

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Reproduction de mécanismes de "conscience" dans un agent logiciel :
Application à un agent d'enseignement intelligent

Document présenté
comme exigence partielle du cours
DIC9410 "Présentation du projet de recherche au
Doctorat en Informatique cognitive"

par
Daniel Dubois

15 avril 2004

Table des matières

Remerciements.....	4
1.Introduction.....	5
2.Présentation sommaire de mon projet de recherche.....	6
2.1.But du projet.....	6
2.2.Objectifs.....	6
2.3.Solution spécifique proposée.....	7
2.4.Intérêts et avantages.....	7
3.État de la Recherche sur les Environnement Intelligents pour l'Apprentissage Humain (EIAH).....	8
3.1.Quelques courants sur la cognition.....	8
3.1.1.Le behaviorisme.....	8
3.1.2.Le cognitivisme	9
3.1.3.L'interactionnisme	9
3.2.Les architectures actuelles d'EIAH : la fin des systèmes d'enseignement linéaires.....	10
3.2.1.Nécessité d'explicitier la connaissance nécessaire à l'adaptation.....	10
i.Modélisation de l'apprenant.	10
ii.Modélisation du domaine.....	11
iii.Modélisation de l'expertise pédagogique.....	11
3.2.2.Nécessité de favoriser la communication avec l'apprenant.....	11
i.L'interaction adaptative en langue naturelle	12
ii.Mises en situation réalistes.	12
3.2.3.Nécessité d'auto-adaptation du système	13
i.Adaptation de l'enseignement au contexte d'apprentissage.	13
ii.Choix de la théorie pédagogique et de la stratégie appropriées.....	13
iii.Adaptation par replanification.....	13
3.3.Quelques exemples d'EIAH actuels.....	14
3.3.1.AutoTutor/Why2000.....	14
3.3.2.TOBIE (Teaching Operators-Based Instructional Environment)	15
4.Problématique identifiée.....	17
5.La conscience.....	17
5.1.La conscience, source inconsciente de solutions pour l'IA.....	17
5.2.Mais qu'est-ce que la conscience?.....	18
5.3.Confusion avec d'autres termes proches.....	19
5.4.Les rôles attribués à la conscience	20
5.5.La théorie de Baars sur la conscience.....	20
5.6.Comment peut-on implanter une conscience dans un être artificiel?.....	21

6.La solution proposée.....	22
6.1.Détail de l'architecture proposée pour un agent d'enseignement "conscient"	24
6.1.1.Modules hérités d'IDA.....	24
i.Le module de "conscience" et les microprocessus attentionnels.....	24
ii.Les tampons pré-conscients et l'atelier.....	25
iii.Le mécanisme de la mémoire parcellaire (sparse distributed memory) : ses "trois" applications....	26
iv.La perception.....	27
6.1.2.Le réseau des actes.....	29
i.Ses composantes et son fonctionnement.....	29
ii.Application du réseau des actes à l'enseignement.....	30
iii.Un première version du réseau des actes.....	31
iv.Exemple plus détaillé d'une séquence d'actes.....	33
v.Exemple d'une chaîne de microprocessus réalisant un acte.....	35
vi.Fonctionnalités prévues dans le réseau des actes.....	35
6.1.3.Les modules spécifiques au ACA et à élaborer.....	37
i.Le modèle de l'apprenant (MA).....	37
ii.Le modèle du domaine (MD).....	38
iii.La métacognition.....	39
iv.La gestion des émotions de l'apprenant.....	39
v.La gestion des émotions de l'agent.....	39
7.Exemple de fonctionnement.....	40
8.Intérêt de la solution.....	43
8.1.Adaptativité élevée.....	43
8.2.Réactivité élevée.....	43
8.3.Apprentissage.....	44
8.4.Proximité cognitive de l'architecture	44
8.5.Clarté et explicitation des structures.....	44
8.6.Facilité pour l'enrichissement et la maintenance de l'agent.....	44
9.Respect des paramètres du Doctorat	44
9.1.Aspects touchant les sciences cognitives : Psychologie, Neurosciences et Pédagogie.....	44
9.2.Aspects touchant l'informatique.....	45
10.Contributions de ma recherche.....	46
11.Échéancier prévisionnel.....	46
11.1.Analyse et réorganisation des programmes d'IDA (Patrick Hohmeyer) : 2 mois.....	47
11.2.Un système d'essai : 2 mois.....	47
11.3.Un premier prototype (simple, mais "complet") : 11 mois.....	47
11.4.Essais avec des apprenants : 3 mois.....	47
11.5.Le plan de travail en plus de détail.....	48
Bibliographie.....	49
Annexes.....	52

Remerciements

L'ordre de mes remerciements n'indique aucunement la préséance de l'un sur l'autre. Chacun des acteurs mentionnés tient un rôle crucial, essentiel dans le chemin que j'ai parcouru jusqu'à maintenant.

Dans un ordre chronologique, je remercie tout d'abord M. Lévesque, qui a su percevoir l'intérêt que je porte aux sciences cognitives et soutenir ma candidature au programme. Je mets tout mon coeur pour remercier mes co-directeurs. Tout d'abord, Roger Nkambou, qui a cru en mon projet, m'a accueilli dans son équipe de recherche et m'a fourni un appui financier appréciable; puis Pierre Poirier, qui n'a jamais compté les heures pour m'aider à traverser les méandres sombres et incertains de la conscience et a toujours su me poser les bonnes questions pour me relancer. Ils ont su permettre la formation, le démarrage et l'orientation de mon projet. Leurs commentaires, leur soutien indéfectible et leur générosité ont fourni le terreau et les "tuteurs" sans lesquels mon projet n'aurait pas survécu.

Il me faut souligner la collaboration toute particulière du professeur Franklin et de l'université de Memphis qui ont accordé au laboratoire GDAC l'accès au code source d'une partie importante de leur agent IDA, fondation essentielle à l'agent que je développe dans ma recherche.

Finalement, quoique cela semble conventionnel, c'est avec une émotion profonde que je remercie mon épouse et ma petite famille pour leur patience et leur compréhension au cours des dernières années, m'ayant permis de me consacrer en priorité à ma recherche.

1. Introduction

La conscience est un mot qui fascine ou qui rebute, cela dépendant souvent de notre champs d'intérêt scientifique : philosophie, neurosciences ou informatique. Quelque soit notre point d'observation, nous comprenons aujourd'hui qu'elle constitue le summum de l'adaptabilité du monde biologique. Le problème, c'est qu'on ne parvient pas facilement à la cerner, à en saisir les contours, à en décrire l'essence. Si on accepte une démarche scientifique exploratoire, on peut poser certaines hypothèses à son sujet puis s'attacher à réaliser le modèle qu'elles supposent, et même implanter ce dernier dans un agent logiciel ou robotique. Au bout de l'effort, on se retrouvera avec une meilleure compréhension de sa réalité et, si les intuitions étaient justes, avec un agent hautement adaptatif. C'est dans ce but qu'une partie de ma recherche s'intéressera à la circonscrire, à identifier les divers aspects qui s'y cachent afin de mieux saisir ceux pouvant nous servir dans la réalisation d'un agent logiciel "conscient" d'apprentissage.

Les chapitres du présent document forment virtuellement quatre blocs. Le premier présente un sommaire du projet : objectifs, solution proposée, intérêts de cette solution. Le deuxième apporte les antécédents qui étayent le projet, présentant un rapide état de la recherche sur les systèmes d'enseignement par ordinateur, y cible une problématique et décrit la solution proposée. Ce deuxième grand bloc décrit aussi un scénario de fonctionnement qui met en évidence le potentiel de l'agent ainsi équipé de "conscience". La troisième section s'intéresse aux paramètres académiques en identifiant la contribution propre de ma recherche et son respect des paramètres du Doctorat en Informatique cognitive. La dernière section décrit les travaux à accomplir et donne une estimation des délais.

2. Présentation sommaire de mon projet de recherche

La compréhension de mon projet requiert le survol initial de plusieurs domaines qui en constituent les fondations : la conscience, les théories pédagogiques et le domaine des EIAH (Environnements Intelligents pour l'Apprentissage Humain). Afin de fournir un fil conducteur dans ce détour et aider à en saisir la pertinence, il m'apparaît judicieux de débiter mon exposé par un sommaire des finalités du projet.

2.1. But du projet

Démontrer la capacité supérieure d'adaptativité que recèlent pour un agent premièrement la conscience, deuxièmement un réseau des actes modifié, et troisièmement leur couplage.

2.2. Objectifs

Dans mon projet, je me propose fondamentalement de poser les jalons principaux d'un Agent Logiciel pour l'Apprentissage Humain¹ hautement adaptatif dont la particularité première tient dans l'utilisation d'une "conscience" artificielle plausible comme l'une des composantes majeures de son adaptativité. Le deuxième objectif, intimement lié au premier, est de développer la deuxième composante de son adaptativité sous la forme d'un réseau d'actes intégrant à la fois une démarche pédagogique et quelques facultés essentielles pour un agent d'apprentissage. Le troisième objectif est de lier cet agent à une théorie pédagogique (et même, d'en faire un ACA multi-théories puisque l'architecture élaborée pour le réseau permettra d'en intégrer et gérer plusieurs, tel que nous le ferons si le temps le permet).

Plusieurs objectifs accompagnent ou découlent de ces objectifs, principalement : création de modèles pour l'apprenant et pour le domaine à enseigner, et traduction de ceux-ci dans des structures compatibles aux mécanismes de la "conscience", au réseau des actes et aux microprocessus; sélection d'une théorie pédagogique et traduction de celle-ci en un réseau d'actes et en microprocessus; élaboration de plusieurs facultés de l'agent en termes de ces mécanismes (dont au moins une stratégie de présentation); finalement, élaboration de microprocessus de métacognition.

¹ Les écrits francophones parlent souvent de "systèmes tutoriels intelligents" (STI, traduction de *Intelligent Tutoring Systems*) -- ou plus récemment de EIAH (Environnement Intelligent d'Apprentissage Humain) dans les textes français, où "apprentissage humain" se situe par opposition à "apprentissage machine". L'appellation EIAH évite la confusion possible avec la stratégie pédagogique du tutorat, qui ne constitue qu'une stratégie de présentation parmi plusieurs possibles. Mon système voulant rester ouvert à une panoplie d'approches, et s'intéressant à reproduire la conscience -- et génériquement l'esprit -- dans un agent logiciel, j'utiliserai un amalgame de ces idées pour produire l'acronyme ALCAH : *Agent Logiciel "Conscient" pour l'Apprentissage Humain*. Mais par soucis de simplicité de lecture, je laisserais tomber les mots "logiciel" et "humain" pour produire l'acronyme final ACA.

2.3. Solution spécifique proposée

Je propose une architecture d'agent logiciel fondée sur trois composantes essentielles : un mécanisme pour une "conscience" fonctionnelle, un réseau d'actes lié à des théories pédagogiques, et des *micro-agents* (dorénavant appelés *microprocessus*). Diverses mémoires, un module de perception et des capacités métacognitives complètent cette architecture.

Une théorie centrale fonde l'architecture de l'agent et une grande partie de son fonctionnement : la théorie de l'*espace global* de Baars (le *Global Workspace*) [Baars 1988, 1997]. Selon celle-ci, l'ensemble de l'esprit humain est mis en oeuvre par un grand nombre de processus simples, presque toujours inconscients, formant des coalitions dont le contenu est diffusé (publié) par la "conscience" à l'ensemble des sous-systèmes afin de recruter les ressources nécessaires au traitement d'une situation. Nous reviendrons à cette théorie dans le texte et nous en détailleront la réalisation. Toutes les facultés de l'agent seront ainsi mises en oeuvre par des microprocessus (et devront avoir été conçues sous cette forme), dans certains cas soumis aux actes du réseau des actes. Ce réseau des actes (RA) reprend une extension réalisée par Stanley Franklin sur l'idée de Pattie Maes pour un mécanisme de sélection des actes [Maes 1989], et va servir à mettre en oeuvre une théorie pédagogiques (peut-être plusieurs) ainsi que plusieurs des facultés de l'agent. Plusieurs théories ou observations expérimentales documentées soutiennent les autres modules et facultés de l'agent.

Notre projet utilise comme tremplin les travaux de l'équipe du professeur Stanley Franklin [Franklin 1998, 2005] ayant résulté en l'agent logiciel "conscient" IDA, un agent d'assignation pour la Marine américaine. En effet, une entente nous a donné accès à plusieurs des modules d'IDA, qui vont fournir à notre agent certains de ses aspects, peut-être directement (mémoires, module de "conscience"), sinon en tant que mécanismes à adapter à nos besoins (réseau des actes, microprocessus, affectivité). Notons ici uniquement que parmi les diverses formes de "consciences" artificielles possibles (tel que nous le verrons plus loin), Franklin a opté pour une mise en oeuvre du type fonctionnel, qui peut se mouler à de multiples applications.

2.4. Intérêts et avantages

L'hypothèse première est que la conscience humaine constitue l'ultime réalisation biologique de l'évolution en ce qui touche à l'adaptabilité de l'être humain [Sloman et Chrisley 2003]; [Seth, Baars et Edelman 2004]). Mon projet de recherche permet d'explorer cette hypothèse ainsi que d'autres concernant sa nature, et certains liens qu'elle entretient avec le reste de l'esprit humain. Peut-être notre projet livrera-t-il un outil pour leur exploration. Mais cela ne constitue pas le centre de ma recherche, qui se veut plus pragmatique. Je vais surtout tenter de démontrer l'efficacité de la conscience en tant que


mécanisme d'adaptation et de régulation, tel qu'exprimé dans les Objectifs. De plus, du point de vue des auteurs d'ACA, elle offre la possibilité d'intégrer un grand nombre de sous-systèmes décentralisés sans requérir un "Grand Organisateur" (par exemple, un module fédérateur comme le *facilitator* du standard FIPA, qui régule les communications dans des contextes multi-agent). Avantage en découlant, elle rend possible l'ajout de nouveaux modules à l'agent sans devoir repenser leur interaction et retoucher l'architecture.

Notre projet veut aussi démontrer que l'ensemble des aspects de "conscience" intégrés dans l'agent et le réseau des actes, individuellement et couplés, fournissent des moyens de planification hautement adaptatif meilleurs à l'idée originale de Maes et globalement supérieur à une majorité de moyens implantés dans les systèmes d'enseignement intelligents actuels. Du point de vue des concepteurs et pédagogues auteurs, l'utilisation d'un réseau des actes offre un cadre de représentation plus intuitif, plus naturel pour l'articulation de la théorie pédagogique qu'un système de production à base de règles.

Je souhaite finalement réussir à démontrer le potentiel de cette architecture pour la réalisation d'un agent artificiel d'enseignement plus "humain" par sa capacité à intégrer un pilotage émotif à l'intérieur du pilotage *pulsionnel et événementiel* de l'agent.

3. État de la Recherche sur les Environnement Intelligents pour l'Apprentissage Humain (EIAH)

3.1. Quelques courants sur la cognition

Avant d'examiner des réalisations de EIAH, nous allons très brièvement mentionner ici quelques thèses sur la cognition à garder en tête lors de l'élaboration d'un environnement d'apprentissage. 

3.1.1. Le behaviorisme

Les théories pédagogiques de Skinner furent les premières élaborées, simples à mettre en oeuvre et efficaces dans des domaines où le taylorisme (principes élaborés vers 1915) avait fait son oeuvre structurante. Tout comme le travail, l'enseignement devenait automatisé grâce au morcellement et à la systématisation du savoir, et grâce à la création de tâches séquentielles permettant un enseignement individuel simple à reproduire. En effet, le behaviorisme ne s'intéresse qu'aux comportements observables et mesurables, sans tenir considération aux processus mentaux cachés et hypothétiques. Le modèle d'enseignement de Skinner (proposé en 1954) était linéaire, son contenu et sa séquence ayant été décidés par les concepteurs du matériel, sans égard aux souhaits, inté-

rêts ou besoins de chaque apprenant spécifique. Des aménagements furent apportés pour permettre plus de souplesse et tenir compte des capacités individuelles.

Cette approche a produit et produit encore de bons résultats pour enseigner des tâches où l'uniformité est requise ou dans des tâches motrices. On pourrait encore l'intégrer avec profit dans un EIAH. Cependant, nous voudrions aller plus loin, en accord avec les recherches et hypothèses plus récentes, qui affirment l'existence de mécanismes cognitifs et d'autres facteurs dont il faut tenir compte pour un meilleur apprentissage. Notons que la théorie pédagogique de Gagné [Gagné 2005] trouve de fortes racines dans le behaviorisme (les conditions de l'apprentissage), mais a évolué pour intégrer des perspectives cognitivistes et l'interactionnistes. On la connaît d'ailleurs davantage en tant que théorie de l'*apprentissage* que de l'enseignement.

3.1.2. *Le cognitivisme*

La thèse cognitiviste cherche à rendre compte de la complexité des opérations cognitives mises en oeuvre dans les stratégies d'apprentissage et de communication [Rézeau 2001]. Ce courant de pensée sur la cognition s'intéresse avant tout au traitement que notre esprit fait des captations appréhendées et des connaissances qui construisent notre historique d'apprentissage. En réaction aux thèses behavioristes, le cognitivisme s'est principalement intéressé aux processus de fonctionnement interne de l'individu, sans référence aux autres individus ni au contexte réel et significatif de son action.

