiques ou différents)

Sachant que dans beaucoup de situations de traçage, il est nécessaire d'observer à la fois l'utilisation d'un environnement côté client et côté serveur, la troisième situation qui est l'association des deux précédentes, sera sans doute la plus courante. Il s'agit d'une situation où les acteurs utilisent une application client-serveur qui constitue une « double » source de traçage. La collecte s'effectue alors à la fois côté serveur (avec un modèle de trace unique) et du côté client (avec des modèles de trace individuelle identiques ou différents selon les acteurs). Pour obtenir une trace conjointe croisée, il est nécessaire de centraliser les traces individuelles collectées côté client et celles collectées côté serveur.

Dans le cas de modèles de trace individuelle identiques, il s'agit de générer une trace individuelle pour chaque acteur à partir de la double collecte (client + serveur), puis de procéder comme précédemment (« fusion » des différentes traces individuelles). Dans le cas de modèles différents, il est nécessaire de passer par un modèle pivot qui réalise l'union des différents modèles de traces individuelles.

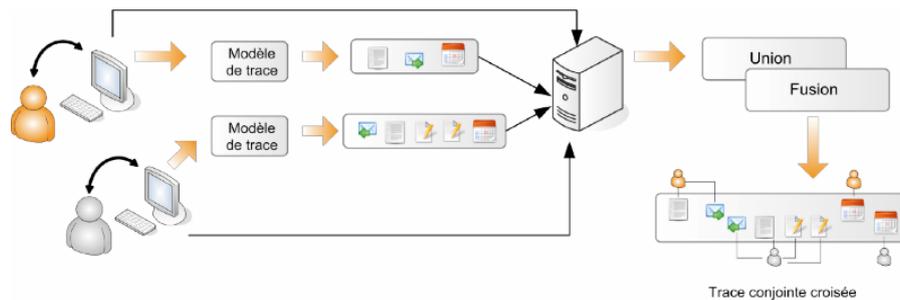


Figure 7.6 : Collecte mixte.

Loin de nous l'idée d'avoir, à travers ces trois situations caricaturales (et leur variantes), couvert l'ensemble des difficultés rencontrées pour la mise en place d'une fonction de collecte répondant aux contraintes imposées par une extension du *framework* informatique des SBTm aux traces modélisées conjointes. Des architectures plus complexes, par exemple *Peer-to-Peer*, demandent encore à être étudiées. Notre objectif n'étant toutefois pas de dresser la liste de ces contraintes techniques mais de souligner la complexité d'une implémentation d'un Système à Base de Traces conjointes, nous allons pour terminer ce chapitre nous intéresser aux multiples usages des traces conjointes que permettent désormais d'instrumenter un Système à Base de Traces conjointes.

7.3 Les potentiels usages conjoints des traces conjointes

Nous avons posé dans les sections précédentes un cadre conceptuel étendu, expliqué rapidement sa formalisation et souligné les conséquences de son extension sur l'implémentation concrète d'un Système à Base de Traces conjointes. Nous pouvons maintenant revenir à ce qui a motivé au départ cette extension à savoir les *usages* des traces conjointes et leur(s) finalité(s). En guise de préambule, nous dirons quelques mots des situations d'analyse des traces conjointes, puis nous aborderons directement le cas des traces conjointes réutilisées *au sein* de l'activité de plusieurs acteurs.

7.3.1 Traces modélisées comme support d'analyse de l'activité conjointe

Globalement, la situation dans laquelle un analyste exploite des traces conjointes pour caractériser ou évaluer l'activité conjointe qui en est à l'origine ne diffère pas fondamentalement de la situation « standard » que nous avons plusieurs fois décrite. L'analyste n'est en principe pas directement impliqué dans l'activité et son intervention n'a pas vocation à modifier la situation d'activité considérée. Là où un changement est sensible c'est dans l'objet visé par l'analyse des traces conjointes que nous visions à instrumenter.

On assiste en quelque sorte à un changement de niveau d'interprétation avec un analyste qui va étudier les comportements collectifs des acteurs dans leur activité et non ceux de ses membres pris individuellement. Qu'apportent dans ce contexte les traces conjointes ? Une *trace conjointe croisée* permet par exemple d'étudier les actions *d'articulation* de l'activité conjointe (pour déterminer le caractère coopératif ou collaboratif de celle-ci) en étudiant l'exploitation, durant la réalisation de l'activité, des outils de *coordination* (agendas partagés, outils de communication, *etc.*), ou encore le partage de docu-

ments, en particulier les documents pour l'action²⁰⁴. L'intérêt des traces modélisées dans ce contexte d'usage serait de permettre à l'analyste d'aborder les phénomènes d'articulation dans leur contexte d'occurrence, voire d'accéder (pour tout ou partie) aux contenus impliqués, et de ne pas se limiter aux habituelles « traces de la collaboration » que délivrent les indicateurs statistiques que nous avons déjà évoqués (chapitre 2)²⁰⁵. Avec une trace *conjointe enrichie*, un analyste peut également se pencher sur l'activité réalisée par un sujet composé de plusieurs acteurs, *i.e.* un groupe ou un sous-groupe, sans distinction individuelle. L'intérêt d'une telle manipulation est de faire apparaître dans la trace des phénomènes d'organisation des interactions qui ne sont observables qu'en tant que comportement d'un groupe.

7.3.2 Traces modélisées comme support de l'activité conjointe

Nous voici arrivé à la situation d'usage qui nous intéresse tout particulièrement, celle de traces conjointes réintroduites dans l'activité conjointe dont elles sont issues. Pour décrire l'ensemble des situations d'usages possibles, prenons un exemple schématique impliquant trois acteurs. Considérons qu'ils réalisent une activité conjointe pour laquelle chacun d'eux est l'utilisateur d'un système observé instrumenté par un Système à Base de Traces conjointes. Considérons enfin que chaque utilisateur génère par ses interactions une trace modélisée individuelle de sa propre activité (Figure 7.7).

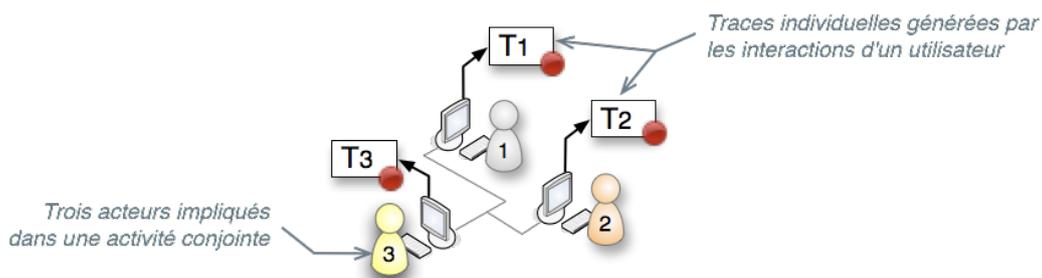


Figure 7.7 : Schématisation d'une activité conjointe type pour laquelle chaque acteur génère une trace modélisée individuelle.

7.3.2.1 Partage concret d'une trace individuelle et/ou conjointe

L'usage conjoint d'une trace conjointe nécessite des *moyens de partage* de cette trace entre les différents acteurs, *i.e.* qu'il soit possible concrètement à un utilisateur 1 d'accéder à la trace d'un utilisateur 2 pour l'interpréter. Dans le cours d'une activité effective réelle ce partage peut se matérialiser sous différentes formes que nous regroupons en deux cas de figure :

- Dans le premier cas de figure que nous qualifions de partage « direct » (Figure 7.8), un utilisateur 1 décide de partager sa trace avec un autre acteur et lui donne accès directement à sa trace individuelle. Soit il transmet sa trace par le réseau pour une consultation à distance, soit l'utilisateur 2 se déplace et vient aux côtés de l'utilisateur 1 pour consulter la trace en question sur la machine de ce dernier (ce qui aurait pu être le cas dans notre expérimentation).

²⁰⁴ Que nous avons déjà pris en compte dans l'analyse de l'activité, voir chapitre 6.

²⁰⁵ Nous pouvons d'ailleurs remarquer au passage que « la collaboration » en elle-même ne peut être tracée : elle n'a aucune consistance, aucune matérialité, elle n'est que la *qualité* d'une activité considérée.

- Dans le second cas de figure, que nous qualifions de partage « indirect » (Figure 7.8), la trace d'un utilisateur 1 est déposée sur un serveur qui centralise l'accès à toutes les traces générées par les trois acteurs. Chaque acteur « publie » en quelque sorte ses traces individuelles. Comme dans le cas de figure précédent, un utilisateur 1 peut inviter une utilisateur 2 à consulter sa trace, mais cette fois l'utilisateur 2 peut également de lui-même chercher et utiliser cette trace de manière indépendante.

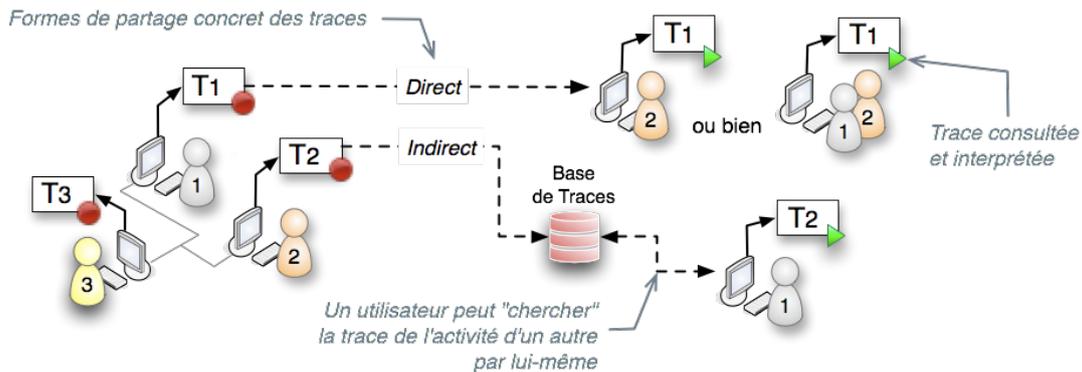


Figure 7.8 : Formes concrètes de partage des traces individuelles et/ou conjointes.

Dans notre exemple schématique ci-dessus, les traces en question sont individuelles. Il en serait de même si les sujets (ici les utilisateurs 1, 2 et 3) représentaient en fait un groupe ou un sous-groupe d'acteurs, et les traces individuelles des traces conjointes. Maintenant que nous avons précisé ce que peut signifier concrètement le partage d'une trace entre deux acteurs.

7.3.2.2 Situations concrètes d'usage de traces individuelles et conjointes

Ayant posé les modalités concrètes de partage des traces modélisées (individuelles ou conjointes) dans un Système à Base de Traces conjointes, nous pouvons revenir aux différentes situations d'usages qui se présentent au sein même d'un activité conjointe lorsque ses acteurs sont en position de réutiliser leurs traces. Pour essayer d'être aussi clair qu'exhaustif, nous avons regroupé tous les cas de figure dans une matrice présentée ci-dessous (Figure 7.9).

Chaque situation d'usage des traces va constituer un contexte d'interprétation particulier de la trace. Dans cette matrice nous avons distingué horizontalement les trois types de traces modélisées pouvant se prêter à l'interprétation : sur la première ligne (A) les *traces individuelles*, sur la seconde (B) les *traces conjointes croisées* et sur la troisième (C) les *traces conjointes enrichies*. Verticalement nous avons distingué les différents contextes d'interprétation d'une trace donnée par un ou plusieurs acteurs, qu'ils soient ou non eux-mêmes sujet de la trace interprétée.

- Concernant la trace individuelle d'un sujet donné tout d'abord. Le sujet de la trace individuelle peut bien entendu lui-même y accéder (A-1), c'est la trace modélisée support d'une activité réflexive telle que nous l'avons décrite jusqu'ici. Cette même trace peut être également partagée, et interprétée avec un autre acteur, soit en présence (A-2), soit à distance (A-3). C'est la situation que nous aurions pu rencontrer par exemple dans le cadre de notre expérimentation où la trace d'un premier développeur aurait pu être discutée à deux. Enfin, il y a également la possibilité pour deux (ou plus) utilisateur de faire une interprétation commune de la trace d'un troisième (A-4).

- Concernant une trace conjointe croisée ensuite, qui (pour schématiser) sera ici l'union de deux traces modélisées (T1 et T2). Cette trace peut être interprétée par un des sujets de la trace (B-1). Le sujet en question pourra ainsi *situer* son activité *relativement* à celle de son (ou de ses) collaborateur(s). L'ensemble des sujets concernés peuvent mener ensemble une interprétation de cette même trace, accompagnés (B-3) ou on (B-2) d'un acteur qui n'est pas sujet de la trace conjointe considérée. Enfin ce troisième acteur peut également se référer à cette trace conjointe (B-4), par exemple pour déterminer qui le quel des sujets concernés a réalisé telle ou telle action au cours de la réalisation de l'activité.
- Concernant une trace conjointe enrichie pour terminer, qui (pour schématiser) sera ici la trace d'un sujet « groupe » (groupe composée des sujets 1 et 2). On retrouve des situations similaires aux précédentes, avec des nuances quant aux interprétations possibles. Par exemple un acteur seul (C-1) pourra utiliser la trace conjointe enrichie du groupe auquel il appartient non plus pour situer son propre travail relativement à celui des autres mais simplement à déterminer par exemple quel a été le travail réalisé durant une période donnée. De même pour une interprétation menée par l'ensemble des sujets concernés par la trace conjointe (C-2), éventuellement accompagnés (C-3) par un acteur extérieur. L'idée reste d'interpréter cette trace comme celle de l'activité réalisée par le groupe dans son entier, sans distinction de la part prise par chaque sujet. Pour le dernier cas (C-4), celui où seul un acteur qui n'est pas sujet de la trace (dans une posture semblable à celle de l'analyste), on peut imaginer un responsable de groupe voulant faire le point sur le travail de son équipe, sans chercher à individualiser son analyse.

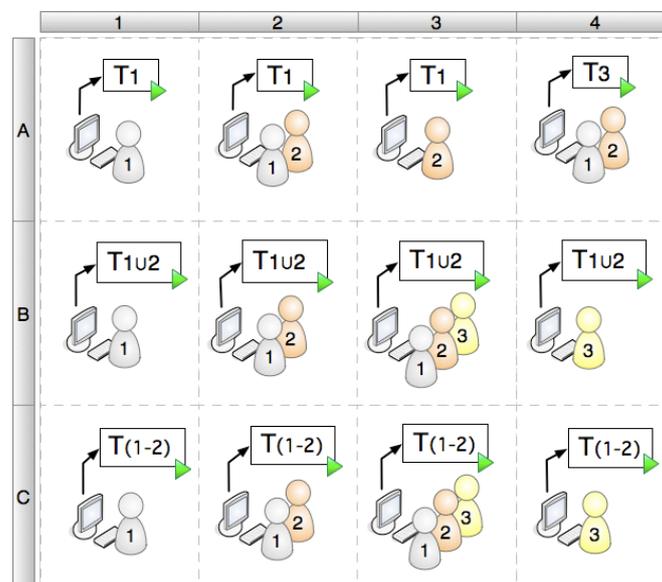


Figure 7.9 : Situations d'usage des traces individuelles et conjointes dans une activité conjointe.

7.3.3 Finalités en situation

Pour dresser une liste des situations d'usages des traces individuelles et conjointes dans une activité conjointe, nous avons raisonné de manière théorique sur des contextes « neutres » avec pour unique souci de décrire toutes les configurations possibles de situations impliquant des acteurs devant interpréter ces traces. Pour terminer nous aimerions redonner à certaines de ces configurations l'épaisseur

d'une situation en contexte, qui permet de mesurer les enjeux et les apports concrets qu'on pourrait attendre de l'usage de ces traces conjointes. Nous allons même voir à travers trois exemples, que les traces conjointes en tant que représentation de l'activité effective d'un sujet peuvent être envisagées comme un véritable instrument du travail collectif, coopératif ou collaboratif.

7.3.3.1 Dans une situation d'activité collective

Dans diverses activités documentaires instrumentées les expériences de réalisation de l'activité ne sont pas nécessairement capitalisées, contrairement au résultat de ces activités qui sont régulièrement archivés, indexés, classés, « documentarisés » pour être éventuellement réutilisés. C'était le cas par exemple sur notre second terrain d'expérimentation : une partie des contenus produits étaient l'objet d'une capitalisation alors que le processus de création d'une formation par des développeurs restait quant à lui dispensé d'une telle attention. On pourrait tout à fait imaginer que des traces modélisées, individuelles comme conjointes, pourrait être laissées en libre accès pour réutilisation. Il ne s'agit pas de savoir qui a fait quoi dans quelles conditions précises, mais de donner une idée de comment une activité donnée a pu être réalisée. L'utilisateur de la trace en question aura pour objectif de réutiliser une façon de faire qu'il pourra adapter à sa propre situation. De ce point de vue on peut considérer qu'une trace modélisée peut tout à fait jouer un rôle dans des activités collectives où il n'y a pas forcément d'échange direct entre les acteurs concernés.

7.3.3.2 Dans une situation d'activité coopérative

Dans une activité coopérative, les acteurs ne sont pas directement dépendants les uns des autres dans leur activité, ce qui ne signifie pas qu'ils ne coordonnent ou n'articulent pas leur travail. Ils le font généralement de manière implicite. De nombreuses situations seraient l'occasion pour un acteur de partager sa propre trace d'activité avec quelqu'un. Un acteur pourrait par exemple utiliser sa propre trace modélisée comme support de démonstration pour illustrer une méthode de résolution de problème ou l'application d'une stratégie de réalisation d'une activité particulière, auprès d'un autre acteur. Mais ce partage d'une trace modélisée pourrait avoir de toutes autres motivations. L'utilisateur en question peut également proposer à un pair de partager sa trace pour en faire une interprétation commune, comme c'est le cas dans ce que nous avons proposé entre tuteur et étudiant dans *eLycée* (voir section 6.4), ou bien encore lui demander son avis sur la méthode employée, en faire une évaluation, un commentaire. De manière générale, une trace modélisée conjointe peut soutenir une articulation implicite de l'activité puisqu'elle rend disponible à un acteur donné ce que font/ont fait les autres acteurs, les ressources (de l'activité) mobilisées, modifiées, *etc.*

7.3.3.3 Dans une situation d'activité collaborative

Dans une activité collaborative, l'articulation explicite entre les acteurs est nécessaire. À nouveau on peut imaginer un usage possible d'une trace modélisée, individuelle ou conjointe. Un acteur, ou un groupe d'acteur pourrait par exemple prescrire une activité à faire grâce à une trace : soit en envoyant une trace modélisée d'une activité passée pour indiquer la marche à suivre (en imitant ou répétant ce qui avait été réalisé dans un autre contexte), soit en envoyant une trace de ce qui a été fait jusqu'à présent pour prescrire ce qui « reste à faire ».

Un autre exemple intéressant nous est donné par les mémoires de projets qui sont d'autant plus difficiles à constituer que l'activité comporte de nombreux acteurs et une collaboration étroite et complexe entre eux. Une trace conjointe de leurs activités effectives durant un projet pourrait être un support tout à fait utile à partir duquel un mémoire de projet pourrait être constitué, ou au moins complété. L'idée est que la trace conjointe sert de récapitulatif de ce qui s'est passé, tant du point de vue de chacun (trace conjointe croisée) qu'au niveau du groupe dans son ensemble (trace conjointe enrichie). De plus, adjointe au mémoire lui-même une trace peut permettre d'indexer – relativement à l'activité effective – les différents événements et réalisations qui ont eu lieu. Si les acteurs se l'appropriait suffisamment on peut même imaginer qu'une trace conjointe pourrait être elle-même le cœur d'un mémoire de projet, qui serait *in fine* simplement une trace conjointe commentée et annotée.

Pour finir, nous pouvons revenir plus modestement à un exemple de situation que nous avons déjà évoqué : celui d'une « transmission » de projet (section 6.3.4.4). Dans la simulation de situation de relève que nous avons créé sur notre second terrain d'application, un développeur devait transmettre un projet de développement à un collègue. Ce projet avait été entamé en collaboration avec un second développeur. Concrètement, durant cet exercice, le développeur devant transmettre son projet s'est évertué à dresser une liste plus ou moins structurée de repères relatifs à l'activité à reprendre, sans pouvoir rentrer dans les détails de ce qui avait été réellement fait au début de la production, et sans pouvoir non plus donner d'éléments quant à la réalisation de son collaborateur. Dans cette situation à nouveau une trace conjointe aurait pu être un support utile à l'échange qui a eu lieu, voire être le moteur de cet échange.

