

---

# Vers un modèle de diagnostic de compétence

É. Delozanne<sup>1,2</sup>, D. Prévité<sup>2</sup>, B. Grugeon-Allys<sup>2,3</sup>, F. Chenevotot-Quentin<sup>3</sup>

<sup>1</sup>LIP6 UPMC-Paris-Universitas, <sup>2</sup>AIDA, 4 Place Jussieu, Boîte Courrier 1210, 75005 PARIS

[Elisabeth.delozanne@lip6.fr](mailto:Elisabeth.delozanne@lip6.fr), [dominique.previt@ac-rennes.fr](mailto:dominique.previt@ac-rennes.fr)

<sup>3</sup>LDAR- Université Paris Diderot - Paris 7, Case 7018, 175, rue du Chevaleret, 75013 PARIS

[brigitte.grugeon@amiens.iufm.fr](mailto:brigitte.grugeon@amiens.iufm.fr), [francoise.chenevotot@lille.iufm.fr](mailto:francoise.chenevotot@lille.iufm.fr)

---

**RÉSUMÉ.** Dans cet article, nous présentons la démarche de recherche mise en œuvre dans le projet Lingot, projet pluridisciplinaire dont l'objectif est d'instrumenter la gestion de la diversité cognitive des élèves pour faciliter l'apprentissage de l'algèbre élémentaire. Ce travail s'appuie sur des cadres théoriques issus de la didactique des mathématiques, de l'ergonomie cognitive et de l'informatique. Nous présentons la méthodologie de prototypage que nous avons mise en œuvre pour explorer les questions de recherche qui nous ont préoccupées et pour développer un outil de diagnostic de compétence. Les résultats de nos travaux concernent (i) l'analyse de raisonnements d'élèves en identifiant plusieurs niveaux d'indicateurs de la compétence algébrique selon les utilisations du diagnostic, (ii) une modélisation assez générale du diagnostic de compétence, et (iii) la mise en œuvre d'artéfacts informatiques pour mettre à l'épreuve et aussi pour faire émerger de nouveaux modèles. Enfin, nous discutons notre approche et nos résultats.

**ABSTRACT.** The Lingot project is a multidisciplinary project that aims to support the management of the cognitive diversity of students in school algebra learning. In this paper, we focus on the research approach based on theoretical frameworks from Computer Science, Cognitive Ergonomics and Didactics of Mathematics. We present our research questions and a cognitive diagnosis tool we implemented. We propose (i) to analyze students' algebraic reasoning with a three layer set of indicators each level adapted to a special usage, (ii) to develop a general diagnosis model based on a model of situations to collect data, to produce indicators from collected data and to display the indicators for different users (iii) to implement artifacts both to test and to suggest the different models.

**MOTS-CLÉS :** Environnements Informatiques pour l'Apprentissage Humain (EIAH), Diagnostic Cognitif, Algèbre Élémentaire, Prototypage, Conception Pluridisciplinaire.

**KEYWORDS:** Technology Enhanced Learning Environments, Cognitive Diagnosis, Elementary Algebra, Prototyping, Pluridisciplinary Design.

---

## 1. Introduction

Le projet Lingot vise la conception d'EIAH (Environnements Informatiques pour l'Apprentissage Humain) qui facilitent la prise en compte par les enseignants de la diversité cognitive des élèves afin de réguler les apprentissages. C'est donc fondamentalement un projet pluridisciplinaire qui regroupe principalement des chercheurs en informatique (du LIUM<sup>1</sup> et du LIP6) et en didactique des mathématiques (du LDAR) ainsi que des enseignants et des formateurs d'enseignants (des IUFM de Rennes, Créteil, Lille et Amiens). Dans le cadre du programme Cognitique du Ministère de la Recherche (appel d'offres 2002, L'apprentissage et ses dysfonctionnements), le projet Lingot a, en outre, bénéficié du concours de psycho-ergonomes (Jeanine Rogalski, CNRS, et son équipe) et d'une linguiste (Sylvie Normand alors à l'université de Rouen) (Delozanne et al. 2005). Dans cet article, nous présentons et discutons la démarche adoptée pour mettre en œuvre, sur le long terme, un tel projet pluridisciplinaire. Nous montrons comment, dans une démarche de prototypage, les modélisations didactiques et informatiques ont évolué et se sont co-construites. Nous nous efforçons de distinguer, autant que faire se peut, les problématiques spécifiquement informatiques, spécifiquement didactiques et celles qui sont inextricablement communes (Grandbastien et Labat 2006 p.19). Nous qualifions ces dernières de problématiques EIAH.

L'équipe du projet Lingot a travaillé sur trois axes de recherche : un axe diagnostic de compétence, un axe apprentissage et un axe instrumentation de l'activité des enseignants de mathématiques. L'axe diagnostic constitue le projet Pépite, c'est l'axe qui a été développé prioritairement et qui est présenté ici. L'objectif est de mettre au point une chaîne logicielle permettant de faire passer des tests à des élèves pour obtenir un profil cognitif pour chaque élève de la classe et, aussi, pour la classe dans son ensemble. Les deux autres axes de recherche concernent l'exploitation des profils cognitifs. En ce qui concerne l'apprentissage, notre objectif est d'associer, à chaque type de profil cognitif, des parcours d'apprentissage susceptibles de le faire évoluer (Grugeon et al. 2003). En ce qui concerne l'axe instrumentation de l'activité de l'enseignant, notre objectif est de concevoir des logiciels facilitant la mise en place d'un enseignement différencié fondé sur un diagnostic de compétence.

Dans la section suivante, nous exposons la problématique de ce projet pluridisciplinaire et nous positionnons ce travail par rapport aux disciplines participantes. Afin de faciliter la compréhension de ce travail, nous présentons dans la section 3 des exemples de diagnostics établis par les prototypes développés, et un

---

<sup>1</sup> LIUM : Laboratoire d'Informatique de l'Université du Maine ; LIP6 : Laboratoire d'Informatique de l'Université Pierre et Marie Curie, Paris Universitatis ; LDAR : Laboratoire de Didactique André Revuz (ex-DIDIREM) ; IUFM : Institut universitaire de formation des maîtres.

scénario de conception qui illustre nos objectifs. Dans la section 4, nous explicitons la démarche de la recherche et précisons l'état d'avancement des réalisations informatiques. Dans la section 5, nous résumons les principaux modèles que nous avons mis au point. Nous discutons de leur validité dans la section 6, avant de conclure sur notre expérience.

## **2. Positionnement scientifique du projet**

Notre projet étant un travail pluridisciplinaire dans le domaine des EIAH, nous situons nos travaux par rapport aux problématiques de recherche dans les différents domaines, avant de préciser la méthodologie mise en place pour ce projet.

### **2.1. Positionnement en didactique des mathématiques**

Du point de vue didactique, nous avons choisi le domaine de l'algèbre élémentaire en fin de scolarité obligatoire car il représente toujours un verrou d'accès à l'enseignement scientifique et, pour beaucoup d'élèves, un obstacle difficilement surmontable. Notre travail se fonde sur la théorie anthropologique du didactique (Chevallard 2003), la dialectique outil-objet (Douady 1986), les travaux de Duval sur les registres de représentation sémiotique (Duval 1992), la théorie des champs conceptuels (Vergnaud 1990) et les travaux en didactique de l'algèbre (Sfard 1991, Kieran 1992). Nous émettons une première hypothèse (Grugeon 1995, 1997) : les réponses des élèves à un ensemble de problèmes recouvrant les différents aspects de la compétence algébrique attendue à ce niveau scolaire révèlent certaines cohérences de leur activité algébrique, et la compréhension de ces cohérences devrait aider les enseignants, dans leur classe, à réguler l'enseignement de l'algèbre, voire à le différencier. Dans notre approche, les réponses des élèves ne sont pas analysées seulement en terme d'absence de savoir (ou de savoir-faire) ou en terme d'erreurs (savoirs ou savoir-faire inadaptés), mais en fonction de critères, dont les valeurs sont repérées par un soixantaine de codes. Ces critères sont organisés selon plusieurs dimensions, à savoir (i) la validité de la réponse, (ii) le type d'usage des lettres, (iii) le type de justification, (iv) le type de traduction d'une représentation en une autre, (v) la nature des transformations algébriques ou numériques mises en œuvre. Nous proposons ensuite de combiner ces critères sur un ensemble de problèmes pour rendre compte du niveau de développement de l'activité algébrique selon trois composantes : usage de l'outil algébrique selon les types de problèmes du domaine algébrique<sup>2</sup>, traduction entre deux registres de représentation, transformation des objets algébriques, expressions, formules, équations, identités

---

<sup>2</sup> Les problèmes du domaine algébrique mettent en jeu différents emplois de l'outil algébrique : modéliser des relations générales entre variables d'un système dans des contextes variés (géométrie, grandeurs), mettre en équation des problèmes, produire des expressions générales et prouver des propriétés numériques ou géométriques.

(Grugeon 1997, Delozanne et al. 2005). Nous définissons le « profil cognitif de l'élève en algèbre élémentaire » comme une description du développement de l'activité algébrique relativement à l'activité attendue par l'institution à ce niveau scolaire : il met en évidence les difficultés de l'élève mais aussi, et c'est une des originalités du travail de recherche, des leviers pour l'apprentissage. L'objectif de notre recherche est ainsi de faciliter la mise en œuvre par les professeurs, dans les classes, de stratégies différenciées d'enseignement. La différenciation, indispensable pour gérer des classes hétérogènes, peut s'appuyer sur les profils cognitifs des élèves pour les faire évoluer.

## 2.2. Positionnement en EIAH

Du point de vue EIAH, nous faisons une deuxième hypothèse : des logiciels pourraient « instrumenter » l'activité des enseignants. D'une part, ils permettraient de rendre des pratiques « d'enseignement sur mesure » écologiquement viables et, d'autre part, ils permettraient de diffuser, dans le corps enseignant, les résultats de recherche, tant au plan français qu'international, qui mettent en évidence des obstacles et des leviers pour l'apprentissage de l'algèbre. Cette hypothèse soulève en particulier le problème du diagnostic cognitif. C'est un processus qui consiste à produire de façon automatique une description des connaissances ou des savoir-faire qu'un système a cru déceler chez un élève en étudiant les traces de son activité. Ce processus pose ainsi deux problèmes interconnectés. Le premier problème, ce qui est souvent appelé modélisation de l'apprenant (Balacheff 2000), est celui de la modélisation des descripteurs à produire (dans notre cas, les éléments du profil cognitif) et des situations pour recueillir des données (dans notre cas les exercices spécifiquement conçus pour le diagnostic). Le second problème, souvent appelé diagnostic cognitif (Py et Hibou 2006) est celui du mode d'inférence des descripteurs à partir des données. Dans notre cas, il s'agit, à partir des réponses des élèves (les données), de produire des indicateurs de la compétence (les critères et les cohérences à étudier). Le modèle de l'apprenant est l'ensemble retenu des descripteurs de l'activité.

Le modèle le plus simple consiste à produire des taux de réussite à un ensemble d'exercices mettant en œuvre un ou plusieurs savoir-faire. On parle alors « d'expertise partielle » (overlay model) : les connaissances de l'élève étant une sous partie des connaissances de référence, les difficultés des élèves sont considérées comme des connaissances manquantes ou des savoir-faire non maîtrisés. Dans ce cas, un diagnostic local consiste à évaluer la validité de la réponse d'un élève à un exercice en la comparant à la bonne réponse, et un diagnostic global compte le nombre de réponses valides. C'est la méthode la plus employée sur les plateformes de formation en ligne ; certains travaux prévoient une analyse des erreurs les plus courantes pour associer une rétroaction adaptée vers l'élève : par

exemple dans la spécification QTI<sup>3</sup> ou dans certains travaux d'évaluation par des tests standardisés administrés à grande échelle et calibrés à partir de techniques numériques (Shute et al. 2008). Dans le projet Lingot, notre objectif est de trouver des cohérences dans l'activité des élèves pour, d'une part, déstabiliser les savoirs ou savoir-faire inadaptés et, d'autre part, pour déterminer les leviers (les *pépites*) sur lesquels s'appuyer pour construire des connaissances adaptées (les *lingots*).

Des modèles plus sophistiqués que l'expertise partielle simulent des fonctionnements cognitifs et appariant les observables avec les sorties de ces simulations. Un exemple emblématique de ce courant de recherches est celui des tuteurs cognitifs du laboratoire de Pittsburg de l'équipe d'Anderson et, particulièrement, Algebra Tutor décrit dans (Koedinger et al. 2006). Ces chercheurs modélisent, par un ensemble de règles de production, chaque pas de résolution d'un problème par un « élève idéal ». Ils modélisent aussi, par des règles erronées, les erreurs courantes sur un pas de résolution. Le diagnostic local à un exercice consiste à mettre en évidence les règles correctes ou incorrectes appliquées : c'est la méthode du modèle de trace (*model-tracing*). Une rétroaction est produite pour l'élève à chaque pas de résolution pour lui signifier que sa réponse est valide ou pour lui signaler une erreur. Puis, un diagnostic global repère l'évolution du savoir-faire entre les exercices (*knowledge tracing*). Ces systèmes sont diffusés sur des milliers d'établissements scolaires, principalement aux États-Unis. Ils ont donné lieu à des évaluations à grande échelle. Une variante de tuteur cognitif est le système MissLinguist (Hefferman et al. 2008) dont l'objectif est de modéliser un tuteur humain dialoguant avec un élève à propos de la mise en équation<sup>4</sup> de problèmes de la vie courante. Ces travaux sont parmi les plus aboutis dans le courant de recherche sur les tuteurs intelligents. Ils sont tout à fait dignes d'intérêt, mais leurs objectifs de recherche sont différents des nôtres. Ils visent à simuler la situation où l'élève est en face-à-face avec un tuteur humain, à modéliser le comportement des tuteurs humains qui cherchent à ramener l'élève dans le raisonnement valide. Dans le projet Lingot, l'objectif est d'outiller un enseignant pour différencier l'enseignement dans une classe. Il ne s'agit pas de simuler un enseignant humain corrigeant pas-à-pas un élève, mais de mettre à la disposition des enseignants des informations sur des cohérences dans l'activité de l'élève qui ne leur seraient pas accessibles sans logiciel.

Dans les études que nous venons de citer, l'évaluation est fondée sur des réponses à choix multiples, sur des réponses préformatées ou sur des réponses courtes (une valeur numérique, éventuellement une expression algébrique, un mot ou un groupe de mots). D'autres chercheurs ont pour objectif de dépasser ces évaluations sur des réponses prédéfinies. (Joosten-ten Brinke et al. 2007) proposent un modèle descriptif pour analyser les productions des élèves, par exemple en

---

<sup>3</sup> QTI : Question and Test Interoperability est une spécification émise par IMS-Global Learning Consortium.