Pour l'IA, cela demande la création d'algorithmes capable de repérer les paramètres signifiants pour les processus mentaux d'une personne (par exemple les interactions bloquant l'apprentissage de la personne) et d'adapter le modèle que le système maintient pour représenter l'apprenant, de moduler conséquemment l'expérience proposée soit en adaptant le contenu et l'interface, soit en sélectionnant d'autre approches pédagogiques ou d'autres types d'activités.

3.1.3. *L'interactionnisme*

Selon cette compréhension des mécanismes de l'apprentissage, l'assimilation de concepts se produit par le truchement de l'interaction. L'interactivité, plus facilement identifiable aux interactions *extérieures au sujet*, se joue d'abord entre interlocuteurs dans un contexte de communication authentique (non artificiellement contrainte), ou plus génériquement dans une interaction du sujet avec le monde. Les difficultés de compréhension et les recherches de sens doivent se résoudre dans la discussion et l'échange de points de vue.

L'agent logiciel d'enseignement peut jouer ce rôle d'interlocuteur s'il sait produire une interaction, un dialogue. Il faudra prévoir des capacités de rétroaction et, dans une version plus audacieuse, des capacités de dialogue.

3.2. Les architectures actuelles d'EIAH : la fin des systèmes d'enseignement *linéaires*

Les architectures récentes d'EIAH sont toutes issues d'un constat fondamental : il importe que l'environnement sache adapter l'interaction pour optimiser les chances d'apprentissage de l'apprenant. Diverses avenues sont possibles. D'un point de vue objectif : considérer le domaine, l'apprenant et l'expertise pédagogique. D'un point de vue relationnel : s'intéresser à l'interface. D'un point de vue subjectif : adapter le comportement de l'agent et moduler l'interaction.

3.2.1. Nécessité d'explicitier la connaissance nécessaire à l'adaptation

i. Modélisation de l'apprenant.

Une étape décisive fut franchie pour l'adaptativité lorsqu'on décida de séparer les types de savoirs des logiciels d'enseignement en règles pédagogiques, modèle du domaine et modèle de l'apprenant. Ce dernier modèle y est crucial, représentant l'apprenant dans ses connaissances et compétences, mais aussi dans son profil psychologique et l'historique de son apprentissage. Ainsi, diverses analyses peuvent être tenues, permettant de déterminer le succès d'une méthode de présentation, d'un type d'activité ou d'une théorie pédagogique; le module répond aussi aux requêtes sur la matière vraisemblablement suffisamment maîtrisée, permettant de déterminer les prochains concepts à aborder. Une méthode populaire pour établir cela est la superposition de modèles (*overlay*), comparant le modèle du domaine au modèle des connaissances de l'apprenant. Cependant, ces connaissances de l'apprenant ne peuvent être établies avec une complète certitude (sauf dans le cas d'une simple mémorisation de termes). Les réseaux bayésiens ont apporté une solution tout-à-fait adéquate à cette situation, permettant d'annoter les connaissances avec une probabilité de possession.

Pour bien encadrer l'apprenant et l'assister dans son apprentissage, il est souvent plus instructif d'identifier ses méconceptions. Les possibilités sont quasiment infinies et excèdent rapidement les capacités et l'imagination d'un concepteur. Même des techniques d'apprentissage automatique (*machine learning*) s'avèrent limitées puisqu'ils ne peuvent excéder la banque d'exemples qui leur sont soumis. [Baffes et Mooney 1996] présente une solution appelée *theory refinement* mise en oeuvre dans le système ASSERT. Le système altère les règles du module expert jusqu'à avoir généré un ensemble de règles productives reproduisant la solution reçue de l'utilisateur. Il sait alors où se trouve l'erreur, de quelle type elle relève et comment elle a modifié la

règle correcte. ASSERT peut de là modifier ses explications et accompagner l'apprenant vers une solution correcte.

ii. Modélisation du domaine

Diverses techniques permettent de représenter le domaine à enseigner. Les ontologies deviennent de plus en plus populaires pour comprendre et décrire le domaine, et des outils sont actuellement en élaboration permettant à un système d'enseignement d'utiliser ces ontologies pour son fonctionnement (dans sa planification ou ses opérations de diagnostic). Il a été jusqu'à maintenant plus fréquent d'utiliser des réseaux sémantiques (non contraints par une ontologie) et des *cadres (frames)* pour modéliser un domaine, parfois en combinant les deux. Lorsque le domaine est plutôt procédural, un système à base de règles s'avère mieux adapté et demeure l'instrument de choix, comme nous l'avons vu au point précédent avec ASSERT.

iii. Modélisation de l'expertise pédagogique

Cette connaissance forme l'essence d'un système d'apprentissage : c'est le savoir-faire de l'enseignant, c'est-à-dire les stratégies dont il dispose et les règles lui permettant de déterminer lesquelles semblent appropriées dans la situation. Ce savoir-faire concerne la planification des séances d'enseignement (choisir le contenu et les stratégies de présentation), le diagnostic des réponses fournies par l'apprenant (expertise qu'on incorpore souvent au modèle de l'apprenant), et la remédiation aux problèmes détectés. Chacun de ces points oriente vers un domaine spécialisé et complexe, ceci démontrant l'importance d'avoir segmenté le thème global. Les systèmes les plus récents s'attachent habituellement à une théorie pédagogique pour guider leur agissement, quoique parfois non explicitement; certains auteurs proposent que le système ait accès à plus d'une théorie afin de pouvoir passer de l'une à l'autre selon les réactions de l'apprenant (qualité de son apprentissage) ou le type de contenu à présenter [Frasson 1998]. Sur cette voie, les ontologies pédagogiques reçoivent de plus en plus d'attention [Bourdeau *et al.* 2004].

3.2.2. Nécessité de favoriser la communication avec l'apprenant

Pour communiquer avec l'apprenant, divers modes sont disponibles : l'apprenant peut indiquer ses intentions dans une interface contrainte (choix de réponses), et le système peut offrir des rétroactions textuelles simples et prédéterminées. Mais de plus en plus on voit des efforts pour améliorer les échanges textuels en langue naturelle, en intégrant dans le système informatique une compréhension de surface par reconnaissance de mots clés, par analyses statistiques, ou par une

véritable analyse sémantique. Encore très peu fréquentes, les recherches avancent pour des interfaces vocaux.

Le mode de communication dépendra principalement du domaine à enseigner. Par exemple, la reconnaissance de la langue naturelle s'avérerait d'utilité limitée dans l'enseignement de règles de logiques, alors qu'elle pourrait aider en résolution de problèmes.

i. L'interaction adaptative en langue naturelle 

La compréhension de la langue naturelle pourrait permettre de véritablement suivre les raisonnements développés par l'apprenant. Le logiciel *Why2-Atlas* s'attache fortement à l'interaction adaptative, et il le fait d'un point de vue résolument constructiviste. Il amène l'étudiant à expliquer en ses mots (en "langue naturelle") les concepts vus parce qu'une telle génération l'amène à véritablement reconstruire le savoir pour lui-même et expose explicitement les compréhensions fausses [Makatchev 2004]. *Why-Atlas* utilisait initialement un classement statistique pour comprendre les réponses de l'apprenant, mais cette approche fut évaluée comme ne produisant pas une compréhension suffisante des raisonnements suivis par l'étudiant [VanLehn *et al.* 2004]. Une analyse syntaxique et sémantique approfondie, répartie sur différents modules, fait maintenant l'analyse des réponses de l'apprenant afin non seulement d'identifier les éléments de réponse corrects, mais aussi pour tenter d'en induire les *croyances* qui les sous-tendent.

ii. Mises en situation réalistes.

Des pistes nouvelles commencent à apparaître pour offrir un environnement d'apprentissage contextuel, c'est-à-dire rapprochant l'apprenant des situations réelles d'où proviennent les connaissances enseignées et où l'apprenant devra à son tour les appliquer. Selon Brown [Brown *et al.* 1989], il importe de reconnaître que la connaissance est *enracinée (situated)* dans l'activité, le contexte et la culture où elle a été créée. En ignorant la nature enracinée de la cognition, l'éducation sape ses propres buts, qui devraient être de transmettre des savoirs utiles, c'est-à-dire utilisables et déployables dans une variété d'applications. Plutôt que par l'écoute passive, apprendre par l'utilisation fournit une richesse de points de vue sur un même savoir, une variété de compréhensions complémentaires provenant du contexte d'utilisation, parfois implicites mais toujours nécessaires à une assimilation profonde et transférable à d'autres domaines. Cet aspect de mise en situation relève avant tout de la création du contenu par le pédagogue, à qui il revient de prévoir des activités signifiantes donnant perspective aux concepts.

3.2.3. Nécessité d'auto-adaptation du système

i. Adaptation de l'enseignement au contexte d'apprentissage.

S'il est à la portée de peu de systèmes informatiques d'agir sur l'environnement physique pour l'adapter aux besoins de l'utilisateur, il est cependant une contextualisation que peut traiter l'IA : celle qui tient compte du contexte personnel de l'apprenant (par exemple, du but qu'il poursuit : l'apprentissage est-il relié à un but professionnel ou de divertissement? L'information désirée est-elle liée à un besoin d'ingénierie ou à un intérêt d'esthétisme?); l'adaptation peut aussi tenir compte du groupe d'interaction en cours ou de sa sélection. Une telle adaptativité se trouve dans les *hypermedia adaptatifs*, évolution des hypermedia statiques. Peter Brusilovsky [Brusilovsky 2001] décrit leur fonctionnement comme d'abord bâtissant un modèle des besoins, buts, préférences et connaissances préalables de chaque apprenant, modèle qui sera utilisé essentiellement pour adapter l'ensemble des interactions. L'adaptation peut modifier les textes (ou leur simple sélection) ou le style des interactions en fonction de la culture nationale d'où origine l'apprenant ou de celle du groupe auquel adhère l'apprenant, de son âge, de son sexe.

ii. Choix de la théorie pédagogique et de la stratégie appropriées

L'agent d'enseignement peut faire varier son comportement de multiples manières et à de multiples niveaux : choix d'une théorie pédagogique, de la stratégie de présentation à appliquer [Frasson 1998], du style à adopter. Il peut faire varier le rythme en multipliant ses interventions, soumettre plus ou moins fréquemment des exercices et des questions, donner préférence à un médium. Ces adaptations doivent être liées à l'apprenant tout en veillant à atteindre les objectifs pédagogiques. Mais la flexibilité décrite n'est souvent rencontrée que dans les écrits des chercheurs car elle demande d'avoir établi le lien entre les divers facteurs et leur solution, et d'avoir développé les capacités annoncées!

iii. Adaptation par replanification

Quoiqu'assez traditionnelle et bien connue, il faut néanmoins mentionner cette forme d'adaptation. Il s'agit ici d'adapter le comportement de l'agent en utilisant un niveau de planification de plus en plus élevé en terme d'ampleur : itération : suggérer un nouvel exercice, reprendre la présentation d'un concept; replanification locale : changer le segment immédiat pour atteindre l'objectif prévu; replanification globale : construire un cheminement nouveau pour passer de l'état actuel au but final.

3.3. Quelques exemples d'EIAH actuels

3.3.1. AutoTutor/Why2000

AutoTutor (<http://www.autotutor.org/>) constitue une réalisation remarquable, fruit d'une recherche interdisciplinaire produite par plus de 35 chercheurs et élément intégrateur d'une série de réalisations élaborées par différents membres du groupe au fil du temps, tel *Why*, *Andes*, *Atlas*, *Coconut* et *APE*. Il possède donc une richesse et une profondeur théorique peu commune particulièrement dans l'utilisation de la langue naturelle. Une hypothèse fondamentale de son développement affirme que c'est la qualité du dialogue entre le tuteur et l'étudiant, et non pas des stratégies tutorielles sophistiquées, qui explique que les tuteurs humains facilitent l'apprentissage. Des études que [Graesser *et al.* 2003] cite montrent que le tutorat permet d'améliorer l'apprentissage d'au moins 1 niveau (de 'C' à 'B', par exemple) avec des tuteurs inexpérimentés. L'article démontre par une étude que ce n'est pas l'utilisation d'un mannequin sophistiqué qui contribue à placer AutoTutor en bonne position, situant ses performances entre le tuteur inexpérimenté et le tuteur d'expérience, mais bel et bien la qualité de ses dialogues. Ces résultats viennent démontrer la validité de leur hypothèse initiale et fondamentale ayant découlé en une stratégie de développement très focalisée, ciblant essentiellement la qualité des dialogues et délaissant les diverses théories pédagogiques -- si ce n'est les théories constructivistes orientant vers l'apprentissage par les questions. L'agent se suffit d'un script pour organiser les problèmes et réponses liés à un domaine. Pour chaque problème, le script possède une réponse idéale, les réponses attendues, les méconceptions, les paires de questions-réponses prévisibles, une liste de concepts importants, et un réseau guidant les tournants conversationnels liés au domaine. L'agent dispose de multiples stratégies touchant l'interaction verbale, mais aucune ne se développe en plusieurs étapes (comme une théorie pédagogique ou une stratégie de présentation)

Ainsi, en dehors de son puissant traitement de la langue naturelle composé de plusieurs techniques, AutoTutor effectue peu de planification et ne modélise pas l'apprenant. Un projet récent reconnaît pourtant l'importance d'autres facteurs et va s'affairer à intégrer un traitement affectif dans l'agent. Peut-être AutoTutor gagnerait-il à considérer davantage l'apprenant et divers aspects de contextualisation.

3.3.2. TOBIE² (Teaching Operators-Based Instructional Environment)

Beaucoup plus "traditionnelle", cette architecture de *coquille* pour un EIAH s'apparente considérablement à mon propre projet. Son noyau essentiel tient dans ses "opérateurs" d'enseignement (OE) inspirés du langage STRIPS (tout comme le sont les noeuds d'actes et les microprocessus du ACA), qui implantent les règles de production d'un graphe ET/OU directionnel multi-niveau avec leurs préconditions, action et effets. À la différence du ACA, les opérateurs d'enseignement prescrivent aussi directement le mode de présentation du matériel (présenter, expliquer, concentrer, rappeler, résoudre un problème, fournir un exercice ou un test) – ils peuvent récursivement faire appel à un niveau inférieur d'organisation de la matière afin de détailler une opérations mathématique, par exemple. Les OE incorporent aussi leurs liens aux opérations diagnostiques gardées dans la bibliothèque des actions pédagogiques. Finalement, un OE spécifie son type d'action pédagogique (par exemple, une explication, un exemple ou un exercice), son niveau de difficulté et le type de médium (texte, graphique, image statique ou animée, son ou vidéo). Tout le savoir de TOBIE se trouve représenté dans un graphe ET/OU directionnel multi-niveau donc les plans créés constituent un sous-ensemble; le graphe global incorpore la décomposition des concepts, la décomposition des buts (tâches) et des étapes de résolution. Les plans produits en constituent un sous-ensemble, tout comme le modèle de l'apprenant. Sur ce dernier point, TOBIE tient compte de plusieurs aspects : préférences pédagogiques de l'apprenant, ses paramètres psychologiques et motivationnels, tels sa capacité de concentration, son niveau de confiance et de ténacité.

La planification de TOBIE s'effectue au départ d'une leçon et permet de spécifier des critères (plan d'enseignement le plus bref, éviter certains concepts, préférer un type, etc.). Le plan élaboré n'est que partiellement spécifié; il n'implique aucune dépendance entre les buts des opérateurs, laissant la mise en oeuvre établir sur le vif quand présenter les concepts afin de saisir les occasions particulières pouvant se présenter (sans toutefois permettre de modifier la teneur du plan, à la différence). Selon le même principe, les diagnostics et remédiations ne font pas partie du plan, se mettant en oeuvre au besoin.

TOBIE partage donc plusieurs bons aspects de mon ACA :

- structure claire du savoir et du savoir-faire de l'agent dans les OE (intégrant les règles d'un graphe ET/OU). Des structures équivalentes devraient se retrouver dans mon agent

2 Il semble que TOBIE se soit arrêté à l'état de prototype, selon les dernières informations que j'ai pu trouver [Vassileva 1997].

d'enseignement : réseau sémantique pour le domaine et le savoir acquis par l'apprenant, réseau des actes pour les savoir-faires.

- planification en plans partiels. Cela réduit la lourdeur de la planification et évite de concevoir des prévisions qui ne s'exécuteront peut-être jamais (suite à des changements dans l'environnement).
- adaptation sur le vif des plans partiels. Seuls les éléments à mettre en oeuvre sont l'objet d'une planification fine. L'ACA effectue pareille adaptation par la construction interactive, adaptée, du plan de présentation (contenu et modes); il le fait aussi au moment de mettre en oeuvre tout élément du plan par le passage quasi obligé par la "conscience" afin de solliciter les ressources appropriées, ce qui produit un plan optimalement adapté à la situation (mais soumis aux valeurs spécifiées pour les paramètres du réseau des actes).
- conservation du plan initial autant que possible. L'ACA obtient un effet similaire par l'accumulation de l'énergie d'activation à l'intérieur d'une même séquence d'actes, qui tend à privilégier la sélection du prochain acte parmi cette séquence (ceci dépendant de l'ajustement des paramètres du réseau des actes).
- Des Opérateurs Diagnostics servent à détecter des situations requérant une intervention, non seulement dans l'exactitude de la solution développée par l'apprenant, mais aussi pour l'accompagner psychologiquement (il est distrait, n'est plus motivé, a déjà eu de telles difficultés, etc.). Ces OD correspondent aux séquences et actes du réseau de diagnostic et du réseau de remédiation dans l'ACA.
- Le modèle de l'apprenant inclut un calque (*overlay*) du domaine et une liste de paramètres psychologiques avec leurs valeurs.