7.4 Conclusion

Les particularités des situations d'usages conjoints des traces modélisées nous ont poussé à repenser leur conceptualisation. Relativement à la classification que nous avons nous même établie au départ de cette thèse, la situation qui pose problème est celle d'une activité conjointe, dans laquelle la finalité des traces modélisées ne peut être clairement l'analyse ou la réflexivité, mais une association complexe des deux postures à partir du moment où la trace est réintroduite dans l'activité du groupe. Nous avons ainsi proposé dans ce chapitre d'étendre le cadre conceptuel des Systèmes à Base de Traces en y intégrant la notion de « trace conjointe ». Dans ce cadre étendu se dessinent de nouvelles problématiques, qui relèvent tout d'abord d'aspects techniques pour le SGBT qui devra être mis en place, mais aussi – et surtout – de perspectives liées aux usages des traces modélisées. Nous avons essayé de poser de manière systématique les bases de ce qui sera un axe de développement majeur de l'approche de Systèmes à Base de Traces, en listant les différentes situations d'usages possibles impliquant des acteurs travaillant conjointement dans des environnements numériques partagés. Nous allons voir dans le chapitre suivant, les perspectives que nous offre ce nouveau cadre pour penser les traces modélisées.

Chapitre 8

Conclusion

L'ultime chapitre de ce manuscrit sera consacré à une conclusion étendue portant sur l'ensemble de notre travail de thèse. Nous dresserons dans un premier temps un *bilan* de notre travail de thèse en mettant en avant les principaux apports de celui-ci (section 8.1), puis dans un second temps nous discuterons de ses *limites* que nous commenterons (section 8.2). Pour finir, nous détaillerons dans un troisième temps les *perspectives* que ce travail offre quant à de futurs travaux de recherche sur les traces modélisées en particulier et sur les traces numériques en général (section 8.3).

8.1 Bilan et apports de la thèse

8.1.1 Apports théoriques et bibliographiques

Présentée grossièrement, la problématique initiale de cette thèse était de déterminer comment créer des traces numériques effectivement réutilisables d'activités documentaires instrumentées. Nous avons porté notre premier effort de recherche sur la notion de « trace numérique » elle-même, sa signification, son lien avec la notion commune de « trace » (Chapitre 1). Nous avons affirmé qu'il ne pouvait y avoir assimilation de la notion de « trace numérique » à une notion de trace en général, et ce au moins en raison du changement de support : le numérique n'offre pas un support permanent à l'inscription de trace comme le fait un environnement matériel. Une trace numérique ne peut être quelque chose de « donné » à l'observateur, il s'agit nécessairement de quelque chose de « construit » car tout enregistrement susceptible de constituer ou de composer une trace numérique est d'une manière ou d'une autre programmé, prédéterminé. Nous avons ensuite caractérisé ce qu'une trace numérique partage fondamentalement avec la notion de trace en général et avons proposé de les considérer comme un ensemble d'indices enregistrés composant le *signe interprétable* qu'une activité interactionnelle a eu lieu, voire comment elle a eu lieu.

Nous avons ensuite établi un large état de l'art des types de *systèmes traçants* c'est-à-dire de tous les dispositifs informatiques visant à réutiliser des enregistrements systématiques d'interaction (Chapitre 2). Nous avons présenté ces systèmes traçants selon leur finalité, en distinguant les approches pour lesquelles les traces numériques d'interaction sont utilisées comme *support d'analyse* et celles où ces traces sont réintroduites dans la situation d'activité pour devenir un *support à l'activité* dont elles sont issues. Dans un effort de synthèse nous avons proposé de considérer toutes ces approches d'une part comme une réponse à un problème *technique* : celui de collecter concrètement des données pour en faire des traces numériques, d'autre part comme une réponse à un problème *théorique* et pratique : celui de « donner du sens » au données collectées.

Nous avons dégagé de la littérature deux points de vue sur cette question, deux visions des traces numériques. Pour le premier, les traces numériques existent déjà dans les systèmes informatiques en fonctionnement, et il suffit de collecter des données de *logfiles* pour en constituer des traces. Pour en obtenir une trace interprétable de l'activité qui a eu lieu il faut alors enrichir *a posteriori* cet ensemble de données d'une sémantique qu'elle n'ont pas au départ (grâce à des techniques de traitement statistique principalement). Pour le second, il faut au enrichir *a priori* la sémantique des données interactionnelles qui seront enregistrées et constitueront la trace numérique. Dans ce second cas de figure, la trace numérique devant représenter une activité qui a eu lieu devient un objet informatique à part entière, manipulable en tant que tel, et nous avons dénommé les systèmes servant à les manipuler les *Systèmes à Base de Trace* du nom de l'approche visant une description générique de ce type de systèmes. L'approche des *Systèmes à Base de Traces modélisées* dans laquelle nous avons choisi d'inscrire ce travail de thèse, en est une *formalisation*.

L'approche des *Systèmes à Base de Traces modélisées* se présente aujourd'hui comme un cadre conceptuel (Chapitre 3). Il est le fruit d'un travail d'élaboration de plusieurs années²⁰⁶ auquel nous avons nous-mêmes participé. Une première approche, l'approche MUSETTE, avait jeté les premières bases en proposant d'instrumenter une modélisation des traces *a priori* pour en tirer des épisodes d'utilisation réutilisables par des agents assistants. L'approche des SBTm qui lui a fait suite correspond à une généralisation de l'approche. Reposant sur le concept de *trace modélisée* définie comme une collection d'observés générée en respect d'un *modèle de trace*, l'approche offre de concevoir de manière générique des Systèmes à Base de Traces modélisées concrets (SGBT). Si l'approche offrait bien un support conceptuel, il lui manquait cependant une démarche systématisée et rationalisée de modélisation des traces. Nous avons donc concrétisé notre problématique initiale dans le cadre particulier de l'approche des SBTm et nous sommes donnés pour objectif de concevoir une méthodologie de modélisation propre à l'approche des SBTm pour en faire une approche complète des traces numériques, tant sur le plan théorique, conceptuel, technique que méthodologique.

8.1.2 Apport méthodologique

Nous avons donc établi une première méthodologie de modélisation des traces en menant un *travail de recherche* à la fois *théorique* et *appliqué*. Le travail théorique a consisté à chercher dans la littérature les démarches de modélisation susceptibles de nous servir de support ou de source d'inspiration (Cha-

²⁰⁶ Travail mené par l'équipe Silex du LIRIS.

pitre 4). Nous avons ainsi exploré le domaine des IHM et en particulier la modélisation de tâche, puis l'IC et la modélisation des connaissances et enfin le *CSCW* et la modélisation des activités instrumentées individuelles et collectives. Nous avons montré en quoi la modélisation de tâche ne pouvait être notre unique source d'inspiration et expliqué la nécessité de faire appel aux approches de modélisation des connaissances. Même si ces dernières étaient plus proches de notre démarche, il a fallu l'apport complémentaire du *CSCW* pour donner à un modélisateur les outils d'une observation fine nécessaire à la modélisation des traces.

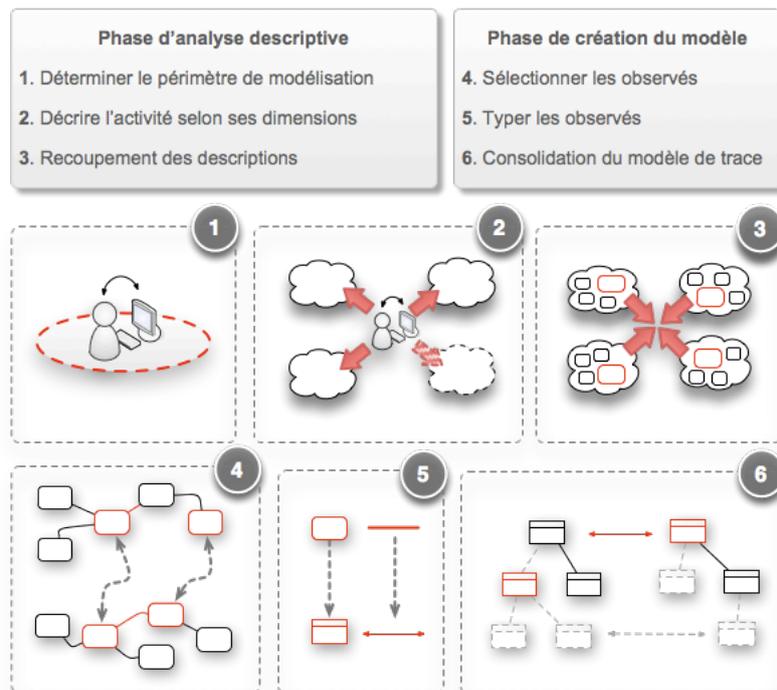


Figure 8.1 : Résumé des six étapes du processus générique de modélisation de trace numérique.

Nous avons tiré de ces sources une méthodologie composée principalement de trois éléments (Chapitre 5) : (a) un processus de modélisation générique des traces numériques, (b) une spécialisation de ce processus en fonction du contexte d'usage des traces et (c) une intégration de la modélisation dans une démarche englobant également la conception et la mise en place d'un SGBT que le modélisateur devra accompagner. Le processus de modélisation générique que nous proposons est lui-même composé de 6 étapes regroupées en deux phases, ce que rappelle le schéma ci-dessus (Figure 8.1).

La méthodologie ainsi présentée est également le fruit d'un travail de recherche appliqué. Nous avons confronté directement notre démarche à la réalité du travail de modélisation sur le terrain (Chapitre 6). Notre expérience sur un premier terrain d'application (une activité d'apprentissage dans une classe virtuelle) nous a permis de constater à quel point il était important d'offrir un cadre structurant au modélisateur de trace, ce dont nous avons tenu compte dans la mise en place des étapes de notre méthodologie que nous avons ensuite éprouvée dans le cadre d'un second terrain d'application. Pour ce deuxième travail de modélisation nous avons été porteur du projet de mise en place d'un SBTm pour l'activité d'une équipe d'ingénieurs pédagogiques concevant des formations (y compris la production des supports). Les traces modélisées y ont été conçues de manière à être le support d'une *réflexivité* de l'activité des développeurs de formation.

8.1.3 Apport expérimental et instrumental

En parallèle du travail de modélisation lui-même, nous avons développé et mis en place sur ce second terrain une *situation expérimentale* visant à valider indirectement notre approche en évaluant le *potentiel des traces modélisées* utilisées en contexte. Le protocole choisi fût une interruption / reprise d'activité. Nous avons pour cela mobilisé deux sujets qui ont entamé, interrompu, puis repris une activité de développement de formation, dans leur environnement de travail dédié. Lors de la reprise, seul l'un des deux sujets bénéficiait d'une trace modélisée de son activité passée. L'hypothèse de départ, de manière grossière, était que la trace modélisée faciliterait la reprise de l'activité en permettant au sujet d'adopter une posture réflexive relativement à son activité. Le bilan que nous tirons de cette expérimentation est double. Tout d'abord la nécessité de raffiner le contexte d'usage : la trace modélisée a permis *d'objectiver* une activité à travers une trace modélisée, sans pour autant se montrer un support décisif pour une reprise effective de l'activité interrompue. Ensuite le constat que la dimension *collective* de réutilisation des traces joue également un rôle qui peut remettre en cause jusqu'au cadre conceptuel adopté au départ.

À l'occasion de ce travail d'expérimentation sur le terrain nous avons également développé et utilisé un certain nombre d'outils. Nous avons fabriqué une chaîne « d'édition assistée » de traces d'interaction collectées à la fois par l'observation directe et par la récupération des *logs* d'application. La mise au point de ce dispositif a nécessité le développement (*java*) d'un « transcodeur » et d'un « éditeur » d'observés visant à produire une trace visualisable dans un outil de visualisation. L'outil en question, la *Timeline* (Simile), a elle-même nécessité un travail d'installation et surtout de paramétrage pour fournir aux utilisateurs un interface d'interaction avec la trace d'une utilisabilité maximum. Nous espérons bien entendu que ce travail pourra être réutilisé et amélioré dans le contexte d'autres utilisations de traces modélisées.

8.1.4 Apport conceptuel

Le dernier apport de cette thèse est conceptuel (Chapitre 7). Nous avons en effet expliqué qu'une situation d'activité collective induisait nécessairement des usages collectifs des traces modélisées. Or, si ces usages ne peuvent se développer sans un long processus d'appropriation des traces d'abord au niveau individuel, puis collectif, de la part des acteurs, l'introduction d'une trace modélisée dans une telle activité expose immédiatement la trace en question à une multitude d'usages possibles qu'il s'agit de pouvoir instrumenter quelles que soient les conditions. Nous avons donc proposé d'étendre le cadre conceptuel des SBTm et d'ajouter un « sujet » au concept de trace modélisée, permettant de distinguer des traces modélisées *individuelles* et des traces modélisées *conjointes*. Ce cadre étendu permet d'envisager, sur un plan informatique, la gestion générique des opérations nécessaires à l'instrumentation d'une trace pour l'analyse aussi bien que pour des usages collectifs de ces traces. Ce nouveau cadre ouvre également de nouvelles perspectives quant au développement de nouvelles applications de l'approche des SBTm. Concernant la méthodologie de modélisation proposée, son processus de modélisation générique qui prenait déjà explicitement en compte la dimension collective de l'activité reste donc solide, mais il faudra plusieurs applications réelles avant de pouvoir affirmer que les contextes d'usages collectifs, coopératifs et collaboratifs ne nécessite pas une spécialisation particulière du processus.

8.2 Limites et commentaires

Le travail mené dans cette thèse comporte des limites dont certaines sont propres à tout travail de recherche exploratoire. Pour quelques unes, nous les avons évoquées au cours de la présentation des différentes parties du travail en question. En cette fin de manuscrit, il nous paraît important cependant de revenir sur les plus importantes d'entre elles.

8.2.1 Critique du travail réalisé

Les premières remarques concerneront le travail proposé dans cette thèse. Tout d'abord il faut être conscient qu'il s'agit d'une première tentative de méthodologie de modélisation de cette ampleur. À ce titre elle peut comporter des défauts communs à tout travail de recherche exploratoire. C'est d'autant plus vrai qu'une méthodologie n'est réellement aboutie qu'après de nombreuses applications. Nous avons fait de notre mieux et travaillé successivement sur deux terrains d'application ce qui nous a permis de réagir aux plus gros problèmes et adapter immédiatement la méthodologie de modélisation en question.

8.2.1.1 La précision de la méthodologie proposée

Parmi les reproches qu'on pourrait faire à notre travail on pourrait évoquer un manque de précision et de prescription du processus de modélisation. Malgré la structuration que nous avons tenu à lui conférer on pourrait objecter que cette méthodologie ne prescrit pas suffisamment le travail du modélisateur en mettant davantage l'accent sur les objectifs à atteindre au cours de la modélisation que sur les outils à utiliser pour le faire, alors que des outils de représentation graphiques sont présents dans certains des domaines convoqués, notamment en Ingénierie des Connaissances. En effet, nous parlons par exemple de schématisation sans prescrire un langage de représentation particulier. Si nous avons effectivement choisi de ne pas prescrire un outil particulier dans une première proposition de méthodologie c'est tout d'abord parce que nous la souhaitons suffisamment générique pour répondre à la très grande diversité des contextes de modélisation des traces modélisées et ensuite parce qu'il faut rester prudent sur ce type d'emprunts. Par exemple le travail de modélisation en *CSCW* notamment rencontre souvent les plus grandes difficultés à se conformer à un langage de représentation ou à une convention graphique pour mener ses analyses. Les chercheurs sont souvent obligés de développer leur propre solution de représentation de manière *ad hoc* sur le terrain et en fonction des contraintes particulières de celui-ci.

Cette remarque sur les outils de représentation graphique peut sans doute être portée à des degrés divers sur l'ensemble de notre proposition. Nous avons mobilisé une large littérature, et empruntés des idées et concepts à de très divers champs de recherche en assumant la pluridisciplinarité de cette démarche, mais on pourrait reprocher à ce travail d'avoir manqué de profondeur dans l'exploitation de ces diverses ressources. En Ingénierie des Connaissances par exemple, les approches de modélisation de résolution de problème n'ont pas été pleinement exploitées, et des outils comme *commonKads* pas assez utilisés. Cette critique est bien entendu recevable, toutefois nous défendons notre démarche comme celle d'un travail exploratoire. Notre objectif était de d'identifier les apports méthodologiques dans les disciplines et domaines pouvant se rapprocher de la méthodologie que nous avons à réaliser. Nous en avons parcouru trois, IHM, IC et *CSCW* dont nous avons tirés des principes et une structuration appliquée dans la méthodologie proposée. La prescription d'outils demanderait plusieurs applica-

tions de la méthodologie de modélisation avec plusieurs d'entre eux, pour pouvoir les évaluer, les comparer et en retenir les meilleurs candidats. Pour en arriver à ce point, un autre travail de recherche doit être mené autour de cet objectif précis et il faudra de plus montrer l'avantage, pour une situation donnée, de l'utilisation d'un outil prescrit en lieu et place d'une approche plus empirique.

8.2.1.2 La périmètre de la méthodologie proposée

Pour terminer on peut également critiquer les limites de la portée de la méthodologie proposée. Nous avons souhaité contribuer à l'approche des Systèmes à Base de Traces modélisées en la dotant d'une méthodologie de modélisation propre. Dans l'absolu, cette méthodologie devrait être applicable quel que soit le type d'activité informatique considérée. Or, pour cette première proposition de méthodologie nous nous sommes concentrés sur une famille particulière d'activités, les activités documentaires instrumentées (ADI). Notre volonté de traiter ces ADI de la manière la plus générique possible n'est sans doute pas suffisante pour garantir *a priori* l'efficacité et la pertinence de notre méthodologie pour les autres types d'activités. Il faudra encore plusieurs applications de cette première proposition, dans des contextes de modélisation très différents, avant de pouvoir prétendre à une méthodologie pleinement aboutie. Ce travail devrait être facilité par la finalisation d'un *framework* informatique qui permettra de soutenir techniquement la mise en place d'un SGBT, donnant toute latitude au modélisateur pour réaliser son travail.

8.2.2 Commentaires sur l'approche proposée

Au-delà des critiques adressables à notre travail, il existe également des limites intrinsèques à l'approche proposée. Nous allons ici en évoquer trois : le *coût* d'un travail de modélisation, la difficulté de *l'évaluation* et le manque d'une approche de la *présentation des traces* numériques.

8.2.2.1 Le coût d'une modélisation de trace

Le travail de modélisation de trace qu'exige la méthodologie mise en place est un travail relativement coûteux pour le modélisateur. Notre proposition repose en effet sur une analyse très fine de l'activité qui peut se révéler fastidieuse. Nous avons fait le choix, pour la sélection des observés, de procéder d'abord à une mise en évidence du plus d'observés potentiels possibles, pour ensuite effectuer une sélection définitive ce qui n'est sans doute pas la solution la plus rapide. C'est une critique commune à d'autres approches faisant appel à des analyses très fines de l'activité comme c'est le cas avec la Cognition Distribuée (section 4.4.3.2). L'investissement demandé au modélisateur apparaît sans doute de manière plus importante encore du fait qu'il se situe en amont des premiers résultats tangibles du côté du système, *i.e.* qu'il faut attendre qu'un long travail d'analyse soit mené à bien avant de voir les premières traces modélisées concrètement générées par un SGBT. Cet investissement est cependant nécessaire pour que la trace modélisée produite soit une représentation qualitative de l'activité et à se titre directement réutilisable.

Dans les approches qui produisent rapidement des traces à l'aide de *logs* récupérés directement dans un système en fonctionnement les traitements peuvent également demander de longues heures de travail sans pour autant pouvoir prétendre à une trace « qualitative ». Il est vrai également que sur le terrain, le modélisateur doit se donner les moyens d'une analyse fine nécessitant l'adhésion et la participation des utilisateurs du système, et doit également gérer les problèmes techniques d'implémentation

d'un SGBT, soit par lui-même, soit en collaboration avec un codeur. Pour justifier un tel investissement le modélisateur doit mettre en place rapidement un modèle de trace opérationnel, mais également les moyens de son évolution de l'activité tracée et celle de l'utilisateur de la trace lui-même. La méthodologie proposée invite à mettre le travail de modélisation en perspective : il ne se termine pas à la fin d'une phase d'amorçage d'un SBTm et il doit accompagner son déploiement et ses premiers usages réels.

