

Extraction de pépites de connaissances dans des réponses d'élèves en langage naturel

Sylvie Normand-Assadi*, Lalina Coulange*
Elisabeth Delozanne**, Brigitte Grugeon***

* IUFM de Créteil, Rue Jean Macé, 94861 BONNEUIL Cedex, France
(sylvie.normand, lalina.coulange)@creteil.iufm.fr

** CRIP5 - Paris V, 45-46 rue des Saints-Pères, 75 006 PARIS, France
elisabeth.delozanne@math-info.univ-paris5.fr
<http://www.math-info.univ-paris5.fr/~delozanne>

*** IUFM d'Amiens 49, boulevard de Châteaudun 80044 AMIENS CEDEX,
brigitte.grugeon@creteil.amiens.fr

Résumé. Le projet Pépite a pour objectif la construction d'un diagnostic des compétences d'élèves en algèbre élémentaire permettant aux enseignants de gérer la diversité cognitive de leurs élèves. Dans cet article, nous présentons une étude pluridisciplinaire (linguistique, didactique et informatique), s'appuyant sur un corpus de productions d'élèves utilisant le logiciel Pépite. Le corpus est analysé selon les points de vue croisés de la linguistique et de la didactique. L'objectif de cette démarche est d'améliorer l'évaluation des réponses d'élèves aux questions ouvertes quand ces derniers répondent avec leurs propres mots. Après avoir situé notre étude, nous présentons la méthodologie retenue et les premiers résultats. Nous montrons ensuite la pertinence de ces résultats avec le point de vue de la recherche sur l'enseignement des mathématiques. Nous terminons par les perspectives ouvertes par ce travail en nous interrogeant sur les apports d'approches EGC à la problématique du diagnostic de compétences.

1 Introduction

L'analyse des réponses d'élèves exprimées dans leurs propres mots est un verrou sur lequel butent de nombreux projets en EIAH (Rosé et al 2003a). Beaucoup de travaux s'intéressent à l'analyse automatique des explications produites en langage naturel et des interactions verbales d'étudiants (avec des pairs, avec l'enseignant), ou se centrent sur l'utilisation de modèles linguistiques dans la conception de logiciels d'apprentissage (Aleven et al 2002), (Arroyo et al 2001), (Rosé et al 2003). Comment prendre en compte la diversité des productions spontanées d'élèves « dans leurs propres mots » dans une analyse informatisée ? En quoi la nature de ces productions peut-elle nous informer sur les apprentissages ou leurs dysfonctionnements ? Notre étude vise à apporter des éléments de réponses à ces questions, dans le contexte très spécifique d'un logiciel d'évaluation diagnostique en algèbre élémentaire. Cette étude se situe dans le cadre du projet Pépite qui vise à concevoir un logiciel pour évaluer des compétences en algèbre élémentaire. Il s'agit d'analyser les erreurs et les cohérences de raisonnement d'élèves en algèbre en tenant compte de leurs réponses à un test couvrant les principales dimensions de la compétence algébrique de base. L'objectif est, à partir de ce diagnostic de proposer des situations d'apprentissage adaptées aux profils d'élèves ainsi repérés. Le logiciel Pépite ne se contente pas de répertorier des erreurs, mais

explicite aussi les conceptions ou les fonctionnements sous-jacents à la production de ces erreurs (Grugeon 1995), (Jean et al 1999). Pour les didacticiens des mathématiques de notre équipe, il s'avère indispensable de recueillir des réponses où les élèves s'expriment dans leurs propres termes, même si le logiciel n'est pas capable à ce jour d'en faire une analyse automatique complète et fiable.

L'étude linguistique présentée dans cet article apporte des éléments afin d'améliorer, le diagnostic de ce type de réponses. Tout d'abord, nous montrons comment cette étude s'inscrit dans l'ensemble du projet. Puis nous présentons notre méthodologie et l'analyse des données recueillies. Nous articulons ensuite les résultats de cette étude linguistique avec un point de vue issu de la didactique des mathématiques. En guise de conclusion, nous dégagons des perspectives pour dépasser les limitations de ce travail en adoptant une approche EGC.

