UN ESSAI DE MODELISATION INFORMATIQUE DE CONCEPTIONS D'ELEVES EN ALGEBRE

David Renaudie
Laboratoire Leibniz-IMAG, Equipe Apprentissage
46, av Félix Viallet, 38031 Grenoble Cedex
France
email:david.renaudie@imag.fr

Walter Olivera-Silvera
aboratoire Leibniz-IMAG, Equipe Apprentissage
46, av Félix Viallet, 38031 Grenoble Cedex
France
email:walter.olivera-silvera@imag.fr

Résumé :

Nous nous intéressons à l'analyse de comportements d'élèves du secondaire qui résolvent des exercices d'algèbre, les données sont recueillies lors de l'utilisation d'un EIAH (Environnement Informatique pour l'Apprentissage Humain) pour l'algèbre, le logiciel Aplusix. Toutes les actions des élèves ont été enregistrées et constituent notre corpus de données initial. En nous basant sur les résultats d'un interpréteur automatique de ces séquences d'actions (le module Anais), et en nous inspirant de la théorie cK¢ de modélisation des conceptions, nous obtenons une description des productions en termes de conceptions de synthèse. Ces dernières, comparées avec des analyses manuelles effectuées par des didacticiens, peuvent être assimilées à un modèle épistémique conceptuel de l'élève.

Mots-clés :

représentation de l'information, conceptions, EIAH, cK¢, apprentissage automatique

1 – INTRODUCTION

L'analyse de comportements d'élèves, étape centrale de toute étude didactique, est souvent manuelle et individuelle. Il en ressort une description «détaillée» des phénomènes mais au prix d'un temps considérable. Ainsi, les outils statistiques (Bronner et al, 2003) apportent une aide précieuse au didacticien en synthétisant le résultat des observations.

De telles automatisations ont été utilisées par le passé dans la modélisation d'élèves. C'est le cas par exemple de l'utilisation de l'apprentissage machine. Ainsi, Sison et Shimura (1998) dresse un éventail des techniques employées et des systèmes développés.

Notre approche s'inscrit dans un projet qui a pour but de concevoir un tuteur artificiel. Un de ses axes de recherche est de chercher à faire émerger de façon automatique des conceptions d'élèves en algèbre, ce qui pose la question de la représentation des connaissances dans un EIAH.

Plusieurs approches se sont succédées qui ont enrichi le savoir faire dans ce domaine. Balacheff (2001) en fait un recensement en trois catégories : d'abord des projets tels que PIXIE [Sleeman (1982)], que Nicaud et Vivet (1988) ont relevé et traduit par « expertise partielle », approche consistant à considérer la connaissance de l'élève comme un sous-ensemble de celle d'un expert. Puis il y eut les projets tels que REPAIR de Brown et VanLehn (Brown et VanLehn 1980) et BUGGY de Brown et Burton (cf Wenger 1987) basés sur le paradigme des « Misconceptions » qui attribue à l'élève une variante fausse d'une connaissance globale et juste. Enfin avec l'introduction par Kayser (1997) de la « représentation des connaissances dans un modèle », la modélisation tente de décrire les croyances de l'élève mais sans chercher à rendre compte d'une réalité mentale, c'est-à-dire un modèle cognitif. Pour répondre à ce dernier point, Balacheff (2001) propose une modélisation basée sur le lien connaissance-problème établi par Vergnaud (1991) : le modèle cK¢ (Soury-Lavergne 2003). Nous rappelons ici, très brièvement, la définition de conception dans ce modèle.

Une conception C est un quadruplet (P, R, L, Σ) avec :

- **P = ensemble de problèmes**.

P doit permettre la description de la sphère de pratique dans laquelle C est opératoire. Une condition minimale pour qu'un problème appartienne à P est qu'il existe une suite de transformations par des éléments de R qui conduise à un problème résolu au sens de C. P est une description ou énumération de problèmes concernés par la conception.

- **R = ensemble d'opérateurs**

Un opérateur (noté O) transforme un problème en un nouveau problème.