8.2.2.2 L'évaluation de la méthodologie

Une autre limite à signaler est celle induite par le problème de l'évaluation de la méthodologie de modélisation que nous venons de proposer. Nous ne disposons en effet pas de moyens directs d'évaluer une telle méthodologie. Nous avons mis en place une expérimentation pour confronter l'utilisateur d'un atelier de production de contenu à sa propre activité. Nous avons utilisé pour cela un outil de visualisation simple que nous avons modifié au mieux. Les conclusions que nous pouvons tirer sur la pertinence ou l'utilisabilité de la trace proposée restent sujettes à discussion. Il est en effet difficile d'établir ce qui est réellement évalué, modèle de trace, visualisation, interaction possible avec la trace, situation dans son ensemble, *etc.* C'est un problème commun aux situations « d'évaluation indirecte » où les critères d'évaluation sont soit superficiels, soit impliquent plusieurs facteurs d'explication. L'évaluation de l'impact des traces sur l'activité d'un utilisateur est déjà en elle-même indirecte. Dans le cas d'une trace modélisée exploitée par un apprenant on ne pourra pas objectivement mesurer la part prise par les « traces induisant des processus méta-cognitifs » dans le succès (ou non) de l'apprentissage qui est en jeu. Il en va de même pour évaluer la méthodologie de modélisation de trace utilisée pour aboutir aux traces modélisées en question.

8.2.2.3 La présentation des traces

La *présentation* des traces modélisées est également un point critique. Nous entendons par présentation la *visualisation* et les possibilités *d'interaction* qu'offre une trace modélisée présentée à son utilisateur. Nous n'avons pas disposé du temps nécessaire dans le périmètre de cette thèse pour consacrer à cette question le travail de recherche qu'elle exigerait par ailleurs. Quel que soit le projet d'usage de la trace, qu'il s'agisse d'analyse ou de réflexivité de l'activité, la qualité de la visualisation de celle-ci est primordiale, tant pour l'analyste qui y cherche des motifs remarquables que pour l'utilisateur qui y cherche la représentation de sa propre activité. Il est évident que même une trace modélisée dotée d'un modèle hypothétiquement parfait resterait inexploitable en situation sans une visualisation permettant de l'exprimer. Nous ne minimisons donc pas l'importance de la visualisation, y compris dans des approches exploratoires comme la notre. En témoigne le soin que nous avons apporté à l'amélioration visuelle de notre *Timeline*. Cependant, nous nous sommes concentrés sur la modélisation elle-même et le choix des observés pertinents pour une activité donnée, en défendant l'idée qu'une visualisation ne peut mettre en valeur une trace modélisée pertinente que si celle-ci dispose à la base des bonnes informations à visualiser.

Parmi les futurs travaux à envisager se dessinent donc deux sujets inévitables, avec d'une part la visualisation de traces d'activités, et d'autre part la dimension interactive de ces traces. Certains travaux de l'équipement ont commencé à aborder le problème d'un point de vue technique (Cram, 2007), mais une étude plus approfondie ne pourra être menée qu'en étroite collaboration avec un travail de modéli-

sation de trace sur le terrain. Il faudra ensuite envisager plusieurs alternatives de visualisation pour évaluer et démontrer les avantages et inconvénients des différentes interfaces possibles. De même pour les interactions, qui ne peuvent être réellement conçues et évaluées que dans des conditions d'utilisation concrètes, ce qui passe par une appropriation de la trace modélisée comme objet à part entière de l'activité de leurs utilisateurs. Ceci implique une utilisation prolongée des traces, ce qui n'a pas encore été le cas dans les applications réalisées jusqu'ici. Ces pistes de recherche font partie des perspectives de prolongement de ce travail de thèse.

8.3 Perspectives

Nous pouvons évoquer pour finir les perspectives qu'offre notre travail à deux niveaux : tout d'abord le prolongement du travail réalisé sur les terrains d'application abordés (section 8.3.1), puis les pistes de développement de la méthodologie proposée (section 8.3.2) et celles de la poursuite de la recherche sur les traces modélisées en général (section 8.3.3).

8.3.1 Prolongement du travail de terrain

Concernant notre premier terrain d'application, c'est-à-dire les activités d'apprentissage à distance sur la plateforme *eLycée*, le projet lui-même est pour le moment gelé à cause de difficultés rencontrées par la société qui commercialise la plateforme. Le travail de modélisation de trace réalisé dans le cadre de ce projet quant à lui pourra être réutilisé et approfondi dans le cadre de projets similaires qui débutent actuellement au sein de l'équipe Silex²⁰⁷. Le contexte reste celui de l'apprentissage grâce à une plateforme d'enseignement à distance et notre objectif restera de créer des traces modélisées susceptibles d'être un support à l'activité réflexive des apprenants.

Concernant le second terrain d'application et l'atelier de création de formation (*Emulsion*), le travail engagé est appelé à se poursuivre concrètement. En effet, l'équipe des créateurs de l'atelier s'est montrée intéressée par notre démarche sur la question des traces de l'activité de création de contenus. En s'appuyant sur le travail que nous avons réalisé de manière *ad hoc* pour les besoins de notre expérimentation, les développeurs du logiciel ont depuis notre intervention codé un module (Java) qui permet de générer des *logs* issus de l'atelier *Emulsion* directement au format *XML* compatible avec la *Timeline* que nous avons utilisée (section 6.3.5.1). Nous envisageons donc de travailler directement avec *Solunea* pour implémenter complètement le modèle de trace prévu sans avoir recours à des corrections manuelles comme nous avons dû le faire pour notre expérimentation. Grâce aux terrains de recherche potentiels que constituent les situations d'usages de l'atelier chez les clients de *Solunea*, nous pensons qu'un projet pourrait être monté pour soutenir de futurs travaux en partenariat.

De plus, une nouvelle version de la *Timeline* a été mise en ligne depuis la fin de notre intervention sur le terrain. Plus rigoureuse dans sa programmation, elle utilise désormais exclusivement du *CSS / XSLT* pour sa mise en forme et surtout intègre des possibilités d'interaction avec la trace présentée comme l'affichage différencié des événements selon des catégories prédéfinies, mais également surlignage de

²⁰⁷ C'est le cas notamment du projet ITHACA (voir <http://liris.cnrs.fr/~ithaca/fr/apropos>).

mots clefs résultats d'une requête effectuée sur la *Timeline*. Ces nouvelles fonctionnalités nous permettent d'envisager d'autres expérimentations de réutilisation des traces avec cette fois des possibilités d'interaction.

Concernant la situation spécifique rencontrée dans le service formation (SF), nous pensons que la principale perspective relève des *traces conjointes*. Un effet de bord de notre travail de modélisation de trace a été de mettre en évidence le fait que le processus de création d'une formation n'était pas capitalisable car non représenté explicitement dans l'environnement numérique de travail. Une trace, même minimale de ce processus effectif serait un support à cette capitalisation. Pour le moment, nous ne disposons pas sur place des moyens d'implémenter un SBTm pour des traces conjointes. Dans un premier temps, nous pensons que cela ne constitue pas un obstacle à la poursuite de notre travail. En effet, les membres de l'équipe ont une pratique reposant fortement sur *l'utilisation du papier*, même si cela change petit à petit.

Nous pensons que l'usage conjoint des traces exige des processus d'appropriation qui peuvent être relativement longs et qui demandent à être *accompagnés*. Il nous paraît donc intéressant dans ce contexte de commencer à introduire les traces modélisées dans l'activité *individuelle* et de proposer un partage de traces modélisées médié par une impression papier. Ce support convient (en tout cas pour le moment) à la culture d'échange direct qui imprègne aujourd'hui les pratiques du service. Un accompagnement de l'introduction des traces conjointes sur ce mode serait une source d'information précieuse pour enrichir notre méthodologie d'une spécialisation aux traces conjointes dans l'optique d'une facilitation du travail collaboratif. Ce type de projet est lié aux perspectives plus larges que nous pouvons envisager pour l'approche des Systèmes à Base de Traces modélisées en général.

8.3.2 Développement de la méthodologie de modélisation des traces

Bien que nous l'ayons directement confrontée à une utilisation concrète sur le terrain durant sa conception, notre méthodologie de modélisation reste une première proposition que nous envisageons bien entendu de renforcer. En premier lieu, elle devra être mise à l'épreuve de plusieurs nouvelles applications concrètes afin de démontrer sa robustesse et/ou de révéler ses faiblesses. À court terme et en complément du prolongement du travail entamé durant cette thèse (voir section précédente), nous allons travailler dans le cadre du projet ITHACA²⁰⁸ à une nouvelle application de notre méthodologie sur un terrain similaire à celui d'*eLycée*. À moyen terme, d'autres applications sont envisagées qui prendront place dans le développement de l'approche des traces modélisées dont nous parlerons ci-après. L'application concrète de la méthodologie n'est toutefois pas la seule voie par laquelle nous comptons contribuer au renforcement de notre méthodologie de modélisation.

En second lieu en effet, nous avons pour projet de renforcer notre méthodologie de modélisation sur des aspects de modélisation plus « techniques ». Nous pensons notamment tester l'utilisation de langages de représentation graphiques empruntés à l'Ingénierie des Connaissances dans le cadre de l'analyse descriptive d'une part et dans la schématisation nécessaire à la sélection des observés (voir section 5.3). Nous pensons tester des langages répandus tels *SADT* et ses « actigrammes » (Marca et MacGowan, 1987) ou *CML* de *CommonKads* (Schreiber et al., 1994) mais également étu-

²⁰⁸ ITHACA (voir <http://liris.cnrs.fr/~ithaca/fr/apropos>).

dier le potentiel de proposition moins prestigieuses comme par exemple *SeeMe*²⁰⁹ qui est destiné à la description d'activité.

Sur un autre plan, nous souhaitons également renforcer les fondements théoriques sur lesquels reposent la *spécialisation du processus de modélisation*, tout particulièrement en ce qui concerne la « réflexivité » de l'activité des utilisateurs. Nous avons en effet expliqué que dans le cas où une trace modélisée est destinée à être exploitée par l'utilisateur qui en est à l'origine, la trace devait servir de support à la réflexivité de l'activité, c'est à dire permettre et susciter des processus mnésiques et métacognitifs. Dans ce cas le modélisateur ne dispose pour fonder son analyse de l'activité, que de ses observations sur le terrain et de ses connaissances théoriques des processus en question (mnésiques et métacognitifs). Il nous paraît important de chercher à *compléter* ces connaissances pour mieux encadrer le travail d'analyse de l'activité et de modélisation des traces. Nous comptons donc dans de futurs travaux approfondir nos connaissances de ces processus psychologique en essayant de réunir des informations récentes et puisées dans des domaines de recherches qui ne sont pas directement liés à l'informatique (Psychologie, Neuro-psychologie, *etc.*), car la majorité des approches concernant les traces pour la réflexivité de l'activité se réfèrent encore aujourd'hui aux premiers travaux fondamentaux sur les processus métacognitifs.

Toujours concernant les traces modélisées pour la réflexivité de l'activité, il nous paraît également important de réinterroger la façon dont le problème se pose habituellement : comment une trace peut-elle représenter au mieux l'activité afin de tendre à l'utilisateur un « miroir » donnant sur sa propre activité. Nous pensons qu'une trace numérique ne doit pas chercher à « produire à la place » de l'utilisateur le résultat des processus que nous cherchons à soutenir. Une trace modélisée ne peut « remplacer » la *remémoration* d'une activité passée ni la *prise de conscience* par utilisateur de son propre raisonnement. Elle doit être pensée pour *permettre* et *susciter* ces processus.

Or, de manière générale, rien ne permet d'affirmer qu'une trace modélisée suscite ou non de tels processus à un moment donné. Face à ce problème, nous affirmons qu'il existe au moins une situation dans laquelle c'est le cas : lorsqu'un utilisateur *modifie lui-même* sa trace d'activité dans le but de la réutiliser. En effet, si un utilisateur modifie sciemment une trace modélisée, s'il la « documentarise » (Zacklad, 2005a ; Yahiaoui et *al.*, 2009), l'enrichie, l'adapte pour *prescrire sa propre interprétation* de la trace pour plus tard alors il s'est *nécessairement approprié* la trace modélisée comme une représentation de son activité effective, ce qui implique qu'il a également mis en œuvre les processus métacognitifs que nous cherchons à soutenir. La possibilité d'une trace modélisée visualisable *et* modifiable par l'utilisateur n'était pas une option à notre disposition lors de notre thèse. Nous espérons donc trouver les moyens de poursuivre nos travaux sur cette piste. Cette possibilité représente également un défi pour notre méthodologie de modélisation de trace qui devra intégrer les contraintes de conception liées aux interactions possibles de l'utilisateur avec une trace modélisée « modifiable ».

²⁰⁹ Voir <http://web-imtm.iaw.ruhr-uni-bochum.de/iug/projekte/seeme/inhalt/index.html> (consulté le 12/03/2007).

8.3.3 Nouvelles applications et nouveaux usages des traces modélisées

8.3.3.1 Les traces modélisées de Recherche d'Information

L'amélioration de notre méthodologie est évidemment un objectif important, mais il n'est pas le seul. Nous avons également en perspective de nouveaux terrains d'application de l'approche des Systèmes à Base de Traces modélisées. Nous avons évoqués les terrains d'application liés à des projets existants, mais nous espérons également ouvrir une nouvelle voie de recherche prenant pour objet la Recherche d'Information. C'est en effet une activité que nous n'avons abordée dans cette thèse que d'un point de vue théorique et non expérimental, situation qui changera prochainement. Le domaine des Sciences de l'Information et de la Communication offre de nombreux modèles comportementaux (descriptifs et/ou opératoires) de la Recherche d'Information. Dans ce contexte, notre idée est d'utiliser les traces modélisées en tant qu'instrument d'analyse « qualitatif-quantitatif ».

Nous souhaitons tout d'abord sélectionner un modèle et créer une trace modélisée qui permette d'implémenter les concepts, de sorte que le SBTm mis en place soit un instrument d'analyse de la Recherche d'Information relativement à ce modèle « comportemental ». Nous voulons démontrer le potentiel de l'approche des SBTm pour conduire une analyse articulant dimension quantitative et qualitative. Les résultats d'une expérimentation sur ce point permettraient peut-être de faire évoluer le modèle de Recherche d'Information exploité. Dans le prolongement de cette exploitation des traces pour l'analyse, notre ambition est de passer par la suite à une trace destinée à la réflexivité de l'activité d'un utilisateur en Recherche d'Information. En d'autres termes, nous souhaitons offrir une trace modélisée soutenant la réflexivité d'une activité de Recherche d'Information en basant l'analyse de l'activité nécessaire sur une première exploitation des traces modélisées, cette fois-ci comme support à l'analyse. Cette démarche pourrait nous conduire à généraliser une approche des traces englobant les deux grands types d'exploitation des traces modélisées que nous avons distingués jusqu'ici.

8.3.3.2 Les traces conjointes

L'autre champ de recherche qui s'ouvre à la fin de cette thèse est lié au cadre conceptuel étendu des Systèmes à Base de Traces conjointes que nous avons proposé. Comme nous l'avons montré, de nouveaux (et nombreux) usages des traces modélisées en situation sont imaginables. En tant qu'objet de recherche les traces conjointes devront mobiliser pleinement deux de domaines que nous avons rapprochés dans cette thèse : l'Ingénierie de Connaissances et le *CSCW*. Le travail sur la modélisation de traces conjointes doit être poursuivi et nous comptons mettre à l'épreuve notre méthodologie sur ce terrain et l'étendre si besoin. Pour concrétiser ce travail nous pensons que le terrain le plus pertinent pourrait être celui des *mémoires de projet*. Les mémoires de projet sont un exemple de sujet sur lequel de nombreux systèmes ont été proposés, sans pour autant fournir un support facilement utilisable pour y inscrire les connaissances accumulées lors de la réalisation d'un projet. Les traces modélisées pourraient là aussi être une alternative intéressante en constituant un support flexible d'inscription de l'expérience d'un acteur ou d'un groupe d'acteur.

Il ne s'agirait plus de constituer, de construire (difficilement) un mémoire de projet qu'il faudra entièrement se réapproprier pour le réutiliser par la suite. Les traces modélisées des acteurs sont à la fois une inscription des interactions qui ont eu lieu et un support d'inscription pour les acteurs qui y inter-

prêtent leur propre activité. Lors de leur réutilisation ces traces ont l'avantage de décrire l'activité qui a eu lieu dans les termes de l'activité effective tout en étant « documentarisable » pour faciliter ou orienter leur interprétation dans ce cadre. De notre point de vue, les traces modélisées représentent pour l'IC un objet de recherche tout à fait important. Notre travail sur les traces numériques répond d'ailleurs, à sa manière, à une préoccupation centrale de l'IC qui « n'aura donc pas d'expérience scientifique, mais des *expériences humaines d'utilisation dont il faudra garder la mémoire* » (Bachimont, 2004b).

Références

- (Aamodt et Plaza, 1994) Aamodt A., Plaza E., *Case-based reasoning: foundational issues, methodological variations, and system approaches*, AI Communications, vol.7, n°1, mars 1994, pp. 39-59.
- (Aamodt, 2001) Aamodt A., *Modeling the knowledge contents of CBR systems*, Actes de 4th International Conference on CBR, Vancouver, 2001, Naval Research Lab Technical Note AIC-01-003, pp. 32-37.
- (Agre et Chapman, 1987) Agre P.E. et Chapman D., *PENGI: A implementation of a theory of activity*. In : Proc. 6th National Conference on Artificial Intelligence. Menlo Park, CA, Morgan Kaufmann, 1987, pp. 268-272.
- (Anderson et al., 1996) Anderson J. R., Reder L.M., Simon H., *Situated Learning and Education*. In : Educational Researcher, vol. 25, n°4, 1996, pp. 5-11.
- (Anjewierden et Efimova, 2006) Anjewierden A., Efimova L., *Understanding weblog communities through digital traces: a framework, a tool and an example*. In : Proc. OTM Confederated International Workshops, Springer LNCS 4278, Montpellier : France, 2006., pp. 279-289.
- (Arana et al., 2004) Arana-Lozano J., Hassas S., Prié Y., *Mazette: multi-agent Musette for sharing and reusing ontologies*. In : WOSE workshop on ontologies, semantics and e-learning, OTM Workshops, Springer LNCS 3292, 2004, pp. 741-752.
- (Aussenac, 1989) Aussenac N., *Conception d'une méthodologie et d'un outil d'acquisition des connaissances expertes*, Thèse de l'Université Paul Sabatier, Toulouse, Octobre 1989.
- (Aussenac-Gilles et al., 1992) Aussenac-Gilles N., Krivine J.P., Sallantin J., Editorial. In : Revue d'Intelligence Artificielle. Hermès, numéro spécial sur l'acquisition des connaissances, vol.6, n°1-2, 1992, pp. 7-18.
- (Aussenac-Gilles et al., 2005) Aussenac-Gilles N., Biébow B., Sulzman S., *Modélisation du domaine par une méthode fondée sur l'analyse de corpus*, In : Ingénierie des connaissances, Paris : L'Harmattan, 2005, p. 49-71.
- (Avouris et al., 2004) Avouris N., Margaritis M., Komis V., *Modelling interaction during small-group synchronous problem-solving activities : The Synergo approach*, In : 2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction, ITS2004, 7th Conference on Intelligent Tutoring Systems, Maceio, Brasil, 2004, pp. 13-18.
- (Avouris et al., 2005) Avouris N., Komis V., Fiotakis G., Margaritis M., Voyiatzaki E., *Logging of fingertip actions is not enough for analysis of learning activities*, In : Proc. Workshop Usage Analysis in learning systems, AIED 2005, Amsterdam, 2005, pp.1-8.
- (Avouris et al., 2007) Avouris N., Dimitricopoulou A., Komis V., Margaritis M., *Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data*. In : Journal of Interactive Learning Research JILR, vol.18, n°2, (Special Issue) Usage Analysis in Learning Systems : Existing Approaches and Scientific Issues, 2007, pp. 231-250.
- (Bachimont, 1993) Bachimont B., *Nature, culture et artefacture : la place de l'intelligence artificielle dans les sciences cognitives*. In : Intellectica, vol.26, n°17, 1993, pp. 213-238.
- (Bachimont, 2000) Bachimont B., *Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances*. In : Charlet et al. (eds.), *Ingénierie des connaissances : évolutions récentes et nouveaux défis*, Paris : Eyrolles, Collection Technique et Scientifique des Télécommunications, 2000, chap..19, pp. 305-323.
- (Bachimont, 2004a) Bachimont B., *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire de HDR, UTC, Compiègne, 2004, 267 p.
- (Bachimont, 2004b) Bachimont B., *Pourquoi n'y a-t-il pas d'expérience en ingénierie des connaissances ?* In : Actes de la conférence Ingénierie des Connaissances, Lyon, 2004. pp. 55-64.
- (Badre et al., 1995) Badre A.N., Guzdial M., Hudson S.E., Santos P.J., *A user interface evaluation environment using synchronized video, visualizations, and event trace data*. In : Software Quality Journal, Springer (ed.), vol.4, 1995, pp. 101-113.