Il se différencie des principes de fonctionnement de mon ACA par cette particularité :

- formalisation centralisatrice des Opérateurs d'Enseignement. Le fonctionnement du ACA s'avère plus répartie, comptant davantage sur les interactions dans sa planification et son adaptation; la plus grande flexibilité permise par mon ACA constitue autant un avantage qu'une difficulté. En effet, l'aspect distribué des capacités demande de mettre en rapport des modèles séparés. Un outil auteur pourrait cependant éliminer cet inconvénient.

Globalement, TOBIE et mon ACA ne se différencient que dans le type fondamental d'agent. TOBIE se veut un cadre de développement symbolique pour un ITS. Mon ACA se veut un agent implantant la théorie de Baars sur l'esprit humain et appliqué à l'enseignement. Cela contraint l'architecture du ACA mais aussi, par la présence d'une "conscience d'accès", porte le potentiel d'étendre à un registre plus large les facultés de l'agent, entre autres une personnalité (partiellement implantée par les pulsions pédagogiques), une affectivité, une métacognition générique, la génération de dialogues, etc., sans avoir à modifier ses principes de fonctionnement ou son architecture. Mon ACA inclut le potentiel d'un agent plus "humain".

4. Problématique identifiée

L'essence de la problématique principale se résume en un mot : complexité. Les EIAH et leurs concepteurs doivent savoir tenir compte d'une multitude d'éléments souvent complexes, de sources multiples et aux formes variées : l'environnement et l'apprenant, l'enseignement et l'apprentissage (deux points de vue complémentaires au contenu différent), le diagnostic des erreurs, et les multiples barrières à la communication homme-machine. Chacun de ces points contient un ensemble important d'aspects, de règles et d'inconnues à traiter, et leurs interactions défient la capacité de planification d'un EIAH autant que la capacité analytique des auteurs. L'adaptativité des EIAH s'en trouve d'autant limitée.

Je crois qu'un EIAH doté, d'une part, d'un mécanisme inspiré de la conscience humaine et, d'autre part, d'un réseau d'actes, recelle la capacité de traiter cette complexité. Mais pour soutenir cette proposition, il faut d'abord bien saisir ce qu'est la conscience humaine et quelles solutions elle apporte à l'adaptativité de l'être humain. Je vais donc d'abord offrir dans les paragraphes qui suivent une description aussi brève que possible de ces points. Par la suite, je pourrai enfin présenter ma proposition et en développer une description détaillée.

5. La conscience

5.1. La conscience, source inconsciente de solutions pour l'IA


L'IA a intégré depuis très longtemps la conscience humaine dans ses réalisations. [Bechtel 1994] rappelle ce fait avec beaucoup d'élégance, expliquant que plusieurs aspects de la conscience semblaient critiques au succès de tout modèle de traitement d'information : traitements automatiques compilés fonctionnant en parallèle sur des sous-systèmes, traitement sériel des résultats importants par le CPU, sélection d'un nombre restreint des stimuli de l'environnement, décisions fondées sur l'examen de


conditions internes au programme, etc.. Tous ces éléments ne pointent vers la conscience que discrètement mais bien réellement. Des auteurs se sont intéressés à déterminer les fonctions de la conscience pour ensuite les implanter dans des algorithmes informatiques [Johnson-Laird 1988].

Des recherches récentes, particulièrement stimulées par les découvertes en neurosciences, s'affairent à intégrer de manière explicite la conscience dans diverses réalisations : modèle de l'esprit humain, modèle de la conscience, ou implantation effective de l'un ou de l'autre dans un agent. Comme nous le verrons au point suivant, plusieurs efforts proposent soit la réalisation d'une "simulation" convaincante des *fonctions* connues la conscience (une récréation *fonctionnelle*), soit la création d'une "authentique" conscience artificielle (*biologiquement plausible* car soutenue par des mécanismes claqués sur la physiologie). Mais avant de présenter ces possibilités, il s'avère nécessaire de mieux cerner la réalité de la conscience.

5.2. Mais qu'est-ce que la conscience?

Parce qu'il s'agit d'un terme couvrant un ensemble de concepts, le même mot évoque des compréhensions différentes et, de là, des opinions souvent divergentes sur son essence, voire son existence même, et, bien sûr, sur son applicabilité à des machines.

Les diverses descriptions proposées par un dictionnaire nent une bonne idée de l'entité dont on parle. Cependant, ces définitions n'expliquent aucunement comment on pourrait la recréer. Il faut plutôt fouiller du côté des écrits plus élaborés de divers philosophes, psychologues et neuroscientifiques qui, s'ils peuvent faire avancer notre étude, offrent un choix de directions assez diverses! Soulignons que les positions vont de l'inutilité des tentatives de la saisir ([McGinn 1991], [Chalmers 1995]), [Searle 2003]), jusqu'à la négation de son existence parce qu'elle n'est qu'une illusion créée par des mécanismes biologiques (philosophie ou neurosciences réductionnistes [Dennett 1991], [Churchland 1989]). On la verra autrement expliquée, par exemple, par la synchronie de divers modules du cerveau autour de 40 Hz [Crick et Koch 1995], dans les rappels occasionnés par les boucles de *réentrence* [Edelman 1989], ou par la connectivité variable de coalitions temporaires [Baars 1988], [Dehaene, Kerszberg et Changeux 1998].

Ned Block a proposé en 1995 [Block 1995] qu'on ne pourra bien discuter de la conscience que si on reconnaît qu'elle recouvre plusieurs phénomènes et mécanismes. Il déclare que le terme cache en fait quatre "types" de conscience  conscience *d'accès* (qui donne accès aux vastes ressources inconscientes), la conscience phénoménale (qui contient la qualité propre à une "expérience"), la conscience *de surveillance* (qui demeure attentive aux processus et états internes), et la conscience de soi (la conscience de posséder notre existence propre et séparée du reste du monde). Cette individuation


des aspects nous permet enfin d'envisager en quoi et comment la conscience peut fournir des modèles de solutions à l'IA, et pourquoi mon projet envisage faire usage de cette fondation inusitée pour un agent. Le domaine des systèmes d'information a surtout utilisé les vues "conscience d'accès" et "conscience de surveillance" parmi toutes ces possibilités, quoique les projets plus récents s'intéressent aussi à la conscience de soi (parce qu'elle fonde l'intégrité de l'agent, la continuité dans ses choix), et même à la conscience phénoménale. Un premier niveau de conscience de soi, sans vouloir ouvrir une épineuse discussion philosophique, peut être considéré implanté dans des processus référant aux buts fondamentaux et préférences de l'agent dans leurs prises de décision (que ces buts et pulsions aient été explicités par le concepteur de l'agent ou laissés implicites dans l'ensemble des programmes le constituant).

Dans le cadre limité de cette présentation de projet, j'évite d'entrer dans le lourd débat tentant de déterminer si vraiment il peut y avoir une "conscience" dans une machine, *quelqu'un* qui y existe vraiment, tel que le disait Descartes. Qu'il suffise de rappeler le célèbre test de Turing, repris dans un écrit récent d'Harnad [Harnad 2003] qui soutient qu'un être recréé par ingénierie, s'il sait nous tromper sur son essence durant sa vie entière, mérite d'être considéré conscient au même titre que n'importe quel de nos congénères : sur quelle base accorderions-nous à ces derniers d'être conscients, par quel examen, sur le repérage de quelle caractéristique immatérielle, et la refuserions-nous à un être artificiel totalement convaincant.


5.3. Confusion avec d'autres termes proches

Mon rapport de stage de l'été 2004 consacre sept pages à séparer plusieurs concepts proches : conscience, intelligence, *awareness*, *sentience*, réflexivité et métacognition. Le document de ma thèse reprendra ces fines distinctions, dont je ne donnerai ici qu'un résumé très succinct. D'abord, qu'intelligence et conscience sont deux réalités séparées, mais en interaction intime, et probablement la présence de l'une est-elle proportionnelle à la présence de l'autre dans un être. **La conscience améliore l'intelligence**, et l'intelligence donne des moyens à la conscience pour exister. Toutes deux poursuivent un but commun : l'adaptativité du système qui les contient. Deuxièmement, sensibilité (ma traduction de *sentience*) et avisence³ (ma traduction d'*awareness*) se ressemblent suffisamment pour ne pas en discuter ici, si ce n'est que pour dire que l'avisence découle de la sensibilité et donne à un être sa connaissance d'un événement ou d'un état. Troisièmement, réflexivité et métacognition sont souvent confondues entre elles, et parfois avec la conscience. En peu de mots, il s'agit de la faculté cognitive qui permet au sujet de réfléchir sur comment il réfléchit. De tels mécanismes désignent également les

3 Devant l'insuffisance du français dans le domaine de la conscience, pour traduire *awareness*, je propose ce néologisme : avisence, composition de 'avis' ("être avisé de quelque chose") et du suffixe 'ence' ("qui a le caractère de")

mécanismes de régulation ou de contrôle du fonctionnement cognitif. La métacognition  est la réflexivité pour diriger l'attention de processus analytiques vers la conscience elle-même. Elle utilise aussi la conscience pour effectuer son travail régulateur en avisant l'ensemble de sous-systèmes de ses observations et inférences. Réciproquement, la conscience sollicite la métacognition par l'information qu'elle publie des états et événements n'ayant pas été traités automatiquement par les processus inconscients. La métacognition n'est donc pas la conscience elle-même; elle est une collection de processus analytiques qui *font usage* de la conscience pour réaliser leurs tâches de supervision et de régulation.

5.4. Les rôles attribués à la conscience

Au-delà des classifications et des attributs  qu'on lui reconnaît, il importe d'établir les rôles qu'on accorde à la conscience et qui lui confèrent sa valeur. Au chapitre dix de sa théorie, Baars en énumère dix-huit : "1. Definitional and Context setting Function; 2. Adaptation and Learning Function; 3. Editing, Flagging and Debugging Function; 4. Recruiting and Control Function; 5. Prioritizing and Access Control Functions; 6. Decision-making and Executive Function; 7. Analogy-forming Function; 8. Metacognitive and Self monitoring Function; 9. Autoprogramming and Self-maintenance Function." [Baars 1988]

Dans plusieurs de ces termes, on peut pressentir l'intérêt que comporte la conscience pour des réalisations informatiques/robotiques : sensibilité au contexte d'application, adaptation et amélioration des processus, apprentissage, épuration des entrées (*debugging*), appel à des ressources compilées et rapides, établissement des priorités, supervision des décisions et de leurs résultats (métacognition), fidélité aux buts fondamentaux (conscience de soi). J'annonce dès maintenant que ces points vont se voir intégrés dans l'architecture proposée par notre projet, tel que décrit à partir du point présentant la solution proposée (page 22). Voyons cependant immédiatement comment la théorie les présente.

5.5. La théorie de Baars sur la conscience

La théorie de l'*atelier global* [Baars 1988] constitue une architecture cognitive comportant un rôle explicite pour la conscience. Selon son auteur, 1) l'esprit peut être vu comme une collection de réseaux spécialisés distribués (processeurs); 2) la conscience est associée à un atelier global dans le cerveau – une capacité de mémoire éphémère dont le contenu est largement diffusé à de nombreux réseaux spécialisés inconscients; 3) inversement, l'atelier global sert aussi à intégrer plusieurs réseaux d'entrée compétitifs et collaborateurs; 4) certains réseaux inconscients, appelés *contextes*, influencent les contenus conscients; 5) de tels contextes collaborent pour contraindre les événements conscients; 6) les motifs et les émotions peuvent être vues comme des contextes de buts; 7) les fonctions supérieures constituent des hiérarchies de contextes de buts [Baars et Franklin 2003]. Détenteur de son doctorat en psychologie, Baars s'est vu décerner le titre de *Senior Fellow in Theoretical Neurobiology* par

l'Institut des Neurosciences de San Diego, reconnaissance de la valeur de ses travaux liés à ce domaine. Sa théorie excède le seul monde de la psychologie et organise dans un cadre cohérent une variété de travaux d'autres chercheurs, dont plusieurs en neurosciences, et les hypothèses qu'il propose se voient de plus en plus soutenues par les études et expériences empiriques (voir par exemple [Baars 2002]).

5.6. Comment peut-on implanter une conscience dans un être artificiel?

Si la conscience a souvent été décrite dans ses attributs et ses fonctionnalités de manières plutôt vagues, plusieurs modèles de la conscience ont néanmoins déjà été élaborés par divers auteurs, souvent à l'intérieur d'un modèle plus complet de l'esprit ([Baars 1988], [Dennett 1991], [Edelman 1989], etc.). Ce qui peut surprendre davantage, c'est que plusieurs réalisations informatiques en ont aussi déjà été réalisées, soit comme créations autonomes de la conscience, soit à l'intérieur d'agents. Cependant, elles n'originent pas toutes du même point de départ, ni ne poursuivent les mêmes objectifs. Elles peuvent avoir les visées pragmatiques de l'ingénieur qui fait feu de tout bois, tout comme elles peuvent plutôt chercher à atteindre divers niveaux de réalisme et de plausibilité biologique. Je propose ici un regroupement simple de ces réalisations : fonctionnelles ou fidèles à la physiologie ("biologiquement plausibles").

5.6.1. Les réalisations *fonctionnelles*

Les réalisations dites *fonctionnelles* ne veulent reproduire que les rôles tenus par la conscience. Dans certains cas, les auteurs respectent une certaine plausibilité, par exemple en se liant à une théorie psychologique de l'esprit ou de la conscience (ex.: Conscious Mattie et IDA, liés à la théorie de Baars; Clarion de Sun, soutenu par diverses données expérimentales et théories cognitives); dans d'autres, seuls sont recherchés les avantages promis par le mode de fonctionnement supposé de la conscience, la manière de les mettre en oeuvre important peu (ex.: Langage réflexif de McCarthy [McCarthy 2002], GLAIR [Hexmoor, Lammens et Shapiro 1993], Introspect [Cazenave 1998], CAMEL [Sabah 1999]). Des architectures fonctionnelles de l'esprit humain, quoique très célèbres, n'entrent pas dans la présente catégorie puisqu'elles ne s'intéressent pas à la conscience, ou seulement d'une manière très secondaire. Je les mentionne pour ne pas laisser d'ambiguïté à leur sujet. ART, de Stephen Grossberg [Grossberg 1999], se veut avant tout un modèle de la cognition humaine et mentionne dans un seul article sa correspondance possible avec la conscience. Similairement, les très célèbres SOAR [Lewis 2001] et ACT-R [Anderson 2004] constituent d'autres modèles éminemment dignes de mention et d'examen, mais qui ne s'intéressent qu'à

modéliser les processus cognitifs ou l'esprit dans son ensemble, sans considération (sauf de manière très accessoire) à la conscience en elle-même.

5.6.2. Les réalisations *biologiques*

Les réalisations que j'appelle "biologiques" veulent non seulement accomplir ce que fait la conscience, mais aussi atteindre ce but en respectant l'architecture des organes biologiques supposés la supporter. Nous pouvons nommer ici CyberChild [Cotterill] et Cog [Brooks]. Il faut immédiatement souligner que les propositions *physiquement* biologiquement plausibles (expliquant le phénomène par la simulation des organes biologiques au niveau cellulaire et physiologique, et élaborées selon le mode de *rétro-ingénierie* de [Harnad 2003]) sont actuellement rares (voir [Kurzweil 1999]; voir aussi les pistes lancées sur le sujet des micro-tubules), et qu'il n'existe présentement que quelques *pistes* pour des propositions biologiquement *véritables* (faisant usage de tissus synthétiques).

6. La solution proposée

Nous disposons maintenant d'un vocabulaire plus précis et de plus nombreux concepts pour décrire la solution proposée. Considérant la foule d'outils que la conscience regroupe pour permettre les meilleurs agissements possibles, il m'apparaît tout naturel de vouloir réaliser l'implantation d'une conscience artificielle comme mécanisme central d'un agent, en l'occurrence un agent d'enseignement. Je viendrai à traiter de l'aspect pédagogique de la solution à l'intérieur de ma discussion sur l'aspect cognitif de l'agent, au point «Application du réseau des actes à l'enseignement» de la page 30.

Tenter une réalisation biologiquement plausible est hors de portée dans le cadre de ma recherche doctorale. Les plus grands chercheurs n'ont encore produit que des réalisations incomplètes fondées sur des souhaits autant que sur des connaissances solides, et la complexité des modèles est redoutable et difficile à mettre à profit pour une application spécialisée. En contre-partie, faire des choix eclectiques encourt le danger de produire un résultat n'ayant que peu de parenté avec la véritable conscience. De plus, on risque de négliger des aspects apparemment secondaires mais effectivement cruciaux, et de ne tirer qu'une partie plus ou moins importante du potentiel accessible.

La voie la plus prometteuse pour développer un agent me semble actuellement de s'attacher à une théorie de la conscience suffisamment détaillée et étayée par des études reconnues. Voilà la voie qu'a poursuivie le professeur Stanley Franklin. Avec son équipe, il a élaboré une lignée d'agents dont les fonctions sont réalisées par des *microprocessus* (*codelets*). Alors que Virtual Mattie n'était pas dotée d'une "conscience"

explicite, et fut suivie de Conscious Mattie, IDA se veut la mise en pratique de la théorie psychologique de Bernard Baars, la théorie du *Global Workspace* qui regroupe dans un ensemble cohérent divers travaux complémentaires issus de nombreux chercheurs (voir le point sur la théorie de Baars). En bref, elle décrit le fonctionnement de l'esprit comme étant le résultat de la collaboration d'un grand nombre de processus simples et spécialisés, la plupart du temps automatiques et inconscients. Ces processus forment des coalitions qui entrent en compétition pour obtenir l'accès à la conscience lorsque des informations (situations, événements) ne peuvent être traités par les processus automatiques et requièrent des ressources supplémentaires.