- **L = langage de représentation**

Il s'agit du langage ou formalisme dans lequel sont décrits P, R et Σ

- **Σ = structure de contrôle**

C'est un mécanisme permettant de décider si oui ou non, le problème est résolu au sens de C. Σ contient aussi des éléments de stratégie guidant les choix de l'élève.

Le contexte de travail : le projet « Cognitique-Algèbre », Aplusix et Anais

Notre travail se déroule dans le cadre du projet « Modélisation cognitive de comportements d'élèves en algèbre », une ACI financée par le ministère de la recherche. Il repose sur l'utilisation d'un EIAH en algèbre, le système Aplusix, dont la présentation détaillée peut être trouvée dans (Bouhineau, 2001). Ce logiciel permet à l'élève de manipuler facilement des expressions mathématiques, tout en « découpant » son travail en plusieurs étapes de calcul (il y en a 3 dans la Figure 1).

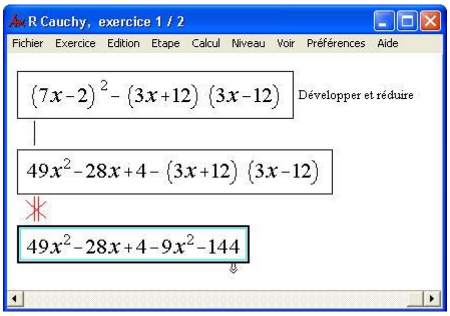


Figure 1 : le logiciel Aplusix

Chaque prestation d'élève utilisant Aplusix est systématiquement enregistrée dans un fichier recueillant la séquence des actions effectuées (frappes de touches, clics souris …).

Dans le cadre du projet, un module a été développé dans le but d'analyser les productions d'élèves : le module Anais. Il dispose d'une base de règles de réécriture (correctes et erronées), qu'il utilise pour proposer une explication, sous forme d'un enchaînement d'une ou plusieurs de ces règles, des transitions dans la résolution d'exercices par l'élève. Ces règles sont décrites elles mêmes par des traits symboliques, et sont le fruit d'un travail d'experts du domaine. Ce module produit donc une modélisation de l'élève, que l'on peut voir comme un modèle épistémique procédural. On peut l'assimiler à un agent qui comme d'autres par le passé (tels que BUGGY) reproduit le même comportement que l'élève mais par des procédures internes qui lui sont propres. Comme ceux-ci, il ne permet pas de déduire immédiatement les capacités cognitives de l'élève : la description des processus par le module Anais n'est pas forcément le reflet de ceux mis en œuvre par l'élève (auxquels nous n'avons pas accès), mais seulement une explication possible.

Notre démarche a pour but d'automatiser la recherche des conceptions d'élèves, à partir des données synthétisées par le module Anais. Celles-ci constituent justement le Système d'Information Complexe que nous nous proposons d'étudier en y recherchant des régularités. Nous souhaitons franchir une étape par rapport à ce qui a été fait auparavant en proposant une modélisation des compétences cognitives. Pour cela, nous proposons d'appliquer le formalisme cK¢ aux productions de l'agent artificiel (Anais), dans l'optique de le caractériser en termes de conceptions artificielles.

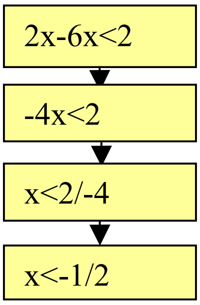
L'idée est de s'appuyer sur un raisonnement par analogie qui permette d'effectuer un mapping de la caractérisation de l'agent artificiel vers l'élève réel dont il est issu. En effet, les descripteurs perceptifs et mathématiques qu'utilise le module Anais sont d'une granularité suffisamment fine pour émettre l'hypothèse que, s'il y a des régularités dans les caractérisations des agents artificiels, nous puissions en déduire par analogie qu'il y a les mêmes régularités chez les élèves réels.

2 – MODELE INFORMATIQUE

2.1 – Quelques définitions

On appelle *expression algébrique* une expression mathématique bien formée. Dans toute la suite, nous ne manipulons que de telles expressions algébriques. Exemple : E= « -4x<2 »

On appelle *transformation* le passage d'une expression algébrique à une autre. On appelle *source* l'expression initiale de la transformation et *résultat* l'expression finale de la transformation.