- (Baker et al., 2003) Baker M.J., Quignard M., Lund K., Séjourné A., *Computer-supported collaborative learning in the space of debate*. In : Wasson B., Ludvigsen S., Hoppe U. (Eds.), *Designing for Change in Networked Learning Environments : Proc. International Conference on Computer Support for Collaborative Learning*, Dordrecht : Kluwer Academic Publishers, 2003, pp.11-20.
- (Balbo et Coutaz, 1993) Balbo S., Coutaz J., *Modèles de tâche : Analyse comparative, utilité et limitations*. In : In Actes des 5^{èmes} journées sur l'ingénierie des interfaces homme/machine, IHM'93, 1993, Lyon, pp.131-137.
- (Balbo, 1994) Balbo S., *Evaluation ergonomique des interfaces utilisateur : un pas vers l'automatisation*. Thèse Informatique, Grenoble, 5 septembre 1994, Université J.Fourier Grenoble 1, 288 p.
- (Baneyx, 2007) Baneyx A., *Construire une ontologie de la pneumologie : aspects théoriques, modèles et expérimentations*. Thèse, Université Pierre et Marie Curie (Paris VI), février 2007. Disponible sur <http://tel.archives-ouvertes.fr/tel-00136937/fr/> ou <http://www.edithese.com/doc/8527>.
- (Bardram, 1998) Bardram J., *Collaboration, Coordination, and Computer Support - An Activity Theoretical Approach to the Design of Computer Supported Cooperative Work*. PhD Thesis, Daimi PB 533, University of Aarhus, Aarhus, 1998.
- (Bardram et al., 2006) Bardram J., Bunde-Pedersen J., Soegaard M., *Support for activity-based computing in a personal computing operating system*. In : Proc. SIGCHI Conference on Human Factors in Computing Systems, New York, ACM, 2006, pp.211-220.
- (Baron, 2002) Baron M., *Intégration d'un modèle de tâche dans une démarche sûre de construction d'interface*. In : actes 14^{ème} Conférence Interaction Homme/Machine, vol.32, New York, ACM, 2002, pp.73-80.
- (Basque et al., 2008) Basque J., Paquette G., Pudelko B., Léonard M., *Collaborative Knowledge Modeling with a Graphical Knowledge Representation Tool: A Strategy to Support the Transfer of Expertise in Organizations*. In : Okada A., Shum S.B., Sherborne T. (eds), *Knowledge Cartography. Mapping Techniques and Software Tools*. Londres, Springer-Verlag, 2008, pp.357-382.
- (Beauvisage, 2004) Beauvisage T., *Sémantique des parcours des utilisateurs sur le Web*. Thèse Sciences du langage. Paris : Université Paris X – Nanterre, 2004, 361p.
- (Benford et al., 1999) Benford, S., Taylor, I., Brailsford, D., Koleva, B., Craven, M., Fraser, M., Reynard, G., Greenhalgh, C., *Three dimensional visualisation of the world wide web*. New-York : ACM Computing Surveys vol.31, n°4, 1999, art.25.
- (Bergman et al., 2005) Bergman L., Castelli V., Lau T., Oblinger D., *DocWizards : a system for authoring follow-me documentation wizards*, In : Proc. 18th annual ACM Symp. on User Interface Software and Technology, UIST2005, 2005, pp.191-200.
- (Berlage, 1994) Berlage T., *A Selective Undo Mechanism for Graphical User Interfaces Based On Command Objects*. In : ACM Transactions on Computer-Human Interaction, vol.1, n°3, New-York, 1994, pp. 269-294.
- (Berners-Lee, 1998) Berners-Lee T., Fielding R., Masinter L., *Uniform Resource Identifiers (URI) : Generic Syntax*. RFC 2396, The Internet Society, August 1998 (en ligne) <http://www.ietf.org/rfc/rfc2396.txt>.
- (Berners-Lee et al., 2001) Berners-Lee T., Handler J., Lassilla O., *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American, 2001.
- (Bouaud, 1989) Bouaud J., *K: un Langage pour l'Implémentation d'Outils de Représentation des Connaissances*. Thèse de doctorat Informatique, Université Paris VII, 1989.
- (Bouaud et al., 1995) Bouaud J., Bachimont B., Charlet J., Zwetgenbaum P. (1995). *Methodological principles for structuring an « ontology »*. In : Proc. IJCAI'95 Workshop on « Basic Ontological Issues in Knowledge Sharing », Montréal, Canada, 1995.
- (Bourguin, 2000) Bourguin G., *Un support informatique à l'activité coopérative fondé sur la Théorie de l'Activité : le projet DARE*, Thèse de doctorat Informatique, n°2753, Université des Sciences et Technologies de Lille, 2000, 210 p.
- (Bourguin et Derycke, 2005) Bourguin G., Derycke A., *Systèmes Interactifs en Co-évolution, Réflexions sur les apports de la Théorie de l'Activité au support des Pratiques Collectives Distribuées*, In : Revue d'Interaction Homme-Machine (RIHM), AFIHM Europa, vol. 6, n°1, 2005, pp.1-31.
- (Bratitsis et Dimitracopoulou, 2006) Bratitsis T., Dimitracopoulou A., *Indicators for measuring quality in asynchronous discussion forums*, In : International conference on Cognition and Exploratory Learning in Digital Era, International Association for Development of the Information Society, Barcelona, 2006.

- (Bressole et al., 1996) Bressole M. C., Decortis F., Pavard B., Salembier P., *Traitement cognitif et organisationnel des micro-incidents dans le domaine du contrôle aérien : analyse des boucles de régulation formelles et informelles*. In : DeTerssac G. et Fridberg E. (eds), *Coopération et conception*, Toulouse, Octarès, 1996, pp. 267-287.
- (Briet, 1951) Briet S., *Qu'est-ce que la documentation ?* Paris: EDIT, 1951, 48 p.
- (Bringay et al., 2004) Bringay S., Barry C., Charlet J., Les documents et les annotations du dossier patient hospitalier, In : Salaun J.M. et Charlet J. (eds), *Le document numérique*. Numéro thématique de la revue Information - Interaction - Intelligence, Cépaduès : Toulouse. vol.4, n°1, 2004, pp. 191-211.
- (Bringay et al., 2005) Bringay S., Barry C. et Charlet J., *Une fonctionnalité d'annotation pour le dossier patient informatisé*. In : Beuscart R. et Brunetaud J.M. (coord.), *Actes des 11^{èmes} Journées Francophones d'Informatique Médicale*, Lille, 2005.
- (Bringay, 2006) Bringay S., *L'annotation pour supporter la collaboration dans le dossier patient électronique*, Thèse de doctorat en Informatique, Université de Picardie Jules Verne Amiens, 2006.
- (Brinkman et al., 2006) Brinkman W.P., Gray P., Renaud K., *Computer-Assisted Recording, Pre-Processing and Analysis of User Interaction Data*, Workshop, In : Brinkman W.P. (eds), *Proc. Human-Computer Interaction Conference*, Londres, 2006.
- (Brown, 1989) Brown J.S., Collins A., Duguid P., *Situated Cognition and the Culture of learning*. In : *Educational Researcher*, vol.18, n°1, 1989, pp.32-42.
- (Brown et Duguid, 2002) Brown J. S. et Duguid P., *The Social Life of Information*, Harvard Business Press, 2002, 330 p.
- (Brown et al., 2003) Brown B., MacColl I., Chalmers M., Galani A., Randell C., Steed A., *Lessons from the lighthouse : collaboration in a shared mixed reality system*, In : Proc. SIGCHI conference on Human Factors in Computing Systems, Ft. Lauderdale, ACM, New York, 2003, pp.577-584.
- (Brunie et al., 2000) Brunie V., Bachimont B., Mrizet-Mahoudeaux P., *Modélisation des connaissances structurées documentaires pour la conception d'un dossier médical hypertextuel*. In : Charlet et al. (eds), *Ingénierie des connaissances : évolutions récentes et nouveaux défis*. Paris : Eyrolles, 2000, chapter 25, pp. 407-422.
- (Bull et Kay, 2008) Bull S. et Kay J., *Metacognition and open learner models*. In : 3^{ème} Workshop on Metacognition and Self-Regulated Learning in Educational Technologies, à ITS2008, 2008, pp.7-20.
- (Bush, 1945) Bush V., *As We May Think*, In : The Atlantic Monthly, July 1945, Reprinted in *Interactions*, vol.3, n°2, 1996, pp.35-46.
- (Button, 1997) Button G., *Cognition in the wild* (review), In : *Computer Supported Cooperative Work : The Journal of Collaborative Computing*, vol.6, 1997, pp.391-395.
- (Button, 2008) Button G., *Against distributed cognition*. In : *Theory, Culture Society*, Sage (ed), vol.25, n°2, 2008, pp.87-104.
- (Cadez et al., 2000) Cadez I., Heckerman D., Meek C., Smyth P., White S., *Visualization of navigation patterns on a web site using model based clustering*. In : Proc. 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp.280-284.
- (Cahier et al., 2004) Cahier, J.-P., Zacklad, M., et Monceaux, A. *Une application du Web Socio Sémantique à la définition d'un annuaire métier en ingénierie*. In : Conférence Ingénierie des Connaissances, IC 2004, Lyon, 2004, pp.29-40.
- (Cahier, 2005) Cahier J.-P., *Ontologies sémiotiques pour le Web socio sémantique : Etude de la gestion coopérative des connaissances avec des cartes hypertextuelles*. Thèse de doctorat Informatique, Université de Technologie de Troyes, Troyes, 2005, 292 p.
- (Calvary, 1998) Calvary G., *Proactivité et Réactivité : de l'Assignment à la Complémentarité en Conception et Evaluation d'Interfaces Homme-Machine*, Thèse de doctorat Informatique de l'Université Joseph Fourier, 1998.
- (Card et al., 1983) Card S., Moran T.P., Newell A., *The Psychology of Human Computer Interaction*. Lawrence Erlbaum Associates, 1983.
- (Card et al., 2001) Card S., Pirolli P., Van Der Wege M., Morrison J.B., Reeder R.W., Schraedley P.K., Boshart J., *Information scent as a driver of Web Behavior Graphs : Results of a protocol analysis method for web usability*. In : Conference on Human Factors in Computing Systems, Seattle, WA., ACM, 2001, pp.498-505.

- (Cardon, 1997) Cardon Dominique, *Les sciences sociales et les machines à coopérer : Une approche bibliographique du Computer Supported Cooperative Work (CSCW)*, Réseaux, n°85, Lavoisier, 1997, pp.13-51.
- (Caroll et al., 1996) Carroll S., Beyerlein S., Ford M., Apple D., *The Learning Assessment Journal as a tool for structured reflection in process education*, In : Proc. Frontiers in Education'96, IEEE, 1996, pp.310-313.
- (Catledge et Pitkow, 1995) Catledge L.D., Pitkow J.E., *Characterizing browsing strategies in the World-Wide Web*. In : Computer Networks and ISDN Systems, vol.27, n°6, 1995, pp.1065-1073.
- (Chabin, 2004) Chabin M.A., *Document trace et document source. La technologie numérique change-t-elle la notion de document ?* In : Revue I3, vol.4, n°1, 2004, pp.141-158.
- (Champin et al., 2001) Champin P.A., Prié Y., Mille A., *Annotating with uses: a promising way to the semantic web*, In : Handschuh S., Dieng R., Staab S. (eds), *Workshop on Knowledge Markup and Semantic Annotation, K-CAP*, Victoria (Australia), 2001, pp.79-86.
- (Champin et al., 2002) Champin P.A., Prié Y., Mille A., *Conteneurs de connaissances : une approche fondée sur les usages pour le Web sémantique*, In : Reconnaissances de Formes et Intelligence Artificielle, Angers (France), 2002, 10 p.
- (Champin et al., 2003) Champin P.A., Prié Y., Mille A., *MUSETTE : Modelling USEs and Tasks for Tracing Experience*, In : WS5, From Structured Cases to Unstructured Problem Solving Episodes For Experience-Based Assistance at ICCBR'03, NTNU, Trondheim (Norway), 2003, pp. 279-294.
- (Champin et Prié, 2003) Champin P.A., Prié Y., *Musette: uses-based annotation for the Semantic Web*. In : S. Handschuh (eds.), *Annotation for the Semantic Web*, IOS Press, Amsterdam, 2003, pp. 180-190.
- (Champin et al., 2004) Champin P.A., Prié Y., Mille A., *MUSETTE : a Framework for Knowledge Capture from Experience*. In : Proc. EGC'04, papier court, vol.1, Clermont Ferrand (France), 2004.
- (Charlet, 2004) Charlet J., *L'ingénierie des connaissances, entre science de l'information et science de gestion*. Version allongée d'un chapitre de livre faisant suite au colloque de Cerisy, Activité, connaissance, organisation, 2004, Disponible sur http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/08/05/index_fr.html.
- (Chartrand, 2003) Chartrand M., *La veille stratégique, un outil de la décision et du changement*. Coup d'œil, 2003, vol.9, n°1, pp. 2-5.
- (Chatterjee et al., 2003) Chatterjee P., Hoffman D. L. et Novak T.P., (2003). *Modeling the Clickstream: Implications for Web-Based Advertising Efforts*. In : Marketing Science, vol. 22, n°4, 2003, pp. 520-541.
- (Chen et al., 2004) Chen J., Sun L., Zaiane O. R., Goebel R., *Visualizing and discovering web navigational patterns*. In : Proc. 7th International Workshop on the Web and Databases : colocated with ACM SIGMOD/PODS. Paris, France, 2004, pp. 13-18.
- (Chevalier et Ciaccia, 2009) Chevalier M., Ciaccia A., *Modélisation utilisateurs et systèmes d'information sur le web : pour une approche centrée usager*. In : actes de H2PTM'09. Paris, Hermès, 2009, pp. 34-46.
- (Chi, 2002) Chi E.H., *Improving web usability through visualization*. IEEE Internet Computing archive, 2002, vol. 6, n°2, pp. 64-71.
- (Chi et al., 1998) Chi E.H., Pitkow J., Mackinlay J., Pirolli P., Grossweiler R., Card S.K. *Visualizing the evolution of web ecologies*. In : Proc. ACM CHI 98 Conference on Human Factors in Computing Systems. ACM Press, Los Angeles, Californie, 1999, pp. 400-407.
- (Choo et al., 2000) Choo C.W., Detlor B. et Turnbull D., *Working the web: An empirical model of web use*. In : I.C. Society (ed), Hawaii International Conference on System Sciences, vol.7, Organizational Systems and Technology, Maui, Hawaii, 2000, 10 p.
- (Choquet et Iksal, 2007) Choquet C., Iksal S., *Modeling Tracks for the Model Driven Reengineering of a TEL System*, Journal of Interactive Learning Research (JILR), 2007, vol.18, n°2, pp. 161-184.
- (Christensen et Bardram, 2002) Christensen H.B. et J.E. Bardram, *Supporting Human Activities – Exploring Activity-Centered Computing*. In : Proc. Ubicomp 2002, Springer Verlag, 2002, pp. 107-116.
- (Clancey, 1997) Clancey W.J., *Situated Cognition: On Human Knowledge and Computer Representations*, Cambridge University Press, 1997, 428 p.
- (Clases et Wehner, 2002) Clases C., Wehner T. *Steps Across the Border – Cooperation, Knowledge Production and Systems Design*. Computer Supported Cooperative Work, 2002, vol.11, Special Issue on activity theory and design, Nardi B. et Redmiles D. (eds), n°1-2, pp. 39-54.

- (Clot, 2002) Clot Y., Avec Vygotski, ouvrage collectif, Paris : La Dispute, 2002, 345 p.
- (Clot et Faïta, 2001) Clot Y., Faïta D., Fernandez G. et Scheller L. *Entretiens en autoconfrontation croisée : une méthode en clinique de l'activité*. Education Permanente, 2001, vol.146, n°1, pp. 17-25.
- (Cockburn et al., 2002) Cockburn A., McKenzie B., JasonSmith M. *Pushing Back: Evaluating a New Behavior for the Back and Forward Buttons in Web Browsers*. International Journal of Human-Computer Studies, 2002, vol.57, n°5, pp. 397-414.
- (Collins et al., 2002) Collins P., Shukla S. et Redmiles D. *Activity Theory and System Design : A View from the Trenches*. Computer Supported Cooperative Work, 2002, vol.11, Special Issue on activity theory and design, Nardi B. et Redmiles D. (eds), n°1-2, pp. 55-80.
- (Colmerauer et Roussel, 1996) Colmerauer A. et Roussel P. *The birth of Prolog*. History of Programming Languages, Bergin T.J. et Gibson R.G. (eds), ACM Press Addison-Wesley, 1996.
- (Cooke et al., 1996) Cooke N.J., Neville K.J., Rowe A.L. *Procedural network representations of sequential data*. Human computer Interactions, 1996, vol.11, pp.29-68.
- (Corcho et al., 2005) Corcho O., Fernandez-Lopez M., Gomez-Pérez A., Lopez-Cima A. *Building Legal Ontologies with METHONTOLOGY and WebODE*. In : Benjamins R., Casanovas P., Gangemi A., Selic, B. (Eds.). Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications., Springer-Verlag, 2005, pp.142-157.
- (Cram et al., 2007) Cram D., Jouvin D. et Mille A. *Visualisation interactive de traces et réflexivité : application à l'EIAH collaboratif synchrone eMédiathèque*, Sciences et Technologies pour l'Éducation et la Formation STI-CEF, 2007, vol.14.
- (Crowston et Kammerer, 1998) Crowston K., Kammerer E.E. *Coordination and Collective Mind in Software Requirements Development*. IBM Systems Journal, 1998, vol.37, n°2, pp. 227-246.
- (Cugini et Scholtz, 1999) Cugini J., Scholtz J. *VISVIP : 3D visualization of paths through web sites*. In : Proc. of International Workshop on Web-Based Information Visualization (WebVis'99), IEEE Computer Society, Florence, Italy, 1999, pp. 259-263.
- (Cuomo, 1994) Cuomo D.L. *Understanding the applicability of sequential data analysis techniques for analysing usability data*. In : J. Nielsen (ed), Behavior and Information Technology, vol.13, numéro spécial, n°1-2. pp. 171-182.
- (Dartois, 2001) Dartois C. *Quelle veille pratiquer sur un site web ?* BBF, Paris, 2001, vol.46, n°2, pp. 88-93.
- (Davallon et al., 2009) Davallon J., Flon E., Tardy C., Jeanneret Y. *Traces d'écriture, traces de pratiques, traces d'identités : de l'inscription à l'échange*. In : Proc. H2PTM'09, Hermès Publication, Paris, 2009, pp. 181-193.
- (Decortis et al., 1997) Decortis F., Noirfalise S., Saudelli B. *Distributed cognition as framework for cooperative work*, EU TMR Network COTPOS WP1, 1997, pp. 56-70.
- (Decortis et al., 2000) Decortis F., Noirfalise S., Saudelli B. *Activity theory, cognitive ergonomics and distributed cognition: three views of a transport company*, International Journal of Human-Computer Studies, 2000, vol.53, pp. 5-33.
- (Demonet, 1994) Demonet-Launay M-L. *Si les signes vous fâchent..., inférence naturelle et science des signes à la Renaissance*. Bul. de l'Association d'étude sur l'humanisme, la réforme et la renaissance, 1994, n°1, pp. 7-44.
- (Desprès et Leroux, 2003) Desprès C., Leroux P., *Le tutorat synchrone en formation à distance : un modèle pour le suivi pédagogique synchrone d'activités d'apprentissage à distance*. In : Environnements Informatiques pour l'Apprentissage Humain, Strasbourg, 2003.
- (Desprès et al., 2004) Desprès C., Coffinet T., *Reflet, un miroir sur la formation*, In : Conférence Internationale sur les Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et l'Industrie, 2004, Compiègne, France, pp. 19-24.
- (Dessus et Sylvestre, 2003) Dessus P., Sylvestre E., *Transposition d'une tâche en activité*. Résonances, 2003, n°5, pp. 8-9.
- (Dillenbourg, 1999) Dillenbourg P. *What do you mean by Collaborative Learning?* In P. Dillenbourg (Ed.) Collaborative Learning : Cognitive and Computational Approaches, Oxford, UK: Elsevier Science, 1999, pp. 1-19.
- (Dimitracopoulou, 2005) Dimitracopoulou A. *State of the Art on Interaction Analysis for Metacognitive Support and Diagnosis*. Report JEIRP. D.31.1.1, Kaleidoscope Network of Excellence, 2005, pp. 6-62.