Le laboratoire GDAC a conclu une entente avec l'université de Memphis donnant accès au code source de plusieurs modules de l'agent IDA. Ma recherche s'appuie donc sur une architecture majoritairement implantée et tout à fait compatible à un réalignement vers un agent d'enseignement. Si certaines fonctionnalités d'IDA devraient pouvoir être utilisées directement (mémoires, mécanismes de la "conscience" d'accès, mécanismes du réseau des actes), d'autres paquetages reçus ne fournissent qu'une mécanique vide, à modifier et instancier selon nos besoins (réseau des actes, microprocessus, perception). Toutes nécessitent un examen approfondi pour comprendre leur fonctionnement exact établir comment les utiliser dans le cadre d'un ACA et déterminer comment réaliser les éléments manquants afin qu'ils forment à nouveau un agent complet. L'ensemble du contenu propre au ACA (modèle de l'apprenant, modèle du domaine, expertise pédagogique) ainsi que la métacognition de l'agent résulteront d'une écriture complète .

Conséquemment, en termes simples, ma recherche tient premièrement dans l'élaboration d'un réseau d'actes comportant toutes les capacités requises pour la réalisation du prototype minimal d'un Agent d'Apprentissage "conscient". Elle tient en même temps dans l'élaboration des microprocessus (*codelets*) mettant en oeuvre les actes du réseau ainsi que les fonctionnalités des autres modules et aspects de l'agent. Des tâches "secondaires" devront aussi être accomplies : élaboration d'un modèle de l'apprenant et du modèle d'un domaine d'enseignement, tous deux dans une forme compatible aux deux architectures centrales ("conscience" et réseau des actes), élaboration des mécanismes perceptuels nécessaires, et élaboration de capacités métacognitives pour l'agent.

6.1. Détail de l'architecture proposée pour un agent d'enseignement "conscient"

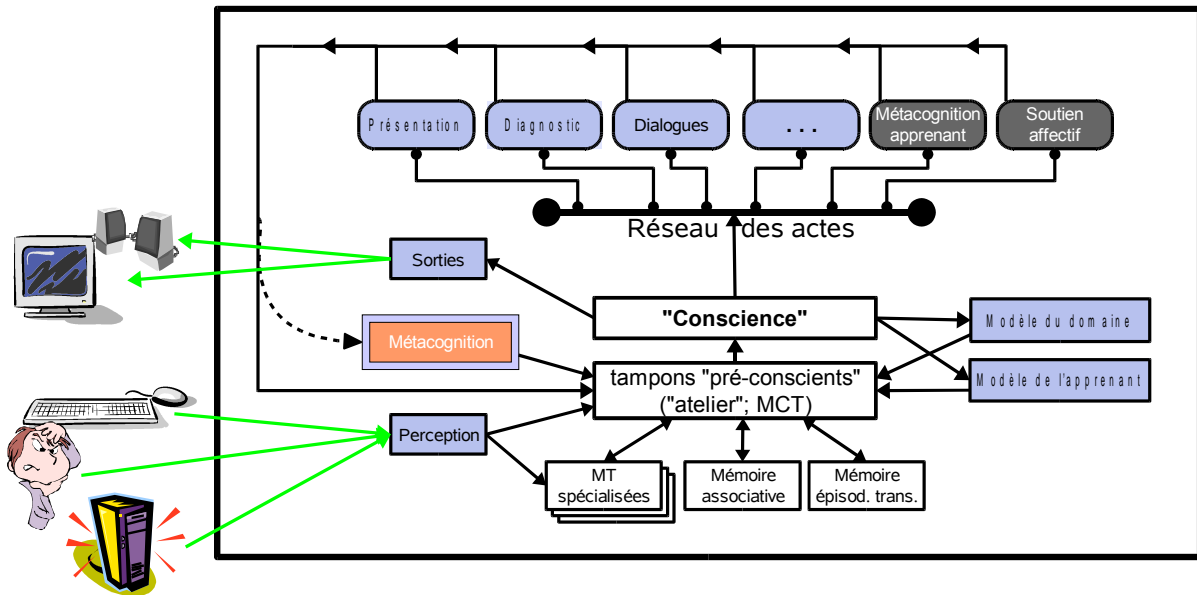


Illustration 1: rchitecture fonctionnelle de l'agent d'enseignement. Les différents "modules" surplombant le réseau des actes sont en fait des sous-réseaux composant le réseau global. Les effets et sorties de ces modules vont soit retourner des informations à la "conscience", soit produire des rétroactions pour l'apprenant à transférer au module des sorties, soit établir des communications bi-directionnelles avec l'extérieur, des banques de données par exemple. La couleur lilas indique les exemples d'aspects à développer, et le gris, des exemples de fonctionnalités souhaitables mais non prévues pour le prototype.

Le diagramme de l' montre une vue de haut niveau de l'architecture fonctionnelle de l'agent d'enseignement. Il est quelque peu difficile de devenir familier avec une telle architecture, son fonctionnement, ses possibilités. Le traitement y est tellement parallèle qu'on en oublie constamment certains aspects et leurs impacts. La simplicité apparente du modèle est trompeuse, et la richesse des interactions s'avère un défi à embrasser. Je vais en décrire tout d'abord les modules d'IDA utilisables directement, sans modification (à première vue) et montrés par des cases claires : mécanisme de la "conscience", atelier, mémoires à long-terme; je toucherai aussi brièvement le module de la perception tel qu'implanté dans IDA, mais dont l'utilité pour mon ACA demeure actuellement indéterminée. Je détaillerai ensuite le réseau des actes et une première implantation. Je proposerai finalement des hypothèses de travail pour les autres modules uniques au ACA (par rapport aux fonctionnalités d'IDA) : modèle du domaine, modèle de l'apprenant métacognition.

6.1.1. Modules hérités d'IDA

i. Le module de "conscience" et les microprocessus attentionnels

Dans cette architecture, en accord avec la théorie de Baars, la conscience est multiforme : d'abord dans le module de "conscience" qui réalise la conscience d'accès (le *global workspace*), mais aussi dans les microprocessus attentionnels autonomes ou existant à l'intérieur de plu-

sieurs tâches spécialisées, microprocessus réalisant la conscience de supervision ainsi que la métacognition. En eux-mêmes, ils ne possèdent qu'un niveau élémentaire de "conscience", la conscience de leur environnement (avisence, *awareness*), qui les amène à se mettre en oeuvre advenant la détection de conditions spécifiques. On trouve aussi la conscience de soi à travers des mémoires, mécanismes décisionnels, "pulsions" et microprocessus métacognitifs.

Informatiquement parlant, le module de "conscience" est réalisé par un gestionnaire des coalitions, un gestionnaire de l'attention (*spotlight controller*), un gestionnaire des publications (*broadcast manager*), et une collection de microprocessus attentionnels capables de reconnaître des situations nouvelles ou problématiques requérant l'intervention de la "conscience". Il est impossible de bien décrire le module de "conscience" sans y relier l'atelier. Au point suivant, nous décriront leur interaction.

En ce qui concerne les microprocessus attentionnels, si génériquement leur tâche est de repérer l'avènement d'un état ou d'un événement particulier, il s'en trouve de divers types et pour diverses applications, et l'information qu'ils recherchent leur est spécifique. Par exemples, certains surveillent l'atelier, où tous les événements et états requérant publication viennent s'écrire; d'autres surveillent le tampon d'entrée (contenant les captations) afin de se mettre à l'oeuvre pour débiter la perception; d'autres supervisent les retours d'informations des mémoires à long-terme pour s'assurer qu'elles ont retourné une information pertinente; d'autres encore écoutent les publications afin de marquer (*to prime*) les séquences du réseau des actes pouvant traiter la situation. Le but de leur repérage est d'amener la connaissance sur la situation dans l'atelier et, ultimement, de faire entrer les informations dans la "conscience" à fin de publication. Ainsi, leur travail de bas niveau collabore avec les structures de conscience de plus haut niveau afin de produire une "conscience" ayant plusieurs liens de parenté avec la conscience humaine.

ii. Les tampons pré-conscients et l'atelier

L'atelier⁴ correspond à peu près à la mémoire à court terme et joue plusieurs rôles. Il recueille toutes les informations nécessitant l'intervention de la "conscience"; il sert de tableau privé où s'élaborent les pistes de solutions; il offre un passage obligé par où passent les perceptions et sollicitent ainsi une réponse des mémoires à long terme. Il est constitué exclusivement de registres prévus pour des types d'informations pré-déterminés.

4 ou workspace, ainsi appelé par Franklin malgré sa similitude au nom de la théorie de Baars, global workspace, qui désigne l'ensemble des mécanismes amenant la conscience à l'esprit.

Les processus y présentent les informations qu'ils souhaiteraient voir publier : perceptions (événements et états observés dans le monde extérieur à l'agent), événements et états internes qui ne peuvent être traités automatiquement. Chaque microprocessus attentionnel surveille l'atelier pour y repérer une situation qu'il connaît. Advenant le cas, ce microprocessus va recueillir autour de lui les microprocessus informationnels qui transportent l'information décrivant la situation. Les gestionnaire des coalitions reconnaîtra ce regroupement ou jugera approprié d'en former une nouvelle coalition; à partir de ce moment, le groupe fait partie de la compétition pour "obtenir l'attention". Plus l'information récupérée correspond à ses critères, plus le microprocessus attentionnel peut élever son niveau d'activation et signaler son importance au gestionnaire de l'attention. Ce dernier examine toutes les coalitions présentes dans l'atelier (incluant des coalitions issues des cycles précédents, plus ou moins en perte d'activation) et sélectionne celle la plus activée, lui accordant l'entrée dans la "conscience" afin de voir son contenu publié. Le gestionnaire des publications accomplit cela.

L'atelier permet aux perceptions (obtenues suite à la reconnaissance des captations) de stimuler les mémoires à long-terme. Certains des registres qui le composent, formant le *focus*, ont une fonction spécifique de lien avec les mémoires à long-terme. Lorsque le module perceptuel dépose la perception réalisée dans les tampons perceptuels de l'atelier, les mémoires associative et épisodique sont automatiquement stimulées et cherchent à repérer les informations qui y sont associées, et retourne dans les tampons de retour ce qu'elles ont récupéré. Ces informations vont venir compléter la coalition qui a commencé à se former durant le rappel des mémoires.

iii. Le mécanisme de la *mémoire parcellaire (sparse distributed memory)* : ses "trois" applications

Dans les écrits de l'équipe de Franklin, les mêmes mémoires reçoivent des appellations diverses, variant d'un écrit à l'autre et au fil du temps, alors que l'architecture des agents sont modifiées. Cela rend difficile une description juste, entre autres pour les mémoires! La mémoire associative reçoit parfois l'appellation de mémoire "déclarative" ou simplement "associative" (dans les écrits plus anciens). Jusque là, pas trop de problème. Un article parmi les plus récents (2004) la désigne maintenant comme soutenant à la fois les fonctionnalités d'une mémoire déclarative et celles d'une mémoire épisodique (que divers écrits nommaient "autobiographique"), produisant une nouvelle appellation "*autobiographical/declarative memory*" [Ramamurthy, D'Mello et Franklin 2004]). Cette mémoire à double fonctionnalités sert à enregistrer les éléments associés dans la "conscience" (par une coalition), qu'ils soient issus de la perception

(concepts, événements et états externes) ou de processus internes (événements et états internes). Un troisième type de mémoire dynamique s'est ajoutée (voir le même article), la mémoire épisodique transitoire (Transient Episodic Memory), aussi mise en oeuvre en utilisant un mécanisme de *mémoire parcellaire*, alors que mémoire épisodique du prédécesseur d'IDA, Conscious Mattie, avait été réalisée dans un mécanisme à base de cas afin de mieux servir comme mécanisme d'apprentissage pour le réseau perceptuel (nouveaux concepts) et le réseau des actes (nouveaux actes). À la différence de la mémoire autobiographique (mémoire épisodique à long-terme), le contenu de la mémoire épisodique transitoire se dégrade (perd son activation dans un délai assez bref, en termes d'heures. Il vise des informations d'une durée utile assez brève servant dans les interactions courantes et l'adaptation relativement immédiate (un exemple tiré de la vie courante : "Où ai-je rangé ma voiture ce matin, dans ce grand stationnement?). Périodiquement, les informations subsistantes (n'ayant pas encore décliné sous un seuil donné) se voient consolidées dans la mémoire associative déclarative.

La *mémoire parcellaire* (*sparse distributed memory*) utilise une version modifiée du principe conçu par Kanerva [Kanerva 1988], [Ramamurthy, D'Mello et Franklin 2004], un mécanisme plutôt idéal pour une mémoire à long-terme parce que ses comportements y ressemblent : il permet d'accéder aux informations à partir d'indices (informations plus ou moins complètes) à la place d'adresses physiques; plus une même information est reçue souvent, plus elle est écrite souvent et renforce sa présence en mémoire, ce qui fait qu'un item bien répété peut être récupéré avec des indices plus petits; à l'opposé, un item non répété voit sa trace progressivement détériorée par les nouvelles écritures ayant des traits de parenté, ce qui peut créer initialement le phénomène du "mot sur le bout de la langue" (la recherche ne parvient pas à converger vers une valeur stable) puis dégénérer vers l'oubli (n'est plus repérable, sauf par chance ou avec un indice très fort).

iv. La perception

Les besoins perceptuels d'un agent dépendent du domaine d'application et du traitement qu'on souhaite effectuer (par exemple, la chauve-souris ne possède pas les mêmes instruments de captation que l'oiseau, bien qu'ils partagent le même espace de chasse!) Similairement, le ACA sera doté de capacités appropriées et suffisantes, et non pas de toutes les capacités possibles. Ainsi, il reste à déterminer s'il sera doté d'un mécanisme complexe pour sa perception, où si un simple tampon d'entrée suffira. S'il est très probable qu'il interagira avec l'apprenant uniquement par le truchement du clavier, de la souris et de l'écran de l'ordinateur, le genre d'informations échangées et leur forme restent à déterminer (choix multiples, cases à cocher ou mots

simples). À un niveau rudimentaire (mais supérieur un seul tampon d'entrée), un tel réseau pourrait servir à identifier le type de communication ayant eu lieu (question de l'apprenant, commentaire, réponse, communication non reconnue, etc.). À un niveau plus évolué, le réseau perceptuel pourrait servir à comprendre la communication grâce à sa toile conceptuelle et aux activations qui se transmettent de noeud à noeud.

La perception du ACA, si elle requiert quelque niveau de "compréhension", sera calquée sur celle dont dispose IDA. Celle-ci (IDA est un prénom féminin!) contient déjà les capacités de repérer des informations textuelles (dans les courriels qu'elle reçoit ou dans les banques de données qu'elle interroge). Elle fait usage pour cela d'un tampon d'entrée, de microprocessus attentionnels, d'une mémoire de travail, d'un réseau perceptuel (le *slipnet*) et d'une banque de formulaires.

Le réseau perceptuel est un réseau sémantique où circule de l'énergie d'activation (selon le principe du *feed-forward*) et dont les noeuds représentent divers niveaux syntaxiques et sémantiques.

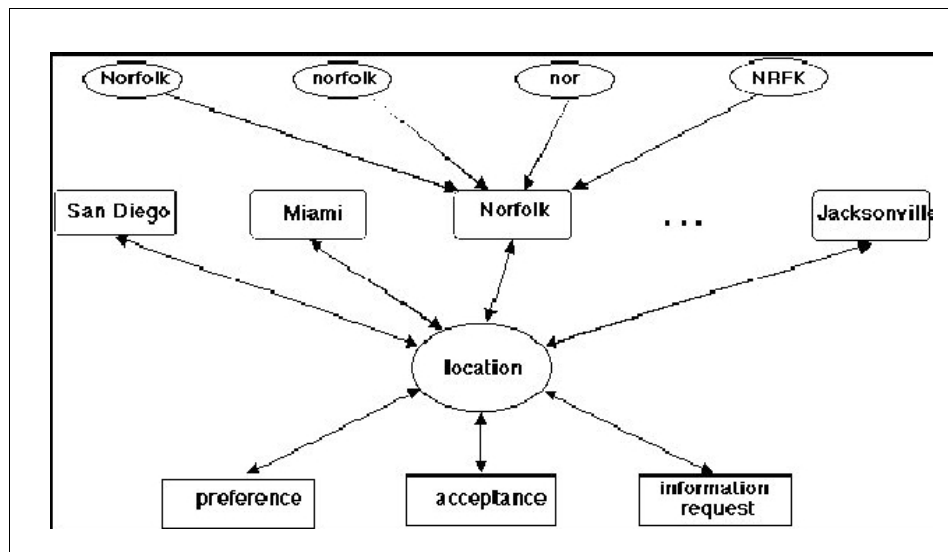


Illustration 2: Fragment du slipnet d'IDA. Les liens entre les noeuds peuvent représenter plusieurs types de relations issues des attributs de chaque noeud parent. Source: [Franklin 2005]

tiques. Les noeuds les plus à la périphérie de la structure contiennent les diverses formes d'un mot, ou les concepts les plus élémentaires. Par exemple, de la périphérie vers le centre, on trouvera : les diverses formes et abréviations d'un mot (ex.: "lun", "lun."), les mots auxquels elles se rattachent (ex.: "lundi"), les idées résumant certains mots (ex.: semaine), et finalement les concepts de plus haut niveau (ex.: préférence pour une journée), obtenus par la mise en rapport de plusieurs idées. Les liens peuvent représenter divers types de relations (exemple, abréviation, partie, instance, relation temporelle, etc.) et peuvent relier un noeud à plusieurs autres. Le

niveau "ponctuel" d'activation des noeuds ne s'estompe que progressivement, jusqu'à retourner à un niveau "fondamental" acquis avec le temps; cette rémanence à deux paliers crée un contexte d'interprétation pour les perceptions subséquentes.