<sup>4</sup> Il s'agit d'équations linéaires du premier degré.

évaluant des portfolios ou bien des productions coopératives. Mais, à notre connaissance, ce modèle reste descriptif et n'a pas été mis en œuvre dans une application informatique.

Dans le projet Lingot, nous nous attaquons au difficile problème de l'évaluation automatique de productions d'apprenants, c'est-à-dire de l'analyse des réponses à des questions ouvertes où l'élève formule ou construit lui-même sa solution. De notre point de vue, ce problème est complexe en raison de trois facteurs principaux. Le premier facteur est le degré de liberté laissé à l'apprenant dans l'expression de ses réponses : les questions à choix multiples ne posent aucune difficulté d'analyse<sup>5</sup>, tandis que les réponses libres sont complexes à analyser. Une façon de diminuer la complexité est de définir des contraintes sur le langage d'expression des réponses : moins le langage est contraint, plus l'analyse est difficile. Le deuxième facteur de complexité réside dans le degré d'ouverture du problème. Un problème est fermé quand il admet une solution unique et ouvert quand il admet de nombreuses voir une infinité de solutions valides. Enfin, le troisième facteur de complexité concerne les dimensions d'analyse de la réponse : l'analyse est plus complexe quand elle ne se limite pas à l'étude de la validité de la réponse mais vise à caractériser la réponse sur plusieurs dimensions. Des recherches dans différents domaines ont été menées pour résoudre ce problème. Les chercheurs utilisent ou construisent un logiciel spécifique pour rechercher des indicateurs témoignant des différentes dimensions d'analyse qu'ils veulent mener (Delozanne et al. 2007). Par exemple, dans le tuteur Logic-ITA, (Merceron et al. 2004) utilise un démonstrateur de formules logiques pour vérifier la validité des transformations logiques effectuées par les étudiants. (Hakem et al. 2005) ont mis au point une interface de saisie et un analyseur construit spécifiquement pour repérer les stratégies développées par des enfants de l'école primaire pour résoudre des problèmes additifs. Pour le tuteur intelligent Andes qui enseigne la physique (VanLehn et al. 2005), (Shapiro 2005) a développé un système de calcul formel qui compare les entrées des étudiants avec un graphe de solutions généré automatiquement. (Nicaud et al. 2004) ont également développé un logiciel de calcul formel sophistiqué pour le micro-monde Aplusix afin, d'une part, de vérifier l'équivalence entre les expressions saisies par l'élève et, d'autre part, de diagnostiquer les règles correctes et erronées qu'il a (probablement) appliquées. Ces logiciels sont très complexes à mettre en œuvre. Ils nécessitent une modélisation fine des processus de résolution les plus couramment observés à un niveau scolaire donné. Ils sont génériques dans le sens où ils permettent d'analyser les réponses à un ensemble assez varié de problèmes, mais ils sont ad hoc en ce sens qu'ils sont construits pour détecter certains types d'indicateurs et sont difficilement réutilisables dans un autre contexte.

Du point de vue EIAH, notre objectif est de concevoir un logiciel qui (i) recueille des données sur l'activité algébrique des élèves à partir, non seulement de réponses à des questions fermées, mais aussi de questions ouvertes, (ii) calcule des

---

<sup>5</sup> Mais elles posent d'autres problèmes pour le diagnostic (Shute et al. 2008)

indicateurs sur plusieurs dimensions pour décrire, à différents niveaux d'abstraction, le développement de l'activité algébrique des élèves à partir des données recueillies, (iii) soutient les enseignants dans l'exploitation de ces descripteurs en leur permettant de contrôler les calculs des différents indicateurs effectués par la machine et, surtout, de mettre en place des parcours d'apprentissage différenciés.

### 2.3. Positionnement en Informatique

Du point de vue informatique, ces objectifs placent l'interaction entre le logiciel et l'utilisateur humain au centre de nos préoccupations. Dans le projet Lingot, nous distinguons trois catégories d'utilisateurs : les élèves, les chercheurs et les enseignants. Du point de vue logiciel et à court terme, l'interaction entre l'élève et la machine a principalement pour objectif de modéliser des caractéristiques de son activité dans un domaine d'apprentissage donné et à un moment donné. Dans cet article, l'objet de la modélisation est la trace de l'activité algébrique laissée lors de l'interaction de l'élève avec le logiciel. Ce sont les chercheurs qui construisent cette modélisation dans un cadre nécessairement pluridisciplinaire. Les chercheurs en sciences humaines et sociales (SHS) produisent des « modèles descriptifs, qui permettent une analyse rétroactive des traces d'une activité » (Nicolle 1996) mais qui ne sont pas interprétables par des machines. Celles-ci ont besoin de modèles pour agir, de « modèles effectifs » (Nicolle 1996) qui spécifient, avec un niveau de détail inhabituel en SHS, les objets, les structures et les processus. En ce sens, l'informatique permet une expérimentation « in silico » (Nicolle 1996) des modèles issus de SHS. De ce point de vue, le premier problème, pour les informaticiens, consiste à mettre au point des modèles formels pour systématiser et rendre interprétables par des machines, les modèles descriptifs. Ceci n'est pas un problème trivial. Le passage du langage naturel ambigu et polysémique à un langage où la syntaxe et la sémantique sont définies entièrement, n'est pas une simple traduction mais produit une nouvelle modélisation. Le fait de rendre actifs les modèles, entraîne des modifications profondes de ces modèles, suggère de nouveaux modèles qui n'existeraient pas sans ce test d'effectivité que constitue le passage d'un modèle descriptif à un modèle exécutable. Ceci pose à la fois des problèmes de représentation de connaissances (connaissances sur le domaine, sur l'activité mathématique des élèves, sur les stratégies pédagogiques), des problèmes algorithmiques et des problèmes de génie logiciel (en particulier des problèmes de maintenance puisque les modèles évoluent, et des problèmes de généricité pour faciliter leur développement). Le second problème, pour les informaticiens, est dual du premier : faire en sorte que ces modèles exécutables, ne soient pas des boîtes noires mais puissent être interprétés, transformés, complétés, modifiés par des humains, dans notre cas, les chercheurs et, aussi, les enseignants et les élèves.

Ce dernier problème relève des méthodes et des techniques de l'Interaction Humains-Machines (IHM). L'interaction entre l'enseignant et le logiciel est à imaginer puisqu'un tel logiciel permet d'accéder à un degré de complexité

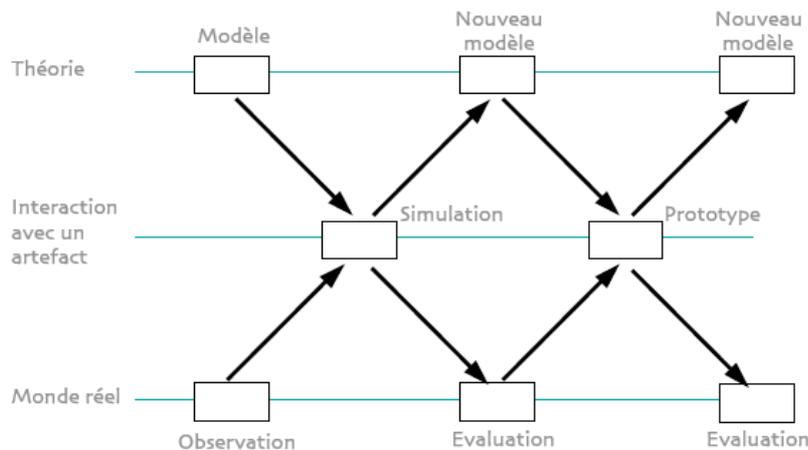
impossible à atteindre sans le logiciel et que de tels logiciels n'existent pas à l'heure actuelle. Le problème n'est donc pas entièrement spécifié à l'avance : c'est l'activité de conception et d'expérimentation qui permet de le comprendre et de le spécifier de façon incrémentale par coadaptation (Mackay et Fayard 1997) entre les modèles qui servent à fonder le logiciel et les usages effectivement mis en place de ces logiciels. De plus, le problème n'est pas défini une fois pour toutes, mais il évolue en fonction des variations de l'environnement technique et aussi institutionnel. Dans les domaines où la recherche en informatique doit faire face à l'imprédictabilité de l'activité humaine, les chercheurs utilisent une méthodologie de recherche itérative, fondée sur l'utilisation de prototypes (Wegner 1995, Mackay et Fayard 1997, Nicolle 1996, 2001). C'est celle que nous avons adoptée dans le projet Lingot.

#### **2.4. Méthodologie**

Le rôle du prototypage dans la conception de logiciels interactifs est bien documenté et permet de mettre en œuvre des démarches de conception itérative (par exemple ISO13407 1999, Beaudoin-Lafon et Mackay 2002, Delozanne 2006). Les prototypes réifient les idées et facilitent la création d'une culture commune entre les différents acteurs et les différentes disciplines d'une équipe de conception. Ils autorisent un travail fructueux avec les futurs utilisateurs. Du point de vue recherche, outre le test d'effectivité que nous venons de mentionner, en objectivant les modèles, ils permettent de les partager, de les discuter avec d'autres chercheurs et utilisateurs et, à partir de là, de créer des modèles plus génériques et de faire surgir de nouvelles questions de recherche. Éventuellement, selon leur niveau de développement, le niveau de déploiement et les questions de recherche, ils permettent d'observer des usages et la pertinence des modélisations par rapport aux objectifs opérationnels sur lesquels se fondent l'analyse et les tests des versions ultérieures.

Nous avons mis en œuvre une démarche itérative qui a comporté plusieurs cycles de recherche comme l'illustre le Schéma 1. Tout d'abord, nous avons posé des hypothèses de travail puis défini les questions de recherche qui nous permettent d'explorer la pertinence de ces hypothèses. Nous avons ensuite élaboré plusieurs modèles à partir de l'étude des travaux dans différents domaines et de l'observation, d'une part des acteurs (élèves ou enseignants) et, d'autre part, des prototypes construits à l'étape précédente. Ces prototypes sont réalisés pour permettre d'interpréter des modèles qui, par ailleurs, continuent d'évoluer au cours de la mise en œuvre informatique. Celle-ci, en effet, permet leur mise au point et, au final, montre leur effectivité. Une fois le prototype réalisé, son comportement interne est étudié : nous regardons les résultats produits sur des corpus de réponses obtenus auprès des élèves et vérifions leur cohérence, d'une part au regard du modèle descriptif et, d'autre part, par une inspection par des experts humains. Puis, le prototype est mis à l'épreuve auprès d'enseignants et d'élèves dont nous observons le comportement à la lueur de nos préoccupations et des cadres théoriques issus des

disciplines des participants. Ces observations soit confortent nos hypothèses, soit suggèrent de nouvelles idées et questions de recherche, soit nous amènent à préciser ou même à restructurer les modèles précédents. Notons que ces mises à l'épreuve ne prennent pas la forme d'expérimentations contrôlées. Leur objectif n'est pas d'évaluer les performances d'un prototype, mais de confronter nos modèles avec les exigences des contextes d'utilisation dans le cadre de l'institution scolaire. Ces mises à l'épreuve permettent aussi de recueillir des corpus d'interaction sur lesquels se fondent l'analyse et les tests des versions ultérieures.



**Schéma 1** : Stratégies de recherche pluridisciplinaire (Mackay et Fayard 1997)

### 3. Quelques exemples et scénarios

Pour illustrer notre propos sur des exemples de productions d'élèves, nous décrivons le diagnostic cognitif établi par le prototype Pépite<sup>6</sup>. Puis, nous présentons un scénario de conception qui décrit des hypothèses sur les utilisations par les enseignants des logiciels sur lesquels nous travaillons.

#### 3.1. Côté élève

Le diagnostic cognitif va au-delà de l'évaluation de la validité d'une réponse. Pour montrer la diversité des raisonnements des élèves, le Tableau 1.a présente les réponses de plusieurs élèves de classe de troisième à un même exercice du test Pépite. Remarquons tout d'abord qu'aucune de ces réponses n'est complètement correcte. Elles témoignent cependant de différences très importantes dans la façon dont ces élèves abordent l'algèbre. Le Tableau 1.b présente, sur ces réponses, l'analyse multidimensionnelle automatiquement menée par le logiciel Pépite en

<sup>6</sup> Téléchargeable gratuitement sur le site <http://pepite.univ-lemans.fr>

s'appuyant sur l'analyse didactique de Grugeon (op. cit.). L'analyse porte sur six *dimensions d'évaluation* qui comportent, chacune, un ensemble de *critères d'évaluation*. Pour repérer ces critères, le logiciel attribue à chaque réponse des *codes* (indiqués sur le Tableau 1.b entre parenthèses<sup>7</sup>). C'est ce que nous appelons le *diagnostic local* ou encore *le codage des réponses des élèves*. Dans un deuxième temps, le logiciel analyse les codes attribués sur d'autres exercices et les agrège pour mettre en évidence des cohérences dans l'activité algébrique de l'élève (ce que nous appelons le diagnostic global). En s'appuyant sur ce diagnostic global, le système émet des conseils pour les faire évoluer. Bien qu'il soit hasardeux de déduire une cohérence dans l'activité d'un élève à partir d'un seul exercice, les paragraphes qui suivent visent à donner une idée générale de la démarche implémentée.

Énoncé			
Un prestidigitateur est sûr de lui en réalisant le tour suivant. Il dit à un joueur : “ Tu penses un nombre, tu ajoutes 8, tu multiplies par 3, tu retranches 4, tu ajoutes ton nombre, tu divises par 4, tu ajoutes 2, tu soustrais ton nombre : tu as trouvé 7 ”. L'affirmation est-elle vraie ? Justifie ta réponse.			
Réponses de 4 élèves			
Khemarak	Nicolas	Karine	Laurent
Soit 5 un nombre $((5+8) \times 3 - 4 + 5) / 4 + 2 - 5 = 7 ?$ $((13) \times 3 - 4 + 5) / 4 + 2 - 5 = 7 ?$ $(39 - 4 + 5) / 4 + 2 - 5 = 7 ?$ $10 + 2 - 5 = 7 ?$ $10 - 3 = 7 ?$ $7 = 7 ?$ Oui donc cela marche	$3 + 8 = 11$ $11 \times 3 = 33$ $33 - 4 = 29$ $29 + 3 = 32$ $32 / 4 = 8$ $8 + 2 = 10$ $10 - 3 = 7$	$x + 8 = 8x$ $8x$ $3 \times 8x = 24 + 3x =$ $27x$ $27x - 4 = 23x$ $23x + x = 24x$ $24x / 4 = 6x$ $6x + 2 = 8x$ $8x - x = 7$	$= [(x+8) \times 3 - 4 + x] / 4 + 2 - x$ $= (3x + 24 - 4 + x) / 4 + 2 - x$ $= 4x + 20 / 4 + 2 - x$ $= x + 5 + 2 - x$ $= 7$

**Tableau 1.a :** Réponses de 4 élèves à un exercice de Pépite

Khemarak n'est pas encore entré dans une démarche d'utilisation de l'algèbre pour prouver puisqu'il se limite à donner un exemple ; toutefois, sa maîtrise des calculs numériques et sa façon d'appréhender de façon globale l'expression numérique et de la transformer par équivalence, est un levier d'apprentissage qui devrait lui permettre d'entrer rapidement dans une démarche algébrique. Après avoir

<sup>7</sup> Les modélisations ayant évolué au cours de l'avancement du projet, le nombre et les noms des dimensions ont varié. Dans le tableau 1.b, nous adoptons la dénomination actuelle tandis que les écrans du logiciel Pépite (Annexe, Figure 2), présentent la dénomination originelle issue des travaux de (Grugeon 1995).

vérifié que ces indicateurs se retrouvent sur d'autres exercices du test, un conseil possible est d'amener Khemarak à comprendre la nécessité de l'usage des lettres. Pour cela, le logiciel conseille de travailler des exercices où l'algèbre est nécessaire comme outil de généralisation et de preuve ou de modélisation (par exemple recherche de formule d'aire d'une figure).