Notation : T = « $E_{source} \rightarrow E_{resultat}$ »

Exemple : T=« -4x<2 → x<2/-4 »

Cette transformation est fausse à cause de l'absence de changement du sens de l'inégalité. Une transformation n'est nullement tenue d'être mathématiquement juste.

On appelle *production* une succession de transformations T_i, i=1..m.

Notation : $P=[T_1,T_2,..T_m]$

Ex : $P_1=[T_1,T_2,T_3]$= « 2x-6x<2→-4x<2 → x<2/-4→x<-1/2 »

L'entité ayant généré cette production peut être aussi bien humaine (on pense bien sûr à l'élève, ou au professeur, ou à tout autre expérimentateur) qu'une machine, un programme. Nous parlerons donc d'*agent* pour désigner indifféremment l'élève ou la machine (i.e. Anais) en tant que générateur de production.

2.2 – Anais et l'analogie avec l'élève réel

Supposons maintenant qu'un agent A_1 fournisse une production P_1 sans explicitation (i.e. sans préciser les opérandes et les choix inhérents à chaque transformation), et qu'un agent A_2 fournit une production P_2 avec explicitation, telle que $P_1=P_2$ ou P_1 « peu différente » de P_2. Notre hypothèse de raisonnement par analogie nous permet d'attribuer à l'auteur de P_1 les conclusions tirées de l'analyse de P_2. Ainsi, vu que chaque production d'Anais est calculée à partir de la production d'un élève réel, nous proposons de caractériser l'élève réel à l'aide des conceptions artificielles découvertes grâce aux traitements décrits dans les parties suivantes.

2.3 – Décomposition d'une production dans notre interprétation de la théorie cK¢

L'objet de cette partie consiste à proposer un lien entre la théorie cK¢ (le quadruplet P,L,R,Σ) et les productions précédemment définies, afin de présenter par la suite notre modèle informatique.

2.3.1. Problème

Il y a Problème (au sens cK¢) dès lors que l'agent est confronté à une difficulté qui l'oblige à mettre en œuvre un raisonnement et des outils afin d'arriver à une solution : cela signifie que la seule donnée de l'expression source ne suffit pas à déterminer entièrement le Problème. Il ne peut donc être déduit qu'a posteriori, en observant l'expression résultat.

Nous considèrerons donc que chaque transformation T_i d'une production est l'expression d'un problème P_i pour l'agent.

2.3.2. Opérateur

L'Opérateur est, par définition dans cK¢, ce qui permet de transformer un problème en un autre problème. Nous considèrerons que l'opérateur O_i assure le passage de P_i à P_{i+1}.

2.3.3. Contrôle

Le contrôle Σ est une validation par l'agent que le problème a été résolu par application de l'opérateur. En première approximation, nous considèrerons qu'il y a eu validation dès lors que la transformation a été avalisée par l'agent.

2.3.4. Langage

Le Langage L est précisément le formalisme que nous proposons dans ce document.

La Figure 2 donne un exemple d'une telle décomposition. Notons qu'il n'est pas encore question de parler de conception à ce stade. En effet, P_i et O_i ne sont que des manifestations (des instances particulières) des conceptions sous-jacentes que nous cherchons à formaliser.