- (Doise et al., 1991) Doise W., Deschamps J-C., Mugny, G. *Psychologie Sociale Expérimentale*. Arman Collin (ed), 1991, 282 p.
- (Dömel, 1995) Dömel P. *WebMap: a graphical hypertext navigation tool*. Computer Networks and ISDN Systems, vol.28, n°1-2, 1995, pp. 85-97.
- (Dragunov et al., 2005) Dragunov A.N., Diettrich T.G., Johnsrue K., McLaughlin M., Li L., Herlocker J.L. *TaskTracer: A Desktop Environment to Support Multi-tasking Knowledge Workers*, In : Proc. 10th International Conference on Intelligent User Interfaces, San Diego, California, 2005, pp. 75-82.
- (Dubois et al., 2000) Dubois J-M., Dao-Duy J-M., Eldika S. *L'analyse des traces informatiques des usages : un outil pour valider la conception d'un site web*. In : Actes des rencontres jeunes chercheurs en Interaction Homme-Machine, Ile de Berder, 2000, pp. 85-89.
- (Dumazeau et Karsenty, 2000) Dumazeau C. et Karsenty L. *Analyse empirique des communications distantes dans le cadre du contrôle aérien*. Rapport de recherche de fin de projet (IRIT/00-39-R). Toulouse : IRIT, 2000.
- (Ellis et Wainer, 1994) Ellis C., Wainer J., *A Conceptual Model of Groupware*, In : Proc. Computer Supported Cooperative Work (CSCW'94); Chapel Hill, North Carolina, USA, 1994, pp. 79-88.
- (El-Ramly et al., 2001) El-Ramly M., Ingiski P., Stroulia E., Sorenson P., Matichuk B., *Modeling the System-User Dialog using Interaction Traces*. In : Proc. 8th Working Conference on Reverse Engineering, IEEE Computer Soc. Press (ed), Los Alamitos, USA, 2001, pp. 208-217.
- (Engeström et al., 1997) Engeström Y., Brown K., Christopher L., Gregory J., *Coordination, cooperation and communication in the courts*. In : Cole M., Engeström Y., Vasquez O. (eds), *Mind, Culture and Activity*. Cambridge University Press, Cambridge, UK, 1997, pp. 369-384.
- (Engeström, 2005) Engeström Y. *Development, movement and agency: an activity-theoretical analysis*. Presented at Symposium *Artefacts et collectifs, action située et théories de l'activité*, INTEFP, Marcy l'Étoile, France, 2005, Disponible sur <http://sites.univ-lyon2.fr/artco/home.html> (consulté le 02/07/09).
- (Ermine, 2000) Ermine J.L., *Les systèmes de connaissances*. Hermès-Lavoisier, Paris, 2^{ème} édition, 2000, 240 p.
- (Escorcía, 2006) Escorcía, D., *La participation des composantes métacognitives à la production d'écrits des étudiants en sciences humaines et sociales*. In : 8^{ème} biennale de l'éducation et de la formation : Expérience(s), savoir(s), sujets(s), Lyon, France, 2006, 7p.
- (Ferrer, 1997) Ferrer D., *Le matériel et le virtuel : du paradigme indiciaire à la logique des mondes possibles*. In : Contat M. et Ferrer D. (eds.), *Pourquoi la critique génétique ? Méthodes, théories*. CNRS Éditions, Paris, 1998, Disponible sur <http://www.item.ens.fr/index.php?id=14019> (consulté le 03/12/2008).
- (Ferraris et al., 2007) Ferraris C., Martel M., Vignollet L., *Helping teachers in designing CSCL Scenarios: a methodology based on the LDL Language*. In: Proc. 8th International conference on Computer Support for Collaborative Learning (CSCL'07), New Brunswick, New Jersey, USA, 2007, pp. 193-195.
- (Flavell, 1977) Flavell J.H., *Cognitive development*, Englewood Cliffs, N.J. : Prentice-Hall Inc, 1977, 286 p.
- (Fogli et al., 2005) Fogli D., Fresta G., Mussio P., Marcante A., Padula M., *Annotation in cooperative work: from paper-based to the web one*. In : Proc. International Workshop on Annotation for Collaboration - Methods, Tools and Practices, Boujut J.F. (Ed.), La Sorbonne, Paris, France, 2005, pp. 1-10.
- (Foss et al., 2001) Foss A., Wang W., Zaïane O.R., *A non-parametric approach to weblog analysis*. In : Proc. of Workshop on Web Mining in First International SIAM Conference on Data Mining, 2001, pp. 41-50.
- (Gabriel, 2004) Gabriel M.A., *Learning together: Exploring group interactions online*, Journal of Distance Education, 2004, vol. 10, n°1, pp. 54-72.
- (Gagnière et al., 2004) Gagnière L., Bétrancourt M., Détienne F., Chabert G., *Developing a reflective incentive to encourage metacognitive activities in a computer-supported collaborative learning environment*, In : EARLI SIG on Metacognition, Amsterdam, Netherland, 2004.
- (Gaussier et Stéfani, 2003) Gaussier E. et Stéfani M-H., *Assistance intelligente à la recherche d'informations*, Hermès Science Pub., Paris, Traité des sciences et techniques de l'information, 2003, 319 p.
- (Gay et Mazur, 1993) Gay G., Mazur J., *The utility of computer tracking tools for user-centred design*. In : Educational Technology, vol.33, n°4, 1993 pp. 45-59.

- (Georgeon et al., 2006a) Georgeon O., Mille A., Bellet T., *Analyzing behavioral data for refining cognitive models of operator*. In : Proc. 17th International Conference on Database and Expert Systems Applications, Krakow, Poland. IEEE Computer Society Press, 2006, pp. 588-592.
- (Georgeon et al., 2006b) Georgeon O., Mille A., Bellet T. *Abstract: un outil et une méthodologie pour analyser une activité humaine médiée par un artefact technique complexe*. In : actes Ingénierie des Connaissances (IC 2006), Lehn R., Harzallah M., Aussenac-Gilles N., Charlet J. (ed). Nantes, France, 2006, Disponible sur : liris.cnrs.fr/publis/?id=2376 (Consulté le 10 novembre 2006).
- (Georgeon, 2008) Georgeon O., *Analyse de traces d'activité pour la modélisation cognitive : Application à la conduite automobile*. Thèse de l'université de Lyon, Février 2008.
- (Giboin, 2004) Giboin A., *La construction de référentiels communs dans le travail coopératif*. In : Hoc J-M. et Darses F. (eds), *Psychologie Ergonomique : tendances actuelles*, PUF, Paris, Le travail humain , 2004, p.119-139.
- (Gibson, 1977) Gibson J.J., *The theory of affordances*. In : Shaw R.E. et Bransford J. (eds.), *Perceiving, acting, and knowing: Toward an ecological psychology*. Hillsdale, NJ: Erlbaum, 1977, pp. 67-82.
- (Ginzburg, 1989) Ginzburg C., *Traces : Racines d'un paradigme indiciaire*. Mythes, Emblèmes, Traces, - Morphologie et histoire, Paris, Flammarion, 1989, pp. 139-180.
- (Gogoulou et al., 2005) Gogoulou A., Gouli E., Grigoriadou M., *Analysing Learner Interaction in an Adaptive Communication Tool*. In : 12th International Conference on Artificial Intelligence in Education (AIED2005) : Workshop on Representing and Analyzing Collaborative Interactions: What works? When does it work? To what extent?, Amsterdam, Netherland, 2005.
- (Gomez-Perez et al., 2004) Gomez-Perez A., Fernandez-Lopez M., Corcho O., *Ontological Engineering*, In : *Methodologies and Methods for Building Ontologies*, Springer, pp. 107-196.
- (Gray et al., 1994) Gray P., England D., et McGowan S. XUAN: Enhancing the UAN to capture temporal relation among actions. Department of Computing Science, University of Glasgow, February 1994. Department research report IS-94-02.
- (Grčar, 2004) Grčar M., *User Profiling: Web Usage Mining*. In : Proc. 7th International Multiconference Information Society IS 2004, Ljubljana, Slovenia, 2004, pp. 79-82.
- (Greenberg et Cockburn, 1999) Greenberg S. et Cockburn A., *Getting Back to Back: Alternate Behaviors for a Web Browser's Back Button*. In : 5th Human Factors and the Web Conference, NIST, Gaithersburg, USA, 1999.
- (Greenberg et Witten, 1988) Greenberg S. et Witten I.H., *How Users Repeat Their Actions on Computers : Principles for Design of History Mechanisms*. In : Proc. CHI'88 Human Factors in Computing Systems Conference. Washington, DC, USA, 1988 pp. 171-178.
- (Greeno et Moore, 1993) Greeno G.J. et Moore L.J., *Situativity and symbols : Response to Vera and Simon*. *Cognitive Sciences*, vol. 17, n°1, pp. 49-61.
- (Gronier et Sagot, 2002) Gronier G. et Sagot J-C., *Confrontation des activités coopératives de conception en présence et assistées par un collectifiel*, In : IHM 2002, Poitiers, France, ACM, 2002, pp. 283-284.
- (Gruber, 1993) Gruber T. R., *A translation approach to portable ontology specifications*. In : *Knowledge Acquisition*, 1993, n°5, pp. 199-220.
- (Grudin, 1994) Grudin J., *Groupware and social dynamics: eight challenges for developers*, ACM Communication, NY, USA, 1994, vol.37, n°1, pp. 92-105.
- (Guéraud et al., 2004) Guéraud V., Adam J-M., Pernin J-P., Calvary G., David J P., *L'exploitation d'Objets Pédagogiques Interactifs à distance : le projet FORMID*, *Revue STICEF*, 2004, vol. 11, pp. 109-164.
- (Guimbretière et al., 2007) Guimbretière F., Dixon M., Hinckley K. *ExperiScope: an analysis tool for interaction data*. In : Proc. Computer Human Interaction (CHI'07). San Jose, California, 2007, pp. 1333-1342.
- (Guin et al., 2002) Guin-Duclosson N., Jean-Daubias S., Nogry S., *The AMBRE ILE: How to Use Case-Based Reasoning to Teach Methods*. In : Proc. ITS'2002, Biarritz, Springer, 2002, pp. 782-791.
- (Guy, 2005) Guy E. S., *"...real, concrete facts about what works..." : integrating evaluation and design through patterns*. In : Proc. International ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA, 2005, pp. 99-108.

- (Guzdial et al., 1996) Guzdial M., Kolodner J., Hmelo C., Narayanan H., Carlso D., Rappin N., Hubscher R., Turns J., Newstetter W., *The collaboratory notebook*, Communications of the ACM, 1996, vol. 39, n°4, pp. 32-33.
- (De Haan, 2000) Haan De G., *ETAG, A Formal Model of Competence Knowledge for User Interface Design*. Informatics PhD thesis, Amsterdam : Vrije Universiteit, 2000.
- (Halverson, 1992) Halverson C.A., *Analyzing a Cognitively Distributed System : a Terminal Radar Approach Control*. Unpublished Masters diss., Cognitive Science Department, University California San Diego, CA, 1992.
- (Halverson, 1994) Halverson, C.A., *Distributed Cognition as a theoretical framework for HCI : Don't throw the Baby out with the bathwater - the importance of the cursor in Air Traffic Control*. Tech Report n°94-03, Dept. of Cognitive Science, University of California, San Diego, 1994.
- (Halverson, 1995) Halverson C.A., *Inside the Cognitive Workplace : New Technology and Air Traffic Control*. PhD Thesis, Dept. of Cognitive Science, University of California San Diego, 1995, .
- (Halverson, 2002) Halverson C.A., *Activity Theory and Distibuted Cognition : Or what does CSCW need to Do with theories*, Computer Supported Cooperative Work, Kluwer Academic (ed), 2002, vol. 11, pp. 243-267.
- (Halvey et al., 2005) Halvey M., Keane M.T., Smyth B., *Time based segmentation of log data for user navigation prediction in personalization*. In: Skowron A., Agrawal R., Luck M., Yamaguchi T., Morizet-Mahoudeaux P., Liu J., Zhong N. (eds), *Web Intelligence*, IEEE Computer Society, 2005, pp. 636-640.
- (Hanson et al., 1984) Hanson S.J., Kraut R.E., Farber J.M., *Interface design and multivariate analysis of UNIX command use*. ACM Transactions on Office Information Systems, 1984, vol.2, n°1, pp. 42-57.
- (Hardy et al., 2004) Hardy J., Antonioletti M., Bates S., *e-Learner Tracking : Tools for discovering Learner Behavior*. In : IASTED International Conference on Web-base Education, Innsbruck, Austria, 2004, pp. 833-834.
- (Hartson et al., 1990) Hartson H.R., Siochi A.C., Hix D., *The uan: a user-oriented representation for direct manipulation interface designs*. In : ACM Trans. Inf. Syst., 1990, vol. 8, n°3, pp. 181-203.
- (Hauser et al., 2009) Hauser J.R., Urban G.L., Liberali G., Braun M., *Website morphing*. Marketing Science, 2009, vol. 28, n°2, pp. 202-223.
- (Hawkey et Inkpen, 2005) Hawkey K. et Inkpen K., *Privacy gradients: exploring ways to manage incidental information during co-located collaboration*. In Proc. CHI 2005, Portland, 2005, pp. 1431-1434.
- (Hawkey, 2006) Hawkey K., *Mission Impossible? Capturing Rich Yet Natural User Behaviour on theWeb*. In : Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, WWW'2006 conference, Edinburgh, Scotland, 2006, pp. 123-132.
- (Hemmecke et Stary, 2004) Hemmecke J., Stary C., *A Framework for the Externalization of Tacit Knowledge Embedding Repertory Grids*. In : Proc. 5th European Conference on Organizational Knowledge Learning and Capabilities (OKLC-2004), Innsbruck, 2004.
- (Héon et al., 2009) Héon M., Paquette G., Basque J., *Méthodologie assistée de conception d'une ontologie à partir d'une conceptualisation consensuelle semi-formelle*. In : Proc. Journées Francophones d'Ingénierie des Connaissances (IC'09). Hammamet, Tunisie, 2009, pp. 61-73.
- (Héraud et al., 2004) Héraud J-M., France L., Mille L., *Pixed : An ITS that guides students with the help of learners' interaction logs*. In : 7th International Conference on Intelligent Tutoring Systems (Workshop Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes), Maceio, 2004, pp. 57-64.
- (Hightower et al., 1998) Hightower R., Ring L., Helfman J., Bederson B., Hollan J., *Graphical Multiscale Web Histories: A Study of PadPrints*. In : Proc. ACM Conference on Hypertext, ACM Press, 1998 pp. 58-65.
- (Hilbert et Redmiles, 2000) Hilbert D., Redmiles F.D., *Extracting Usability Information from User Interface Events*. ACM Computing Surveys, 2000, vol. 32, n°4, pp. 384-421.
- (Hill et al., 1992) Hill W.C., Hollan J.D., Wroblewski D., McCandles T., *Edit Wear and Read Wear*. In : Proc. SIGCHI conference on Human factors in computing systems, Monterey, California, US, 1992, pp. 3-9.
- (Hill et Hollan, 1993) Hill W.C. et Hollan J.D., *History-enriched digital objects*. In : Proc. Third ACM Conference on Computers, Freedom and Privacy, San Francisco, US, ACM, pp. 917-920.
- (Hoc, 1987) Hoc J-M., *Psychologie cognitive de la planification*, P.U.G., Sciences et Technologies des connaissances, 1987, 197 p.

- (Hoc, 1998) Hoc J.-M., *L'ergonomie cognitive : un compromis nécessaire entre des approches centrées sur la machine et des approches centrées sur l'homme*. In : Proc. 2^{èmes} Journées Recherche et Ergonomie, Quéinnec Y. (ed), Toulouse, 1998.
- (Hoc et Darse, 2004) Hoc J.-M. et Darses F., *Psychologie ergonomique : tendances actuelles*. Paris : PUF, Le Travail Humain, 260 p.
- (Hutchins, 1994) Hutchins E., *Comment le "cockpit" se souvient de ses vitesses*. Sociologie du Travail, 1994, numéro spécial Travail et Cognition, vol.4, pp. 451-473.
- (Hutchins, 1995a) Hutchins E., *How a cockpit remembers its speeds*. Cognitive Science, 1995, vol.19, pp. 265-288.
- (Hutchins, 1995b) Hutchins E., *Cognition in the Wild*, Cambridge, MIT Press, 1995, 408 p.
- (Hutchins et Klausen, 1996) Hutchins E et Klausen T., *Distributed cognition in an airline cockpit*. In : Engeström Y. et Middleton D. (eds), *Cognition and communication at work*, NY, Cambridge University Press, 1996 pp. 15-34.
- (Hyams et al., 2003) Hyams J., Sellen A., *Gathering and Sharing Web-Based Information : Implications for "ePerson" concepts*. HP Labs : Tech. Report: HPL-2003-19, Feb. 2003, 42 p.
- (Ihadjadene, 2004) Ihadjadene M., *Les systèmes de recherche d'informations : modèles conceptuels*. Paris, hermes science publications, 2004, 216 p.
- (Ivory et Hearst, 2001) Ivory M.Y. et Hearst M.A., *The state of the art in automating usability evaluation of user interfaces*, ACM Computing Surveys (CSUR), vol.33, n°4, 2001, pp. 470-516.
- Iqbal et Bailey, 2007) Iqbal S.T., Bailey B.P., *Understanding and developing models for detecting and differentiating breakpoints during interactive tasks*. In : Proc. SIGCHI conference on Human factors in computing systems. San Jose, California, USA, 2007, pp. 697-706.
- (Jaczynski et Trousse, 1998) Jackinski M., Trousse B., *WWW Assisted Browsing by Reusing Past Navigations of a Group of Users*. In : 4th European Workshop on Case-Based Reasoning, LNCS vol.1488, 1998, pp.160-171.
- (Jansen et Pooch, 2000) Jansen B. J., Pooch U. W., *A Review of Web Searching Studies and a Framework for Future Research*. Journal American , Society of Information Science, 2000, vol.52, n°3, pp. 235-246.
- (Jambon, 1996) Jambon F., *Erreurs et interruptions du point de vue de l'ingénierie de l'interaction homme-machine*. Thèse Informatique de l'Université J.Fourirer, Grenoble1, 5 décembre 1996.
- (Järvelin et Wilson, 2003) Järvelin K. et Wilson T., *On conceptual models for information seeking and retrieval research*, Information Research, 2003, vol. 9, n°1, paper 163.
- (Jermann et al., 2001) Jermann P.R., Soller A. et Mühlenbrock M., *From mirroring to guiding: A review of state of the art technology for supporting collaborative learning*. In : Proc. European Perspectives on Computer-Supported Collaborative Learning. Bergen, Norway, 2001, pp.324-331.
- (Jermann, 2004) Jermann P.R., *Computer support for interaction regulation in collaborative problem-solving*. PhD thesis, CRAFT, EPFL, Lausanne, Suisse, 2004.
- (Johansen, 1991) Johansen R., *Groupware : Future Directions and Wild Cards*, Journal of Organizational Computing, vol.2, n°1, 1991, pp. 219-227.
- (Jonhson et Jonhson, 1991) Jonhson D.R. et Jonhson R.T., *Learning together and alone, Cooperative, Competitive and Individualistic learning*. Englewood Cliffs : Prentice Hall, 1991, 272 p.
- (Jonassen et al., 1995) Jonassen D., Davidson M., Collins M., Campbell J., Haag, B., *Constructivism and computer-mediated communications in distance education*, The American Journal of Distance Education, vol. 7, n°2, pp. 7-26.
- (Jones et al., 2004) Jones J.A., Orso A., Harrold M.J., *GAMMATELLA: visualizing program-execution data for deployed software*, Information Visualization, 2004, vol. 3, n°3, pp.173-188.
- (Kassel, 2002) Kassel G., *Ontospec : une méthode de spécification semi-formelle d'ontologies*. In : 13^{ème} journées francophones d'Ingénierie des Connaissances (IC'02), Rouen, France, 2002, pp. 75-87.
- (Katz et al., 1992) Katz S., Lesgold A., Eggan G., Gordin M., *Modelling the student in Sherlock II*, Artificial Intelligence in Education, 1992, vol. 3, n°4, pp. 495-518.