Dimensions	Exemples de critères et leur code			
	Khemarak	Nicolas	Karine	Laurent
<b>Justification</b>	Par l'exemple (J2)	Par l'exemple (J2)	De type formel scolaire avec application de règles fausses (J31)	Par l'algèbre (J1)
<b>Utilisation des lettres</b>	Pas d'utilisation des lettres (L5)	Pas d'utilisation des lettres (L5)	Utilisation des lettres avec utilisation de règles fausses (L3)	Utilisation correcte des lettres (L1)
<b>Traduction</b>	Traduction par une expression globale parenthésée (T1)	Traduction par expression partielle (T2)	Traduction par expression partielle enchaînée en succession d'opérations (T4)	Traduction par expression globale parenthésée (T1)
<b>Signe =</b>	Relation d'équivalence (E1)	Annonce un résultat (E2)	Annonce un résultat (E2)	Relation d'équivalence (E1)
<b>Écritures algébriques</b>	Correctes (EA1)	Correctes (EA1)	Identification incorrecte de + et × (assemble les termes)(EA42) règles fausses : $x + a \rightarrow x a$ $a x \pm b \rightarrow (a \pm b) x$ $a x - x \rightarrow a - 1$	Erreur de parenthèse avec mémoire de l'énoncé (EA31)
<b>Validité</b>	Réponse invalide (V3)	Réponse invalide (V3)	Réponse invalide (V3)	Réponse invalide (V3)

**Tableau 1.b :** Diagnostic automatique de Pépite sur les réponses de quatre élèves

Nicolas utilise lui aussi une preuve par l'exemple et maîtrise ces calculs simples. Par contre, il utilise une démarche arithmétique en indiquant une suite de calculs où le signe égal « annonce » un résultat. Là encore, après avoir vérifié sur d'autres exercices, le logiciel conseille de le faire travailler sur les différentes écritures d'un même nombre, par exemple, avec « la calculatrice défectueuse » logiciel développé par l'équipe de G. Lemoyne de l'université de Montréal (Lemoyne 2005).

Karine utilise la lettre  $x$ , mais sa résolution est qualifiée de justification par le « formel scolaire » car elle utilise des règles fausses bien identifiées par les enseignants et les chercheurs en didactique. Par cette expression, nous désignons les élèves pour qui, faire des mathématiques, c'est appliquer des règles vides de sens. Pour faire progresser ce type d'élèves (nous disons les élèves qui ont le même *stéréotype*), il ne suffit pas de leur répéter les règles correctes ou de corriger leurs erreurs. Un long travail est nécessaire, d'une part pour déstabiliser les règles incorrectes et, d'autre part, pour donner du sens à l'usage des lettres. On peut leur proposer des exercices de reconnaissance d'identités (où l'égalité est mobilisée comme symbole d'équivalence et non comme annonce de résultat ce qui est le propre des démarches arithmétiques) et, aussi, les exercices proposés à Khemarak.

Laurent est entré dans une démarche algébrique et son raisonnement prouve qu'il sait traiter des expressions équivalentes. Comme beaucoup d'élèves à ce niveau (classe de troisième), il ne maîtrise pas encore parfaitement l'utilisation des parenthèses mais il garde le sens des opérations. L'utilisation d'un logiciel comme Aplusix (Nicaud et al. 2004) peut lui permettre de maîtriser les techniques de calcul.

Ces exemples ont pour objectif de montrer la diversité des réponses des élèves, la largeur du spectre des réponses analysées automatiquement et des exemples de cohérences que le logiciel détecte dans l'activité des élèves. Remarquons que l'analyse automatique des réponses à cet exercice (Prévit 2002) ne se prétend pas exhaustive, mais analyse toutes les réponses qui correspondent à des indicateurs (les critères) dans l'analyse de la compétence établie par les didacticiens. Elle produit, en particulier, automatiquement les codages indiqués dans le Tableau 1.b.

### **3.2. Côté enseignant**

Illustrons maintenant l'utilisation envisagée des outils développés dans le projet Lingot et leur intégration dans les classes. Une méthode de conception utilisée pour centrer les projets sur les besoins des utilisateurs et leur activité consiste à mettre au point des scénarios (Beaudoin-Lafon et Mackay 2002, Carroll et al. 2001, Delozanne 2006). Les scénarios sont des histoires qui mettent en scène l'utilisation d'un futur logiciel interactif. La création de scénarios est une technique qui permet d'impliquer les utilisateurs dans la vision du système à venir, de prendre en compte le contexte d'utilisation avant de spécifier les fonctionnalités du système. Elle permet en outre à l'équipe de conception de se référer à ces scénarios. Dans le projet Lingot, des scénarios ont d'abord été écrits par des enseignants, puis ils ont été retravaillés au

sein de l'équipe pour incorporer des idées issues des nombreuses observations d'enseignants en classe ou en formation que nous avons pu mener (Delozanne et al. 2003). Nous présentons ici un de ces scénarios de conception de Pépite.

Clémence est une enseignante expérimentée de mathématiques en classe de seconde. Avant de commencer la première séquence d'algèbre, Clémence veut savoir où en sont ses élèves et jusqu'où sont allés ses collègues de l'année précédente. Elle choisit un test de niveau « début de classe de seconde » dans la banque de tests Pépite et fait passer le test individuellement à chaque élève de cette classe. Elle imprime, pour chaque élève, un bilan individuel (Figure 3 en Annexe). Elle imprime également un bilan de l'ensemble de la classe (section 4.3) mettant en évidence trois groupes d'élèves et proposant une liste de savoir-faire à travailler. Pour les trois groupes d'élèves, le logiciel conseille de travailler sur la « reconnaissance d'expressions algébriques » (c'est-à-dire repérer les opérations et les opérandes qui les constituent). Elle décide de travailler cette compétence avec les logiciels CIME ou AILE (Grugeon et al. 2003, Delozanne et al. 2005) : CIME permet de travailler sur la relation entre un énoncé de problème et sa mise en équation. AILE permet d'associer une expression algébrique à son expression en français ou à une représentation arborescente.

Actuellement, un certain nombre de prototypes ont été développés qui permettent le déroulement de ce scénario. Nous en présentons une partie dans la section suivante.

#### **4. La démarche de recherche**

Le projet Lingot s'est structuré autour d'une démarche de recherche itérative fondée sur la réalisation de prototypes pour mettre à l'épreuve des hypothèses, produire des résultats et formuler de nouvelles questions de recherche. Dans cette section, nous présentons les différents cycles de la recherche en précisant, à chaque itération, les hypothèses de travail, les questions de recherche, puis les modèles et les prototypes mis au point et, enfin, une discussion portant sur les résultats et les nouvelles questions soulevées. Nous terminons cette section par une description de l'état actuel du projet.

##### **4.1. Cycle de recherche 1 : un outil de diagnostic papier-crayon**

Le projet Lingot s'appuie sur une étude épistémologique, cognitive, sémiotique et didactique de l'enseignement et de l'apprentissage de l'algèbre à la fin de la scolarité obligatoire (Grugeon 1995, 1997, Grugeon-Allys 2008). Nous ne présentons ici que ce qui est utile à la compréhension de notre travail. Le *problème* posé aux chercheurs peut être formulé ainsi : comment des élèves parmi les meilleurs de lycées professionnels, peuvent-ils se retrouver en échec en arrivant en seconde de lycée d'enseignement général ? Grugeon a posé l'*hypothèse* que (H1)

des décalages dans les attentes des différentes institutions étaient une des explications de ces dysfonctionnements. Pour tester la pertinence de cette hypothèse, deux *questions de recherche* principales ont été formulées : Q1.1 : Est-il possible de définir une structure d'analyse de la compétence algébrique élémentaire, une référence, indépendante des institutions ? Q1.2 : Est-il possible d'identifier à la fois les difficultés des élèves mais aussi des leviers pour l'apprentissage ?

#### 4.1.1. *Une grille d'analyse multidimensionnelle de la compétence algébrique à la fin du collège*

S'appuyant sur des travaux théoriques et expérimentaux menés en didactique des mathématiques (Chevallard 2003, Douady 1986, Duval 1992, Kieran 1992, Sfard 1991, Vergnaud 1990), et aussi sur une étude de l'activité en algèbre élémentaire d'une cohorte d'élèves sur une période de plusieurs années, Grugeon a établi une grille d'analyse multidimensionnelle de la compétence algébrique attendue des élèves en fin de collège. La compétence et les difficultés des élèves en algèbre y sont analysées selon trois composantes :

- le type de problème. L'algèbre est considérée comme un outil de résolution de problèmes, un outil de généralisation, de preuve ou de modélisation de situations ;
- les objets de l'algèbre. En particulier la grille identifie l'utilisation des lettres, la production et l'interprétation des expressions algébriques ;
- le calcul formel en articulant les aspects sémantiques, syntaxiques, techniques et sémiotiques.

#### 4.1.2. *Un prototype papier-crayon d'un outil de diagnostic*

Afin de situer les élèves par rapport à cette grille, Grugeon a mis au point un outil de diagnostic papier-crayon composé de trois parties. Tout d'abord, elle propose un *test* : un ensemble d'une vingtaine d'*exercices* comportant des questions fermées et des questions ouvertes qui recouvrent l'ensemble des *types de problèmes* du domaine algébrique à ce niveau scolaire. Puis, une analyse didactique a priori de chaque exercice et une *grille de codage* des réponses des élèves, établies à partir de la grille d'analyse multidimensionnelle de compétence, permettent à l'enseignant (ou au chercheur) d'interpréter les réponses des élèves aux différentes questions du test en leur attribuant un code (diagnostic local). Les différentes *dimensions* considérées sont : la validité de la réponse selon les types de problèmes, l'utilisation des lettres (inconnue, variable, nombre généralisé, abréviation ou étiquette), la pratique du calcul algébrique, la traduction entre différents registres sémiotiques (graphique, géométrique, algébrique, langue naturelle), les types de rationalité. Cette grille de codage, ainsi que l'analyse didactique a priori des exercices, sont présentées sous forme de tableaux et décrites en langue naturelle.

Ensuite, un calcul s'appuyant sur des seuils de fréquence des codes, permet d'agréger les différents codes, pour construire le profil cognitif de l'élève en algèbre. Dans ces premiers travaux, un profil cognitif en algèbre est une description sur trois axes de la compétence algébrique des élèves :

- Une description quantitative exprimée en termes de taux de réussite et de traitements algébriques maîtrisés,
- Une description qualitative exprimée en termes de modes de fonctionnement : utilisation des lettres, calcul algébrique, conversion (c'est-à-dire l'articulation entre le registre algébrique et d'autres registres sémiotiques), type de preuve,
- Une description de l'articulation entre les différents cadres et registres (graphique, algébrique, géométrique, langage naturel) sous la forme d'un diagramme.

#### 4.1.2. Discussion

Ce travail a, de notre point de vue, produit des résultats importants. Ce test papier-crayon a été administré à 600 élèves de fin de collège en 1995. Ce corpus a été étudié dans le cadre d'une recherche en didactique des mathématiques (Lenfant 1997) qui a montré que la grille d'analyse de la compétence algébrique s'est avérée assez robuste pour prédire les types de raisonnements observés sur le corpus recueilli. Cette grille d'analyse a été utilisée par la suite pour mettre en évidence des décalages entre les rapports institutionnels à l'algèbre dans les transitions entre institutions (professionnel/général, classe de troisième/seconde), pour étudier l'évolution des programmes de collège, pour étudier les manuels scolaires et, enfin, elle a été adaptée pour un diagnostic en algèbre en classe de cinquième et quatrième (Grugeon-Allys 2008, Chenevotot-Quentin et al. 2008). Dans le cadre de formation d'enseignants, l'outil de diagnostic est utilisé pour permettre à des enseignants de définir des parcours d'apprentissage différenciés. Cependant, cet outil est apparu trop complexe et trop lourd pour être utilisé dans le cadre habituel du travail en classe sans un outillage adapté. Cette nécessité de produire un logiciel de diagnostic pour instrumenter la différenciation a été le point de départ des projets Pépite et Lingot.

## 4.2. Cycle de recherche 2 : Un logiciel de diagnostic assisté

Un des thèmes de recherche qui préoccupe les chercheurs en Intelligence Artificielle depuis ses débuts est la conception de modèles des connaissances des élèves calculables par une machine et exploitables par des enseignants (Py et al. 2006). La modélisation « semi-formalisée » proposée par Grugeon paraissait un point de départ intéressant pour une collaboration visant à relever le défi d'automatiser l'outil de diagnostic papier-crayon. De plus, le corpus des réponses obtenues auprès de 600 élèves constituait une base solide pour spécifier les interfaces et pour tester a priori les procédures d'analyse des réponses.

Ainsi, est né le projet Pépite. Ce premier travail a principalement été mené par Stéphanie Jean dans le cadre de sa thèse (Jean 2000). Il avait pour objectif de tester trois hypothèses : (H2.1) il est possible à l'aide d'un ordinateur de collecter des données riches sur la compétence des élèves ; (H2.2) il est possible, à partir de ces données, d'obtenir automatiquement (ou semi-automatiquement), un profil cognitif

des élèves ; (H2.3) les profils cognitifs élaborés aident les enseignants à prendre des décisions pour leurs élèves.

Les membres de l'équipe se posaient principalement trois questions de recherche : Q2.1 : L'utilisation d'un clavier et d'une souris limiterait-elle l'expression des élèves, que ce soit en langue naturelle ou en algèbre ? Q2.2 : Comment analyser automatiquement les réponses des élèves en leur laissant la possibilité de produire eux-mêmes des réponses, l'étude de la production par les élèves d'expressions algébriques étant indispensable au diagnostic ? Q2.3 : Comment les enseignants accepteraient-ils un tel outil qui leur donne accès à des informations inhabituelles sur les élèves ?