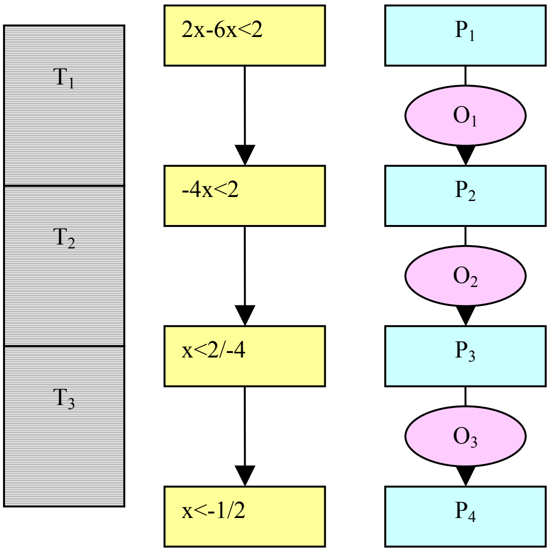


Figure 2 : exemple de décomposition de production

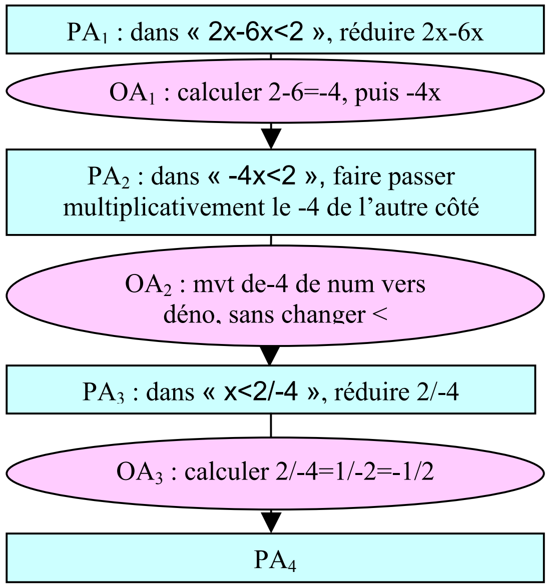


Figure 3 : exemple d'explicitation de production par un agent

3.3 – Modèle informatique

La seule donnée des expressions constituant chaque transformation est insuffisante pour connaître P et O ; tout au plus pouvons-nous prêter à l'agent des intentions, mais en aucun cas nous ne sommes capables de savoir exactement quelles raisons l'ont mené à agir comme il l'a fait. Cela constitue une sérieuse limite si on veut construire un modèle épistémique de l'agent : cela suppose un accès au moins partiel à une explication des mécanismes mis en jeu lors de la genèse de chaque transformation. Supposons maintenant que l'agent explicite lui-même chacune des transformations qu'il a produite : à partir de ce renseignement supplémentaire, il devient possible d'exprimer P et O dans un formalisme que nous allons détailler. Le module Anais effectue cette explicitation pour chaque production. Nous en déduisons des *Problèmes A* et des *Opérateurs A* (A pour « *Artificiel » ou « Anais »*), comme dans l'exemple montré en Figure 3.

Description de CPA et COA

Chaque Problème Artificiel, tel qu'il est décrit, n'est qu'une instance d'une classe de problèmes plus générale, que nous noterons CPA (pour *Classe de Problème Artificiel*). De même pour les opérateurs : il nous faut abstraire ces entités afin de nous placer à un niveau de généralité suffisant pour détecter des régularités comportementales chez notre agent.

Anais fournit des données sous forme de vecteurs de traits pour chaque transformation. Parmi ces traits, nous avons choisi de sélectionner un mélange d'informations d'ordre mathématique (par ex. le type d'équation ou d'argument) et perceptives (par ex. la position de l'argument) dans le but de pouvoir caractériser les circonstances pouvant être à l'origine de différences de comportements. On fait l'hypothèse que la finesse de ces traits permet de modéliser les contextes de mobilisation de conceptions.

Voici les traits (avec exemples) décrivant une CPA et une COA (resp. tableau de gauche et de droite):

Trait	Exemple : CPA_2
Type d'équation	Inégalité
Mouvement à faire	Multiplicatif
Position de l'argument	Gauche
Signe de l'argument	Négatif
Type de l'argument	Numérique

Trait	Exemple : COA2
Modif du type de relation	ChangePasType
Mouvement fait	Multiplicatif
Orientation verticale	NumérateurVersDénom
Changement de signe	ChangePasSigneArg
Changement de sens de l'inégalité	ChangePasSens

Constitution du jeu de données

Le corpus analysé dans ce travail est constitué par l'ensemble des transformations de mouvement (les seules pour lesquelles les traits fournis par Anais sont à ce jour assez développés) effectuées par un agent, toutes productions confondues. Nous disposons donc de n couples (CPA_i, COA_i), i=1..n, pour un agent donné.

Filtrage des instabilités

Nous traitons ici les cas où l'agent a réagi de manière différente face à un même problème, en regroupant les différents CPA pour un même COA et en observant leur répartition.