Une multitude de microprocessus attentionnels (du type perceptuel) surveillent l'arrivée d'un courriel pour immédiatement s'affairer à y faire le repérage des mots (chacun sachant reconnaître un ou quelques mots et formes). Lors de la détection d'un tel mot clé, un microprocessus perceptuel active le noeud du réseau perceptuel qui lui correspond. Chacun d'entre eux fera de même, et divers noeuds du niveau inférieur (en périphérie) de ce réseau se trouveront ainsi activés. Cette activation sera immédiatement transmise à des noeuds plus génériques, cette généralisation se poursuivant jusqu'au niveau final (vers le centre du réseau) où existent des noeuds représentant les types reconnus de messages. Cela crée la reconnaissance d'une série de mots et concepts qui, lorsque placés dans un formulaire de validation, permet de confirmer un type de communication et le contenu pertinent qu'elle contient. Un premier niveau de perception aura ainsi été accompli, et sera complété par l'appel aux mémoires associatives en déposant les mots dans les registres perceptuels de l'atelier. Tout ce travail aura été effectué sans l'intervention du module de "conscience", celui-ci n'étant sollicité que par la suite, après la réponse des mémoires et la formation d'une coalition avec toute ces informations.

6.1.2. Le réseau des actes

C'est ici que se concentrent les capacités de réponse au monde extérieur. Le "réseau" des actes est en fait le collage de réseaux construisant les divers aspects des savoir-faire de l'agent. Par exemple, un premier réseau le rendra apte à effectuer la planification et la présentation de la matière; un deuxième, à diagnostiquer en profondeur les difficultés observées chez l'apprenant, et un troisième, à engager un processus de soutien affectif. Le réseau, donc (et je laisse dorénavant tomber les guillemets!), constitue en même temps le moyen de sélectionner les actes qui seront mis en oeuvre par l'agent.

i. Ses composantes et son fonctionnement

Maes a proposé en 1989 un mécanisme hybride (à la fois symbolique et connexionniste) devant permettre à un agent d'agir dans un environnement complexe et hautement imprévisible : le Mécanisme de Sélection des Actes [Maes 1989]. Il s'agit d'un réseau de noeuds qui représentent les divers actes formant le répertoire de l'agent. Chaque noeud spécifie non seulement son travail, mais aussi ses préconditions et les effets qu'il devrait avoir sur l'environnement. Ce qui rend le réseau plus intéressant encore, c'est l'énergie qui le parcourt et le rend dynamique. Les liens,

allant de post-conditions (effets) à pré-conditions (et réciproquement), indiquent les possibilités de transmission de l'énergie, et des liens d'inhibition existent aussi entre noeuds dont l'action viendrait faire perdre des états souhaités et acquis. Mentionnons que plusieurs "liens" peuvent exister entre deux noeuds, et un noeud peut aussi être lié à plusieurs autres noeuds, éventuellement parties d'autres séquences. Les noeuds sont de deux types principaux : noeuds d'actes et noeuds de buts. Dans la version modifiée par l'équipe de Franklin, les actes sont organisés en plusieurs chaînes, chacune soumise au but qu'ils permettent d'atteindre, transformant la structure en une hiérarchie de plans partiels (que j'appellerai dorénavant "séquences").

Permettant un traitement hautement parallèle, l'architecture perçoit simultanément les diverses sources d'influences, transmues en une "monnaie commune", l'énergie d'activation. D'abord, les "pulsions" de l'agent (ses buts fondamentaux) insufflent de l'énergie par la haut du réseau, vers les buts (et leur enchaînement) qui les satisfont; ils insufflent une énergie variant selon l'évolution des situations. Ensuite, les événements et états viennent faire sentir leur influence en injectant de l'énergie directement aux noeuds dont ils sont les préconditions. Toute cette énergie coule de noeud en noeud, d'abord "vers l'arrière", d'un but vers l'acte initial de l'enchaînement qu'il représente, puis vers l'avant à partir du moment où un acte devient "exécutable" (toutes ses préconditions sont satisfaites). En transmettant à rebours de l'énergie, un noeud cherche à favoriser les actes qui vont créer les conditions dont il a besoin pour s'exécuter. En transmettant son énergie vers l'avant, un noeud tend à favoriser la poursuite du plan en cours et l'atteinte du but actuellement poursuivi. La planification s'effectue donc à la fois pour satisfaire les buts fondamentaux de l'agent, mais en même temps en réaction aux états et événements observés, tant à l'intérieur qu'à l'extérieur de l'agent.

Le réseau constitue un mécanisme de planification et de *sélection* des actes. De cela découle, dans l'architecture modifiée par Franklin, que les noeuds d'actes *ne sont pas effectifs par eux-mêmes*, s'en remettant aux microprocessus qui leur sont rattachés à chacun des noeuds d'actes et effectuant le travail prévu par l'acte. Les noeuds d'actes jouent un rôle de *sémaphore* pour leurs microprocessus, auxquels ils transmettent à la fois le signal d'entrer en action et leur énergie d'activation (pour qu'ils puissent compétitionner avec les autres microprocessus déjà sur le terrain de jeu ou toujours actifs dans le système). Un noeud représente un but local que doivent atteindre leurs microprocessus.

ii. Application du réseau des actes à l'enseignement

Si nous acceptons de nous contenter d'un tuteur au comportement mécanique, préfixé, comme dans les systèmes d'enseignement assisté par ordinateur des années '80, il suffirait de relier les

plans partiels d'action entre eux et de suivre le franchissement des étapes; mais cela présenterait peu d'intérêt, et ne ferait que reproduire ces antiques systèmes de CBT et de CAI, avec toutes les limitations et défaillances qu'on leur reconnaît aujourd'hui (même s'ils pouvaient s'appuyer sur des réseaux. Nous souhaitons au contraire un mécanisme capable de sélectionner la prochaine suite d'actes en fonction des diverses sollicitations.

Pour représenter le comportement d'un enseignant à l'aide d'un réseau, on peut assigner chacun des les actes qu'il peut entreprendre à un noeud. Comme l'enseignant poursuit habituellement l'atteinte de buts plutôt que de s'intéresser à la réalisation d'actes détachés, on groupera similairement les actes en enchaînements, chacun soumis à l'atteinte d'un but.

Le contenu du réseau d'actes pédagogiques (RAP) proposé repose actuellement principalement sur la théorie pédagogique de Robert Gagné [Gagné 2005]. Celle-ci décrit bien les étapes à franchir pour accomplir un enseignement, ainsi que les conditions nécessaires pour les franchir. Précisons immédiatement que le réseau ne requiert évidemment aucunement l'affiliation à une théorie pédagogique pour se construire. De fait, je *souhaite* construire le réseau ultime de manière à ce qu'il soit relativement indépendant d'une théorie spécifique, capable d'en mettre plusieurs en oeuvre. Cela pourrait s'accomplir en élaborant une version "riche" du réseau incorporant les actes de plusieurs théories, inhibés ou activés selon la théorie sélectionnée, jusqu'à ce qu'une analyse métacognitive révèle qu'il soit préférable de passer à une autre théorie pour tenter d'améliorer les performances de l'apprenant. Plutôt que d'envisager une structure énorme, on pourrait aussi envisager un ensemble de réseaux propres à chaque théorie, qu'on charge lorsqu'approprié et amène à jour avec la situation actuelle.

iii. Une première version du réseau des actes

L'espace ne me permet pas d'afficher le réseau complet, même s'il s'agit d'une version simplifiée pour un prototype initial. Je me contenterai de montrer une portion laissant voir son architecture générale (Illustration 3).

Dans le présent cas, la théorie de Gagné a servi principalement à déterminer plusieurs buts locaux du réseau : "avoir obtenu l'attention de l'apprenant", "avoir informé l'apprenant des objectifs de la séance", "avoir stimulé le rappel des prérequis", "avoir présenté la matière prévue", "avoir fourni des indications et repères", "avoir fait manipuler les concepts", "avoir fourni une rétroaction", "avoir fait appliquer un transfert", et "avoir effectué un approfondissement". Ces appellations constituent des adaptations des étapes décrites par Gagné. D'autres buts doivent être ajoutés afin de franchir certaines autres étapes (il faut tout détailler à un agent artificiel!) :

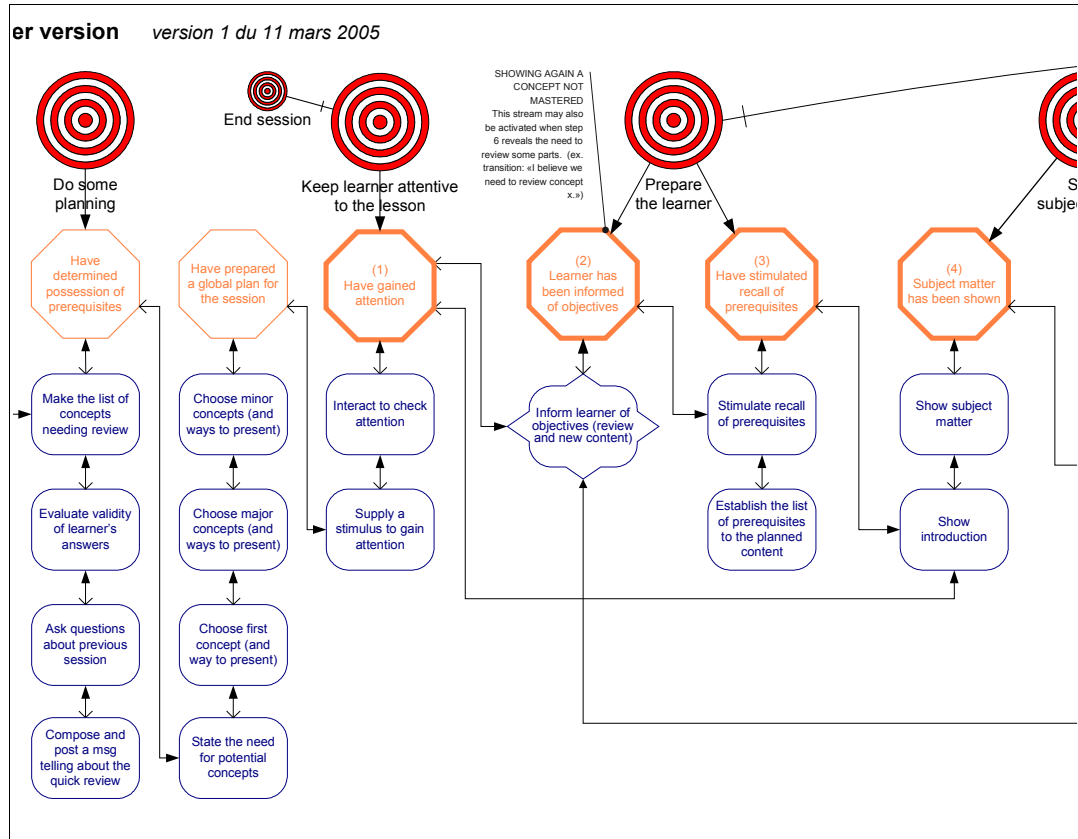


Illustration 3: Fragment du réseau des actes du prototype. On peut observer dans cette première version du réseau les éléments essentiels d'un réseau d'actes : noeuds de buts (octogones), noeuds d'actes (rectangles arrondis), pulsions (cibles) et liens de transfert d'énergie; deux liens réciproques y sont simplifiés par un lien bidirectionnel (flèche aux deux extrémités du segment). Ces liens entre les pulsions et les buts sont unidirectionnels puisqu'un but ne transfère pas d'énergie vers l'avant à une pulsion. Y apparaissent aussi partiellement des liens d'inhibition; ils proviennent d'une pulsion de niveau supérieur, chapeautant d'autres pulsions (on en aperçoit une réplique à la gauche de la pulsion «Keep learner attentive to the lesson»). Quoiqu'aucun lien d'inhibition n'apparaît entre actes ou buts, il se peut que l'étude en révèle le besoin. En ce qui concerne les pulsions, elles apparaissent ici toutes "pédagogiques", mais l'agent peut posséder d'autres pulsions liées à sa propre maintenance et à sa survie, ainsi que des pulsions "personnelles" advenant qu'on le dote d'aspects de personnalité supplémentaires à celles de l'enseignant.

Les étapes de Gagné sont repérables en tant que buts numérotés.

"avoir accueilli l'apprenant", "avoir validé la possession des prérequis", et "avoir préparé un plan de session". Les actes eux-mêmes seront inspirés autant de Gagné que des travaux d'autres chercheurs du monde des ACA ayant produit taxonomies et listes d'actes pédagogiques ([Murray 1996], par exemple).

Pour illustrer, supposons qu'à un moment de la séance, le ACA devrait "vouloir" déterminer si l'étudiant maîtrise les notions prérequis pour une recevoir une explication ou se faire présenter le sujet prévu, l'intensité relative de sa pulsion de planification sera alors plus élevée, transmettant à ses buts davantage d'énergie que les autres pulsions ne le font vers les leurs. Ceci devrait permettre de mettre en branle l'enchaînement des actes servant ce but (c'est-à-dire concourant à son atteinte).

Dans cette illustration, deux groupes d'éléments forts importants n'apparaissent pas : les pré- et post-conditions détaillées de chaque noeud (quoique les liens de transmission d'énergie en suggèrent l'existence), et les microprocessus dont dépendent les actes pour effectivement produire leurs effets. En effet, le réseau des actes sert de mécanismes de contrôle au déclenchement des microprocessus les accomplissant; lorsqu'un acte est sélectionné parmi tous les actes du réseau, alors seulement ses microprocessus sautent-ils sur le terrain de jeu, dotés du niveau d'énergie de leur noeud d'acte. Nous verrons au point suivant un exemple plus détaillé de segment (une chaîne d'actes), ainsi qu'une chaîne de microprocessus réalisant un acte.

Il m'apparaît important de souligner que, quoique cette version simple semble livrer un enseignement presque exclusivement linéaire, plusieurs mécanismes dans l'architecture rendent l'enseignement livré par le ACA non linéaire. Tout d'abord, le jeu de l'énergie d'activation peut amener un comportement de l'agent effectivement non linéaire, tenant compte d'effets prévus non atteints et exigeant la reprise d'étapes antérieures. Ensuite, les actes ici montrés demeurent de très haut niveau et font en réalité appel à des mécanismes eux-mêmes non linéaires. Premièrement, la sélection et le séquençage de la matière par l'entremise de "délibérations" "conscientes" rend aussi non linéaire l'enchaînement des sujets, à la différence d'un plan hiérarchique. Deuxièmement, la sélection des stratégies de présentation s'effectuant aussi au cours des délibérations et tenant compte de l'évolution de la situation, ne contraint pas à l'avance l'interaction. Au total, même cette chaîne apparemment quasi linéaire offre déjà une réactivité essentiellement non linéaire.

iv. Exemple plus détaillé d'une séquence d'actes

Cette séquence forme un plan partiel ayant pour but d'avoir présenté une notion à un apprenant. Le niveau de détail et la longueur de cette présentation dépendra de l'interaction mise en branle par le troisième acte («Construire le texte présentant une notion», en partant du coin inférieur gauche). Tel qu'actuellement prévu et décrit dans l'exemple de fonctionnement plus loin, ce processus implique la collaboration de plusieurs modules (modèle des connaissances de l'apprenant, modèle psychologique de l'apprenant, modèle du domaine, "conscience") afin de tenir compte de plusieurs aspects pertinents dans la planification d'une présentation (contenu et stratégie). Un noeud d'acte constitue une opération de haut niveau, un but local requérant la mise en oeuvre de plusieurs actions plus simples réalisées par les microprocessus issus autant de l'acte lui-même que d'autres fonctionnalités de l'agent. Ces actions doivent être détaillées acte par acte. Je montre un exemple d'enchaînement d'actions (microprocessus) au point suivant.