#### 4.2.1. Le premier prototype *Pépité*

Du point de vue informatique, ce premier travail a abouti à la réalisation du logiciel *Pépité* dont la mise en œuvre permettait de tester ces hypothèses et d'étudier ces questions. Ce logiciel est composé de trois modules : *PépiTest*, *PépiDiag* et *PépiProf*. *PépiTest* propose 22 exercices inspirés de ceux de l'outil de diagnostic papier-crayon et recueille les réponses des élèves aux exercices (Annexe, Figure 1). *PépiDiag* interprète les réponses des élèves à chaque exercice de *PépiTest* en appliquant des heuristiques dérivées de la grille d'analyse issue de l'analyse didactique. Comme l'outil de diagnostic papier-crayon, il associe chaque réponse d'élève à un code permettant de qualifier la réponse sur plusieurs dimensions. *PépiDiag* analyse automatiquement les réponses aux questions fermées (42 % des items) et les expressions algébriques simples (23 % des items). Les réponses en langage naturel (3% des items) sont très partiellement analysées par recherche de mots clés. Les raisonnements algébriques (32 % des items), par exemple du type présenté dans le Tableau 1.b, n'étaient pas analysés par cette première version, mais l'ont été par la suite (Prévit 2002). *PépiProf* est le logiciel destiné aux enseignants. Il établit le profil de l'élève par une analyse transversale des codes attribués aux réponses à chaque exercice et le présente au professeur. L'enseignant peut modifier le codage s'il n'est pas d'accord avec le codage automatique (Annexe, Figure 2). .

#### 4.2.2 Discussion

Ce premier prototype s'est avéré essentiel pour nos recherches en dépit d'un module de diagnostic assez simple qui assure un diagnostic semi-automatique. Par rapport aux logiciels d'évaluation traditionnels, cet outil a deux caractéristiques originales. D'une part, il ne se réduit pas à une analyse en termes de réussite/échec ou de listes d'erreurs, mais il donne une description qualitative de la compétence algébrique des élèves. D'autre part, si l'analyse des réponses porte en partie sur des questions fermées, elle traite néanmoins des questions ouvertes où les réponses sont produites par les élèves ce qui apparaît essentiel pour les didacticiens.

Ce premier prototype de *Pépité* a été testé dans des contextes très différents auprès d'élèves et d'enseignants, dans des classes, dans des ateliers d'enseignants en formation initiale et continue ou lors d'ateliers de chercheurs (Tableau 2). De plus le

logiciel étant téléchargeable sur Internet, nous avons reçu des retours de la part d'enseignants qui l'ont testé. La mise en place de ces tests utilisateurs a pris des formes très diverses : observation du comportement des élèves passant le test dans des établissements scolaires, entretiens avec des enseignants ou avec des élèves autour du profil établi par le logiciel, entretiens avec des enseignants ayant utilisé Pépite avec leur classe, observation de stagiaires-enseignants pendant et après des formations, atelier de travail pour recueillir l'avis de chercheurs et de formateurs d'enseignants.

Contexte	Situation	Utilisateurs observés	Données collectées
Test d'élèves	Classes (3 <sup>ième</sup> , 2 <sup>nde</sup> )	Élèves (N=400)	Réponses d'élèves Questionnaires Observations Rapports
Recherche	Recherche de régularités dans les profils d'élèves, fouille de données	Chercheurs (N=10)	Problèmes d'utilisation ou bogues Définition de classes de profils
Formation d'enseignants, de formateurs	Étude de la compétence en algèbre (IUFM Créteil, Rennes, Montpellier)	Stagiaires et professeurs en responsabilité, formateurs (N=140)	Observations Questionnaires Entretiens Analyse de discours
Session pilote	Classes (Aide individualisée et évaluation)	Enseignants (N=6)	Observations Rapports oraux ou courriels Cassettes audio
Utilisations spontanées	Classes	Enseignants (N=9)	Rapports oral ou courriel

**Tableau 2.** *Les différentes mises à l'essai de Pépite*

Un corpus numérique riche de 400 réponses d'élèves au test a ainsi été constitué pour servir de corpus de référence dans des études ultérieures, que ce soit pour vérifier les algorithmes diagnostic ou pour restructurer les descripteurs du profil cognitif. En rendant actif le premier modèle papier-crayon, ce premier Pépite a permis tout d'abord de montrer la faisabilité de la démarche (Jean 2000), puis de tester le modèle proposé : de le systématiser, de l'affiner, de le compléter et d'étudier sa robustesse en mettant à l'épreuve nos hypothèses auprès des élèves et des enseignants dans différents contextes, mais aussi auprès d'autres chercheurs et

ainsi de reprendre et reformuler nos questions de recherche (Delozanne et al. 2002, 2005).

Les principales leçons que nous avons tirées de ces mises à l'épreuve sont les suivantes.

PépiTest recueille des données, les réponses des élèves, qui peuvent être ensuite utilisées pour le diagnostic (Jean 2000). Premièrement, comme nous l'avions anticipé, les élèves ont rencontré des problèmes pour écrire les expressions algébriques « linéaires » avec clavier-souris mais ces difficultés ne les empêchent pas d'en produire. Deuxièmement, à partir de ces données, les chercheurs en didactique peuvent appliquer la grille d'analyse et construire « manuellement » le profil cognitif de l'élève (Grugeon et al. 2003). Troisièmement, l'ensemble des réponses aux questions ouvertes, recueillies par Pépite, couvre complètement le spectre des réponses prévues par l'analyse didactique a priori. Cela signifie que le logiciel ne diminue pas l'éventail de réponses repérées par l'analyse a priori papier-crayon. Quatrièmement, malgré des différences locales, des chercheurs ou des enseignants expérimentés identifient des cohérences de fonctionnement de leurs élèves comparables à celles identifiées avec l'outil papier-crayon (Lenfant 1997). Ces résultats constituent, à notre sens, une démonstration de la faisabilité du projet en ce qui concerne le recueil des données sur l'élève. Ils confortent en particulier les hypothèses H2.1 et H2.2.

Du côté des enseignants, dans les sessions pilotes et les utilisations spontanées, lors d'entretiens que nous avons menés, les enseignants ont indiqué avoir repéré chez certains élèves des capacités ou encore des fragilités qu'ils n'avaient pas remarquées auparavant chez leurs élèves. Pépite agit comme un « révélateur ». C'est une des raisons pour laquelle des formateurs d'enseignants de mathématiques pensent que c'est un outil utile pour aider les enseignants à prendre en compte les différents aspects de la compétence algébrique. Ainsi, depuis 2000, Pépite est régulièrement utilisé pour la formation des enseignants de mathématiques dans les centres de formation des maîtres à Paris, Créteil, Amiens et Rennes.

Cependant, notre troisième hypothèse concernant les utilisations par les enseignants en classe, est loin d'être confirmée. Des problèmes techniques liés à la récupération des données devront être surmontés pour rendre le logiciel plus opérationnel. Plus profondément, en dehors de l'utilisation en formation d'enseignants, les premières mises à l'essai ont montré que, si l'utilisation du logiciel de test était peu problématique et ont donné lieu à des usages supplémentaires à ceux prévus lors de la conception, l'utilisation des profils cognitifs laissait souvent les enseignants perplexes. Voici les principaux retours que nous avons reçus.

1. Aucune utilisation spontanée de l'interface de modification ou de vérification du diagnostic local (codage des réponses) n'a été observée. À de rares exceptions près, les enseignants ont réclamé un diagnostic automatique des réponses et non un diagnostic assisté comme nous l'avions prévu dans Pépite.

2. Si l'analyse des réponses d'élèves (diagnostic local) a été peu mise en cause, les enseignants ont jugé le profil cognitif (diagnostic global) trop complexe et même rebutant.
3. Ils ont suggéré des outils pour exploiter ce profil. En particulier,
  - ils ont souhaité que le logiciel propose des parcours d'apprentissage adaptés au profil cognitif.
  - certains enseignants ont ainsi suggéré la création d'« une géographie cognitive de la classe » et la proposition de parcours d'apprentissage qui pourraient s'appliquer à des groupes d'élèves.
4. Les enseignants qui ont utilisé le test ont souhaité disposer de tests similaires à différents moments de l'apprentissage pour évaluer les évolutions.
5. Le test a été jugé trop long par les élèves comme par les enseignants.
6. Les enseignants ont demandé un retour à destination de l'élève sous la forme d'un corrigé ou d'un bilan de compétence plus accessible que le profil de PépiProf et qui les incitent à réfléchir sur leurs points forts et leurs points faibles.

Ce premier travail a, tout d'abord, montré la faisabilité d'un diagnostic automatique (Jean 2000). De plus, l'existence d'un prototype a permis d'observer les difficultés des utilisateurs à apprivoiser dans leurs pratiques un outil novateur (observations 2, 3 et 4). Au-delà des insuffisances de ce premier prototype, il nous est apparu que la méthode de diagnostic mise en œuvre par Pépite entraine en conflit avec les procédures usuelles de diagnostic de nombreux enseignants. Nous avons donc mis sur pied une nouvelle recherche, pour analyser les pratiques usuelles de diagnostic afin d'élaborer des outils conceptuels et logiciels qui entrent en résonance avec celles-ci. Ce travail fait l'objet de la section suivante. Nous revenons sur les autres observations (1, 5, 6 puis 7) dans les sections suivantes.

### **4.3. Cycle de recherche 3 : Bilan de compétence individuel et bilan de la classe**

Une équipe de psychologues-ergonomes a mené une étude approfondie de l'activité d'enseignants utilisant Pépite (Rogalski 2005). Cette étude a identifié des obstacles à l'utilisation courante de logiciels du type de Pépite et à leur intégration dans les pratiques enseignantes. En se plaçant dans le cadre d'une théorie de l'activité et de son organisation, cette étude considère l'enseignement comme « la gestion d'un environnement dynamique ouvert, à long délai de réponse et à tempo d'action multiple (du temps d'un exercice en classe à celui d'un cycle d'enseignement) avec un accès indirect<sup>8</sup> au processus géré, et une diversité des dynamiques particulières des apprentissages individuels dans une même classe. »

---

<sup>8</sup> « Les informations sur les rapports des élèves et des contenus mathématiques ne sont pas directs et doivent être construites par des dispositifs ad hoc (c'est d'ailleurs un des objectifs du logiciel Pépite) et, en fait, inférées à partir des performances produites par les élèves » (ibid.)

(Rogalski 2005, p. 93). La psychologie cognitive a montré un lien étroit entre diagnostic et répertoire d'actions : « Le processus de diagnostic n'existe pas pour lui-même et n'est pas nécessairement identifiable au préalable d'une décision d'action ». Enfin, la problématique de l'instrumentation de l'activité de diagnostic amène à repérer comment Pépite peut, d'une part, s'insérer dans l'activité de diagnostic et, d'autre part, modifier celle-ci. Cette étude s'est appuyée sur des analyses de discours d'enseignants dans des entretiens, sur l'observation et l'analyse de séances en classe et de séances de formation professionnelle.

Les principaux résultats de cette étude, de notre point de vue de chercheuses en EIAH, concernent la diversité des représentations des enseignants sur l'algèbre, mais aussi des invariants dans leurs pratiques de diagnostic.

L'activité de diagnostic est de deux types : l'un dominant orienté vers la classe et, l'autre différentiel orienté vers les élèves. Lors d'une séance dans la classe, le diagnostic est dynamique et lié soit avec les buts visés, soit avec les ressources et les actions possibles. Il vise la conduite de la classe et concerne ainsi davantage l'ensemble de la classe ou des groupes d'élèves que des individus. Ce diagnostic local est lié au thème en cours et porte sur des questions courtes. Il s'exprime souvent en termes d'erreurs à corriger ou de réussite/échec sur les attendus des programmes scolaires. La gestion de la dynamique de la conduite de la classe peut masquer les difficultés conceptuelles de certains élèves que, par ailleurs, l'enseignant repère en temps différé.

Un diagnostic global sur des acquisitions en algèbre est effectué plus ponctuellement sur des devoirs plus longs. Il porte sur les objets de l'algèbre (par exemple : puissances, équations, factorisation) et non pas sur des dimensions épistémologiques d'analyse comme celui de Pépite (par exemple : dimension outil/objet, rupture arithmétique/algèbre, type de tâches). Dans les entretiens, les catégorisations se font en termes de niveaux (par exemple « moyens faibles, moyens rapides ») ou de capacités générales (par exemple « les élèves qui savent/ne savent pas travailler ou ce qu'il faut retenir »).

Ces différents travaux ont ainsi amené l'équipe à formuler plusieurs hypothèses : (H3.1) il est indispensable d'associer diagnostic et exploitation du diagnostic, en particulier en définissant des parcours d'apprentissage adaptés au diagnostic ; (H3.2) l'exploitation du diagnostic dans le cadre d'une classe nécessite d'associer un niveau de diagnostic individuel à un niveau de diagnostic collectif ; (H3.3) il est nécessaire de faciliter l'entrée dans le diagnostic de Pépite en établissant des correspondances avec les entrées diagnostiques usuelles des enseignants (en particulier le niveau de l'élève, les erreurs et les objets de l'algèbre) et les entrées épistémologiques des didacticiens (type de tâche, dialectique outil/objet, jeux de cadre et conversion de registres). Ces hypothèses nous ont amenées à formuler les questions de recherche suivantes. (Q3.1.) Est-il possible de diminuer la complexité du modèle de profil cognitif et la variété des profils d'élèves obtenus en regroupant des profils voisins pour leur associer des objectifs d'apprentissage communs ? En d'autres termes, est-

il possible de définir des stéréotypes représentant des classes d'équivalence de profils cognitifs pour la relation «bénéficiaire du même objectif d'apprentissage prioritaire»? (Q3.2) Quelles modélisations des stéréotypes et des parcours définir pour permettre leur détermination automatique par un logiciel? (Q3.3) Dans quelle mesure les classes d'équivalence de profils, que nous appelons des stéréotypes, constituent-elles un pont conceptuel entre l'analyse didactique et les entrées habituelles des enseignants pour leur permettre de lier diagnostic et actions sur les apprentissages des élèves?

#### 4.3.1. Des profils de Pépite aux stéréotypes

Pour avancer sur ces questions, trois didacticiens ont mené, de façon indépendante, une analyse des profils construits par Pépite sur le corpus de réponses recueilli. Ils ont chacun classé les élèves selon les cohérences qu'ils repéraient dans leur activité algébrique. Puis, ils ont obtenu un consensus sur leur classement, explicité leurs critères de classement et proposé un algorithme de classement (Tableau 3).