Nous faisons l'hypothèse de déterminisme suivante : parmi la gamme de réponses possibles, il y en a une privilégiée, que nous considèrerons être la plus représentative de ses connaissances. Ainsi, nous ne conservons pour la suite des traitements, pour chaque CPA, que la COA majoritaire (constatée dans plus de 50% des cas), et éliminons ainsi le bruit dû à des réponses occasionnelles. Une prise en compte plus fine de ces fréquences de réponses permettrait d'améliorer notre modèle, nous comptons développer cet aspect de notre approche.

Agrégation en Conceptions Artificielles

Nous avons mis au point un mécanisme d'agrégation de couples (CPA, COA), qui consiste à partitionner les n' couples restant après filtrage par paquets de COA égaux. Pour chaque COA, nous agrégeons les CPA ainsi regroupés en une CPA plus générale, en calculant leur intersection ensembliste. Exemple : pour COA_2 ci-dessus, supposons avoir le groupe constitué de CPA_2 ci-dessus, ainsi que de la CPA valant : [Inégalité; Multiplicatif;Gauche;Positif;Numérique]

Alors l'agrégation de ces deux CPA donne : [<vide>;Multiplicatif;Gauche;<vide>;Numérique]

ce qui se lit « la classe de problème artificiel où il s'agit de faire un mouvement multiplicatif d'un argument numérique situé à gauche », sous-entendant « cet argument est indifféremment positif ou négatif, et cela dans une égalité ou une inégalité ». Chaque agrégat (CPAG, COA) ainsi constitué constitue ce que nous appelons une *conception artificielle*.

4 – RESULTATS

Nous avons étudié une classe de 34 élèves, pour laquelle nous disposons de 1213 transformations de mouvement, soit une moyenne de 35,6 transformations/élève. Ce chiffre est satisfaisant car il témoigne d'un bon niveau de couverture des différents problèmes, et cela limite les généralisations basées sur un trop faible nombre de cas rencontrés. Notons que 90.3% de ces transformations sont mathématiquement correctes, ce qui témoigne d'un bon niveau de compétences algébriques.

A partir de ces données, nous obtenons en réitérant sur chaque élève le processus décrit en 3.3, 23 Classes de Problèmes Artificiels, 13 Classes d'Opérateurs Artificiels, 61 couples (CPA,COA) et 27 Conceptions Artificielles. Ces chiffres attestent que notre formalisme est suffisamment général, car seulement 23 CPA et 13 COA permettent de rendre compte de 1213 transformations ; mais assez fin pour différencier un nombre raisonnable de cas de traitements. De plus, alors que chaque élève est décrit (en moyenne) par 35,6 couples (CPA, COA), il est caractérisé par un faible nombre de Conceptions Artificielles), ce qui concorde avec les analyses manuelles effectuées par les didacticiens.

Voici quelques exemples de Conceptions Artificielles, sur un cas particulier : nous avons un élève caractérisé par C1, et C6. C1 correspond selon les didacticiens à une conception sur les mouvements additifs, correcte d'un point de vue mathématique, et très générale car s'appliquant à tout type/signe d'argument, ainsi qu'aux equations et aux inéquations. Exemples : « $2x+3=0 \rightarrow 2x=-3$ », « $4>-5x \rightarrow 4+5x>0$». C6 est définie par : (SCPG2, SCR2) tels que décrits plus haut. On peut y associer l'interprétation « ne jamais changer le sens de l'inégalité lors de mouvements multiplicatifs d'arguments situés à gauche. ». Exemples : « $-4x<2 \rightarrow x<2/-4$ », « $5x<1 \rightarrow x<1/5$ »

Nous retrouvons ici exactement l'analyse manuelle du didacticien, qui, ayant travaillé sur les mêmes données que nous, déclare que cet élève applique au domaine des inéquations sa conception du mouvement multiplicatif jusqu'ici appliquée (avec succès) aux équations. Cela corrobore que notre hypothèse de raisonnement par analogie est opérationnelle, et conforte notre méthodologie consistant à se baser sur les productions d'Anais en tant que reflet des productions d'élève, notamment en ce qui concerne l'identification des erreurs d'élèves.