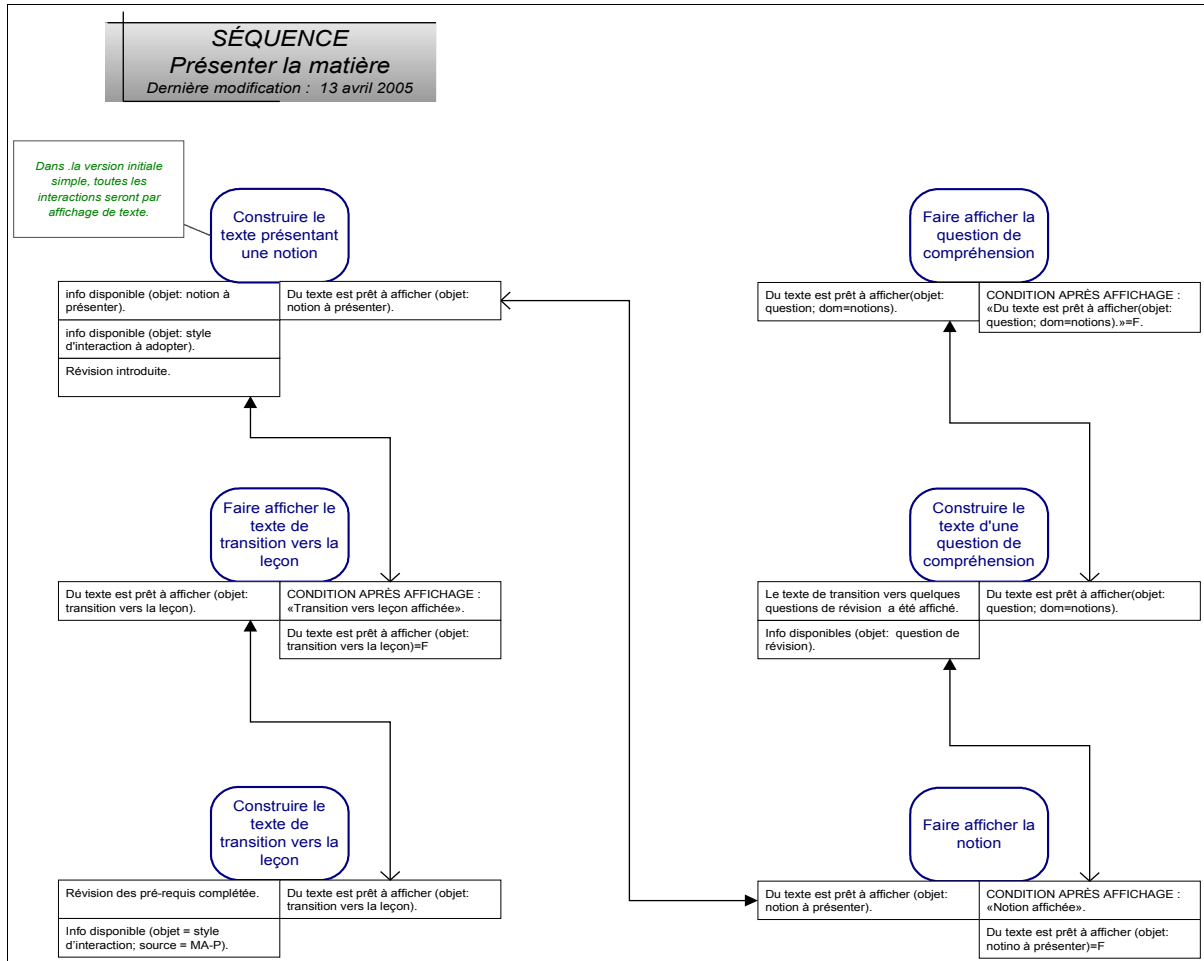


Illustration 4: Exemple d'une séquence d'actes : «Présenter la matière». Cette séquence pourra servir à répétition durant une même leçon pour présenter plusieurs concepts. Les actes, qui constituent en fait un "but local" pour les microprocessus qui le réalisent effectivement, apparaissent dans les rectangles arrondis; les préconditions (à gauche) et les post-conditions (à droite) apparaissent dans les rectangles sous les actes. Des liens de transmission d'énergie vont de post-conditions d'un acte à préconditions d'un autre, et réciproquement (d'où des flèches bi-directionnelles). Quoiqu'ici présenté comme partie intégrante de ce plan partiel, la présentation d'une question de validation sera probablement isolée dans une séquence séparée, améliorant ainsi la flexibilité de l'enseignement pour tenir compte des particularités de chaque apprenant.

Tel que mentionné précédemment, ces actes sont rarement détaillés ou mentionnés clairement dans quelque écrit que ce soit puisque les théories veulent donner des directions, suggérer des pistes à adapter à la situation, toujours plus complexe qu'il pourrait apparaître au néophyte. Néanmoins, ils s'appuient globalement sur les descriptifs fournis dans des livres tels [Reigeluth1999] et [Gagné et al. 2005], respectant donc d'assez près la théorie pédagogique annoncée. Cependant, il se peut que mon réseau doive intégrer plus d'une théorie d'une manière indifférenciée afin de combler certains vides théoriques. L'examen plus avancé du dernier volume de Gagné indiquera cette nécessité.

v. Exemple d'une chaîne de microprocessus réalisant un acte

L'illustration 5 montre les microprocessus rattachés à l'acte visant à obtenir le nom de l'apprenant, en début de séance. Cette information servira à récupérer les informations propres à l'apprenant⁵ (le modèle de l'apprenant qui lui est propre : état de ses connaissances, profil psychologique, préférences, etc.), ainsi, bien entendu, qu'à utiliser son nom dans les interactions. Cet enchaînement de microprocessus constitue actuellement plus une hypothèse qu'un engagement quant au fonctionnement effectif à ce niveau de détail.

vi. Fonctionnalités prévues dans le réseau des actes

Le réseau offre une structure modulaire facile à modifier et à enrichir. Il s'agit à la fois d'un "collage" de réseaux dont chacun détaille une habileté de l'agent, et d'une hiérarchie d'enchaînements à l'intérieur de chacun d'eux. Quoique principalement juxtaposés, des liens peuvent exister entre certains de leurs actes, mais il n'ont pas à être "planifiés" puisque les événements et états de l'environnement constituent des réalités intermédiaires entre chacun des actes. Les noeuds ne s'interrogent pas l'un l'autre, mais scrutent l'environnement ou écoutent les publications faites par la "conscience" pour y percevoir les états nécessaires à les rendre "exécutables", leurs préconditions. Donc, il n'y a qu'à spécifier les préconditions et effets, et l'architecture fonctionnelle de l'agent s'occupe du reste. Il s'avère donc facile d'insérer une nouvelle fonctionnalité dans l'agent. Un principe important d'ingénierie se trouve respecté : couplage faible, intégration forte!

La modification d'un réseau existant se révèle pareillement "simple", d'une part pour la raison tout juste présentée, mais aussi parce qu'une fonctionnalité est découpée en plans partiels qui montrent bien la structure logique suivie; de plus, étant chapeautés chacun par un noeud de but, chaque séquence constitue une "boîte noire" pour les séquences qui en dépendent : il leur suffit de savoir que le but précédent, précondition à leur premier acte, a été atteint.

5 La manière de gérer plusieurs apprenant en n'utilisant que les ressources prévues par l'architecture n'a pas encore été déterminée (comment s'effectuera la connexion à un modèle de l'apprenant spécifique ou sa récupération dans un modèle vierge).

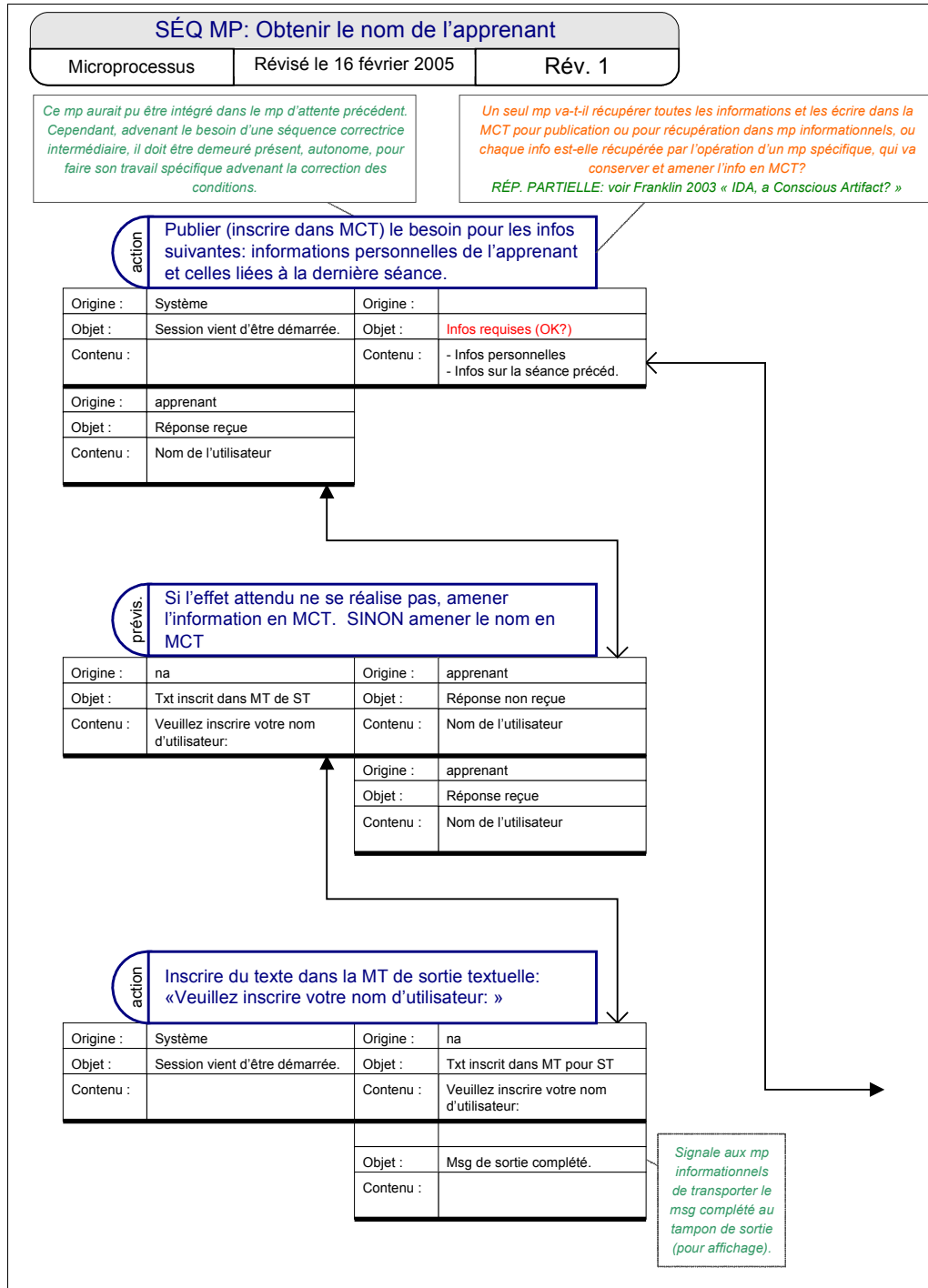


Illustration 5: Enchaînement de microprocessus réalisant l'acte «Obtenir le nom de l'apprenant». Cet acte est le tout premier noeud de la première séquence d'actes devant entrer en jeu dans une séance, celle servant à accueillir l'apprenant. La structure des enchaînements de microprocessus apparaît similaire à celle des séquences, quoique plus détaillée. En effet, si les actes s'intéressent, pour leur déclenchement, aux conditions de l'environnement (états et événements), les microprocessus ont des tâches et domaines plus stricts et directionnels; leurs préconditions s'intéressent non seulement à des événements, mais aussi à leur provenance, leur type, etc.

Ces raisons font baisser l'exigence de planification globale initiale visant à élaborer une architecture complète. On peut facilement enrichir le répertoire de l'agent par la suite. Ceci dit, la puissance de l'architecture, sa souplesse rendent possible de prévoir un grand nombre de fonctionnalités pour l'agent, et l'esprit s'enflamme rapidement! J'en propose ici une liste, qui correspond surtout à une énumération de souhaits puisque si leur insertion s'avère simple, leur élaboration demeure artisanale et limitée par les compétences de l'artisan... Disons que la voie sera ouverte pour des projets de maîtrise et de doctorat ultérieurs! Le premier prototype incorporera donc un nombre limité de fonctionnalités, elles-mêmes initialement rudimentaires puis progressivement enrichies.

Fonctionnalités prévues pour le prototype : Planification du contenu et de la présentation, Présentation, Exercices, Validation des compétences (diagnostic élémentaire), Remédiation élémentaire, Dialogues, Métacognition.

Autres fonctionnalités intéressantes, mais non retenues pour le prototype : diverses stratégies de présentation (socratique, en spirale, description d'exemples, exploration libre, etc.), Interactions en langue naturelle, Diagnostic avancé, Remédiation avancée, Compagnon d'apprentissage, Offre d'indices, support à la métacognition de l'apprenant, etc.

La métacognition de l'agent fait partie de fonctionnalités prévues, mais externes au réseau des actes.

6.1.3. Les modules spécifiques au ACA et à élaborer

i. Le modèle de l'apprenant (MA)

La modélisation de l'apprenant devrait comprendre deux aspects : son profil psychologique et préférences (MA-P), et son état cognitif (MA-C). Si cette hypothèse de travail tient, le MA-P indiquera son type d'apprenant (le type de présentation typiquement le plus efficace pour son profil d'apprenant), ses aptitudes naturelles et préférences (aptitudes pour les mathématiques, les langues, la géographie, etc.; ses préférences pour le type d'interactions, le type de matériel [textes, images, vidéos, audio], etc.), le niveau de diverses habiletés génériques acquises (vitesse de lecture, rapidité et fidélité de mémorisation et de rappel, persistance de la rétention, etc.), sa persévérance (et donc la durée typique à prévoir pour une séance, et ainsi de suite. Certaines informations ne seront pas utilisées dans les premières version du tuteur; d'autres proviendront de questions directes à l'apprenant (et peut-être aussi d'une petite section de l'interface utilisateur où il serait possible pour l'utilisateur d'inscrire son appréciation suite à diverses interactions), de tests initiaux, d'observations et d'inférences menant éventuellement à modifier

certaines valeurs. Le MA-C, pour sa part, pourrait être un réseau bayésien répliquant dans sa structure celle du modèle du domaine à partir d'un niveau sémantique significatif; il devra permettre de connaître le niveau d'expertise ou de maîtrise d'un concept ou habileté, et de là rendre possible l'extrapolation de sa compétence sur les concepts à voir. Nous allons aussi explorer des questions complémentaires et avenues parallèles : 1) quelle est la relation entre le modèle de l'apprenant et les informations enregistrées dans les mémoires parcellaires; 2) la possibilité et les avantages de modéliser les connaissances de l'apprenant avec une mémoire parcellaire (la *sparse distributed memory* décrite à la page 26);

Ces éléments entreront activement dans la planification du contenu et de la présentation. Ils permettront de répondre aux questions touchant au séquençement des concepts (combien de concepts par séance et lesquels pour cette séance-ci, combien de notions plus faciles au début, etc.) et à la présentation à effectuer (stratégie de présentation adopter, quel matériel privilégier, à quels points prévoir des questions ou des exercices, etc.). Ceci dit, j'ajoute que, bien que la modélisation de l'apprenant fonde la capacité de l'agent à adapter son enseignement, le but du projet n'est pas d'atteindre le meilleur résultat possible pour chacune des composantes du ACA (si même une telle idée était possible!), mais avant tout de démontrer la capacité de l'architecture à intégrer un nombre important de facteurs dans son adaptativité ainsi que de créer des plans tout aussi sensibles à de multiples paramètres. L'enrichissement des modèles à la source de ces paramètres pourra suivre, une fois démontrée la puissance adaptative de l'architecture.

De nombreux travaux existent dans le domaine des modèles de l'apprenant. Plusieurs travaux ont aussi déjà été effectués sur ce sujet par des membres du laboratoire GDAC [Nkambou et Gauthier 1996], [Tchétagani, Joséphine M.P. et Roger Nkambou 2002]. Nous allons tenter de tirer profit de ces travaux comme bases d'une instantiation.

ii. Le modèle du domaine (MD)

Un petit modèle d'un domaine sera initialement construit afin de permettre le déroulement d'au moins une leçon. Le domaine retenu court de fortes chances de porter sur les lentilles en Optique. Dans le livre de [Reigeluth 1983], plusieurs auteurs de théories pédagogiques ont en effet proposé leur version d'une leçon portant sur quelques concepts spécifiques de ce domaine, fournissant une utile référence sur l'application des diverses théories, ainsi qu'un étalon pour éventuellement évaluer la validité des comportements de mon ACA. Des recherches m'ont de plus permis de repérer une bonne quantité de matériel, textuel, graphique et animé, récupérable pour mon projet sur l'Internet ainsi que dans un ensemble de manuels sur la Physique.

Je me propose présentement de construire une ontologie partielle du domaine, que j'implanterai dans un *slipnet* (la technologie utilisée pour construire le réseau perceptuel d'IDA). Mais je vais tout d'abord examiner les nombreux travaux existants dans ce domaine, et en particulier ceux ayant été produits par les membres du GDAC [Nkambou, Frasson et Gauthier 2003].

iii. La métacognition

Développer une capacité métacognitive chez l'agent constituerait une extension significative de l'architecture puisque, si elle fut développée et implantée dans Conscious Mattie, Franklin la déclare insatisfaisante et nous a suggéré d'en rebâtir une nouvelle. Ses objectifs, si une telle entreprise trouve le temps d'être intégrée au projet, viseraient à améliorer les performances de l'agent, parfois d'éviter la répétition d'actes infructueux.

iv. La gestion des émotions de l'apprenant

En autant que les informations sur les états affectifs de l'apprenant soient générées par un agent externe, en tenir compte ne constitue pour l'ACA que l'ajout d'une source d'influence supplémentaire pour le réseau des actes. En effet, pour l'ACA, ces constats font tout simplement partie des états de l'environnement, et l'architecture les traite tout naturellement. Attendues par un microprocessus prévisionnel, les réactions affectives viendront dans l'atelier (la "mémoire à court-terme" pour se faire associer à l'acte récent qui les a provoquées – cependant, établir ce lien causal risque d'être épineux! Une fois associées aux concepts ou événements en mémoires à long-terme, elles deviennent des informations ramenées par les mémoires en même temps que le reste des associations.