Suite à ce travail d'identification des stéréotypes par les didacticiens, un travail de conception a été mené par toute l'équipe pour présenter différemment *le profil cognitif d'un élève*. Nous avons ainsi défini différents descripteurs de la compétence algébrique, principalement le *stéréotype* et des *caractéristiques personnelles* de l'élève (Schéma 2). Pour assurer leur compatibilité avec les attentes des enseignants, le stéréotype indique un degré de compétence et les caractéristiques personnelles de l'élève sont exprimées sur chaque composante en termes de :

- *Taux de réussite* sur les exercices mettant en jeu la composante ;
- *Leviers* explicitant les acquis sur lesquels s'appuyer pour faire évoluer l'élève ;
- *Fragilités* explicitant les erreurs repérées et récapitulant les indicateurs négatifs ou plutôt négatifs ;
- Liste des *erreurs* repérées avec une référence à la situation dans laquelle elles ont été repérées (référence à l'exercice et à l'interprétation dans le contexte de l'exercice du critère d'évaluation associé à la réponse de l'élève).

Un logiciel PépiStéreo a été mis au point pour calculer les stéréotypes des élèves d'une classe et pour afficher les bilans de compétence individuels (Figure 3 en Annexe), mais aussi pour présenter la «géographie de la classe» et proposer un parcours d'apprentissage associé à chacun des stéréotypes (Vincent et al. 2005).

#### 4.3.2. Calcul des stéréotypes

Comme pour le prototype précédent, la conception du logiciel a modifié aussi bien l'algorithme de calcul que le modèle des descripteurs à produire. Classer un élève dans un stéréotype revient à lui attribuer un degré de compétence sur une échelle qui, dans Pépite, comporte trois composantes: usage de l'algèbre (composante UA), traduction d'une représentation algébrique en une autre (TA), calcul algébrique (CA). Le calcul du degré de compétence s'appuie sur une analyse

transversale des codages des réponses de l'élève au test. Ce codage des réponses est le niveau de diagnostic le plus fin mis en œuvre.

Composante	Nom	Objectif de diagnostic	Degré de compétence et commentaire
Usage de l'algèbre	UA	Disponibilité de l'outil algébrique et capacité à le mobiliser dans des situations de modélisation (production de formules ou mise en équation) et de preuve.	<p>1 : Disponibilité de l'outil algébrique et mobilisation adaptée.</p> <p>2 : Mobilisation de l'outil algébrique et traduction algébrique non adaptée</p> <p>3 : Mobilisation de l'outil algébrique sans cohérence entre le modèle et la situation</p> <p>4 : Non disponibilité de l'outil algébrique pour généraliser, prouver ou modéliser et démarches arithmétiques persistantes.</p>
Traduction d'une représentation algébrique en une autre	TA	Capacité à traduire une expression d'un registre à un autre et flexibilité de l'interprétation	<p>1 : Traduction correcte.</p> <p>2 : Traduction pas toujours adaptée.</p> <p>3 : Au moins une traduction sans cohérence entre le modèle et la situation.</p>
Calcul algébrique	CA	Capacité à calculer algébriquement.	<p>1 : nombre suffisant d'exercices traités ; interprétations appropriées des expressions.</p> <p>2 : certains exercices sont traités mais trop d'erreurs dans les calculs algébriques avec un appui insuffisant sur l'interprétation des expressions.</p> <p>3 : les calculs ne sont pas correctement menés : soit le rôle des opérateurs n'est pas maîtrisé, soit utilisation de règles de calcul fausses</p>

**Tableau 3 :** *Caractérisation des stéréotypes sur les trois composantes de l'activité algébrique.*

Puis, au second niveau, des indicateurs intermédiaires sont construits à partir du codage et un degré de compétence est attribué sur chaque composante. Ces indicateurs ont une valence positive s'ils témoignent de points forts de l'élève, et négative s'ils témoignent de points faibles. Par exemple : l'indicateur : « Justification par le numérique prééminente » est à valeur négative et est fondé sur la fréquence du critère « justification par l'exemple ». Les indicateurs intermédiaires servent à produire les caractéristiques personnelles de la compétence de l'élève ;

c'est le niveau du profil cognitif établi par PépiProf, mais reformulé pour faciliter sa compréhension par les enseignants.

Logiciels	Données	Sorties
PépiTest	Exercices indexés (Schéma 3)	Réponses de l'élève
PépiDiag (2000, 2002, 2008)	-Pour chaque exercice, une grille d'analyse multidimensionnelle : validité, utilisation des lettres, calcul algébrique, traduction d'une représentation à une autre, type de justification - Les réponses de l'élève à l'exercice	Diagnostic local à l'exercice À chaque réponse un vecteur codant la réponse sur plusieurs dimensions et une liste (éventuellement vide) de règles de réécriture (correctes ou erronées)
PépiProf (2000)	Codage des réponses à l'ensemble du test  Seuils paramétrables de réussite selon les catégories d'exercice (Schéma 3) Calcul d'indicateurs intermédiaires : Modalités de fonctionnement, Types de traitements	Diagnostic individuel global : -taux de réussite (global, questions traitées, type d'exercices, type de traitements privilégié) -description qualitative (cohérences de fonctionnement) -articulation entre représentations
PépiStéréo (2005)	Codage des réponses à l'ensemble du test Calcul de : Indicateurs intermédiaires (taux de réussite par catégories d'exercices et Agrégation de codes)	Diagnostic individuel global : Sur chacune des 3 composantes, <i>caractéristiques personnelles</i> de l'élève : taux de réussite, leviers (indicateurs positifs), faiblesses (indicateurs négatifs), liste des erreurs
	Indicateurs intermédiaires	Diagnostic collectif Stéréotype : Degré sur chaque composante de la compétence

**Tableau 4.** Différents niveaux d'abstraction pour la modélisation de l'élève dans le projet Pépite

Au troisième niveau, ces indicateurs intermédiaires sont combinés pour attribuer un degré de compétence sur chaque composante. Ce troisième niveau d'abstraction permet de situer l'élève par rapport à la classe et par rapport à la compétence de référence (Tableau 4). Il permet aussi de regrouper des élèves qui témoignent de

cohérences similaires dans leur activité algébrique. Ces groupes peuvent ensuite être utilisés pour organiser la différenciation de l'enseignement.

### 4.3.3. Discussion

Dans cette étape nous avons défini un niveau supplémentaire de modélisation permettant de diminuer la complexité de l'exploitation du diagnostic en simplifiant la description des cohérences recherchées (Q3.1). Au lieu d'une infinité de profils cognitifs possibles, 36 stéréotypes sont identifiés. Mais, dans la pratique, seuls treize stéréotypes ont été observés dans le corpus de 400 élèves. A chaque stéréotype, a été associé un parcours d'apprentissage.

Une mise à l'épreuve des stéréotypes et des parcours a été organisée dans deux classes de troisième et, en test de positionnement, à l'entrée de quatre classes de seconde auprès de 180 élèves et de 6 enseignants. La grande variété des contextes et des paramètres de la vie de la classe rend difficile d'affirmer que ces parcours sont efficaces. Cependant, ces premiers tests n'ont pas mis en évidence les réactions négatives suscitées par la complexité des écrans de PépiProfil. Les enseignants de seconde nous ont, par la suite, affirmé que les stéréotypes étaient très prédictifs de la réussite des élèves au baccalauréat trois ans plus tard. Ces mises à l'épreuve ne constituent pas des validations mais prouvent la faisabilité de l'approche et conforte sa cohérence par rapport à l'analyse didactique.

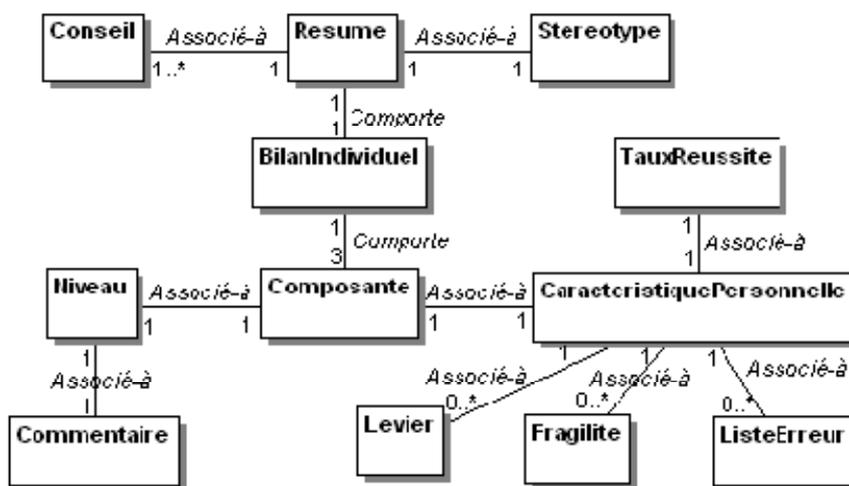


Schéma 2 : Modèle conceptuel d'un bilan individuel

Après avoir étudié les questions Q3.1 et Q3.2, reste à étudier comment les professeurs peuvent se servir de cet outil. D'une part, l'état de développement de ces prototypes ne permet pas une utilisation à large échelle et, d'autre part, nous avons observé qu'une ou deux années étaient nécessaires aux enseignants pour qu'ils

mettent à l'épreuve cette approche nouvelle, prennent confiance et se l'approprient. Un projet pour intégrer ces outils dans un environnement numérique de travail débutera en 2010.

D'un point de vue informatique, cette formulation des profils cognitifs en deux niveaux (stéréotype et caractéristiques personnelles) est assez générale et nous a amenés à définir des Design Patterns sur l'évaluation de compétence (Delozanne et al. 2007). Ainsi, du point de vue de l'élève, le stéréotype permet de situer son degré de compétence sur les trois composantes de la compétence algébrique, en dégageant des cohérences dans son activité décrites par ses caractéristiques personnelles. Du point de vue de la gestion de la classe, un stéréotype regroupe des élèves de profils cognitifs voisins et qui peuvent bénéficier du même type d'exercices.

Cependant, pour définir des parcours, les études de cas que nous avons menées (Grugeon et al. 2003), nous amènent à penser que le type d'exercice étant choisi, pour s'adapter au profil de l'élève, il est nécessaire de descendre au niveau des caractéristiques personnelles pour adapter l'énoncé et les rétroactions. C'est un des axes de travail actuel de l'équipe. Pour illustrer ce point, revenons sur l'exemple de la section 2.2 : selon le stéréotype, un type d'exercice serait proposé aux élèves (par exemple un exercice du logiciel CIME où la consigne est de faire correspondre un énoncé de problème et sa mise en équation en « bouchant des trous ») (Grugeon et al. 2003). CIME pourrait fixer les variables didactiques contraignant les énoncés sur lesquels faire travailler les élèves en fonction de leurs propres leviers et fragilités, par exemple la structure du problème, le nombre et la situation des trous. Des prototypes (CIME, AILE, EXPLOREXP<sup>9</sup>) ont ainsi été développés pour éliciter de telles variables didactiques. Mais un gros travail reste à faire pour passer des études de cas à des modèles computationnels. En particulier, la difficulté est de faire correspondre les variables didactiques et l'indexation des ressources pédagogiques disponibles pour faire travailler les élèves (de la Passardièrre et Jarraud 2005, Auzende et al. 2009). Un deuxième problème consistera à étudier si les langages de modélisation des scénarios pédagogiques sont suffisamment riches pour décrire les parcours mis au point par les didacticiens (Pernin 2006).

En parallèle avec le travail sur les stéréotypes, nous avons mis au point un outil auteur, PépiGen permettant de générer des clones de Pépite afin de permettre d'effectuer des diagnostics à différents moments de la scolarité et à d'autres niveaux scolaires. Ce travail est décrit en détail par ailleurs (Prévit 2008, Chenevotot-Quentin et al. 2008). Pour résumer, nous avons adopté une démarche ascendante pour généraliser les exercices de PépiTest en produisant un modèle de classes paramétrées d'exercices de diagnostic. Puis, en s'appuyant sur les théories de l'analyse syntaxique et sémantique des expressions algébriques (Aho et al. 1991) et la théorie des règles de réécriture (Dershowitz et al. 1990, Gélis 1994), D. Prévité a

---

<sup>9</sup> EXPLOREP est un logiciel développé par Gélis (Delozanne et al. 2005) et qui permet de travailler sur différentes représentations d'une expression algébrique (langue naturelle, écriture algébrique, écriture linéaire, écriture plane ou arborescente).

développé un logiciel de calcul formel pour générer, d'une part des expressions algébriques équivalentes pour créer les énoncés des clones et, d'autre part, pour permettre le diagnostic automatique des réponses aux exercices générés. Elle a, ensuite, mis au point une chaîne logicielle permettant la génération automatique ou assistée des clones d'exercices de Pépite. Notons que la conception de ce logiciel de calcul formel a permis d'améliorer considérablement la fiabilité du diagnostic. Pour les questions ouvertes, sur un corpus de 360 réponses au test, Pépite code actuellement entre 70 et 92 % des réponses selon les exercices. Les réponses non codées sont très majoritairement celles où l'élève utilise le langage naturel en même temps que d'expressions algébriques ce qui perturbe l'analyse ; certaines ne sont pas codées, car elles ne sont pas prédites dans l'analyse a priori. Pour les réponses codées par le logiciel, sur notre corpus, l'accord est presque total avec les experts car le diagnostic a été mis au point sur ce corpus.

#### **4.4. Vers un modèle générique de diagnostic de compétence**

Ces travaux nous ont amenées à définir une architecture d'un système de diagnostic en trois modules.

- Un *module de test* qui expose l'élève à différentes situations problèmes et trace l'activité de l'élève qui résout ces problèmes. Dans Pépite, il s'agit d'un logiciel qui pose des exercices à l'élève, mais il peut s'agir d'un simulateur ou d'un module d'enseignement en ligne. Il existe deux versions de ce module de tests téléchargeable : une version en Delphi et une version en Java. Une version destinée à être accessible sur un environnement numérique de travail est en cours de développement.
- Un *module de diagnostic* qui fonctionne en trois étapes :
  1. Premier niveau de traitement : il interprète les actions de l'élève de manière isolée en fonction de critères d'évaluation sur plusieurs dimensions. Dans le cas de Pépite, c'est ce que nous avons appelé le codage des réponses qui s'appuie sur un logiciel de calcul formel capable de résoudre les problèmes selon les stratégies correctes et erronées détectées par une analyse didactique et de comparer les solutions anticipées par cette analyse aux réponses de l'élève.
  2. Deuxième niveau de traitement : il interprète l'activité de l'élève en construisant des indicateurs de plus haut niveau par agglomération des valeurs des critères précédents (repérées par leur code). Dans le cas de Pépite, cette étape produit les caractéristiques personnelles de l'élève.
  3. Troisième niveau de traitement : il situe l'élève sur une échelle pour chaque composante de la compétence. Dans le cas de Pépite, nous classons l'élève dans un des stéréotypes.