D'autres conceptions artificielles erronées ont été relevées, comme des conceptions artificielles de mouvement additif sans changement de signe de l'argument, ainsi que d'autres plus exotiques et plus rares. Elles constituent autant d'éléments d'étude intéressants pour le didacticien.

5 – CONCLUSIONS ET PERSPECTIVES

Notre interprétation de la théorie cK¢ d'un point de vue informatique nous a permis de faire émerger des conceptions artificielles de mouvement dans le cas particulier de productions en algèbre. Le raisonnement par analogie nous permet, suite aux traitements effectués à partir de l'explicitation de productions par Anais, de proposer une caractérisation des élèves en termes de « conceptions artificielles d'élèves» validée par le didacticien.

Nous pensons que cette direction de recherche est prometteuse, car nous produisons automatiquement, avec un modèle encore améliorable, des analyses de haut niveau des compétences algébriques des élèves. L'objectif est de pouvoir se servir de ces conclusions pour proposer un enseignement adapté aux conceptions qui auront été automatiquement diagnostiquées chez l'élève, cela en vue de faire évoluer ses connaissances lors d'une phase ultérieure de remédiation (actuellement en cours de développement dans le projet). Par exemple, un élève chez qui on aura diagnostiqué C6 décrite précédemment, pourra se voir proposer une série d'exercices sur des inéquations impliquant des mouvements d'arguments positifs et négatifs, accompagnés d'un tutorat sur l'influence du signe de l'argument sur le changement de sens de l'inégalité.

Le modèle informatique proposé ici est grandement améliorable. D'autres mécanismes d'agrégation sont envisageables. De plus, il faudrait raffiner notre mécanisme de filtrage des instabilités, car la loi de la majorité est peu adaptée aux cas de quasi-égalité (par exemple 51%/49%, le 51% sera choisi mais n'est pas « vraiment » très majoritaire).

Par ailleurs, la description d'élèves en termes de conceptions artificielles reste encore trop « bas niveau » pour les besoins du didacticien. En effet, pour décrire plus succintement l'élève, il faudrait trouver une description de plus haut niveau qui « résumerait » ces conceptions artificielles et faciliterait la remédiation ultérieure.

Références bibliographiques

Balacheff N. (2000). Les connaissances, pluralité de conceptions (le cas des mathématiques). In: Tchounikine P. (ed.) *Actes de la conférence Ingénierie de la connaissance* (IC 2000, pp.83-90). Toulouse, 10-12 mai.

Bouhineau, D., Nicaud, J.F., Pavard, X., Sander, E. (2001), Un micromonde pour aider les élèves à apprendre l'algèbre. *6e journées francophones Environnements Interactifs d'Apprentissage avec Ordinateur*. EAIO'2001

Bronner A., Bouhineau D., Chaachoua H., Huguet T. (2003). Analyse didactique de protocoles obtenus dans un EIAH en algèbre. In Desmoulins, C., Marquet, P. & Bouhineau, D. *EIAH2003 Environnements Informatiques pour l'Apprentissage Humain*. Actes de la conférence EIAH 2003, Strasbourg. Paris : INRP, 79-90

Brown, J.S. & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.

Kayser D. (1997) La représentation des connaissances. *Paris : Hermès*

Nicaud J.-F., Vivet M. (1988) Les Tuteurs intelligents : réalisations et tendances de recherche. *Revue techniques et science informatiques7*

Sison, R., Shimura, M. (1998). Student modeling and machine learning (1998), *International Journal of Artificial Intelligence in Education, 9,pp 128-158* .

Sleeman D. H., (l982). Inferring (mal) rules from pupil's protocols. In the *Proceedings of the l982 European AI Conference,* pp 160-164.

Soury-Lavergne S. (2003) : Baghera Assessment Project, designing an hybrid and emergent educational society, *Cahiers Leibniz n°81*, Avril 2003

Vergnaud G. (1991) : « La théorie des champs conceptuels », Vol 10/2.3, *La pensée sauvage*, Grenoble, 1991

Wenger, E. (1987) *Artificial Intelligence and Tutoring Systems*. Los Altos, California : Morgan and Kaufmann Publisher

Remerciements: Bisson G, Gordon M.B, Nicaud JF, Chaachoua H., Pellegrino F.