2 Cadre de l'étude

Le logiciel pépite (Jean et al 1999), (Delozanne et al 2000), se basant sur des travaux de l'enseignement de l'algèbre (Grugeon 1995), (Artigue et al 2001), propose une modélisation cognitive multidimensionnelle de la compétence algébrique élémentaire. Il est composé de trois modules PépiTest, PépiDiag et PépiProf. PépiTest propose aux élèves un test et recueille l'ensemble de leurs réponses. Le système de diagnostic PépiDiag analyse automatiquement la plupart de ces réponses. PépiProf permet à l'enseignant d'accéder au diagnostic ainsi établi, de le compléter, de le modifier. Le logiciel construit enfin un profil cognitif qui permet d'obtenir un regard d'ensemble sur la compétence algébrique construite par l'élève concerné.

Suite à une première amélioration du système de diagnostique permettant l'analyse de réponse faites de plusieurs expressions algébriques, nous cherchons à développer l'analyse des réponses faites en langage mathurel. Par mathurel, nous entendons langage utilisé par les élèves, mélangeant le langage naturel et l'emploi de termes spécifiques aux mathématiques (Rasseneur et al 2002). Cette perspective s'inscrit dans la suite des travaux menés en didactique dans le cadre du projet. Nous exposons à présent notre méthodologie et le matériau utilisé puis nous illustrons notre démarche d'investigation à partir d'une tâche analysée.

2.1 Méthodologie

Dans le cadre de cette étude, nous apprécions les actes de langage produits par les élèves, qui sont conditionnés par le contexte de résolution d'une tâche, et la force illocutionnaire des énoncés qui y sont associés. Nous nous appuyons sur des marqueurs et des structures linguistiques formelles utilisés par les élèves que nous interprétons afin d'octroyer aux justifications des élèves une orientation qui les inscrit dans un registre discursif.

Nous adoptons une méthodologie fondée sur une étude empirique d'un corpus de réponses d'élèves et une mise en relation des marqueurs linguistiques avec une catégorisation des différentes approches de la tâche à effectuer.

L'exercice 2 de PépiTest, sur lequel nous avons travaillé, propose aux élèves de « valider ou invalider l'égalité de deux expressions algébriques par l'assertion vrai/faux puis justifier cette assertion ». Cet exercice est constitué de trois questions et pour chaque question, la réponse de l'élève est composée d'un choix (vrai/faux) et d'une justification (correspondant aux arguments donnés pour justifier le choix vrai/faux). Dans cet article nous ne présentons le traitement que d'une question sur les trois (question 2 : $a^2=2a$). Notre corpus est constitué

de 52 productions d'élèves de 3^{ième} et 2nd sur les 168 recueillies au total (les autres justifications relèvent de l'algébrique ou de l'absence de justification).

Nous avons dans un premier temps distingué deux groupes d'élèves en fonction de l'exactitude de leurs choix :

- Groupe 1 : élèves ayant fait des choix corrects « vrai/faux » pour les trois questions de l'exercice (24 élèves)

- Groupe 2 : élèves ayant fait au moins un choix incorrect « vrai/faux » en réponse à une des questions posées (28 élèves).

Pour ces deux groupes, nous avons codé le caractère « correct » (C) ou « incorrect » (I) du choix entre vrai et faux de l'élève ainsi que le caractère « correct » (C), « partiel » (P) ou « incorrect » (I) de sa justification. Nous obtenons ainsi des combinaisons telles que : choix correct / justification correcte (CC), choix correct / justification partielle (CP), choix correct / justification incorrecte (CI), choix incorrect/justification incorrecte (II). Le tableau 1 synthétise la répartition de ces combinaisons au sein des deux groupes pour la question 2.

	CC	CP	CI	II	Total
$a^2 = 2a$	GROUPE 1				
	14	6	2	0	23
	GROUPE 2				
	8	5	1	10	24

TAB.1 - *Combinaisons relatives à la question $a^2 = 2a$*

2.2 Analyse de données

Dans cette étude, nous cherchons à dégager les critères linguistiques qui peuvent nous apporter des informations, relativement aux compétences en algèbre des élèves. A partir de ces critères nous identifions des modes de discours. Nous faisons l'hypothèse que l'emploi de ces modes révèle différents niveaux d'entrée dans l'algèbre (conceptuel, contextuel, formel scolaire) qui peuvent être corrélés avec la réussite des élèves à la tâche, voire au développement de leurs compétences en algèbre.