Deux versions de ce module ont été développées : la version Delphi (Jean 2000, Prévité 2002) et une version Java développée par (Prévité 2008) qui est en cours de finalisation.

- Un module d'exploitation du diagnostic pour donner des conseils de travail ou pour étudier le bilan d'un élève particulier (formé par le stéréotype et les caractéristiques personnelles), pour étudier le profil d'une classe ou d'un groupe d'élèves. Dans notre projet, deux versions ont été développées, une par (Jean 2000) et une version développée par (Vincent et al. 2005) s'appuyant sur les retours sur la première version et sur une étude ergonomique.

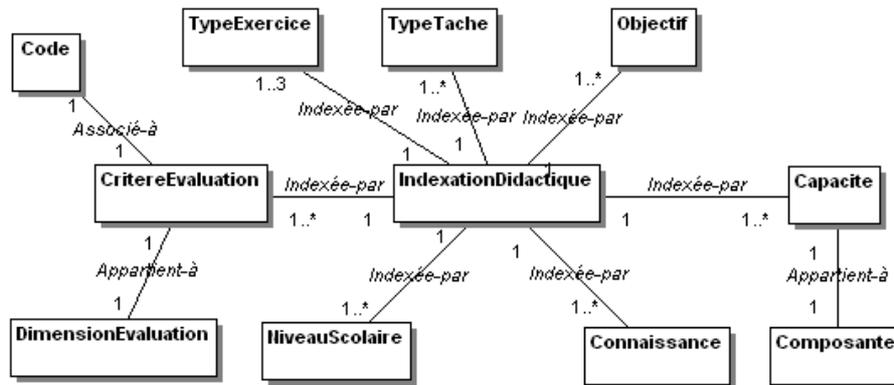


Schéma 3 : Données d'indexation didactique d'un exercice

Pour chacun de ces modules nous avons mis au point un modèle de données qui est général. Nous faisons l'hypothèse qu'il peut s'appliquer à d'autres domaines de diagnostic de compétence, ce qui n'est pas le cas des modèles de traitement des données qui sont spécifiques à l'algèbre à ce niveau scolaire. Par exemple, la détermination du stéréotype d'un élève s'appuie sur le modèle des situations de test. La description complète de ces situations figure dans (Prévité 2008) mais nous donnons ici un extrait qui concerne les données que nous avons appelées « données d'indexation didactique » (Schéma 3). Nous avons vu que la grille d'analyse des réponses met en jeu des *critères* repérés par un *code* sur plusieurs *dimensions* d'évaluation (par exemple ici, la validité de la réponse, le type de justification). Les exercices sont répartis en plusieurs *types* (ici, exercices de mathématisation, de reconnaissance et techniques) et mettent en œuvre des types de tâches (par exemple ici, produire une expression littérale, résoudre une équation). L'*objectif* est une description en langue naturelle de ce que vise à tester l'exercice. Les *connaissances* et les *capacités* sont des informations redondantes mais qui permettent de présenter les bilans (et les exercices) dans les termes des programmes officiels de l'Éducation Nationale. Les *composantes* désignent les composantes de la compétence mises en jeu par l'exercice. Le *niveau* scolaire énumère les classes pour lesquelles l'exercice

est pertinent. Un travail débute pour utiliser ce modèle pour un diagnostic de compétence en géométrie.

## 5. Résultats

Les sections qui précèdent décrivent la façon dont a progressé une recherche pluridisciplinaire en EIAH dans le cadre des projets Pépite et Lingot. L'objectif de notre recherche consiste à produire des artefacts qui rendent compte de processus d'apprentissage humain et ce, dans un domaine particulier, celui de l'algèbre dans l'enseignement secondaire. Du point de vue informatique, pour avancer dans la résolution de ce problème, la méthode scientifique que nous avons adoptée, est la production de modèles par prototypage. Parmi les nombreuses acceptations du terme de modèle, par exemple (Favre et al. 2006) en Génie Logiciel ou (Baker 2000) en EIAH, celles envisagées par Nicolle (Nicolle 2001) nous paraissent pertinentes pour décrire nos travaux car cette dernière s'intéresse aux modèles utilisés par les informaticiens dans le cadre de travaux pluridisciplinaires. Elle distingue : (i) des modèles pour raisonner formellement, (ii) des modèles pour négocier des spécifications et (iii) des modèles pour programmer. Dans le projet Lingot, nous nous sommes intéressés à la conception de ces trois types de modèles. En effet, nous avons, de façon itérative, défini plusieurs modèles qui ont été mis en œuvre dans des prototypes afin d'être testés, examinés, critiqués, raffinés et utilisés à des fins diverses : pour comprendre (modèles de type ii), pour calculer (modèle de type i) et pour développer des logiciels (modèles de type iii). De façon plus précise, nous proposons principalement cinq modèles.

Un *modèle des indicateurs* à calculer pour rendre compte de la compétence. En partant de la « grille d'analyse multidimensionnelle » de (Grugeon 1995), nous avons mis au point un modèle de compétence qui a été affiné, spécifié, précisé et systématisé afin d'être interprété automatiquement. Ce modèle, issu au départ de la recherche en didactique, a aussi été complété et modifié pour correspondre à la vision de différents acteurs et être généralisé à d'autres niveaux scolaires voire à d'autres domaines de compétence. C'était au départ un modèle de type (ii) exprimé par des tableaux et c'est sous cette forme qu'il est négocié, modifié et transformé par des humains. Une transformation assez facile permet de passer de tableaux à des fichiers XML spécifiés par des schémas. Différents prototypes l'ont mis en œuvre (modèles de type (iii)).

Un *modèle des situations de recueil des données* nécessaires au calcul des indicateurs. Un premier modèle ad hoc, mis en œuvre dans PépiTest, a été développé dans un premier temps pour montrer la faisabilité de l'approche. Par la suite, un modèle générique (en ce sens qu'il permet de générer des instances de situations) a été mis au point puis mis en œuvre dans PépiGen. Les informaticiens ont été moteurs dans la construction de ce modèle, construit par abstraction à partir d'exemples produits par les didacticiens en s'appuyant sur des langages de modélisation utilisés en informatique (UML, XML). Ici encore, la négociation sur le

modèle s'est faite sous la forme de tableaux (modèles de type (ii)) qui sont ensuite transformés en fichiers XML spécifiés par des schémas (modèles de type (i)). Le module de diagnostic et le module de génération de la grille d'analyse des exercices raisonnent formellement sur ces modèles.

Un *modèle des descripteurs* de la compétence qui présente les indicateurs pour différents acteurs. Le niveau des critères d'évaluation est un niveau trop détaillé pour prendre des décisions. Nous avons été amenés à définir des modèles de présentation à plusieurs niveaux d'abstraction : niveau critères d'évaluation (local à un exercice), niveau profil cognitif (taux de réussite, modalités de fonctionnement, articulation entre registres sémiotiques), niveau stéréotype (degré de compétence sur une échelle) et niveau caractéristiques personnelles (taux de réussite, fragilités et leviers sur plusieurs dimensions). La mise au point de ces modèles a été décrite en détail dans cet article et ceux-ci sont exprimés comme les deux modèles précédents.

Un *modèle de traitement des données* recueillies pour construire les indicateurs a aussi été conçu. Ce module est décrit dans (Prévit 2008). La construction s'est faite par généralisation à partir d'un module de diagnostic ad hoc, PépiDiag, pour proposer un modèle de diagnostic général fondé sur les théories de l'analyse syntaxique et sémantique des expressions algébriques et la théorie des règles de réécriture. Ce dernier modèle a permis la mise en œuvre du logiciel de calcul formel Pépinière. C'est donc un modèle de type (iii) mais avec une forte composante de type (i) puisque ces deux logiciels conduisent des raisonnements formels sur les expressions algébriques et sur les raisonnements des élèves.

Un dernier type de modèle est l'objet de nos recherches actuelles qui visent à modéliser *l'exploitation des indicateurs* pour associer des parcours d'apprentissage à des stéréotypes (des séquences de types d'exercices), les caractéristiques personnelles de l'élève permettant de fixer des variables didactiques (par exemple le type d'énoncé, la complexité de l'exercice ou le type d'expression) en jeu dans les exercices.

Ces modèles n'ont pas été définis indépendamment les uns des autres mais ont été élaborés en parallèle. Ils ont pour objectif à la fois de mieux comprendre l'évolution de la compétence et de concevoir des artefacts ; ce double objectif de compréhension du problème et de conception est d'ailleurs une des caractéristiques des recherches sur les systèmes interactifs (Mackay et Fayard 1997).

Nous avons déjà mentionné que les modèles de SHS<sup>10</sup> sont considérés comme des modèles descriptifs, pour comprendre, en opposition aux modèles pour agir des informaticiens. Dans le cadre de notre projet, l'équipe de didactique avec laquelle nous travaillons est préoccupée par l'action didactique et leurs modèles sont déjà orientés vers l'action ce qui facilite la coopération. La réification de ces modèles dans des artefacts a permis de produire des données qui permettent de les faire évoluer et, enfin, de partager ces modèles avec d'autres chercheurs. C'est également

---

<sup>10</sup> SHS : Sciences Humaines et Sociales

cette réification qui a permis aux informaticiens de définir des modèles plus abstraits et génériques pour étendre le domaine de validité des modélisations originelles. Ainsi, la production d'artéfacts n'est pas seulement un moyen de mettre en œuvre des modèles mais c'est aussi un moyen d'en créer (Baker 2000).

## 6. Discussion

Le problème de la validation de tels modèles est une question philosophique majeure. Wegner, qui défend un changement de paradigme de l'informatique avec le passage des machines de Turing aux machines interactives, estime que « The goals of formal models of interaction must inevitably be more modest than those for algorithmic models: proofs focus on the existence of correct behavior (type-1 correctness) rather than on the nonexistence of incorrect behavior (type-2 correctness) » (Wegner 1995). Notre objectif, modestement, consiste à mettre à l'épreuve nos modèles pour avancer sur les questions de recherche qui nous préoccupent. Selon une démarche habituelle, nous distinguons validation interne et externe.

Sur le plan interne, une première validation consiste à tester si le modèle est effectivement interprétable par des machines. Cette étape, qui permet de passer d'un « modèle contemplatif » à « un modèle productif » (Favre et al. 2006), est la première difficulté à surmonter ; nos modèles satisfont cette étape et leur effectivité est prouvée par la réalisation de prototype ce que Nicolle 1996 appelle l'expérimentation « in silico ». Nous revendiquons aussi une certaine « validité par construction » ou « développementale » (Tchounikine et al. 2004), du fait que nous nous appuyons, d'une part, sur des cadres théoriques bien établis en didactique des mathématiques et en informatique et, d'autre part, sur des évaluations permanentes en cours de conception. À l'intérieur de chaque cycle, les modèles ont été âprement discutés entre tous les participants de l'équipe (chercheurs et praticiens) et confrontés à des personnes extérieures à l'équipe, chercheurs ou praticiens. Les algorithmes ont été testés sur des corpus de réponses d'élèves et comparés avec les résultats de l'analyse des réponses par des experts. Entre chaque cycle de recherche, des mises à l'épreuve ont été organisées dans des classes ou auprès d'enseignants par nous-mêmes ou par des chercheurs extérieurs. Enfin nous avons fait appel à des spécialistes extérieurs à l'équipe (ergonome, linguiste) pour nous assister ponctuellement.

Sur le plan externe, nous envisageons deux types de validation : comparer nos travaux à ceux qui sont publiés et étudier l'usage des artéfacts produits.

La comparaison avec les travaux existants est essentiellement différentielle et consiste à mettre en évidence les points forts et les points faibles selon des critères qui dépendent des objectifs de recherche de chacun des travaux. Une étude plus complète a été réalisée dans (Prévit 2008) ; nous résumons ici deux exemples de telles comparaisons : les projets Aplusix (Nicaud et al. 2004) et Algebra Tutor

(Koedinger et al. 2006) qui travaillent sur le même domaine d'enseignement mais avec des objectifs différents. Algebra Tutor et Aplusix sont davantage développés : ce sont des produits commerciaux fiables alors que nous avons construit des prototypes de recherche. Ce sont des logiciels destinés à être utilisés par des élèves pour s'entraîner au calcul algébrique. Aplusix est plus spécialisé sur l'apprentissage du calcul algébrique et permet sur ce point un diagnostic plus fin que celui de Pépité mais sur des types de tâches plus restreints (Chachouaa et al. 2007). Ce diagnostic comporte, comme dans Pépité, deux phases : un diagnostic local à une réponse et un diagnostic global sur un ensemble d'exercices qui s'attache à définir des indicateurs de plus haut niveau. Il est exprimé en termes didactiques (théorèmes-en-actes) et vise à étudier, sur des populations très larges, les connaissances des élèves et leur évolution. (Renaudie 2003) et (Rodrigo et al. 2008) ont aussi étudié les traces d'activité d'un ensemble d'élèves pour établir des classes de profils mais leur classification porte sur des styles d'apprentissage (lents/rapides, communicants/non communicants) ou des styles d'utilisation (souris/clavier) et non sur des savoir-faire mathématiques comme les classifications en stéréotype que nous proposons. Pépité a davantage développé le diagnostic sur d'autres dimensions de la compétence algébrique (usage de l'algèbre, traduction du registre algébrique dans d'autres registres) et sur les différents niveaux de modélisation pour aborder la mise au point de parcours d'apprentissage adaptés. Dans les parcours que nous mettons au point, nous envisageons de proposer des séances de travail sur Aplusix (section 3.2, scénario 2). Algebra Tutor couvre aussi plusieurs composantes de l'algèbre. Il modélise les résolutions correctes et des déviations locales par application de règles erronées alors que Pépité traite des démarches complètes correctes ou inadéquates et ne se limite pas à détecter des règles fausses.

Un autre critère de comparaison concerne le domaine de validité des modèles proposés. Nos modèles se sont montrés efficaces pour prédire les observations sur les corpus recueillis en algèbre élémentaire en fin de collège ou début de lycée. La réutilisation possible, dans d'autres projets de recherche, des modèles mis au point et les conditions d'une telle réutilisation est une des questions que nous étudions actuellement. Au sein du projet européen DPULS, nous avons travaillé avec d'autres équipes à écrire un ensemble de Patrons de Conception qui capitalise notre expérience dans l'évaluation des apprentissages (Delozanne et al. 2007). Nous démarrons actuellement une recherche avec d'autres chercheurs pour étudier dans quelle mesure ils pourraient adapter notre modèle pour le situer par rapport aux normes et standards en cours de spécification pour les plateformes de e-learning qui, jusqu'à présent, ne prennent en compte que les taux de réussite et ainsi ne facilitent pas l'évaluation multidimensionnelle (par exemple pour caractérisation des erreurs ou des démarches). Un autre travail débute pour établir un diagnostic en géométrie.