- (Kaasten et Greenberg, 2001) Kaasten S., Greenberg S., *Integrating Back, History and Bookmarks in Web Browsers*. In : Ext. Abstracts CHI'01, 2001, pp. 379-380.
- (Kaptelinin, 2003) Kaptelinin V., *UMEA: translating interaction histories into project contexts*, In : Proc. SIGCHI conference on Human factors in computing systems. Ft. Lauderdale, Florida, USA, 2003, pp. 352-360.
- (Kieras et John, 1994) Kieras D. et John B., *The GOMS Family of Analysis Techniques : Tools for Design and Evaluation*, Tech. Report CS-94-181, Carnegie Mellon University Pittsburgh, PA, USA, 1994.
- (Kieras, 1996) Kieras D., *Task analysis and the design of functionality*. In : Handbook of Computer Science and Engineering, Tucker A. (ed), CRC Press Inc., 1996, pp. 1401-1423.
- (Kiss et Quinqueton, 2004) Kiss A. et Quinqueton J., *Uniscript: a model for persistent and incremental knowledge storage*. In : Proc. First ACM workshop on Continuous archival and retrieval of personal expériences (CARPE'04), ACM, NY, 2004, pp. 66-73.
- (Komlodi, 2004) Komlodi A., *Task management support in information seeking: a case for search histories*, Computers in Human Behavior, 2004, vol. 20, pp. 163-164.
- (Koper et al., 2003) Koper R., Olivier B., Anderson T., *IMS Learning Design Information Model*. IMS Global Consortium, 2003.
- (Kort et DePoot, 2005) Kort J., DePoot H., *Usage analysis: combining logging and qualitative methods*, In : Proc. Human factors in computing systems (CHI'05), 2005, Portland, OR, USA, pp. 2121-2122.
- (Kröner et al., 2006) Kröner A., Heckmann D., Wahlster, W., *SPECTER : Building, Exploiting, and Sharing Augmented Memories*. In : Workshop on Knowledge Sharing for Everyday Life 2006 (KSEL06), Kogure K. (ed), Kyoto, Japan, 2006, pp. 9-16.
- (Kurlander et Feiner, 1992) Kurlander D. et Feiner S., *A history-based macro by example system*. In : Proc. 5th annual ACM symposium on User interface software and technology, Monterey, California, 1992, pp. 99-106.
- (Kuuti, 1991) Kuutti K., *The concept of activity as a basic unit of analysis for CSCW research*. In : Proc. ECS-CW'91, Kluwers Academics Publishers, 1991, pp. 249-264
- (Kuuti, 1993) Kuutti K., *Notes on systems supporting "Organisational context" – An activity theory viewpoint*, COMIC European project, deliverable D1.1, 1993, pp 101-117.
- (Kuuti, 1996) Kuutti K., *Activity Theory as a Potential Framework for Human-Computer Interaction Research*, In : Nardi B.A., Context and consciousness: activity theory and human-computer interaction, 1996, pp. 17-44.
- (Laflaquière, 2003) Laflaquière J., *L'expérience d'utilisation d'un espace documentaire pour améliorer l'assistance à l'exploitation de l'information dans le cadre de la veille*, Informatique, Mémoire de DEA, Lyon : Université Claude Bernard Lyon I, 2003, 31 p.
- (Laflaquière et Prié, 2003) Laflaquière J., Prié Y., *Modélisation d'utilisation de système pour une assistance à base de trace : une application de MUSETTE à la tâche de veille documentaire*. In : Workshop Traces, interactions, co-constructions collectives et relations à la cognition. CoMETE, RTP38-CNRS STIC, Paris, 2003.
- (Laflaquière et al., 2005) Laflaquière J., Champin P.A., Mille A., *Approche de modélisation de l'expérience d'utilisation de systèmes complexes pour l'assistance aux tâches de veille informatiquement médiées*, In : Actes de ISKO-France 2005, David A. (eds), INIST/CNRS Nancy, 2005, pp. 209-230.
- (Laflaquière et Ciaccia, 2005) Laflaquière J., Ciaccia A., *Facilitation de tâches informatiquement médiées : une approche centrée sur la réflexivité de l'utilisation* (poster). In : 6^{ème} Colloque des jeunes chercheurs en sciences cognitives, Bordeaux, 2005.
- (Laflaquière et al., 2007) Laflaquière J., Settouti L.S., Prié Y., Mille A., *Traces et Inscriptions de Connaissances*, (Poster). In : Actes Ingénierie des Connaissances (IC'07), Grenoble, 2007.
- (Laflaquière et al., 2008) Laflaquière J., Prié Y., Mille A., *Ingénierie des traces numériques d'interaction comme inscriptions de connaissances*. In : Actes journées francophones d'Ingénierie des Connaissances (IC'08), Juin 2008, Nancy, pp. 183-195.
- (Laflaquière et Prié, 2009a) Laflaquière J., Prié Y., *Musette : Modéliser les Usages et les Tâches pour Tracer l'Expérience*, Réutilisation de l'expérience : modèles et applications, Hermès, 2009 (à paraître).
- (Laflaquière et Prié, 2009b) Laflaquière J., Prié Y., *Des traces modélisées pour la médiation de l'activité réflexive dans un environnement numérique*, Revue d'Anthropologie des Connaissances, Paris, 2009, (à paraître).

- (Laflaquière et Prié, 2009c) Laflaquière J., Prié Y., *L'expérience tracée des activités conjointes instrumentées*, In : atelier Interaction, Contexte, Traces (ICT2009), Caen, France, 2009 (à paraître).
- (Laperrousaz, 2006) Laperrousaz C., *Le suivi individuel d'apprenants engagés dans une activité collective à distance, TACSI : un environnement informatique support aux activités du tuteur*. Thèse de doctorat, Université de Maine, 2006.
- (Latour, 2007) Latour B., *Beware, your imagination leaves digital traces*, Times (Higher Literary Supplement), 6 avril 2007, pp. 129.
- (Lave, 1988) Lave J., *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge, UK : Cambridge University Press, 1988, 232 p.
- (Lave et Wenger, 1991) Lave J. et Wenger E., *Situated Learning. Legitimate peripheral participation*, Cambridge, UK : Cambridge University Press, 1991, 138 p.
- (LeCalvez et al., 2003) LeCalvez F., Giroire H., Duma J., Tisseau G., Urtasun M., *Combien? a Software to Teach Students How to Solve Combinatorics Exercises*. In : Ext. Proc. 11th International Conference on Artificial in Education (AIED'03), Sydney, Australia, 2003, pp. 447-454.
- (Leleu-Merviel, 2004) Leleu-Merviel S., *Effets de la numérisation et de la mise en réseau sur le concept de document*, Paris, Cépaduès Editions, Revue I3 : Information, Interaction, Intelligence, A Journal in the Sciences of Information Engineering, 2004, vol. 4, n°1, pp. 121-140.
- (Leplat et Hoc, 1983) Leplat J. et Hoc J.M., *Tâche et activité dans l'analyse psychologique des situations*, Cahiers de psychologie cognitive, 1983, vol.3, pp. 49-63.
- (Leplat, 1991) Leplat J., *Organization of activity in collective tasks*. In : J. Rasmussen, B. Brehmer, J. Leplat (eds.) Distributed decision making: Cognitive models for cooperative work. Chichester : Wiley, 1991, pp. 51-73.
- (Lemaire et Moore, 1994) Lemaire B. et Moore J., *An improved interface for tutorial dialogues: browsing a visual dialogue history*, In : Proc. Conference on Human Factors in Computing Systems. Boston, Massachusetts, US, 1994, pp. 16-22.
- (LeMoigne, 1994) LeMoigne J-L., *Le constructivisme : Les fondements*, vol. 1, Paris, ESF, Communication et complexité, 1994, 252 p.
- (Lesgold et al., 1992) Lesgold A., Lajoie S., Bunzo M., Eggan G., *Sherlock A Coached Practice Environment for an Electronics Troubleshooting Job*. In : Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches, Larkin J, Chabay R. (eds), Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1992, pp. 36.
- (Lettkeman et al., 2006) Lettkeman A.T., Stumpf S., Irvine J., Herlocker J., *Predicting task-specific webpages for revisiting*. In : Proc. 21st National Conference on Artificial Intelligence (AAAI-06). Boston, Ma, US, 2006, vol. 2, pp. 1369-1374.
- (Lieberman, 1995) Lieberman H., *Letizia : An Agent that Assists Web Browsing*. In : Proc. International Joint Conference on Artificial Intelligence (IJCAI'95), Morgan Kaufmann. Montreal, Quebec, 1995, pp. 924-929.
- (Lieberman et al., 1999) Lieberman H., Van Dyke N.Y., Vivacqua A.S., *Let's browse: a collaborative web browsing agent*. In : Proc. 4th international conference on Intelligent user interfaces (IUI '99), ACM, NY, USA, 1999, pp. 65-68.
- (Lieberman, 2001) Lieberman H., *Interfaces that Give and Take Advice*, In : Carroll J. (Ed), Human-Computer Interaction for the New Millenium, ACM Press/Addison-Wesley, 2001, pp. 475-485.
- (Little et al., 2007) Little G., Lau T.A, Cypher A., Lin J., Haber E.M., Kandogan E., *Koala: Capture, Share, Automate, Personalize Business Processes on the Web*. In : Proc. ACM symposium on User interface software and technology CHI'07, Victoria, BC, Canada, 2007, pp. 203-212.
- (Lomicka, 1998) Lomicka L.L., *To gloss or not to gloss : an investigation of reading comprehension online*, Language Learning and Technology, 1998, vol. 1, n°2, pp. 41-51.
- (Longchamp, 2003) Longchamp J., *Le Travail coopératif et ses technologies*, Paris : Hermès science, Lavoisier, 2003, 319 p.
- (Lortal et al., 2006) Lortal G., Lewkowicz M., Todiracu-Courtier A., *Des activités d'annotation : De la glose au document*. In : Salembier P. et Zacklad M. (eds), Annotations dans les documents pour l'action, Hermes Publishing, Londres-Paris, 2006, pp. 153-171.

- (Lucquiaud et al., 2002) Lucquiaud V., Scapin D., Jambon F., *Outils de modélisation des tâches utilisateurs : exigences du point de vue utilisation*. In : conférence francophone sur l'Interaction Homme-Machine (IHM'02), Poitiers, France. ACM Press, 2002, pp. 243-246.
- (MacLeod et Renger, 1993) MacLeod M. et Rengger R., *The Development of DRUM: A Software Tool for Video-assisted Usability Evaluation*. In : Proc. Human Computer Interaction (HCI '93), Amsterdam, The Netherlands, pp. 293-309.
- (MacGavin et al., 2006) MacGavin M., Wright T., Marshall S., *Visualisations of execution traces (VET): an interactive plugin-based visualisation tool*. In Proc. Australasian User Interface Conf. (AUIC'06), vol. 50, Australian Computer Society, Inc., 2006, pp. 153-160.
- (Magnusson, 2000) Magnusson M.S., *Discovering Hidden Time Patterns in Behavior : T-Patterns and their Detection*, Behavior Research Methods, Instruments and Computers, 2000, vol. 32, n°1, pp. 93-110.
- (Malone et Crowston, 1990) Malone T.W. et Crowston K., *Towards an interdisciplinary theory of coordination*. Tech. Rep. n°120, MIT, Center for Coordination Science, Cambridge, 1990.
- (Malone et Crowston, 1994) Malone T.W. et Crowston K., *The interdisciplinary study of coordination*. ACM Computing Surveys, 1994, vol. 26, n°1, pp. 87-120.
- (Marca et MacGowan, 1987) Marca D.A. et McGowan C.L., *SADT: structured analysis and design technique*. McGrawHill Inc., NY, 1987, 392 p.
- (Marty et al., 2004) Marty J-C, Heraud J-M, Carron T., France L., *A quality approach for collaborative learning scenarios*, Learning Technology Newsletter of IEEE Computer Society, 2004, vol. 6, n°4, pp. 46-48.
- (Marty et al., 2007) Marty J-C, Heraud J-M, France L., Carron T., *Matching the Performed Activity on an Educational Platform with a Recommended Pedagogical Scenario: a Multi Source Approach*, Journal of Interactive Learning Research (JILR), 2007, vol. 18, n°2, pp. 267-283.
- (Maturana et Varela, 1987) Maturana H.R. et Varela, F., *The tree of knowledge : The biological roots of human understanding*, Boston: New Science Library, 1987, 269 p.
- (May et al., 2008a) May M., George S., Prévôt P., *A Closer Look at Tracking Human & Computer Interactions in Web-Based Communications*, International Journal of Interactive Technology and Smart Education (ITSE), 2008, vol. 5, n°3, pp. 170-188.
- (May et al., 2008b) May M., George S., Prévôt P., *Tracer, analyser, et visualiser les interactions des apprenants pendant les activités de communications médiatisées*. In : actes Colloque JOCAIR : Journées Communication et Apprentissage Instrumentés en Réseau, Hermès Sciences, Lavoisier, Amiens, France, pp. 251-263.
- (Mazza et Dimitrova, 2003) Mazza R., Dimitrova V., *CourseVis: Externalising Student Information to Facilitate Instructors in Distance Learning*. In : Proc. International Conference on Artificial Intelligence in Education (AIED 2003), Sydney, Australia, pp. 279-286.
- (Mazza et Milani, 2005) Mazza R., Milani C., *Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations*, Communication in the Workshop on Usage analysis in learning systems, the twelfth International Conference on Artificial Intelligence in Education, Amsterdam, 2005, the Netherlands, pp. 65-72.
- (Milic-Frailing et al., 2004) Milic-Frailing N., Jones R., Rodder K., Smyth G., Blackwell A., Sommerer R., *Smartback : Supporting Users in Back Navigation*. In : Proc. Int. Conf. World Wide Web, NY, 2004, pp 63-71.
- (Mille et Prié, 2006) Mille A., Prié Y., *Une théorie de la trace informatique pour faciliter l'adaptation dans la confrontation logique d'utilisation/logique de conception*. Communication aux Journées de Rochebrune, Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels, Rochebrune, Megève, ENST, 2006.
- (Mille et al., 2006) Mille A., Caplat M., Philippon M., *Faciliter les activités des utilisateurs d'environnements informatiques : quoi, quand, comment ?*, Intellectica, 2006, vol. 2, n°44, pp. 121-143.
- (Mobasher et al., 2000) Mobasher B, Cooley R., Srivastava J., *Automatic Personalization Based On Web Usage Mining*, Communication of ACM, vol. 43, n°8, 2000, pp. 142-151.
- (Modugno et Myers, 1994) Modugno F. et Myers B.A., *Pursuit: graphically representing programs in a demonstrational visual shell*. In : Conference Companion of CHI'94 Boston, Ma, US, 1994, pp. 455-456.
- (Morse et Steves, 2000) Morse E. et Steves M., *CollabLogger: A Tool for Visualizing Groups at Work*. In: Proc. of WETICE2000, Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Gaithersburg, MD, 2000, IEEE Computer Society, pp. 104-109.

- (Moyle et Cockburn, 2003) Moyle M., Cockburn A., *The Design and Evaluation of a Flick Gesture for 'Back' and 'Forward' in Web Browsers*. In : Proc. Australasian User Interface Conference, AUIC 2003, pp. 39-46.
- (Myers et al., 1997) Myers B.A., McDaniel R.G., Miller R.C., Ferency A.S., Faulring A., Kyle B.D., Mickish A., Klimovitski A., Doane P., *The Amulet Environment : New Models for Effective User Interface Software Development*, IEEE Transaction on Software Engineering, 1997, vol. 23, n°6, pp. 347-365.
- (Nakamura et Igarashi, 2008) Nakamura T., Igarashi T., *An application-independent system for visualizing user operation history*, In : Proc. Symposium on User Interface Software and Technology UIST 2008. Monterey, CA, USA, 2008, pp. 23-32.
- (Neal et Simons, 1983) Neal A.S. et Simons R.M., *PLAYBACK: A method for evaluating the usability of software and its documentation*. In : Proc. SIGCHI conference on Human Factors in Computing Systems (CHI'83). Boston, Ma, ACM, 1983, pp. 78-82.
- (Newell, 1982) Newell A., *The Knowledge Level*, Artificial Intelligence, 1982, vol. 18, n°1.
- (Nonaka, 1999) Nonaka I. *L'entreprise créatrice de savoir*, Harvard Business Review, Paris, pp. 35-63.
- (Nonaka et Takeuchi, 1995) Nonaka I. et Takeuchi H., *La connaissance créatrice. La dynamique de l'entreprise apprenante*, De Boeck Université, Bruxelles, 1995, 303 p.
- (Nunes et al., 2007) Nunes M., Greenberg S., Carpendale S., Gutwin C., *What Did I Miss? Visualizing the Past through Video Traces*, In : Proc. European Conference on Computer Supported Cooperative Work (ECSCW'07), Bannon L., Wagner I., Gutwin C., Harper R., Schmidt K. (eds.), Limerick, Ireland, 2007, pp. 24-28.
- (Oard et Kimn, 2001) Oard D.W. et Kim J., *Modeling Information Content Using Observable Behavior*. In : Proc. Annual Conference of the American Society for Information Science and Technology, Washington, 2001, pp. 481-488.
- (Obendorf et al., 2007) Obendorf H., Weinreich H., Herder E., Mayer, M., 2007. *Web page revisitation revisited : Implications of a long-term click-stream study of browser usage*. In : Proc. SIGCHI Conference on Human Factors in Computing Systems, Rosson M.B. et Gilmore D.J., (eds), ACM Press, NY, pp. 597-606.
- (O'Brien et al., 2006) O'Brien M., Keane M.T., Smyth B., *Predictive modeling of first-click behavior in web-search*. In : Proc. International Conference on World Wide Web 2006. ACM, NY, USA, 2006, pp. 1031-1032.
- (Ollagnier, 2006). Ollagnier-Beldame M., *Traces d'interactions et processus cognitifs en activité conjointe : Le cas d'une co-rédaction médiée par un artefact numérique*. Thèse en Sciences Cognitives de l'Université Claude Bernard Lyon 1, 2006, 247 p.
- (Paquette, 2002) Paquette G., *Modélisation des connaissances et des compétences : un langage graphique pour concevoir et apprendre*. Sainte-Foy, France, Presses de l'UQ., 2002, 388 p.
- (Paris et al., 2003) Paris C., Lu S., Vander Linden K., *Environments for the construction and use of task models*. In : Handbook of Task Analysis for Human-Computer Interaction, Diaper D. et Stantonvol N. (eds), Lawrence Erlbaum Assoc. 2004, pp. 467-482.
- (Paterno, 1996) Paterno F., *A methodology for a task-driven modelling of interactive systems architectures*. In : Benyon D. et Palanque P. (eds), Critical issues in user interface systems engineering, Springer-Verlag, 1996, pp. 93-108.
- (Paterno, 1999) Paternò F., *Model-Based Design and Evaluation of Interactive Applications*. Springer Verlag, 1999, 192 p.
- (Paterno, 2002) Paternò, F., *Tools for Task Modelling : Where we are, Where we are Headed*. In : Proc. TAMODIA 2002, , INFOREC. Bucharest, 2002, pp. 10-17
- (Pédauque, 2005) Pédauque R.T., *Le texte en jeu, Permanence et transformations du document*, STIC-SHS-CNRS, 2005, [En ligne] disponible sur http://archivesic.ccsd.cnrs.fr/docs/00/06/26/01/PDF/sic_00001401.pdf. (consulté le 3/04/07).
- (Pédauque, 2006) Pédauque R.T., *Le document à la lumière du numérique. Forme, texte, medium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité*, Caen, C&F Éditions, 2006, 218 p.
- (Pédauque, 2007) Pédauque R.T., *La redocumentarisation du monde*. Toulouse : Cepaduès éditions, 2007, 213 p.
- (Perry, 1997) Perry M.J., *Distributed cognition and computer supported collaborative design : the organization of work in construction engineering*, Phd Thesis, Brunel University, UK, 1997.