Pour chaque question, nous relevons des caractéristiques de l'égalité d'un point de vue mathématique qui sont susceptibles d'influencer la nature de ces critères. Puis, pour chaque catégorie de réponses codées, nous mettons en avant les formes linguistiques utilisées par les élèves, qui permettent d'élaborer une typologie des justifications produites. Relativement à la question 2 de Pépitest, $a^2 = 2a$, cette égalité est fautive et ne ressemble à aucune règle classique donnée dans les cours d'algèbre. Pour cette égalité, nous réalisons l'analyse suivante dont nous présentons les principaux éléments sous la forme d'un tableau.

2.2.1 Analyse linguistique

Pour les principaux résultats de cette analyse nous renvoyons le lecteur au tableau 2, analyse des réponses associées à l'égalité $a^2 = 2a$ (ci-dessous).

2.2.2 Interprétation de l'analyse d'un point de vue didactique

Le fait que l'égalité soit fautive et non « semi-calquée » sur une règle d'algèbre du cours a pour conséquence que le registre descriptif est relativement peu utilisé, l'élève n'a pas de référent sur lequel s'appuyer. On note également pour les mêmes raisons une quasi-absence du registre légal en ce qui concerne les justifications de réponses incorrectes qui restent dans un mode explicatif. En revanche, un nombre plus important d'élèves ont recours à un mode de discours de type argumentatif. Ces derniers justifient leur choix en mettant en relation les deux membres de l'égalité en spécifiant le type de relation (l'opposition) ou à travers

Extraction de pépites de connaissances

l'emploi d'une conjonction de coordination. Ceci témoigne d'un niveau d'abstraction élevée que l'on peut corrélérer à une progression dans la pensée algébrique.

Code	Registre	Niveau	Exemple	Marqueurs	Nombre
CC	<i>Argumentation opposition</i>	Conceptuel	a^2 signifie $a \times a$ alors que $2a$ signifie $a \times 2$	<i>tandis que , et non pas, alors que</i>	11 (sur 22 CC) 9 Gr1, 2 Gr2
L'élève utilise une phrase complexe construite à partir de deux propositions articulées par une locution conjonctive d'opposition, l'élève marque explicitement la relation d'opposition entre les deux propositions qu'il formule.					
CC	<i>Argumentation coordination</i>	Conceptuel	car le premier ça fait a fois a et le deuxième ça fait 2 fois a	<i>Et, donc</i>	9 (sur 22 CC) 5 Gr1, 4 Gr2
L'élève utilise une phrase complexe construite à partir de deux propositions articulées par une conjonction de coordination, le lien entre les propositions est établi mais non spécifié.					
CP	<i>Descriptif</i>	Contextuel	c'est « $a+a$ » qui est égal à $2a$.	<i>C'est, ça fait, c'est, est égal à</i>	5 (sur 11 CP) 3 Gr1, 2 Gr2
Dans cette catégorie, le lien avec un membre de l'égalité devient implicite : seul un des deux membres est considéré. L'élève ne porte son intérêt que sur l'expression algébrique de ce membre et décrit des expressions algébriques équivalentes en les introduisant par « c'est », « ça fait ».					
II	<i>Explicatif</i>	Formel	c'est vrai car la lettre a qui est élevé au carré donne $2a$ ($a \times a = 2a$).	<i>Car, c'est vrai car</i>	6 (sur 10 II) 6 Gr2
L'élève a recours à la causalité marquée dans son discours par l'utilisation de « car » ou « c'est vrai car ».					
CC, CP, CI, II hors catégories			2 CC (sur 22), 6 CP (sur 11), 3 CI (sur 3), 4 II (sur 10), 6 non justifiés		

TAB.2 - réponses associées à l'égalité $a^2 = 2a$

3 Perspectives

Cette étude exploratoire ouvre des perspectives selon trois directions :

Tout d'abord, il s'agit de tester certaines des hypothèses de travail *a priori* sur lesquelles nous nous sommes appuyées tant d'un point de vue didactique que linguistique : nous devons confronter les hypothèses sur l'avancée en algèbre d'élèves que nous formulons à partir de l'analyse linguistique de leurs discours, aux profils cognitifs en algèbre établis par le logiciel Pépite. Une confrontation systématique entre réussite au test, profils cognitifs et analyse linguistique des productions d'élèves permettra d'attester de la pertinence de notre étude.