En ce qui concerne l'usage des artefacts que nous avons construits à partir de notre travail de modélisation, nous avons mené des études préalables sur les ajustements possibles entre nos modèles et les pratiques usuelles des enseignants. Mais nous ne pouvons pas affirmer, dans l'état actuel du projet, que Pépité est validé

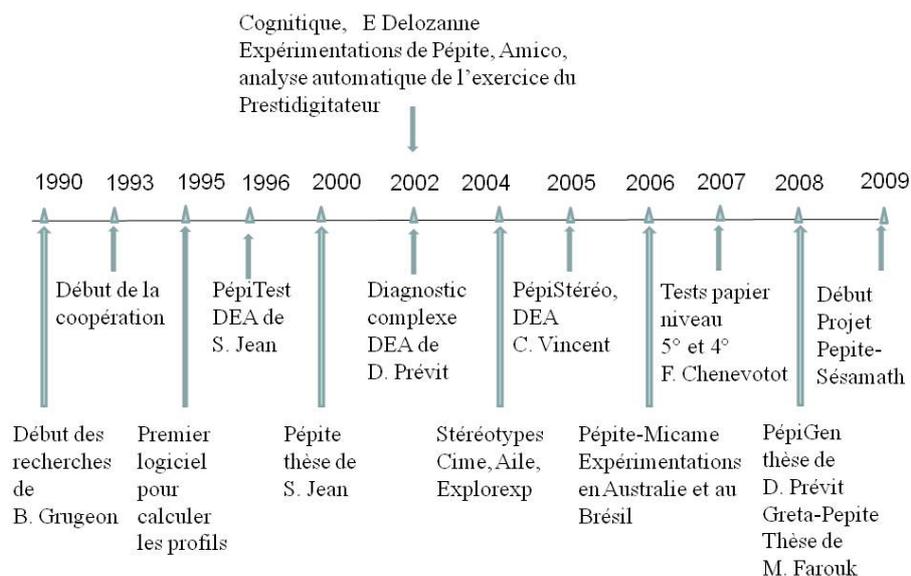
sur ce point ; c'est notre priorité actuelle, d'autant que certains facteurs de blocages possibles identifiés lors du début du projet en 1995 ont évolué depuis. Par exemple, les programmes scolaires se rapprochent du modèle de connaissances de Pépite et les approches d'enseignement par compétence se diffusent en France. Enfin, les enseignants de mathématiques utilisent de plus en plus de ressources en ligne avec leurs élèves comme en témoigne le succès du site de MathEnPoche développé par une communauté de pratique de professeurs de mathématiques. Un projet est en cours pour adapter Pépite sur cette plateforme d'enseignement en ligne très utilisée par les enseignants de mathématiques.

## 7. Conclusion

Notre démarche de recherche est fondamentalement une démarche de modélisation s'appuyant sur la création de prototypes. Dans le projet Lingot, c'est le moteur de notre recherche qui a conduit à la co-construction des modèles mis au point dans le cadre de ce projet. Notons que des phases de co-construction ont alterné avec des phases de collaboration asymétrique (Baker 2000). Le Schéma 4 résume les grandes étapes du projet. Dans cet article, nous avons montré différents cycles de prototypage sur le long terme, en mettant en évidence, à chaque étape, les questions de recherche suscitées par les étapes précédentes. Ces nouvelles questions ont permis à l'équipe du projet Lingot d'évoluer au cours du temps. A côté des informaticiens et didacticiens fondateurs du projet, elle a intégré des chercheurs d'autres disciplines lorsque nous avons besoin d'un regard extérieur qui permette de trianguler les approches (par exemple pour étudier les pratiques usuelles de diagnostic des enseignants). Enfin, la participation sur le long terme de deux enseignants et, ponctuellement, de nombreux autres, a été déterminante pour rendre opérationnels les outils mis au point mais aussi pour proposer de nouvelles pistes. Par exemple, les enseignants ont demandé une géographie cognitive de la classe ouvrant la voie à la définition des stéréotypes, ils ont réclamé des banques d'exercices pour étudier l'évolution des profils sur la durée à l'origine de PépiGen (un logiciel pour cloner les exercices), ils ont souhaité un retour vers l'élève pour le faire réfléchir sur un bilan de sa compétence, bilan à l'origine du projet PépiMicame (Farouk et al. 2007) et d'études actuelles. Les enseignants ont aussi un rôle très important dans le choix des termes pour représenter les concepts des modèles et imposer le vocabulaire métier en lieu et place des concepts didactiques ou informatiques qui semblent si naturels aux autres intervenants. Par exemple, (Grugeon 1995) se référait au concept de « conversion » d'un registre sémiotique dans un autre issu des travaux de (Duval 1992). Les enseignants consultés lui préférèrent le terme de « traduction » qui ne recouvre pas toute la subtilité du concept de conversion de Duval mais est utilisé en classe pour demander aux élèves de produire des relations algébriques exprimant une situation de la vie courante ou une représentation géométrique ou graphique. De même, le terme de « bilan de compétence » semble mieux compris que celui de « profil cognitif », encore que le terme compétence soit polysémique et source de malentendus. Bien sûr, un rôle

primordial des enseignants dans la recherche est celui de donner aux chercheurs des occasions de mettre à l'épreuve les prototypes dans des classes, ce qui nécessite que ces prototypes aient une acceptabilité suffisante pour les enseignants concernés.

Dans les phases exploratoires, les prototypes développés sont assez simples et visent à aider à construire une solution ad hoc pour préciser la formulation du problème et vérifier que les différents acteurs en ont la même compréhension. Ensuite, les techniques de modélisation du génie logiciel s'avèrent indispensables pour construire des modèles robustes et plus génériques, c'est-à-dire des systèmes capables de résoudre des classes de problèmes et non plus un problème spécifique. C'est la voie sur laquelle nous sommes engagées avec la transposition du test Pépite au niveau des classes de cinquième et quatrième (Chenevotot-Quentin et al. 2008) et avec la thèse de D. Prévit (Prévit 2008). Dans ces phases plus formelles, la co-conception tourne à la coopération où les utilisateurs qui n'ont pas de formation informatique, se limitent au rôle plus classique de testeurs. Cependant, une des raisons de la productivité de l'équipe Lingot est certainement la double compétence informatique et didactique en mathématiques (ou, au moins, en enseignement des mathématiques) de la majorité de ses membres.



**Schéma 4 :** Chronologie des projets Pépite et Lingot

Pour conclure, nous estimons que la complexité des problèmes d'enseignement et d'apprentissage, impose l'intervention de nombreuses disciplines et des projets réunissant des chercheurs sur le long terme. Notre projet, grâce à son caractère pluridisciplinaire et sa durée, s'est attaqué au difficile problème de l'évaluation multidimensionnelle de réponses à des questions ouvertes. Nous avons mis au point différents prototypes qui sont originaux dans la recherche en EIAH : principalement

un logiciel de calcul formel qui analyse les raisonnements des élèves et nous proposons un assistant qui, à terme, pourrait permettre à des enseignants de différencier plus efficacement les apprentissages dans les classes. Nous proposons plusieurs niveaux de modélisation explicite et un modèle générique d'exercices de diagnostic. Nos apports à la recherche en informatique sont de type appliqué : nous proposons une architecture logicielle et des modèles exécutable pour recueillir, analyser et exploiter des données sur la compétence en algèbre des élèves du secondaire. Enfin, c'est un travail sur le long terme dans les classes et les centres de formation qui permet aujourd'hui d'envisager la diffusion de nos résultats de recherche sur une plateforme largement utilisée par les enseignants et leurs élèves.

### Références

- (Aho et al. 1991) Aho A., Sethi R., Ullman J., *Compilateurs : Principes techniques et outils*, InterEditions, Paris, 1991
- (Auzende et al. 2009) Auzende O., Giroire H., Le Calvez F. , Using Competencies to Search for Suitable Exercises, *Proceedings of IEEE International Conference on Advance Learning Technologies (ICALT 2009)*, 2009
- (Baker 2000) Baker M., The roles of models in Artificial Intelligence and Education research: a prospective view, *International Journal of Artificial Intelligence and Education*, 11, 122-143, 2000
- (Balacheff 2000) Balacheff N., Les connaissances, pluralité de conceptions (le cas des mathématiques), in Tchounikine P. (Ed.) *Actes de la conférence Ingénierie de la connaissance (IC 2000)*, 83-90, Toulouse, 2000.
- (Beaudouin-Lafon et Mackay 2002) Beaudouin-Lafon M., Mackay W., Prototyping Tools and Techniques, in Jacko J.A. and Sears A. (Eds.), *Human Computer Interaction Handbook*, Lawrence Erlbaum Associates, 1006-1031, 2002.
- (Caroll et al. 2001) Caroll J.M. , Chin G., Rosson M. B., Neale D. C., The development of Cooperation : Five years of Participatory Design in the Virtual School, In Caroll J.M. (Ed.), *Human Computer Interaction in the new Millennium*, Addison Wesley, 373-396, 2001.
- (Chaachoua et al. 2007) Chaachoua H., Croset M. C, Bouhineau D., Bittar M. , Nicaud, Description et exploitations des traces du logiciel d'algèbre Aplusix, *Revue Sciences et Techniques de l'Information et de la Communication pour l'Éducation et la Formation* , INRP, Lyon, 14, 347-388, 2007.
- (Chenevotot-Quentin et al. 2008) Chenevotot-Quentin F., Grugeon-Allys B., Delozanne É , Diagnostic cognitif en algèbre élémentaire à différents niveaux de la scolarité, *Actes du colloque DIDIREM*, Paris, 2008.
- (Chevallard 2003) Chevallard Y., Approche anthropologique du rapport au savoir et didactique des mathématiques, in Maury S., Caillot M. (Eds.), *Rapport au savoir et didactique*, Éditions Fabert, Paris, 81-104, 2003.
- (Delozanne et al. 2007) Delozanne É., Le Calvez F., Merceron A., Labat J.-M., Design Patterns en EIAH : vers un langage de Patterns pour l'évaluation des apprenants, *Revue Sciences et Techniques de l'Information et de la Communication pour l'Éducation et la Formation*, INRP, Lyon, 14, 45-80, 2007.

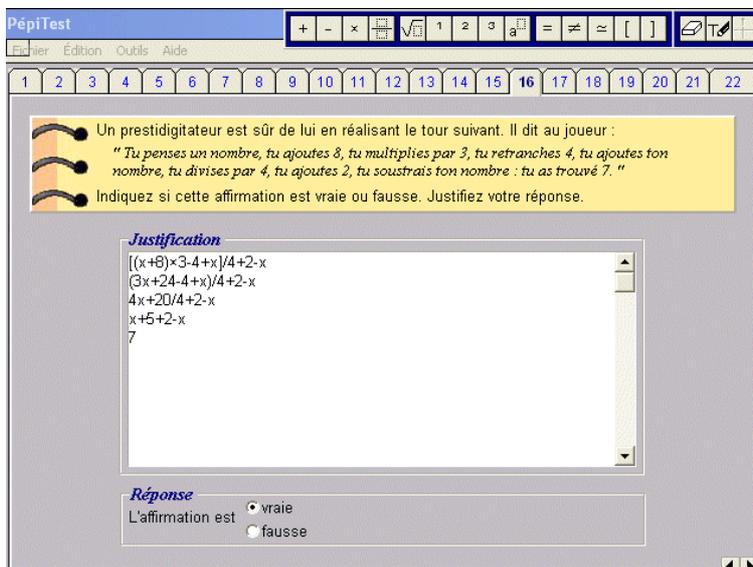
- (Delozanne 2006) Delozanne É., Interfaces en EIAH, in Grandbastien M., Labat J.-M. (Eds.), *Environnements informatiques pour l'apprentissage humain*, Collection IC2, Hermes-Lavoisier, Paris, chapitre 10, 223-248, 2006.
- (Delozanne et al. 2005) Delozanne É., Chenevotot F. (Eds.), Modélisation et mise en œuvre d'environnements informatiques pour la régulation de l'apprentissage, le cas de l'algèbre avec le projet LINGOT, *Projet Cognitique 2002, École et sciences cognitives : Les apprentissages et leurs dysfonctionnements*, rapport de fin de projet, mars 2005.
- (Delozanne et al. 2003) Delozanne É., Prévôt D., Grugeon B., Jacoboni P., Scénarios d'utilisation et conception d'un EIAH, le cas du diagnostic dans Pépité, *Actes du Colloque Intégration des Technologies à l'Enseignement des Mathématiques*, Reims, 2003.
- (Delozanne et al. 2002) Delozanne É., Grugeon B., Jacoboni P., Analyses de l'activité et IHM pour l'éducation, *Proceedings of IHM'2002, International Conference Proceedings Series, ACM*, Poitiers, France, 25-32, 2002.
- (Dershowitz et al. 1990) Dershowitz N., Jouannaud J.-P., Rewrite Systems, in J. van Leeuwen, (Ed.), *Handbook of Theoretical Computer Science*, North-Holland, volume B, ch. 6, 243-320, 1990.
- (Douady 1986) Douady R., Jeux de cadres et dialectique outil/objet, *Recherches en didactique des mathématiques*, La pensée sauvage éditions, 7(2), 5-32, 1986.
- (Duval 1992) Duval R., Sémiosis et noésis, *Actes de la conférence A.P.M.E.P.*, IREM, 1992.
- (Farouk et al. 2007) Farouk M., Réty J.-H., Delozanne É., Grugeon B., Bensimon N., Martin J.-C., Stratégies d'utilisation de la direction du regard en situation de communication interpersonnelle enseignant-élève, *Revue Sciences et Techniques de l'Information et de la Communication pour l'Éducation et la Formation*, INRP, Lyon, 14, 265-288, 2007.
- (Favre et al. 2006) Favre J.-M., Estublier J., Blay-Fornarino M., (Eds.) *L'ingénierie dirigée par les modèles, au-delà du MDA*, Hermes-Lavoisier, Paris, 2006.
- (Gélis 1994) Gélis J.-M., Éléments d'une théorie cognitive et computationnelle de l'algèbre. Application au cas de la factorisation d'expressions polynomiales, *Thèse de doctorat, Université de Nantes*, 1994.
- (Grandbastien et Labat 2006) Grandbastien M., Labat J.-M., Introduction, *Environnements informatiques pour l'apprentissage humain*, Collection IC2, M. Grandbastien, J.-M. Labat (Eds.), Hermes-Lavoisier, Paris, 6- 19, 2006.
- (Grugeon et al. 2003) Grugeon B., Coulange L., Larue V., Familles de situations d'interactions en algèbre élémentaire : deux exemples, *Actes du Colloque Intégration des Technologies à l'Enseignement des Mathématiques, ITEM 2003*, Reims, 2003.
- (Grugeon 1997) Grugeon B., Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire, *Recherches en didactique des mathématiques*, La pensée sauvage éditions, 17(2), 167-210, 1997.
- (Grugeon 1995) Grugeon B., Étude des rapports institutionnels et des rapports personnels des élèves à l'algèbre élémentaire dans la transition entre deux cycles d'enseignement : BEP et Première G, *Thèse de doctorat, Université Paris 7*, 1995.
- (Grugeon-Allys 2008) Grugeon-Allys B., Quelques apports de l'analyse multidimensionnelle des activités des élèves en mathématiques et des pratiques des professeurs de mathématiques ; vers une modélisation, *Habilitation à diriger les recherches, Université Paris-Diderot-Paris 7*, 2008.