Ici encore, les membres du GDAC ont produit des travaux. Un agent, Émilie-2 a été décrit et un prototype fut réalisé, et pourrait éventuellement agir comme agent externe et source d'interprétation pour l'ACA [Faivre, Frasson et Nkambou 2003]

v. La gestion des émotions de l'agent

L'architecture d'IDA incorpore une gestion émotionnelle pour l'agent; IDA "éprouve" des émotions et agit en conséquence. Ses émotions s'intègrent aux informations publiées par la conscience et enregistrées dans les mémoires à long-terme. Elles produisent aussi un effet sur la force de l'inscription des informations en mémoire, tout comme cela se passe chez les humains (et probablement chez d'autres animaux). Cet aspect affectif de l'agent le rend plus proche encore des humains et peut permettre de rendre ses comportements encore plus convaincants, réduisant d'autant les barrières psychologiques de l'utilisateur envers la "machine".

Cependant, les mécanismes impliqués sont à l'étude et sont l'objet d'une reconception. Il me faudra me pencher sur l'aspect émotionnel un peu plus tard.

7. Exemple de fonctionnement

Avant d'entreprendre une session d'enseignement, l'agent doit en déterminer le contenu et les activités appropriés, c'est-à-dire tenant compte de tous les paramètres pertinents : l'étudiant avec ses multiples caractéristiques, et les objectifs poursuivis (ceux du cours ainsi que ceux de l'étudiant). Pour illustrer com-

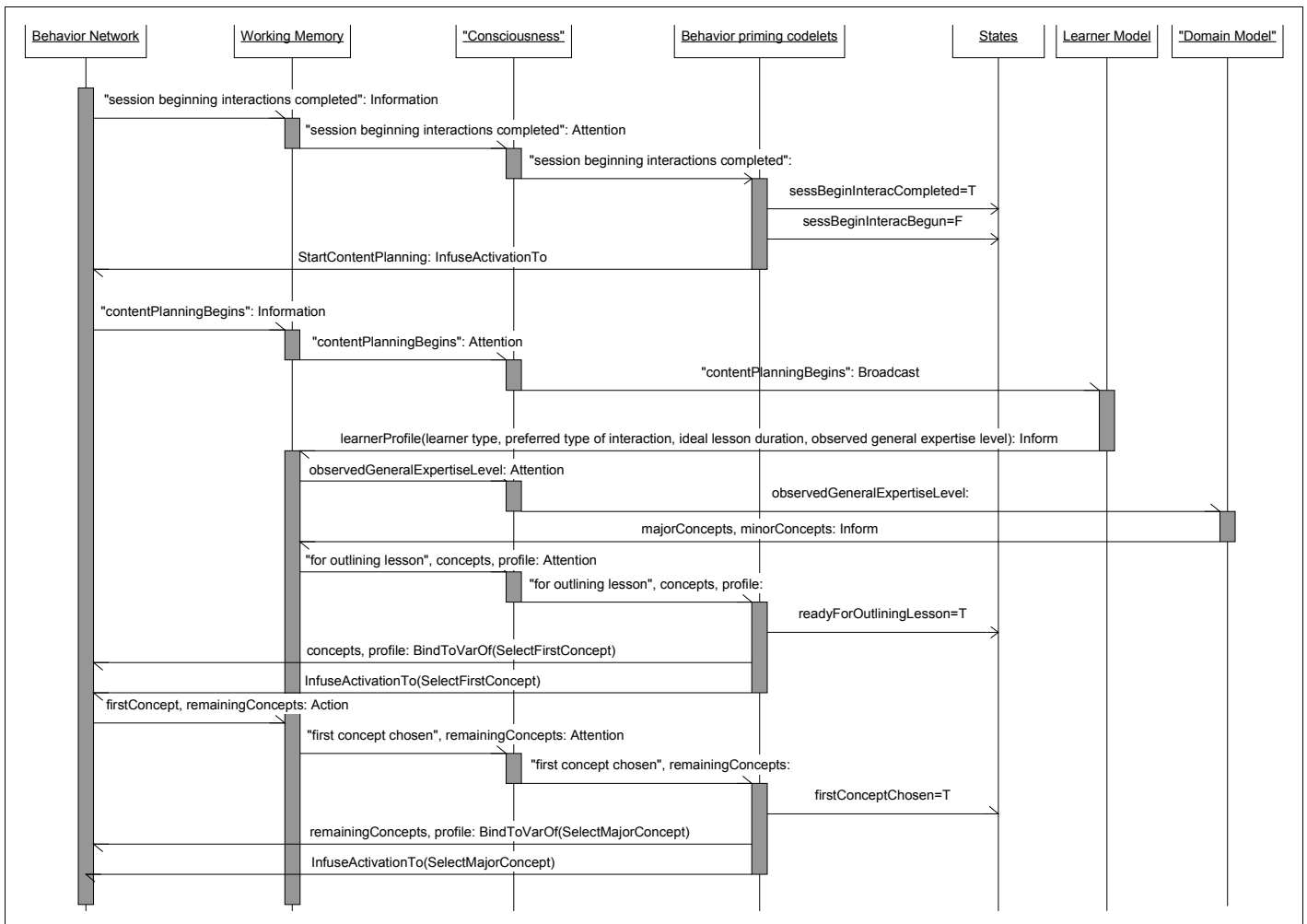


Illustration 6: Diagramme de séquence partiel pour un exemple de planification délibérative. Ce diagramme montre les interactions entre certaines composantes majeures de l'agent visant à élaborer la planification d'un plan de contenu et de présentation. Il importe de considérer les deux aspects simultanément parce que la stratégie de présentation retenue pour un élément a un impact sur le temps à prévoir pour l'interaction, et donc sur le nombre de concept qu'on pourra insérer à l'intérieur de la séance. Le module de "conscience" sert de place publique permettant à chaque composante concernée d'être avisée de l'état d'avancement du plan en élaboration, et ainsi d'intervenir dans le processus.

ment l'agent devrait être en mesure d'atteindre ces objectifs, nous allons ici examiner une portion du processus menant au plan d'enseignement pour une session. L'illustration 6 décrit graphiquement une telle portion. Nous allons supposer que la maîtrise des concepts prérequis a été confirmée à la fin de la ses-

sion précédente. Alternativement, dans le cas où ce savoir fut reconnu incertain et enregistré comme tel dans le modèle MA-C de l'apprenant (parce que non corrigé dans cette précédente session), des évaluations préliminaires pourraient être planifiées et enclenchées à ce moment, menant éventuellement à une révision adaptée. Afin de limiter la taille de cet exemple et d'en assurer la clarté, nous allons examiner la planification pour une séance "idéale". Dans un tel cas, notre petit scénario aura été précédé par les interactions d'accueil entre le système et l'apprenant, après que ce dernier ait sollicité le déclenchement d'une session.

À la toute fin de la séquence d'actes traitant cet événement (le démarrage d'une séance), un microprocessus d'action aura amené dans l'atelier (le *workspace*, à peu près équivalent à la mémoire à court-terme chez l'humain) l'information avertissant que cette étape vient de se compléter. Cette information est repérée par un microprocessus attentionnel dont la tâche est d'observer l'atelier pour y repérer l'apparition d'une telle information révélant la complétion d'une séquence (l'atteinte d'un but), pour l'amener dans la zone de "conscience" (équivalent au rôle de la conscience d'accès). Il accomplit ainsi une tâche rattachée à la "conscience de supervision". Le microprocessus attentionnel se met donc en frais de rassembler toute l'information pertinente qu'il peut repérer dans l'atelier selon son centre d'intérêt spécifique; dans ce cas, l'information est monolithique et entièrement présente dans un seul microprocessus attentionnel avec qui il forme une coalition tout en ayant élevé son niveau d'activation afin de faire entrer sa coalition dans la compétition pour l'attention de la "conscience". Si elle contient la plus importante ou la plus urgente information parmi toutes les coalitions présentes dans l'atelier à ce moment, elle se verra élue et pourra entrer dans la zone de "conscience" d'accès afin que son contenu soit publié (*broadcasted*). Cette diffusion d'information constitue la tâche de la conscience d'accès, qui permet aux ressources appropriées de réagir à la situation. Le contenu de la publication est remarquée par un ou plusieurs microprocessus de marquage (ou *d'amorçage* : *behavior-priming* dans l'appellation originale quelque peu défectueuse) bien spécifiques (tous ceux de leur genre demeurent toujours à l'écoute des publications, à la recherche d'informations qui les concernent). Chaque mp de marquage stimulé va créer les états d'environnement appropriés, formant ainsi les préconditions nécessaires aux actes du RA qu'ils ont marqués; dans le cas présent, il s'agit du premier acte de la séquence de planification. En plus de marquer les actes, les mp de marquage infusent aux noeuds qui leur correspondent une énergie d'activation proportionnelle à la pertinence des informations qu'ils ont repérées. Le premier acte de la séquence, en supposant que toutes ses préconditions existent, qu'il a atteint ou dépassé le niveau d'énergie requis pour son activation, et que l'acte soit le plus excité dans le réseau parmi tous les actes exécutables, s'instancie vers le Sommet (le *Skybox*) (en même temps qu'il instancie ses mp vers la "salle d'attente" (le *sideline*)) où il peut se mettre à l'oeuvre. Sa tâche est d'annoncer que le processus de planification débute, et il dépose dans l'atelier l'information indiquant

cela. Ces premières interactions montrent déjà bien comment le mécanisme de la "conscience" d'accès interagit avec le RA, et comment une séquence peut être mise en action.

L'éventuelle publication de la nouvelle information (si cette information émerge de la compétition pour l'attention) sera "entendue" (ou "vue") par tous les sous-systèmes de l'agent, dont les mémoires à long-terme, les mp de marquage, et divers mp autonomes (mp attentionnels permanents, mp de métacognition, etc.). Comme toujours, seuls certains de ces sous-systèmes vont réagir à l'information publiée. Par exemple, la MA-P (Modèle de l'Apprenant, volet Profil psychologique) va envoyer dans la mémoire de travail (MT) pour la planification, des informations pertinentes au sujet de l'apprenant (par exemple, ses préférences pour les types d'interactions, la durée idéale d'une leçon, et son niveau général d'expertise). À moins que d'autres informations plus importantes aient été présentées dans l'atelier par d'autres sources, les informations provenant du MA-P vont être publiées par la "conscience" d'accès (selon le processus déjà décrit). Réagissant à cette nouvelle publication, le MA-C (le Modèle de l'Apprenant, volet Cognitif), va envoyer à la MT les concepts qu'il juge comme pouvant être considérés dans la planification de la séance (les concepts d'un niveau de difficulté approprié, aucune ne se situant au-dessus des capacités de l'apprenant, à moins qu'une indication à ce sujet ait été publiée, bien sûr); il va aussi envoyer les données indiquant leur durée de présentation prévue (ceci ne donnant qu'une indication générique puisque le type de présentation viendra par la suite indiquer une valeur plus probable), ainsi que leur niveau de priorité (concepts majeurs, concepts secondaires). Jusqu'à ce que nous soyons rendus à assembler du texte à afficher à l'écran, "envoyer les concepts" veut en faire dire "envoyer les liens au matériel correspondant".

Vous pouvez commencer à vous demander où se cache le modèle pour une séance, celui qui permettra d'assembler les concepts à présenter. Il n'y en a pas, du moins aucun explicite. Nous nous trouvons à l'intérieur d'un mode de planification émergente qui s'adapte totalement à l'apprenant. Il n'y aucun plan de cours prédéterminé, si ce n'est celui de très haut niveau, dissimulé dans le RA qui indique les étapes à suivre pour effectuer la planification. L'élaboration suit un ordonnancement "naturel" issu de la structure du domaine et de la considération de divers aspects provenant du Modèle de l'Apprenant; l'auteur du domaine pourra bien entendu avec établi des contraintes supplémentaires pour le séquençement des concepts du domaine. Dans l'état actuel de la situation, toute l'information nécessaire a été accumulée dans la MT, permettant de construire une première version d'un plan pour la séance.

La dernière publication a avisé l'acte suivant que l'acte en cours a terminé sa tâche. Puisque le successeur voit toutes ses préconditions satisfaites, il signale à ses mp de se mettre en oeuvre (ils attendaient dans la salle d'attente) et leur transmet son niveau d'énergie ainsi que les informations dont il dispose. Le travail du premier de ses mp est de trouver le concept le plus approprié pour démarrer la séance. Les règles que le mp possède vont guider son choix. Par exemple, il va considérer que la présentation devrait d'abord intégrer les concepts majeurs (tel que décidé par le concepteur du domaine), mais d'autre part,

considérant les informations provenant du MA-P, il devrait en sélectionner un dont le niveau de difficulté est un échelon en-dessous du niveau général d'expertise de l'apprenant : ce dernier a d'abord besoin de construire une confiance en ses moyens. S'ils sont tous du même niveau, il en choisira un au hasard. Éventuellement, le mp va en sélectionner un et le déposer dans la MT comme premier concept à présenter dans la séance, ce qui, pour le moment, n'engage aucunement le plan final. Ce dépôt sera noté par un mp attentionnel qui fera publier le fait que la tête de la présentation est maintenant en place. Cela amènera un état à être déclaré à cet effet, fournissant la précondition requise pour le prochain mp d'action, qui attend pour entreprendre sa tâche d'ajouter un concept majeur au plan. Tant qu'il se trouve des mp informationnels non assignés dans la mémoire de travail, un mp attentionnel va les assembler en coalition avec lui et tenter de faire publier cette information. Ceci va redonner la précondition attendue par notre ami bâtisseur de plans, qui se verra ainsi à nouveau sollicité pour accomplir sa tâche de sélectionner et ajouter un concept majeur au plan. Lorsqu'il n'y a plus de concepts majeurs, où lorsqu'un mp attentionnel indique que la durée prévisionnelle accumulée excède déjà la durée idéale pour une séance, ce sera le tour de l'acte suivant de s'enclencher (ceci étant toujours soumis aux règles régissant le RA et aux aléas des événements).

Ce petit scénario, quoiqu'incomplet, montre essentiellement comment le RA et diverses composantes de la "conscience" interagissent. Cependant, la planification réelle pourra se révéler bien plus complexe. Par exemple, elle devrait intégrer la sélection immédiate du type de matériel (texte, animation, voix, séquence vidéo) et de la stratégie de présentation puisque ces deux éléments influent sur la durée prévisible de chaque élément du plan en construction.

8. Intérêt de la solution

8.1. Adaptativité élevée

8.1.1. Sensibilité omni-directionnelle du réseau des actes (parallélisme, variété des sources)

8.1.2. Priorisation effectuée par la conscience

8.1.3. Délibération pour le développement et l'ajustement progressif des plans

8.2. Réactivité élevée

8.2.1. Rapidité de réaction grâce à la planification et la mise à jour continues

8.2.2. L'organisation des actes pédagogique en plan partiels favorise deux attributs souhaitables : l'engagement minimal, et la poursuite d'un plan en cours à moins d'événements de grande importance.

8.3. Apprentissage

L'architecture permet l'auto-apprentissage de plusieurs manières [Franklin 2005] :

8.3.1. modification de la force des liens entre nœuds (réseau perceptuel, réseau des actes)

8.3.2. apprentissage de nouveaux nœuds perceptuels et de nouveaux liens

- i. mémoire de cas;
- ii. fusion de concepts proches;
- iii. création de concepts par fusion de coalitions fréquentes.

8.3.3. mémorisation des événements et états (Mémoire parcellaire) et création de nouveaux liens

8.4. Proximité cognitive de l'architecture

Puisqu'elle cherche à calquer l'esprit humain, l'architecture présente une similarité intuitive avec la structure mentale des concepteurs d'applications pédagogiques. Cela ne peut qu'aider à sa manipulation dans la conception d'applications pédagogiques.

8.5. Clarté et explicitation des structures

La formalisation des règles de production en un réseau d'actes, de buts et de pulsions rend claire la logique des facultés de l'agent.

8.6. Facilité pour l'enrichissement et la maintenance de l'agent

L'ajout de nouvelles structures et la modification des existantes peut s'effectuer en tout temps et sans lourdes conséquences sur la structure – aucun besoin de reconcevoir l'architecture ni de reprogrammer des modules autres que celui en modification.

9. Respect des paramètres du Doctorat

Je pense qu'il ne faudra pas une longue discussion pour positionner mon projet comme intégrant parfaitement les deux aspects du DIC, sciences cognitives et informatique!

9.1. Aspects touchant les sciences cognitives : Psychologie, Neurosciences et Pédagogie

D'une part, l'architecture fondamentale de l'agent s'enracine dans la théorie psychologique de Baars décrivant le fonctionnement de l'esprit humain, théorie qui elle-même intègre de nombreuses références aux recherches neuroscientifiques. Dans mon projet, je tente de respecter de la même manière que Franklin la fidélité aux mécanismes concernés. Il serait intéressant d'aboutir à un agent pouvant fournir un outil d'étude et d'analyse des mécanismes de l'apprentissage.

D'autre part, la substance de mon projet concerne l'enseignement et l'apprentissage, leurs mécanismes, leurs étapes, leurs conditions favorables. Il me faut étudier plusieurs théories pédagogiques pour les reproduire dans le réseau des actes ainsi que dans les interactions impliquant la conscience : adaptation et réactivité, planification. Ce travail demande de décortiquer les deux processus (enseignement et apprentissage) pour en isoler les diverses conditions favorables, les préconditions à chaque acte, la manière dont chaque acte doit être mis en oeuvre par une suite d'actions simples correspondant soit à des interactions avec l'apprenant, soit aux microprocessus tenus dans l'esprit d'un enseignant. Il faut établir comment on peut accomplir la réactivité de l'agent à l'aide de notions telles que les pulsions/motivations, lesquelles s'appliquent, comment et où, comment un professeur planifie son enseignement et son adaptation aux étudiants. Une recherche intense, autant dans les livres que par introspection s'avère nécessaire.