- (Philippon et al., 2005) Philippon M., Mille A., Caplat. G., *Aide à l'utilisateur : savoir quand intervenir*. In : conférence francophone sur l'Interaction Homme-Machine (IHM'05), Toulouse, France, ACM, 2005.
- (Pirolli, 2000) Pirolli P., *A Web site user model should at least predict something about users*. Internetworking, vol. 3, n°1. Disponibles sur http://www.sandia.gov/itg/newsletter/mar00/critique_max.html
- (Pirolli et al., 2002) Pirolli P.L., Fu W., Reeder R., Card S.K., *A user-tracing architecture for modeling interaction with the World Wide Web*. In : Proc. Advanced Visual Interfaces 2002, Trento; Italy, 2002, ACM pp. 75-83.
- (Pirolli et Fu, 2003) Pirolli P. et Fu W., *Snif-act : A model of information foraging on the World Wide Web*. In : Proc. International conf. on User Modeling, Johnstown, US, 2003, LNCS n°2702, pp. 45-54.
- (Plaisant et al., 1996) Plaisant C., Milash B., Rose A., Wido S., Shneiderman B., *Lifelines : visualizing personal histories*. In : Proc. SIGCHI conference on Human factors in computing systems, ACM Press, NewYork, NY, USA, 1996, pp. 221-227.
- (Plaisant et al., 1998) Plaisant C., Mushlin R., Snyder A., Li J., Heller D., Shneiderman B., *LifeLines: Using visualization to enhance navigation and analysis of patient records*, American Medical Informatics Association 1998 Annual Fall Symposium. Orlando, AMIA, Bethesda MD, 1998, pp. 76-80.
- (Plaisant et al., 1999) Plaisant C., Rose A., Rubloff G., Salter R., Shneiderman B., *The design of history mechanisms and their use in collaborative educational simulations*. In : Proc. Computer Support for Collaborative Learning confrence (CSCL'99). Palo Alto, CA, Stanford University, 1999, pp. 348-359.
- (Prié, 1999) Prié Y., *Modélisation de documents audiovisuels en Strates Interconnectées par les annotations pour l'exploitation contextuelle*, PhD Dissertation, INSA Lyon, 1999, 270 p.
- (Prié et al., 1999) Prié Y., Mille A., Pinon J.M., *Modèle d'utilisation et modèles de tâches pour l'assistance à l'utilisateur basée sur l'expérience : le cas d'un système d'information audiovisuelle*. In : Proc. Ingénierie des Connaissances (IC'99). Palaiseau, France, 1999, pp. 21-30.
- (Prié et al., 2000) Prié Y., Mille A., *Reuse of knowledge containers: a "local semantics" approach*. In : Workshop on Flexible Strategies for Maintaining Knowledge Containers, 14th European Conference on Artificial Intelligence ECAI 2000, Mirjam Minor (Ed.), Berlin, 2000, pp. 38-45.
- (Prince, 1996) Prince V., *Vers une Informatique cognitive dans les organisations : le rôle central du langage*. Paris, Editions Masson, 1999, 200 p.
- (Quéré, 1999) Quéré L., *Action et cognition situées*. Conférence Publique 17 juin 1999, Montpellier. Disponible sur : <http://serinf2.univ-montp3.fr/mefadis/ancien/conferencequere.htm> (consulté le 12/01/2006).
- (Rabardel, 1995) Rabardel P., *Les hommes et les technologies: Approche cognitive des instruments contemporains*. Paris, Armand Colin, 1995, 239 p.
- (Raby et Borges, 2000) Raby F., Borges M. K. *Un exemple d'une démarche ergonomique appliquée à l'étude des usages des TICES dans l'enseignement des langues vivantes*. In : Estrella et J. ferreira (eds.), Actes du 10^{ème} colloque international de L'AFIRSE. Lisbonne, Portugal, 2000.
- (Rauterberg et Fjeld, 1998) Rauterberg M., Fjeld M., *Task analysis in human-computer interaction - supporting action regulation theory by simulation*. Zeitschrift fuer Arbeitswissenschaft, 1998, vol. 3, n°98, pp.152-161.
- (Renié, 2000) Renié D., *Apport d'une trace informatique dans l'analyse du processus d'apprentissage d'une langue seconde ou étrangère*. In : Duquette et Laurier (dirs) Apprendre une langue dans un environnement multimédia. Outremont, Canada, 2003, pp. 281-301.
- (Reiss, 2003) Reiss S.P., *Visualizing Java in action*. In : Proc. ACM symposium on Software visualization. San Diego, California, 2003, pp. 57-65.
- (Resnick, 1989) Resnick L.B., *Knowing, Learning and Instruction*, LEA, Hillsdale N.J., 1989.
- (Rich et Sidner, 1997) Rich C. et Sidner C.L., *Segmented Interaction History in a Collaborative Interface Agent*. In : Proc. International Conference on Intelligent User Interfaces. Orlando, FL, 1997, pp. 23-30.
- (Richard et al., 2009) Richard B., Prié Y., Calabretto S., *Helping users to recover from interruptions in audiovisual active reading: An interface for reflexive use of traces*. Article Soumis à International Conference on Intelligent User Interface 2010.
- (Ricoeur, 1999) Ricoeur P., *Ontologie*. Encyclopaedia Universalis, 1999.

- (Ringel et al., 2003) Ringel M., Cutrell E., Dumais S.T., Horvitz E., *Milestones in Time: The value of landmarks in retrieving information from personal stores*. In : Rauterberg M., Menozzi M., Wesson J. (eds), IOS Press, INTERACT2003. Zurich, Switzerland, 2003, pp. 184-191.
- (Ritter et Latkin, 1994) Ritter F.E. et Latkin J.H., *Developping process models as summaries of HCI action sequences*, Human-Computer Interactions, 1994, vol.9, pp. 354-383.
- (Rix et Lièvre, 2005) Rix G., Lièvre P., *De l'analyse de pratique au partage d'expérience : le cas d'une expédition polaire à ski*, In : Colloque International ARIS-EDPM-AFRAPS, Louvain La Neuve, Belgique, 2005.
- (Rogers, 1992) Rogers Y., *Ghosts in the Network: Distributed Troubleshooting in a Shared Working Environment*. In : Proc. Conference on Computer-Supported Cooperative Work, Turner J. and Kraut R. (eds). ACM, New York, 1992, pp. 346-355.
- (Rogers, 1993) Rogers Y., *Coordinating computer-mediated work*, Journal of Computer-Supported Cooperative work, 1993, vol. 1, pp. 295-315.
- (Rogers et Ellis, 1994) Rogers Y., Ellis J., *Distributed cognition: an alternative framework for analysing and explaining collaborative working*, Journal of Information Technology, 1994, vol. 9, pp. 119-128.
- (Rogers, 1997) Rogers Y., *A brief introduction to Distributed Cognition*. Discussion Paper Interact Lab, School of Cognitive and Computing Sciences, University of Sussex, 1997, 7p.
- (Romero, 2004) Romero M., *Métacognition dans les EIAH*. Rapport de recherche, LIUM Le mans, Juin 2004, 34p.
- (Rose et al., 2000) Rose A., Salter R., Keswani S., Kositsyna N., Plaisant C., Rubloff G., Shneiderman B., *Simulation based learning environments and the use of learning histories*. In : extended abstracts on Human factors in computing systems, CHI 2000 Conference, The Hague, The Netherlands, pp. 2-3.
- (Rossi et al., 2005) Rossi F., Lechevallier Y., El Golli A., *Visualisation de la perception d'un site web par ses utilisateurs*. In : Actes des 5^{ème} journées Extraction et Gestion des Connaissances (EGC 2005), Revue des Nouvelles Technologies de l'Information (RNTI-E-3), Paris, France, Pinson S., Vincent N. (eds), Cépaduès-Éditions, 2005, vol. 2, pp. 563-574.
- (Rot, 2005) Rot G., *Le Knowledge Management et l'économie du partage des connaissances. Propos sur un désinvestissement de forme*. Economie et société, 2005, vol.4, pp.675-698.
- (Rouncefield et al., 2006) Rouncefield M.F., Crabtree A., French A., Greenhalgh C., *Developing Digital Records: Early Experiences of Record and Replay*. Computer Supported Cooperative Work, 2006, vol. 15, n°4, pp. 281-319.
- (Roussel et al., 2006) Roussel N., Tabard A., Letondal C., *All you need is log*. In : Proc. WWW2006 Workshop on Logging Traces of Web Activity : The Mechanics of Data Collection. Edinburgh, UK, 2006, 4p.
- (Salembier, 1996) Salembier P., *Cognition(s) : Situées, Distribuée, Socialement Partagée, etc*. Bulletin du LCPE, vol. 1, Ecole Normale Supérieure, Paris, 1996.
- (Salembier, 2002) Salembier P., *Cadres conceptuels et méthodologiques pour l'analyse, la modélisation et l'instrumentation des activités coopératives situées*. Systèmes d'information et Management (SIM), 2002, vol. 7 n°2, pp. 37-56.
- (Sanderson et Fisher, 1994) Sanderson P., Fisher C., *Exploratory sequential data analysis : foundations*. In : Human Computer Interaction, 1994, vol. 9, n°3, pp.251-317.
- (Sanderson et al., 1994) Sanderson P., Scott J.T., Mainzer J., Watanabe L., James J., *MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA)*. International Journal of Human-Computer Studies, 1994, vol. 41, pp. 633-681.
- (Scapin et Pierret-Golbreich, 1989) Scapin D. et Pierret-Golbreich C., *Toward a method for task description: MAD*. In : Proc. Work with display units Conference, Montréal, Canada, 1989, pp. 371-380.
- (Schmidt et Bannon, 1992) Schmidt K., Bannon L., *Taking CSCW seriously : Supporting articulation work*, International journal of Computer Supported Cooperative Work, 1992, vol. 1, no. 1, pp. 7-40.
- (Schmidt, 2002) Schmidt K. *Remarks on the Complexity of Cooperative Work*, In : Salembier P. Benckekroun T.H. (eds). *Cooperation and Complexity in Sociotechnical Systems*, Lavoisier: Paris, 2002, pp. 443-483.

- (Schneider et al., 2005) Schneider M., Bauer M., Kröner A. *Building a Personal Memory for Situated User Support*, In : Prante T., Meyers B., Fitzpatrick G., Harvel L.D. (Eds.). Proc. 1st International Workshop on Exploiting Context Histories in Smart Environments ECHISE 2005, Munich, Germany, 2005.
- (Schreiber et al., 1994) Schreiber A., Wielinga B.J., Akkermans J.M., Van de Velde W., Anjewierden A. *CML: The CommonKADS Conceptual Modelling Language*. In : Proc. European Knowledge Acquisition Workshop: EKAW'94, Springer-Verlag, 1994, p. 283-300.
- (Schreiber et al., 2000) Schreiber G., Akkermans H., Anjewierden A., De Hoog H., Shadbolt N., Van de Velde W. *The CommonKADS Methodology. Knowledge Engineering and Management*. MIT Press, 2000, 455 p.
- (Settoui et al., 2007) Settoui L-S., Prié Y., Marty J-C., Mille A. *Vers des Systèmes à Base de Traces modélisées pour les EIAH*. STICEF, numéro spécial *Analyses des traces d'utilisation dans les EIAH*, 2007.
- (Shannon, 1948) Shannon C.E. *A mathematical theory of communication*. Bell System Technical Journal, 1948, vol. 27, pp. 379-423 et pp.623-656.
- (Shirai et al., 2006) Shirai Y., Yamamoto Y., Nakakoji K. *A History-Centric Approach for Enhancing Web Browsing Experiences*. In : Extended Abstracts of CHI'2006, 2006, pp.1319-1324.
- (Shneiderman, 1998) Shneiderman B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* Third Edition, Addison-Wesley (eds), 1998, 448 p.
- (Siochi et Ehrich, 1991) Siochi A.C., Roger E.W. *Computer Analysis of User interfaces Based on Repetition in Transcripts of User Sessions*, ACM Transactions on Information Systems, 1991, vol. 9, n°4, pp. 309-335.
- (Slavin, 1990) Slavin R.E. *Cooperative Learning. Theory, Research and Practice*. Allyn & Bacon (eds), 1990, 208 p.
- (Soller et al., 2005) Soller A., Martinez A., Jermann P., Muehlenbrock M. *From mirroring to guiding : A review of state of the art technology for supporting collaborative learning*. International Journal of Artificial Intelligence in Education, 2005, n°15, pp. 261-290.
- (Sonnewald et al., 2004) Sonnenwald D.H., Maglaughlin K.L., Whitton M.C. *Designing to support situation awareness across distances: An example from a scientific collaboratory*, Information Processing & Management, 2004, vol. 40, n°6, pp. 989-1011.
- (Soulier, 2003) Soulier E. *Techniques de Storytelling pour le partage de connaissances dans les communautés de pratique*. Thèse en Informatique, Paris: Université Paris VI, 2003, 397 p.
- (Soulier, 2007) Soulier E. *La narrativité face à la temporalité : le traitement informatisé des histoires*. In : in Reber B. et Brossaud C. *Humanités numériques. Nouvelles technologies cognitives et concepts des sciences sociales*, Hermes Publishing, Londres-Paris, 2007.
- (Sowa, 1984) Sowa J., *Conceptual Structures : Information Processing in Mind and Machine*, Addison-Wesley Publishing Company, Reading MA.
- (Spiliopoulou et al., 2003) Spiliopoulou M., Mobasher B., Berendt B., Nakagawa M. *A Framework for the Evaluation of Session Reconstruction Heuristics in Web Usage Analysis*. In : INFORMS Journal of Computing, 2003, vol. 15, n°2, pp. 171-190.
- (Stephanov et Stephanova, 2005) Stefanov K. et Stefanova E. *Analysis of the usage of the Virtuoso system*. In : 12th International Conference on Artificial Intelligence in Education (AIED 2005), Workshop on Usage Analysis in Learning Systems, Amsterdam, Netherlands, 2005, pp. 97-104.
- (Srivastava et al., 2000) Srivastava J., Cooley R., Deshpande M., Tan P.-N. *Web usage mining : Discovery and applications of usage patterns from web data*. In : Proc. SIGKDD Explorations, ACM (eds), vol. 1, n°2, 2000, pp. 12-23.
- (Stumpf et al., 2008) Stumpf S., Sullivan E., Fitzhenry E., Oberst I., Wong W-K., Burnett M.M. *Integrating rich user feedback into intelligent user interfaces*. In : Proc. 13th international conference on Intelligent User Interfaces IUI'2008, ACM, Gran Canaria, Spain, 2008, pp. 50-59.
- (Suchman, 1987) Suchman L. *Plans and situated actions : The problem of human machine communication*. Cambridge: Cambridge Press, 1987, 220 p.
- (Suchman, 1993) Suchman L. *Response to Vera and Simon's situated action : A symbolic interpretation*. Cognitive Sciences, vol. 17, °1, pp. 71-77.

- (Systä, 2000) Systä T. *Understanding the Behavior of Java Programs*, In : Proc. 7th Working Conference on Reverse Engineering (WCRE'00), 2000, pp. 214.
- (Tauscher et Greenberg, 1997) Tauscher L., Greenberg S., *How People Revisit Web Pages : Empirical Findings and Implications for the Design of History Systems*. International Journal of Human Computer Studies, Special issue on World Wide Web Usability, 1997, vol. 47, n°1, pp.97-138.
- (Tchounikine, 2002) Tchounikine P. *Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain*. I3 Information Interaction Intelligence, 2002, vol. 2, n°1, pp. 59-95.
- (Terressac et Chabaud, 1990) de Terressac, G. et Chabaud C. *Référentiel opératif commun et fiabilité*. In : J. Leplat, & G. de Terressac (eds.), *Les facteurs humains de la fiabilité*. Marseille: Octarès, 1990, pp. 111-139.
- (Terveen *et al.*, 2000) Terveen L.G., McMackin J., Amento B., Hill W. *Specifying Preferences Based On User History*. In : Proc. Computer Human-Interaction (CHI'2002) ACM Press, Minneapolis, 2002, pp. 315-322.
- (Theureau, 2009) Theureau J. *L'observation des cours d'action, des cours de vie relatifs à une pratique et de leurs articulations collectives*. In : B. Cahour & C. Licoppe, *L'apport de la confrontation aux traces de sa propre activité*, Telecom-Paritech, Paris, 2009.
- (Timini, 2000) Timini I. et Rouault J.. *La veille sur Internet* [En ligne]. Grenoble : Univ. Grenoble III, 2000. Disponible sur : <http://www.u-grenoble3.fr/les_enjeux/2000/Timini-Rouault/Timini-Rouault.pdf> (Consulté le 16/05/03).
- (Tollmar, 2001) Tollmar K. *Towards CSCW design in the Scandinavian Tradition*, Thèse Informatique, Department of Numerical analysis and Computer Science, Stockholm University, Stockholm, Suède, 2001.
- (Tricot et Nanard, 1998) Tricot A. et Nanard J. *Un point sur la modélisation des tâches de recherche d'informations dans le domaine des hypermédias*. Hypertextes et Hypermédias, 1998, Hors série, pp. 35-56.
- (Tricot, 2007) Tricot A. *Apprentissages et documents numériques*. Collection Psychologie, Éditions Belin, Paris, 2007.
- (Uschold et Gruninger, 1996) Uschold M. et Gruninger M. *Ontologies : Principles, methods and applications*. In : Knowledge Engineering Review, 1996, vol.11, n° 2, 69 p.
- (Vera et Simon, 1993) Vera A.H. et Simon H. *Situated action : Reply to Suchman*. Cognitive Sciences, 1993, vol. 17, n°1, pp. 117-135.
- (Vortac *et al.*, 1994) Vortac O.U., Edwards M.B., Manning C.A. *Sequences of actions for individual and teams of air traffic controllers*, Human-Computer Interaction, vol.9, n°3-4, 1994, pp. 319-343.
- (Weinreich *et al.*, 2006a) Weinreich H., Obendorf H., Herder E., Mayer M. *Off the Beaten Tracks : Exploring Three Aspects of Web Navigation*. In : *World Wide Web Conference 2006*, Edinburgh, UK, 2006, pp.133-142.
- (Weinreich *et al.*, 2006b) Weinreich H., Obendorf H., Herder, E. (2006). *Data Cleaning Methods for Client and Proxy Logs*. In : *World Wide Web 2006*, Workshop Logging Traces of Web Activity : The Mechanics of Data Collection, Edinburgh, UK, 2006, 4 p.
- (Wexelblat et Maes, 1997) Wexelblat A. et Maes P. *Footprints: History-rich Web browsing*. In : Devroye L., Chriment C. (eds), *Conference on Computer-Assisted Information Retrieval (RIAO)*, 5th International Conference, McGill University, Montreal, Canada, 1997, pp. 75-85.
- (Wexelblat , 1998) Wexelblat, A. *History-Rich Tools for Social Navigation*, In : Proc. CHI'98 Summary, ACM, NY, 1998, pp. 359-360.
- (Wexelblat et Maes, 1999) Wexelblat A. et Maes P., *Footprints: History-rich tools for information foraging*. In : Proc. ACM Conference on Human Factors in Computing Systems (CHI'99), 1999, pp. 270-277.
- (Wisner, 1995) Wisner A. *Réflexion sur l'ergonomie (1962-1995)*. Toulouse : Octares, 1995, 158 p.
- (Woodruff *et al.*, 2002) Woodruff A., Szymanski M.H., Grinter R.E., Aoki P.M. *Practical strategies for integrating a conversation analyst in an iterative design process*, In : Proc. Conference on Designing Interactive Systems (DIS 2002). London UK, NY: ACM, 2002, pp. 255-264.
- (Wortham, 2001) Wortham S. *Interactionally situated cognition: a classroom example*. Cognitive Science, 2001, vol. 25, pp. 37-66.
- (Wright *et al.*, 2000) Wright P., Fields B., Harrison M. *Analysing Human Computer Interaction as Distributed Cognition : The Resources Model*. Human Computer Interaction, 2000, vol.15, n°1, pp.1-42.

- (Yahiaoui et al., 2009) Yahiaoui L., Prié Y., Boufaïda Z. *The redocumentation process of computer mediated activity traces : a general framework*. In : Hypertext 2009, Turin, Italie, pp.363-364.
- (Youssefi et al., 2004) Youssefi A.H., Duke J. D., Zaki M. J., Glinert E. P. *Visual web mining*. In : Proc. 13th International World Wide Web Conference (poster session). New York, NY, 2004, pp. 394-395.
- (Zacklad, 2004) Zacklad M., *Transférabilité des connaissances : une re-conceptualisation de la distinction tacite/explicite*. In : actes du colloque 1^{er} colloque luxembourgeois sur l'économie de la connaissance dans une perspective européenne : En route vers Lisbonne. Luxembourg, 2004.
- (Zacklad, 2005a) Zacklad M., *Processus de documentarisation dans les Documents pour l'Action (DOPA) : statut des annotations et technologies de la coopération associées*. In : Le numérique : Impact sur le cycle de vie du document pour une analyse interdisciplinaire, 2004, Montréal, Editions de l'ENSSIB. Disponible sur : <http://archivesic.ccsd.cnrs.fr/>
- (Zacklad, 2005b) Zacklad. M., *Introduction aux ontologies sémiotiques dans le Web Socio Sémantique*. In : Jaulent M.C., Proc.16^{èmes} journées francophones d'Ingénierie des Connaissances, Grenoble: PUG, disponible sur : <http://archivesic.ccsd.cnrs.fr/> (consulté le 20/08/2009).
- (Zacklad, 2007a) Zacklad, M., *Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information*. In : Arsenault C. et Dalkir K. (dir.), Actes du 35e Congrès annuel de l'Association Canadienne des Sciences de l'Information (CAIS/ACSI 2007). Partage de l'information dans un monde fragmenté : Franchir les frontières. Montréal, Canada, 2007.
- (Zaher et al., 2006) Zaher L., Cahier J-P, Zacklad M., *Vers la recherche ouverte d'information* (Poster). In : actes 17^{èmes} journées francophones d'Ingénierie des Connaissances (IC'06), SdC2006, Nantes, France, 2006.
- (Zaher et al., 2007) Zaher L., Cahier J-P, Zacklad M., *De la recherche d'information à la recherche ouverte d'information*. In : Actes de la 4^{ème} conférence internationale des Sciences Electroniques, Technologies de l'Information et des Télécommunications (SETIT), Bouhleb M.S. et Solaiman B. (eds.), 2007.