Nous envisageons également de tester la typologie élaborée sur un autre corpus de données. Il s'agit de proposer un test comportant des questions relatives au même type de tâches (*valider ou invalider l'égalité de deux expressions algébriques*), élaborées en modulant les variables dégagées par l'étude (égalité vraie/fausse, complexité des expressions algébriques, etc.). Notre objectif est ici de vérifier nos hypothèses sur l'influence du type d'égalité mathématique sur les réponses des élèves.

Enfin il s'agit d'améliorer le système de diagnostic automatique de Pépité. Sur la base de nos premiers résultats, nous pouvons envisager d'implémenter un système capable premièrement d'évaluer le caractère correct/incorrect des justifications produites en langage mathématique par les élèves. Deuxièmement en étudiant plus en profondeur la justifications à l'aide des critères linguistiques dégagés (marqueurs et structures grammaticales) il deviendra possible d'assigner un niveau de rationalité algébrique pour compléter le profil cognitif existant.

Références

- Aleven V. et al. (2002) Pilot-Testing a Tutorial Dialog System That Supports Self-Explanation, Proceedings of ITS'2002, Biarritz (France), 5-8 juin 2002. Cerri S., Gouardères G., Paraguaçu F. (eds.), Springer-Verlag, pp. 344-354.
- Arroyo I. et al (2001) An Adaptive Web-Based Component for Cognitive Ability Estimation, in J.D. Moore et al. (Eds.), Artificial Intelligence in Education, IOS Press, 2001, pp. 456-466.
- Artigue M. et al. (2001) Teaching and Learning Algebra : approaching complexity through complementary perspectives, In Helen Chick, Kaye Stacey, Jill Vincent, John Vincent (Eds), The future of the Teaching and Learning of Algebra, Proceedings of 12 th ICMI Study Conference, The University of Melbourne, Australia, December 9-14, 2001, pp.21-32.
- Delozanne E. et al (2000) Assessing Students' Competence in Algebra, Workshop "Learning Algebra with the Computer, a transdisciplinary Workshop", ITS 2000, Montreal, June 2000.
- Grugeon B. (1995) Etude 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 VII.
- Jean S. et al. (1999) A diagnostic based on a qualitative model of competence in elementary algebra, in S. Lajoie, M. Vivet, AI&ED'99, IOS Press, Amsterdam, , Le Mans (1999) pp.491-498.
- Rasseneur D. et al.(2002) Learning with virtual agents: Competition and Cooperation in AMICO, Proceedings of ITS'2002, Biarritz (France), 5-8 juin 2002. Cerri S., Gouardères G., Paraguaçu F. (eds.), Springer-Verlag, pp. 61-70.
- Rosé C.P. et al. (2003) A Hybrid Text Classification Approach for Analysis of Student Essays, Proceedings of the HLT-NAACL 2003 Workshop on Educational Applications of NLP.
- Rosé C.P. et al. (2003) Overcoming the Knowledge Engineering Bottleneck for Understanding Student Language Input, in Ulrich Hoppe, Felisa Verdejo, Judy Kay (eds.) Proceedings of Artificial Intelligence in Education, Sydney, July 2003, IOS Press, Amsterdam, pp.315-322.

Summary

We describe an exploratory empirical study to investigate whether some linguistic markers can improve the assessment of students when they answer questions in their own words. This work is part of a multidisciplinary project, the Pépité project, that aims to give math teachers software support to assess their students in elementary algebra. We first set this study within the context of the project and we compare it with related work. Then we present our methodology, the data analysis and how we have linked linguistic markers to discursive modes and then these discursive modes to levels of development in algebra thinking. The conclusion opens onto promising perspectives.