- (Hakem et al. 2005) Hakem K., Sander E., Labat J-M., DIANE : a diagnosis system for arithmetical problem solving, *Proceedings of international Conference on Artificial Intelligence in Education (AIED2005)*, 258-265(2005).
- (Heffernan et al. 2008) Heffernan N., Koedinger K., Razzaq L., Expanding the model-tracing architecture: A 3rd generation intelligent tutor for Algebra symbolization, *International Journal of Artificial Intelligence in Education*, IO-Press Amsterdam, 18(2), 153-178, 2008.
- (Jean 2000) Jean S., PEPITE : un système d'assistance au diagnostic de compétence, *Thèse de doctorat, Université du Maine*, 2000.
- (Joosten-ten Brinke et al. 2007) Joosten-ten Brinke D., van Bruggen J., Hermans H., Burgers J., Giesbers B., Koper R. and Latour I., Modeling assessment for re-use of traditional and new types of assessment, *Computers in Human Behavior*, 23( 6), 2721-2741, 2007.
- (Kieran 1992) Kieran C., The learning and teaching of school algebra, In Douglas A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning*, New York Macmillan, 390-419, 1992.
- (Koedinger et al. 2006) Koedinger K., Corbett A. Cognitive Tutors: Technology bringing learning science to the classroom, In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*, Cambridge University Press, 61-78, 2006.
- (Lemoyne et al. 2005) Lemoyne G., Giroux J., René de Cotret S., Brouillet F., Environnement informatique pour l'enseignement du calcul réfléchi : un travail orienté par la théorie des situations didactiques, In Salin M.-H., Clanché P. et Sarrazy B. (Eds.), *Sur la théorie des situations didactiques, Questions, réponses, ouvertures. Hommage à Guy Brousseau*, La Pensée Sauvage éditions, 279-296, 2005.
- (Lenfant 1997) Lenfant A., Étude sur la transposition d'un outil de recherche destiné aux enseignants, *Mémoire de DEA de didactique des mathématiques*, Université Paris 7, 1997.
- (Mackay et Fayard 1997) Mackay W., Fayard A-L., HCI, Natural Science and Design : A Framework for Triangulation Across Disciplines, *proceedings of Designing Interactive Systems*, ACM Library, Amsterdam, 223-234, 1997.
- (Merceron et al. 2004), Merceron A., Yacef K., Mining Student Data Captured from a Web-based Tutoring Tool: Initial Exploration and Results, *Journal of Interactive Learning Research Special Issue on Computational Intelligence in Web-Based Education*, 15(4), 319-346, 2004.
- (Nicolle 1996) Nicolle A., L'expérimentation et l'intelligence artificielle , *Intellectica*, 22(1),9-19, 1996.
- (Nicolle 2001) Nicolle A., La question du symbolique en informatique, in Paugam-Moisy H., Nyckees V. and Caron-Pargue J. (Eds), *La cognition entre individu et société*, Hermès, Paris, 345-358, 2001.
- (Nicaud et al. 2004) Nicaud J.F., Bouhineau D., Chaachoua H., Mixing microworld and CAS features in building computer systems that help students learn algebra, *International Journal of Computers for Mathematical Learning*, Springer, 9(2), 169-211, 2004.
- (de la Passardière et Jarraud 2005) De la Passardière B., Jarraud P. « ManUeL, LOM et l'indexation de ressources scientifiques, Vers les bonnes pratiques pour l'Université en ligne », *Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH 2005)*, Montpellier, 57-68, 2005.

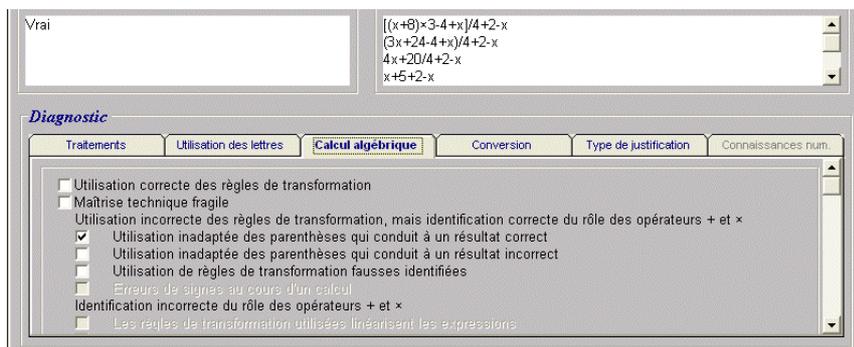
- (Pernin 2006) Pernin J.-P. , Normes et standards pour la conception, la production et l'exploitation des EIAH, in Grandbastien M., Labat J.-M. (Eds.), *Environnements Informatiques pour l'Apprentissage Humain*, Collection IC2, Hermes-Lavoisier, Paris, 201-222, 2006.
- (Prévit 2008) Prévit D., Génération d'exercices et analyse multicritère automatique de réponses ouvertes, *Thèse de doctorat, Université du Maine*, 2008.
- (Prévit et al. 2004) Prévit D., Delozanne É., Grugeon B., Modélisation cognitive en algèbre élémentaire : une conception itérative, *Actes de la Conférence Technologies de l'Information et de la Communication pour l'Enseignement Supérieur*, Compiègne, 138-145, 2004.
- (Prévit 2002) Prévit D., Vers un diagnostic de compétence inspectable par différents types d'utilisateurs, *Mémoire de DEA, Université du Maine*, septembre 2002.
- (Py et Hibou 2006) Py D., Hibou M., Représentation des connaissances de l'apprenant, in Grandbastien M., Labat J.-M. (Eds.), *Environnements Informatiques pour l'Apprentissage Humain*, Collection IC2, Hermes-Lavoisier, Paris, 97-116. 2006.
- (Renaudie 2003) Renaudie D., Application de techniques de classification non supervisée à l'étude de comportements d'élève en algèbre, *Actes de la conférence RJCIA'2003, plate-forme AFIA*, Laval, 181-194, 2003.
- (Rodrigo et al. 2008) Rodrigo M., Anglo E., Sugay J., Baker R., Use of Unsupervised Clustering to Characterize Learner Behaviors and Affective States while Using an Intelligent Tutoring System, *Proceedings of International Conference on Computers in Education*, 49-56, 2008.
- (Rogalski 2005) Rogalski J., Rapport d'activité sur l'axe instrumentation de l'activité des enseignants, Modélisation et mise en œuvre d'environnements informatiques pour la régulation de l'apprentissage, le cas de l'algèbre avec le projet LINGOT, *Projet Cognitique 2002, École et sciences cognitives: Les apprentissages et leurs dysfonctionnements*, rapport de fin de projet, 91-107, mars 2005.
- (Sfard 1991) Sfard A., On the Dual nature of mathematics conceptions: reflections on processes and objects as different side of the same coin, *Educational Studies in Mathematics*, 22, 1-36, 1991.
- (Shapiro 2005) Shapiro J.A., An Algebra Subsystem for Diagnosing Students' Input in a Physics Tutoring System, *International Journal of Artificial Intelligence in Education*, 15, 205-228, 2005.
- (Shute et al. 2008) Shute V., Hansen E., Almond R., You Can't Fatten A Hog by Weighing It - Or Can You? Evaluating an Assessment for Learning System Called ACED. , *International Journal of Artificial Intelligence in Education*, IO-Press Amsterdam, 18 (4), 289-316, 2008.
- (Tchounikine et al. 2004) Tchounikine P. (Ed.), « Platon-1 : quelques dimensions pour l'analyse des travaux de recherche en conception d'EIAH », *Rapport de l'Action Spécifique 51 du RTP39 Fondements théoriques et méthodologiques de la conception des EIAH* , Département STIC du CNRS, 2004.
- (VanLehn et al. 2005) VanLehn K., Lynch C., Schulze K., Shapiro J.A., Shelby R., Taylor L., Treacy D., Weinstein A., Wintersgill, M. The Andes Physics Tutoring System: Lessons Learned, *International Journal of Artificial Intelligence in Education*, IO-Press Amsterdam, 15, 147-204, 2005.

- (Vergnaud 1990) Vergnaud G., La théorie des champs conceptuels, *Revue de Didactique des Mathématiques*, La pensée sauvage éditions, 10 (2), 133-170, 1990.
- (Vincent et al. 2005) Vincent C., Delozanne É., Grugeon B., Gélis J.-M., Rogalski J., Coulange L., Des erreurs aux stéréotypes : Des modèles cognitifs de différents niveaux dans le projet Pépite, *Actes de la Conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH2005)*, Montpellier, 297-308, 2005.
- (Wegner 1995), Wegner P., Interaction as a Basis for Empirical Computer Science, *ACM Computing Survey*, 27 (1), 45-48, March 1995

**Annexe**



**Figure 1.** Réponses de Laurent à un exercice de PépiTest



**Figure 2.** PépiProf, vérification par l'enseignant du codage des réponses

- 2 -		Date du test : 27/02/2006	Classe : 3eme5
		Questions traitées : 48 %	 Imprimer ce profil
		Taux de réussite aux questions traitées : 58 %	
<a href="#">Résultats détaillés de l'élève</a> <a href="#">Accès au groupe de l'élève</a>		<a href="#">En savoir plus sur les stéréotypes</a> <a href="#">Type d'activités adaptées</a>	
Stéréotype et commentaires	Caractéristiques personnelles		
<p><b>Usage de l'algèbre Niveau 3</b></p> <p>L'élève ne résout pas assez d'exercices avec la démarche algébrique. Ses justifications sont mal assurées :</p> <ul style="list-style-type: none"> <li>- soit elles sont trop rarement correctes</li> <li>- soit elles comportent des arguments non algébriques ou incomplets.</li> </ul> <p><b>UA3</b></p>	<p><b>Exercices de mathématisation :</b> Taux de réussite : 36 %</p> <ul style="list-style-type: none"> <li>● Leviers</li> <li>● Début d'utilisation de l'algèbre pour prouver</li> <li>● Fragilités</li> </ul> <p>Justification de type scolaire prééminente et l'outil algébrique n'est pas bien maîtrisé</p> <ul style="list-style-type: none"> <li>● En particulier...</li> </ul> <p><b>Justification donnée par un exemple numérique</b></p> <ul style="list-style-type: none"> <li>✓ Preuve donnée par un exemple numérique à l'exercice 4e</li> <li>✓ Justification avec exemple numérique à l'exercice 16</li> </ul> <p><b>Justification de type scolaire reposant sur l'application de règles incorrectes</b></p> <ul style="list-style-type: none"> <li>✓ Exercice 4c</li> <li>✓ Exercice 4d</li> </ul> <p><b>Justification par des propriétés énoncées en français</b></p> <ul style="list-style-type: none"> <li>✓ Exercice 2b</li> <li>✓ Exercice 2c</li> <li>✓ Exercice 4a</li> </ul> <p><b>Justification s'appuyant sur des formulations d'ordre légal</b></p> <ul style="list-style-type: none"> <li>✓ Preuve avec utilisation de règles au niveau de la forme de l'écriture à l'exercice 2a</li> </ul>		
<p><b>Traduction Niveau 2</b></p> <p>Les compétences de l'élève pour passer d'une représentation à une autre sont encore fragiles</p> <ul style="list-style-type: none"> <li>- soit ses erreurs pour exprimer algébriquement des relations entre les variables en jeu concernent plus de la moitié des exercices traités</li> <li>- soit ses réussites dans les traductions sont inférieures aux erreurs et absences de réponse.</li> </ul> <p><b>T2</b></p>	<p><b>Exercices de reconnaissance :</b> Taux de réussite : 27 %</p> <ul style="list-style-type: none"> <li>● Leviers</li> <li>● Aucune traduction abrégative</li> <li>● Fragilités</li> </ul> <p>Trop d'erreurs dans les traductions et/ou nombre de réponse insuffisant</p> <ul style="list-style-type: none"> <li>● En particulier...</li> </ul> <p><b>Traduction correcte non attendue</b></p> <ul style="list-style-type: none"> <li>✓ Résolution et calculs par morceaux à l'exercice 16</li> </ul> <p><b>Traduction incorrecte</b></p> <ul style="list-style-type: none"> <li>✓ Expression non parenthésée ou confusion aire - périmètre à l'exercice 3</li> <li>✓ Confusion produit - somme à l'exercice 5a</li> <li>✓ Erreur de reconnaissance produit - somme à l'exercice 5b</li> <li>✓ Confusion inverse-opposé à l'exercice 11p1c</li> </ul>		
<p><b>Calcul algébrique Niveau 3</b></p> <p>L'élève ne réussit pas à mener des calculs :</p> <ul style="list-style-type: none"> <li>- soit il ne maîtrise pas le rôle des opérateurs</li> <li>- soit il s'est construit des règles de calcul fausses</li> </ul> <p><b>CA3</b></p>	<p><b>Exercices techniques :</b> Taux de réussite : 10 %</p> <ul style="list-style-type: none"> <li>● Fragilités</li> </ul> <p>Rôle des opérateurs non maîtrisé</p> <ul style="list-style-type: none"> <li>● En particulier...</li> </ul> <p><b>Utilisation inadaptée des parenthèses mais qui conduit toutefois à un résultat correct</b></p> <ul style="list-style-type: none"> <li>✓ Confusion entre <math>-3^2</math> et <math>(-3)^2</math> à l'exercice 1</li> <li>✓ Expression non parenthésée. <math>a+3(a+b)</math> est différent de <math>(a+3)(a+b)</math> à l'exercice 3p2</li> </ul> <p><b>Utilisation inadaptée des parenthèses et qui conduit à un résultat incorrect</b></p> <ul style="list-style-type: none"> <li>✓ <math>3+5a=8a</math> à l'exercice 2</li> </ul> <p><b>Utilisation de règles de transformation fausses identifiées</b></p> <ul style="list-style-type: none"> <li>✓ Développement et résolution équation faux à l'exercice 9d</li> </ul> <p><b>Identification incorrecte de x et + : Assemblage des termes</b></p> <ul style="list-style-type: none"> <li>✓ Règles fausses <math>a^n+a^p=a^{n+p}</math> avec <math>4a^3+3a^2=7a^5</math> à l'exercice 4b</li> </ul>		

Figure 3. Copie d'écran d'un bilan individuel produit par PépiStéréo (Vincent et al. 2005)