Finalement, en autant que le projet livre un produit suffisant dans les délais impartis, il sera intéressant (et nécessaire) d'en évaluer les performances avec de véritables apprenants. La collaboration avec des professionnels en pédagogie ainsi qu'en psychologie pourrait alors s'avérer requise afin de spécifier les essais pertinents et les protocoles appropriés.

9.2. Aspects touchant l'informatique

Traduire les théories pédagogiques et les stratégies d'enseignement en mécanismes informatiques a toujours constitué un défi et des travaux ont tenté d'amoindrir l'ampleur de la tâche. Entre autres, les outils auteurs veulent aider les auteurs en soutenant l'articulation des théories, des stratégies d'enseignement, etc. [Murray 2003]. L'ingénierie ontologique tente aussi d'apporter une partie de la solution en proposant les bases pour une description uniforme, concensuelle et explicite des théories, stratégies et autres concepts s'y rattachant [Bourdeau *et al.* 2004], [Psyché *et al.* 2005]. Nous allons examiner la question afin d'établir les meilleurs moyens pour atteindre le plein potentiel de flexibilité et d'adaptativité de l'architecture. IDA nous fournit une coquille pour soutenir l'élaboration de notre réseau d'acte mais n'aide aucunement à déterminer quoique ce soit au sujet de l'application de cette technologie à un agent d'enseignement. Il faut mettre de l'avant un effort d'ingénierie pour séparer les niveaux, comprendre et déterminer quelle doit être la teneur de chacun, pour transférer en modèles l'expertise pédagogique, et voir comment cet ensemble hautement complexe peut parvenir à un fonctionnement harmonieux et efficace. Par exemple, la paramétrisation du réseau des actes risque de requérir un important travail d'analyse et d'élaboration théorique pour comprendre et décrire l'impact de chaque facteur, et obtenir les résultats escomptés. Un autre travail d'ingénierie va concerner la modélisation des interactions entre les multiples facteurs d'influence entrant dans l'adaptation et la planification (états et

événements externes et internes, pulsions, processus métacognitifs. La programmation à elle seule constitue un travail redoutable.

10. Contributions de ma recherche

À ma connaissance, l'application à un système d'enseignement d'une architecture fortement et explicitement inspirée de l'esprit humain est une première. Impliquer la conscience dans le fonctionnement d'un agent en interaction avec un environnement aussi complexe que l'enseignement en est une autre; à tout le moins, ce projet peut fournir la "preuve de concept" souhaitée par le professeur Franklin pour son architecture – son application à un agent d'assignation s'est révélée trop peu complexe pour vraiment mettre en valeur et démontrer la supériorité de l'architecture.

La mise en service du prototype devrait permettre d'évaluer l'utilité et l'importance de savoir tenir compte d'un grand nombre de facteurs dans l'adaptativité et la planification d'un système d'enseignement – ce nombre de facteur pourra ultérieurement être poussé encore plus loin! Nous croyons que l'agent saura démontrer une plus grande capacité d'adaptation, une plus grande précision dans son choix d'actes (grâce à l'accès à davantage de critères et à une vue plus complète sur la situation.

Étudier et décrire le rôle de chaque sous-processus et de certaines composantes dans le processus d'enseignement pourrait apporter des éléments nouveaux à la réflexion en pédagogie, ou une perspective nouvelle. Peut-être même l'agent attirera-t-il l'attention de philosophes intéressés à la conscience. L'ACA pourrait constituer pour eux un mini-laboratoire d'étude de la conscience.

11. Échéancier prévisionnel

Pour la réalisation concrète du ACA, je ferai équipe avec un étudiant à la maîtrise en Informatique, Patrick Hohmeyer. Nous travaillerons conjointement dans la recherche et l'analyse de solutions sur des aspects spécifiques. Son rôle spécifique se situera bien évidemment dans la mise en programmes des modèles conceptuels et structures logiques, mais aussi en l'élaboration de standards, de la sélection de structures de données appropriées, pour ne nommer que celles-ci. La première étape de son travail s'occupera d'étudier les programmes reçus de l'université de Memphis, de les documenter et de les préparer à recevoir les nouveaux modules propres au ACA. Pendant ce temps, tel qu'il sera souvent le cas au cours du projet, je travaillerai en parallèle à élaborer les aspects théoriques de l'architecture. Étant donné que je prévois un travail de recherche théorique au moins aussi laborieux que les travaux purement informati-

ques, le plan de travail ci-dessous tiendra essentiellement compte de mes tâches propres pour estimer la durée des étapes.

Nous prévoyons quatre grandes étapes :

11.1. Analyse et réorganisation des programmes d'IDA (Patrick Hohmeyer) : 2 mois

Ce qui a été reçu de l'université de Memphis n'est qu'une partie décousue d'IDA, certains de ses modules ne s'avérant pas pertinents pour notre projet. Il faut examiner le code, comprendre ses structures, reficeler le tout (réorganiser les paquets en un tout cohérent). Ces travaux relèveront principalement de Patrick. Au cours de cette période, j'entreprendrai en parallèle des travaux théoriques devant livrer les chaînes de microprocessus liés au réseau des actes.

11.2. Un système d'essai : 2 mois

Mes travaux théoriques devraient avoir permis d'avoir détaillé un nombre suffisant d'actes et de chaînes de microprocessus pour pouvoir tenter une mise en marche du code original complété des ajouts essentiels. Il ne s'agira alors de tenter de mettre en route un ensemble minimaliste.

11.3. Un premier prototype (simple, mais "complet") : 11 mois

Nous allons ici compléter l'agent pour le doter d'un ensemble minimal de fonctionnalités, elles-mêmes dans une version simpliste – il s'agit de mettre en place les fonctionnalités. Ces capacités toucheront la génération de textes, la génération du plan d'une séance d'enseignement (contenu et stratégies), sélection et présentation d'exercices, évaluation des réponses de l'apprenant, remédiation élémentaire, un modèle plus élaboré du domaine, et peut-être les premiers pas d'une métacognition.

11.4. Essais avec des apprenants : 3 mois

À la fin du projet, notre agent d'apprentissage ne sera vraisemblablement encore qu'un prototype doté des fonctionnalités essentielles développées à un niveau insuffisant pour pouvoir se comparer avec des outils commerciaux ou à ceux d'autres chercheurs ayant oeuvré depuis des années sur leur recherche. De plus, le modèle du domaine n'étant ni complet, ni équivalent avec celui d'un autre EIAH, il s'avérera tout à fait impossible d'établir des tests comparatifs. Il faudra plutôt penser à d'autres types d'évaluation, telles : 1) comparaison des décisions du système avec celles qu'aurait pris un pédagogue : furent-elles bonne? optimales? 2) recueillir les impressions générales des utilisateurs (se sentaient-ils mieux supportés qu'avec un autre système qu'ils ont déjà utilisé? Se sentaient-ils plus à l'aise avec ses réactions?) 3) après une brève formation, demander à des concepteurs d'EIAH ce qu'ils pensent de la structure de description des processus (le RA); leur demander d'écrire une séquence d'ac-

tes. Nous espérons en trois mois avoir pu compléter une série d'évaluations, entrecoupées de séances d'optimisation (pour trouver les paramètres optimaux du réseau).

11.5. Le plan de travail en plus de détail

Vous pourrez consulter en annexes l'échéancier prévisionnel dans son état actuel.

Bibliographie

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. 2004. "An integrated theory of the mind". In *Psychological Review*, vol. 111, no 4, pp. 1036-1060.
- Baars, Bernard J.. 1988. *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
- Baars, Bernard J.. 1997. "In the theatre of consciousness: Global Workspace Theory, A Rigorous Scientific Theory of Consciousness." In *Journal of Consciousness Studies*, vol. 4, no. 4, pp. 292-309.
- Baars, Bernard J.. 2002. "The conscious access hypothesis: origins and recent evidence". In *TRENDS in Cognitive Sciences*, vol. 6, pp. 47-52.
- Baars, Bernard J. et Stanley Franklin. 2003. "How conscious experience and working memory interact". In *TRENDS in Cognitive Sciences*, no 7, pp. 166-172.
- Baffes, P. et R. Mooney. 1996. "A novel application of theory refinement to student modeling." In *American Association for Artificial Intelligence*, pp. 403–408.
- Barsalou, L.W.. 2003. "Abstraction in perceptual symbol systems". In *Philosophical Transactions of the Royal Society of London: Biological Sciences*, no 358, 1177-1187.
- Barsalou, L.W. et D.R. Sewell. (1984). "Constructing representations of categories from different points of view", In *Emory Cognition Project Technical Report #2*, Emory University.
- Bechtel, W. (1995). "Consciousness: Perspectives from symbolic and connectionist AI", In *Neuropsychologia*. no 33, pp. 1075-1086.
- Block, Ned. 1995. "On a Confusion about a Function of Consciousness". In *The Behavioral and Brain Sciences*, no 18.
- Bourdeau, Jacqueline, Riichiro Mizoguchi, Valéry Psyché et Roger Nkambou. 2004. "Selecting Theories in an Ontology-Based ITS Authoring Environment (2004)". In *ITS 2004*. J.C. Lester et al. (Eds.), LNCS 3220, pp. 150-161.
- Cazenave, Tristan (1998). "Machine Self-Consciousness More Efficient Than Human Self-Consciousness?", In *Proceedings of EMCSR 98*. <http://www.ai.univ-paris8.fr/~cazenave/emcsr98.pdf>
- Chalmers, David J.. 1995. «Facing Up to the Problem of Consciousness». In *Journal of Consciousness Studies*, vol. 2, no 3, pp. 200-219.
- Churchland, P.M.. 1989. *A neurocomputational Perspective*. Cambridge (MA) : MIT Press.
- Crick, Francis and Christof Koch. 1995. "Why neuroscience may be able to explain consciousness". In *Scientific American*. vol. ?, pp. 84-85.
- Dehaene, S., M. Kerszberg et Jean-Pierre Changeux. 1998. "A neuronal model of a global workspace in effortful cognitive tasks". In *Proceedings of the National Academy of Sciences USA*, 1995, pp.14529-14534.
- Dennett, Daniel. 1991. *Consciousness Explained*. Boston : Little, Brown.
- Edelman Gerald. 1989. *The remembered present: a biological theory of consciousness*, Paris: Basic Books.
- Faivre, Jessica, Claude Frasson et Roger Nkambou. 2003. "Toward Empathetic Agents in Tutoring Systems". In *Proceedings of the Florida Artificial Intelligence Research Society (FLAIRS'2003)*, AAAI Press.
- Franklin, Stanley (2005). "A "Consciousness" Based Architecture for a Functioning Mind". In *Visions of mind*. Darryl N. Davis (éd.). PA: IDEA Group, Inc.

- Frasson, Claude. 1998. "Using cognitive Agents for Building Pedagogical Strategies in a Multistrategic Intelligent Tutoring System", In *Deuxième journée Acteurs, Agents et Apprentissage*, Bayonne, September 1998.
- Gagné, Robert M., Walter. W. Wager, et al.. 2005. *Principles of instructional design (fifth edition)*. Belmont (CA): Wadsworth/Thomson Learning, 387p.
- Graesser, Arthur C., Kristen Moreno, Johanna Marineau et al.. (2003). "AutoTutor Improves Deep Learning of Computer Literacy: Is it the Dialog or the Talking Head?", In *Proceeding of AIED '03*, pp.47-54.
- Grossberg, Stephen. 1999. "The Link between Brain Learning, Attention, and Consciousness", In *Consciousness and Cognition*, vol. 8, pp. 1-44.
- Harnad, Stephen. 2003. "Can a Machine Be Conscious? How?", In *Journal of Consciousness Studies*, vol. 10, nos 4-5, pp. 69-75.
- Hexmoor, Henry H., Johan M. Lammens Stuart C. Shapiro. 1993. "An autonomous agent architecture for integrating «unconscious» and «conscious», reasoned behaviors", In *Computer Architectures for Machine Perception*, IEEE Computer Society Press, pp. 328-336.
- Jacobs, Mindy. 1999. «Situated cognition». Site du cours *IT7000 Foundations of Instructional Technology*, Georgia State University. [en ligne]. <http://www2.gsu.edu/~mstsw/courses/it7000/papers/situated.htm>. Consulté le 15 septembre 2004.
- Johnson-Laird, P. N.. 1988. "A computational analysis of consciousness". In *Consciousness in Contemporary Science*. A. J. M. E.Bisiach (Ed.), Oxford: Clarendon Press.
- Kanerva, Penty. 1988. *Sparse Distributed Memory*, Cambridge MA : The MIT Press.
- Kurzweil, Raymond. 1999. "The Coming Merging of Mind and Machine", In *Scientific American*. oct. 1999 [s.n.], [s.p.].
- Lewis, R.L.. 2001. "[Cognitive theory, Soar](#)". In *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Pergamon (Elsevier Science)
- Maes, Pattie. 1989. "How to Do the Right Thing." In *Connection Science Journal*, vol. 1, no 3, pp.291-323.
- Makatchev, Maxim, Pamela W. Jordan et Kurt VanLehn. 2004. «Abductive Theorem Proving for Analyzing Student Explanations to Guide Feedback in Intelligent Tutoring Systems», In *Journal of Automated Reasoning*, no 32, pp. 187–226.
- McCarthy, John. 2002. "Making Robots Conscious of their Mental States". [en ligne] <http://www-formal.stanford.edu/jmc/consciousness.ps>
- McGinn, Colin. 1991. *The problem of consciousness*. Oxford University Press.
- Mizoguchi, Riichiro et Jacqueline Bourdeau. 2000. «Using Ontological Engineering to overcome common AI-ED Problems», In *International Journal of Artificial Intelligence in Education*, no 11, pp. 107-121. [en ligne].
- Murray, Tom. 1996a. Special Purpose Ontologies and the Representation of Pedagogical Knowledge". In *Proceedings of the International Conference on the Learning Sciences*, AACE, Charlottesville (VA).
- Murray, Tom. 1996b. "ITS Conceptual Vocabulary: Primitive Tutorial Actions". In *Center for Knowledge Communication*.
- Nkambou, Roger et Gilles Gauthier. 1996. "Un modèle de représentation du curriculum dans un système tutoriel intelligent". In *Intelligent Tutoring Systems*, LNCS no. 1086, pp. 420-430. Springer-Verlag.

- Nkambou, Roger, Claude Frasson et Gilles Gauthier. 2003. "CREAM-Tools : An Authoring Environment for Knowledge Engineering in Intelligent Tutoring Systems". In : *Authoring Tools for Advanced Technology Learning Environments : Toward cost-effective adaptative, interactive, and intelligent educational software*. Murray, T., Blessing, S. and Ainsworth, S. (Eds). Kluwer Publishers. pp. 93-138.
- Psyché, Valéry, Jacqueline Bourdeau, Roger Nkambou et Riichiro Mizoguchi. 2005. "Making Learning Design Standards Work with an Ontology of Educational Theories". *To appear in the Proceedings of AIED'2005*.
- Ramamurthy, Uma, Sidney K. D'Mello, et Stanley Franklin. 2004. "Modified Sparse Distributed Memory as Transient Episodic Memory for Cognitive Software Agents". In *IEEE International Conference on Systems, Man and Cybernetics - SMC2004*, La Haye, Pays-Bas.
- Rézeau, Joseph. 2001. «Médiatisation et médiation pédagogique dans un environnement multimédia : Le cas de l'apprentissage de l'anglais en Histoire de l'art à l'université». Thèse de doctorat, Bordeaux, France, Université Bordeaux II, 696 p. [en ligne]. <http://perso.wanadoo.fr/joseph.rezeau/recherche/these.htm>.
- Reigeluth, Charles M. (Ed.). 1983. *Instructional-Design theories and models: A New Paradigm of Instructional Theory*. Mahwah (NJ): Lawrence Erlbaum Associates. 487p.
- Reigeluth, Charles M. (Ed.) 1999. *Instructional-Design theories and models: A New Paradigm of Instructional Theory. Vol.2*. Mahwah (NJ): Lawrence Erlbaum Associates. 715p.
- Sabah, Gérard. 1999. "The Respective Roles of Conscious and Subconscious Processes for Interpreting Language and Music". In *CSNLP-8*, Galway.
- Schoppers, Marcel J. (1987). "Universal plans for reactive robots in unpredictable environments". In *Proceedings of IJCAI-87*, pp. 852-859.
- Searle, John. 2003. *The Rediscovery of the Mind*. MIT Press, 288 p.
- Seth, Anil K., Bernard J. Baars, et al. 2004. "Criteria for consciousness in humans and other mammals." In *Consciousness and Cognition*, vol. 14, no. 1, pp. 119-139.
- Slovan, Aaron et Ron Chrisley. (2003). "Virtual machines and consciousness." In *Journal of Consciousness Studies*, vol.10, nos (4-5), pp.133-172.
- Tchétagni, Joséphine M.P. et Roger Nkambou . 2002. "Hierarchical Representation and Evaluation of the Student in an Intelligent Tutoring System". In *Intelligent Tutoring Systems*, S.A. Cerri, G. Gouarderes and F. Paragacu (Eds), LNCS 2363. Berlin: Springer-Verlag. pp. 708-717.
- VanLehn, Kurt et al.. 2002 . «The Architecture of Why2-Atlas: A Coach for Qualitative Physics Essay Writing». In *Actes de 6th International Conference on Intelligent Tutoring Systems*. pp.158-167.

Annexes