Table des illustrations

Figure 1.1 : Parmi les multiples significations de la trace nous retenons la trace comme <i>indice</i>	11
Figure 1.2 : La trace comme signe, isolée ou composée toujours interprétée.	12
Figure 1.3 : Plateforme proposée par <i>eLycée</i> , composée d'un outil de visioconférence (<i>Marratech</i>), d'un navigateur <i>Web</i> permettant la co-navigation (<i>eMediathèque</i>) et d'un site Web dédié ici affiché dans le navigateur.....	18
Figure 1.4 : Interface d' <i>Emulsion</i> , atelier de production collaborative de contenus.	20
Figure 2.1 : Exploitation des traces comme support d'analyse.	26
Figure 2.2 : Les finalités des traces d'interactions supports d'analyse.	27
Figure 2.3 : Copie d'écran de <i>Playback</i> - Tiré de (Neal et Simons, 1983, p.81).....	29
Figure 2.4 : Interface d' <i>ExperiScope</i> - Tiré de (Guimbretière <i>et al.</i> , 2007, p.1333).	30
Figure 2.5 : Exemple de visualisation de l'utilisation des fenêtres actives d'un navigateur, permettant la mise en évidence de « <i>patterns of web browsing</i> » - Tiré de (Hawkey <i>et al.</i> , 2005 p.1444)	33
Figure 2.6 : Répartition des actions de navigation - Tiré de (Weinreich <i>et al.</i> , 2006, p.135)	33
Figure 2.7 : Visualisation sous forme de graphe de parcours <i>Web</i> . Tiré de (Beauvisage, 2004, p.144).	34
Figure 2.8 : Arbre couvrant minimal du site de l'INRIA - Tiré de (Rossi <i>et al.</i> , 2005, p.573).	35
Figure 2.9 : <i>Protocol transcript</i> - Tiré de (Pirolli, 2002, p.5).	37
Figure 2.10 : Architecture en six étapes de l'approche de modélisation de comportement proposée par M. Rauterberg et M. Fjeld - Tiré de (Rauterberg et Fjeld, 1998, p.155)	38
Figure 2.11 : Exemple de visualisation proposée par <i>CourseVis</i> : il s'agit ici de montrer pour chaque élève quels sont les concepts maîtrisés ou simplement abordés dans le cours de leur activité.	40
Figure 2.12 : Visualisation (a) d'un indicateur de collaboration et (b) de l'évolution du degré de participation dans <i>Synergo</i> - Tiré de (Avouris <i>et al.</i> , 2004, p.17).	42
Figure 2.13 : Interface <i>Replay tool</i> - Tiré de (Rouncefield <i>et al.</i> , 2006, p.38)	43
Figure 2.14 : Exploitation des traces numériques d'interaction au sein même de l'activité.	44
Figure 2.15 : Finalités des traces numériques utilisées comme support à l'activité.....	45
Figure 2.16 : Architecture globale du système de recommandation. Tiré de (Mobasher <i>et al.</i> , 2000 p.145).	47
Figure 2.17 : Exemple de « <i>Attributed-mapped scrall bar</i> » dans <i>Zmacs</i> , montrant des spots de lecture et d'écriture sur certaines zone d'un code - Tiré de (Hill <i>et al.</i> , 1992, p.4).	49
Figure 2.18 : Historique visuel et annoté (à gauche), et symboles d'annotation (à droite) Tiré de (Nakamura et Igarashi, 2008, p.23).	50
Figure 2.19 : Les historiques de navigation dans <i>Webmap</i> (à gauche) et <i>Padprint</i> (à droite) Tirés respectivement de (Domel, 1995, p87) et (Hightower <i>et al.</i> , 1998, p.121).....	51
Figure 2.20 : Arbre (à gauche) et carte (à droite) de l'historique de navigation avec <i>Footprints</i> . Tiré de (Wexelblat et Maes, 1999, p.7 et p.9).	52

Figure 2.21 : Présentation d'une page <i>Web</i> visitée à la croisée de deux axes temporels. Tiré de (Shirai <i>et al.</i> , 2006, p.1323).	53
Figure 2.22 : La présentation sous forme de ligne de temps de <i>Stuff I've Seen</i> . Tiré de (Ringel <i>et al.</i> , 2003, p.186).	53
Figure 2.23 : Interface principale de <i>Slife</i> .	55
Figure 2.24 : Interface utilisateur de <i>UMEA</i> (à gauche) et le panel de <i>Slife</i> (à droite). Respectivement tirés de (Kaptelinin, 2003, p.356) et d'une capture d'écran de l'application (2009).	56
Figure 2.25 : Interfaces du système <i>SPECTER</i> - Tiré de (Kröner <i>et al.</i> , 2006, p.5).	58
Figure 2.26 : Interface de l'assistant conseiller à la navigation du système <i>Letizia</i> . Tiré de (Lieberman, 1995, p.1).	59
Figure 2.27 : Réseau de notion de <i>PIXED</i> - Tiré de (Héraud <i>et al.</i> , 2004)	61
Figure 2.28 : Interface de <i>SimPLE</i> - Tiré de (Plaisant <i>et al.</i> , 1999, p.349)	62
Figure 2.29 : Journal de navigation – Tiré de (Hawkey, 2006, p.2).	66
Figure 3.1 : Génération d'une trace primitives structuration en <i>états</i> et <i>transitions</i> .	73
Figure 3.2 : Extraction d' <i>épisodes significatifs</i> , sur la base de <i>signatures de tâches</i> .	74
Figure 3.3 : Réutilisation(s) des épisodes significatifs par un agent assistant.	75
Figure 3.4 : Généralisation des systèmes traçants.	76
Figure 3.5 : Architecture générale d'un SBTm.	79
Figure 3.6 : Détail du processus de collecte des traces modélisées dans un SGBT.	80
Figure 3.7 : Détail du processus de présentation des traces modélisées dans un SBTm.	81
Figure 3.8 : UTL partie représentation (1) et trace (2). Tiré de (Choquet et Iskal, 2006-07, p.10-11)	84
Figure 4.1 : Processus de modélisation de tâche.	91
Figure 4.2 : Un exemple d'utilisation de la notation QOC. Les flèches pleines pointent les critères favorables à l'option en regard, et les flèches en pointillés pointent les critères non favorables. Tiré de (Balbo, 1994, p.20).	92
Figure 4.3 : Extrait de l'arbre logico-temporel pour une tâche de messagerie. Tiré de (Balbo, 1994, p.57).	93
Figure 4.4 : Description <i>UAN</i> de la sélection d'un objet - Tiré de (Hartson <i>et al.</i> , 1990, p.188).	93
Figure 4.5 : Relations temporelles dans <i>UAN</i> .	94
Figure 4.6 : La Matrice de Johansen permet de positionner les outils en fonctions des caractéristiques spatio-temporelles de leurs situations d'utilisation – Tiré de (Johansen, 1991).	107
Figure 4.7 : Théories issues des Sciences Humaines mobilisées dans le <i>CSCW</i> , dont les cinq que nous avons retenues et mises en évidence. (* <i>Actor Network Theory</i> , ** <i>Socio-économie des Transactions Coopératives</i>)	108
Figure 4.8 : Niveaux hiérarchiques de l'activité humaine dans la Théorie de l'Activité.	109
Figure 4.9 : Structure d'une activité dans la Théorie de l'Activité.	109
Figure 4.10 : Structuration d'un système d'activités selon Engeström (1997).	111
Figure 4.11 : Classification des supports de travail de K. Kuutti - Tiré de (Bourguin, 2000, p.53)	111
Figure 4.12 : Exemple de schéma utilisé dans l'analyse micro (échange entre deux agents, entendu et réutilisé par d'autres agents dans le même espace) - Tiré de (Decortis <i>et al.</i> , 2000, p.22).	114
Figure 4.13 : Exemple d'un système sociotechnique avec ses entrées/sorties.	116

Figure 4.14 : Catégories de dépendances dans la Théorie de la Coordination.	122
Figure 5.1 : Dimensions de description d'une Activité Documentaire Instrumentée.	127
Figure 5.2 : L'organisation des interactions pour une activité donnée n'apparaît qu'à un niveau de description particulier que le modélisateur cherchera à atteindre avec une trace modélisée.	131
Figure 5.3 : Illustration d'un périmètre de modélisation.	134
Figure 5.4 : Dimensions de l'activité utilisée comme grille d'analyse.	135
Figure 5.5 : Les dimensions formant la grille de l'analyse descriptive.	137
Figure 5.6 : Sélection des observés par schématisation de l'activité.	138
Figure 5.7 : Résumé du processus générique de modélisation de trace avec ses étapes et les principaux outils méthodologiques d'analyse mobilisés.	141
Figure 5.8 : Contextes de modélisation possibles dans l'optique d'une traces modélisées exploitée comme support d'analyse : (1) le modélisateur travaille seul ou (2) collabore avec d'autres acteurs.	142
Figure 5.9 : Contextes de modélisation possibles dans l'optique d'une trace modélisée pour la réflexivité de l'activité : (1) le modélisateur implémente seul ou (2) collabore avec d'autres acteurs.	144
Figure 5.10 : Les deux temps de conception et de déploiement d'un SBTm.	146
Figure 5.11 : La méthodologie de modélisation des traces proposée combine les apports méthodologiques et théoriques des domaines de l'IHM, de l'IC et du <i>CSCW</i>	148
Figure 6.1 : Acteurs impliqués dans les activités de <i>eLycée</i>	151
Figure 6.2 : Structure générale de l'activité des participants à la classe virtuelle.	152
Figure 6.3 : Les outils formant l'environnement proposé par <i>eLycée</i>	153
Figure 6.4 : Contexte de modélisation autour de l'activité de la classe virtuelle.	155
Figure 6.5 : Exemple d'activité pédagogique : de traduction de bande dessinée.	155
Figure 6.6 : Modèle de trace simple pour une activité pédagogique de traduction d' <i>eLycée</i>	157
Figure 6.7 : Module de visualisation « temps réel » dans <i>eMediathèque</i> . Tiré de (Cram, 2007, p.26).	158
Figure 6.8 : Les membres de l'équipe impliqués dans la production de formation.	162
Figure 6.9 : Structure générale de l'activité de l'équipe d'ingénierie pédagogique.	163
Figure 6.10 : Périmètre de modélisation de trace pour l'activité des développeurs chef-produit.	164
Figure 6.11 : Système observé centré sur <i>Emulsion</i> (atelier de production de contenus multimédia).	166
Figure 6.12 : Organigramme simplifié de l'équipe.	167
Figure 6.13 : Cycle de vie des contenus de formation produits (OPS).	168
Figure 6.14 : Les principales « unités documentaires » manipulées par les développeur lors de la création d'une formation et leur type d'indexation.	169
Figure 6.15 : Principales opérations possible depuis l'interface d' <i>Emulsion</i> TM	171
Figure 6.16 : Aperçu d'une description globale de la réalisation d'une activité de développement d'un module de formation (OPS).	172
Figure 6.17 : Données enregistrées lors de la réalisation de l'activité et leurs formats.	176
Figure 6.18 : Organisation et enregistrement d'un exercice de relève.	177
Figure 6.19 : Étapes et outils nécessaires à la présentation d'une trace modélisée en <i>Timeline</i>	178
Figure 6.20 : Interface de l'éditeur d'évènements et exemples d'icônes créées spécialement.	179
Figure 6.21 : Extrait de la <i>Timeline</i> proposée à MC comme aide à la reprise d'activité.	181

Figure 7.1 : Schématisation de la modélisation et de la réutilisation d'une trace modélisée dans le cadre d'une activité collective selon l'approche des Systèmes à Base de Traces modélisées.....	188
Figure 7.2 : Distinction des finalités des traces selon leur nature individuelle ou collective.....	189
Figure 7.3 : (a) Deux traces modélisées individuelles, (b) une trace conjointe croisée, (c) une trace conjointe enrichie.....	193
Figure 7.4 : Collecte serveur.....	194
Figure 7.5 : Collecte client.....	194
Figure 7.6 : Collecte mixte.....	195
Figure 7.7 : Schématisation d'une activité conjointe type pour laquelle chaque acteur génère une trace modélisée individuelle.....	196
Figure 7.8 : Formes concrètes de partage des traces individuelles et/ou conjointes.....	197
Figure 7.9 : Situations d'usage des traces individuelles et conjointes dans une activité conjointe.....	198
Figure 8.1 : Résumé des six étapes du processus générique de modélisation de trace numérique.....	203
Figure 1 : La Matrice de Johansen permet de positionner les outils en fonctions des caractéristiques spatio-temporelles de leurs situations d'utilisation – Tiré de (Johansen, 1991).....	235
Figure 2 : Trèfle des systèmes multi-utilisateurs.....	236
Figure 3 : Transferts possibles entre différents types de connaissances. Tiré de (Nonaka et Takeuchi, 1995).....	237

Annexes

1. Matrice de Johansen et Trèfle de Ellis

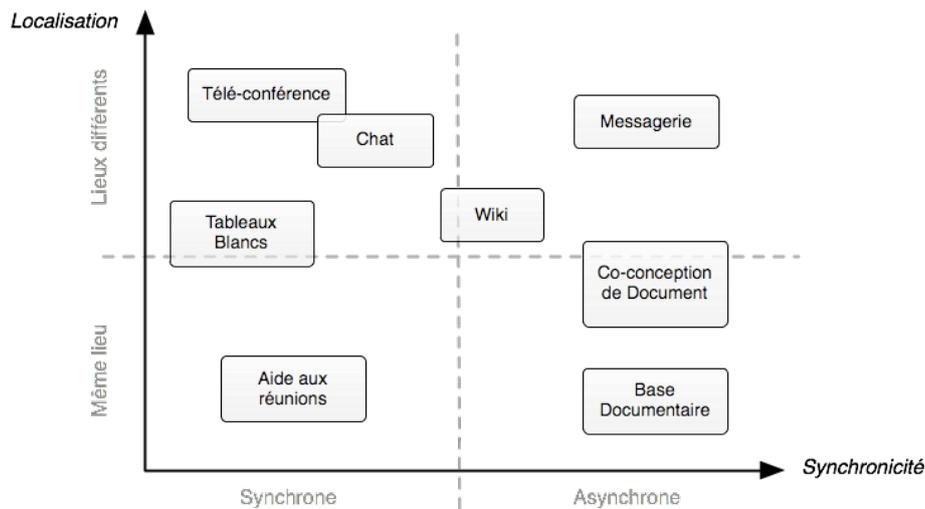


Figure 1 : La Matrice de Johansen permet de positionner les outils en fonctions des caractéristiques spatio-temporelles de leurs situations d'utilisation – Tiré de (Johansen, 1991).

- *Activité synchrone et en présence* : l'activité collective s'appuie sur des outils tels que les vidéo-projecteurs et les tableaux blancs.
- *Activité synchrones et distante* : l'activité collective s'appuie sur des outils de réunion virtuelle (*Real-Time Conferencing*), de vidéoconférence, de chat, de partage d'application ou encore d'éditeurs synchrones ou de tableaux blancs virtuels. La gamme des outils se fait ici plus large. On tente en général de reproduire à distance les outils qui sont à disposition des utilisateurs en face-à-face (tableau blanc par exemple).
- *Activité asynchrone et dans un même lieu* : l'activité collective s'appuie sur certaines solutions de *co-rédaction*, des logiciels de gestion d'argumentation, qui le cas échéant s'intègrent à un intranet, mais on trouve également les mémoires de projet ou les kiosques électroniques.
- *Activité asynchrone et distante* : sans doute la partie de la matrice la plus riche, où l'activité coopérative s'appuie bien entendu sur Internet, les forums, les *Wikis*, les systèmes d'édition collective (CMS), mais également les courriers électroniques.

Bien que cette classification soit *a priori* assez simple, elle n'est pas forcément facile à utiliser. En effet, de nombreux outils cumulent plusieurs dimensions, ce qui rend les distinctions moins pertinentes. De plus, celle-ci ne met pas en avant les *aspects fonctionnels* des outils, ce que propose en revanche une autre classification des outils, celle du « trèfle des systèmes multi-utilisateurs » (Figure 2)

aussi appelé « trèfle de Ellis ». Cette classification propose de regrouper en trois « espaces » les différentes fonctionnalités d'un système coopératif (Gronier et Sagot, 2002).

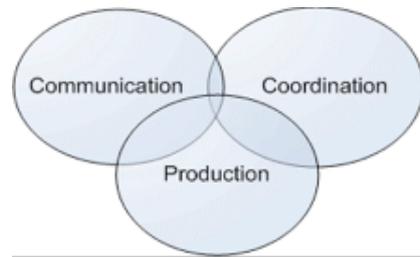


Figure 2 : Trèfle des systèmes multi-utilisateurs.

- *Production* : Les collecticiels sont généralement destinés à supporter un travail de production quelque soit la nature de celle-ci. En d'autres termes, la production correspond à l'exécution effective de la tâche par un groupe d'utilisateurs. Par exemple, lorsqu'il s'agit de la rédaction d'un article avec un éditeur de texte multi-utilisateurs, les fonctionnalités d'édition de texte font partie de l'espace de production.
- *Coordination* : La production collective nécessite une articulation des actions de la part des acteurs, *i.e.* qu'ils coordonnent leurs actions pour progresser dans la réalisation de la production de façon organisée et cohérente. Les fonctionnalités relevant de la coordination englobent par exemple la gestion des conflits et concurrences ou le maintien de la cohérence des actions (que se passe-t-il si deux utilisateurs essaient de modifier simultanément le même mot du texte ?), le séquençement des tâches, la gestion temporelle, etc.
- *Communication* : Enfin, la communication qui est en réalité une communication « Homme-Homme médiatisée » est la dernière feuille du trèfle. Le besoin de communiquer lors de l'exécution d'une tâche coopérative est évidemment très forte, notamment lorsqu'il s'agit de mettre en oeuvre des stratégies de coordination. Pour répondre à ce besoin les outils n'ont apporté dans un premier temps que des possibilités limitées, comme par exemple la possibilité d'utiliser le courrier électronique, puis d'annoter du texte dans la marge. Un peu plus tard ce fût le tour des communications synchrones (Chat) et enfin audio et vidéo, analogiques d'abord, numérique ensuite. Dans tous les cas le contenu sémantique de la communication est étranger au système.

2. Connaissances tacites, connaissances explicites

En soulignant que « *we can know more than we can tell* », I. Nonaka et H. Takeuchi proposent de rendre compte des processus de *transfert possibles* entre ces deux types de connaissances (Figure 3). La *combinaison* tout d'abord, qui est un processus de création de connaissances explicites à partir de la restructuration d'un ensemble de connaissances explicites acquises par différents canaux de communication. L'*internalisation* est un processus d'incorporation de connaissances explicites créant des connaissances tacites qui elles mêmes peuvent se transmettre en partie par l'observation, l'imitation et la pratique constituant le processus de *socialisation*. Enfin, l'*externalisation* est l'explicitation de connaissances tacites en concepts explicites, mobilise le dialogue et l'échange, a recours à la métaphore, à l'analogie ou bien encore à la narration (Soulier, 2003) pour exprimer des concepts difficilement

explicitables. Ce type de caractérisation donne une vision pragmatique et opérationnelle de la connaissance sans en définir le fondement théorique.

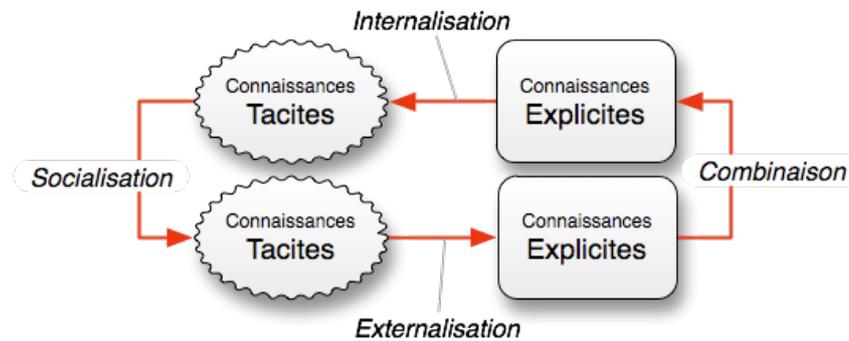


Figure 3 : Transferts possibles entre différents types de connaissances.
Tiré de (Nonaka et Takeuchi, 1995).

3. Donnée, information, connaissance

Le rapport, un peu différent, qu'établit la sémiotique entre ces trois notions est le suivant. La *connaissance* y est vue comme le processus de transformation d'une *donnée*, considérée comme un *signifiant* (fondé sur le symbole ou le signe) en *information*, elle-même considérée comme le *signifié* des données transmises. L'information dans ce cas sera fondamentalement dépendante du codage/décodage des données d'une part, et des connaissances permettant leur interprétation d'autre part (Prince, 1996). Le signe 1500 est une *donnée* dans la mesure où il peut être l'objet de traitements computationnels (mathématiques). Interpréter qu'il s'agit du nombre de pages de ce manuscrit serait une *information* et dans le cadre d'une activité de relecture l'*interprétation* de cette *information*, qui conduirait sans doute le relecteur à aménager son emploi du temps, serait une *connaissance* ayant permis d'agir de la sorte.