Linguistic Markers to Improve the Assessment of students in Mathematics: an Exploratory Study

Sylvie Normand-Assadi[1], Lalina Coulange[1, 2], Élisabeth Delozanne[3], Brigitte Grugeon[1, 4]

[1]IUFM de Créteil, Rue Jean Macé, 94861 BONNEUIL Cedex, France
{sylvie.normand, lalina.coulange,elisabeth.delozanne}@creteil.iufm.fr
[2] DIDIREM - Paris VII, 2, Place Jussieu, 75 251 PARIS Cedex 05, France
[3] CRIP5 - Paris V, 45-46 rue des Saints-Pères, 75 006 PARIS, France
[4]IUFM d'Amiens 49, boulevard de Châteaudun 80044 AMIENS CEDEX, France
brigitte.grugeon@amiens.iufm.fr

Abstract. 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 Pépite 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.

1. Introduction

In this paper we present an exploratory empirical study to improve the diagnosis of students' answers in the Pépite system when answers are articulated in students' language[1]. In previous papers [6, 9] we presented the Pépite project that aims to provide math teachers with software support to assess their students in elementary algebra. We basically assume that students' answers to a set of well-chosen problems show not only errors but also coherences in students' algebra thinking. Like in [13], we are not only interested in detecting errors but also in detecting students' conceptions that produce these errors. We have adopted an iterative design methodology. Our study is the beginning of the second iteration.

At the first design stage, it was important for the educational researchers in our team to have the students answer in their own words even if the software was not able to analyze and understand them completely. So far Pépite software analyses MCQ and answers to open questions when they are expressed using algebraic expressions [6,9].

1 This research was partially funded by the "Programme Cognitique, école et sciences cognitives, 2002-2004" from the French Ministry of Research and by the IUFM of Créteil. Numerous colleagues from the IUFM of Créteil and teachers are acknowledged for testing Pépite in their classes.

Therefore, in order to have a full diagnosis, the system needs the teacher's assessment for answers expressed in "mathural" language such as in Figure 1. By "mathural", we mean a language created by students that combines mathematical language and natural language. The formulations produced by students in this language are often incorrect or not completely correct from a mathematical point of view. But we assume that they demonstrate an early level of comprehension of mathematical notions.

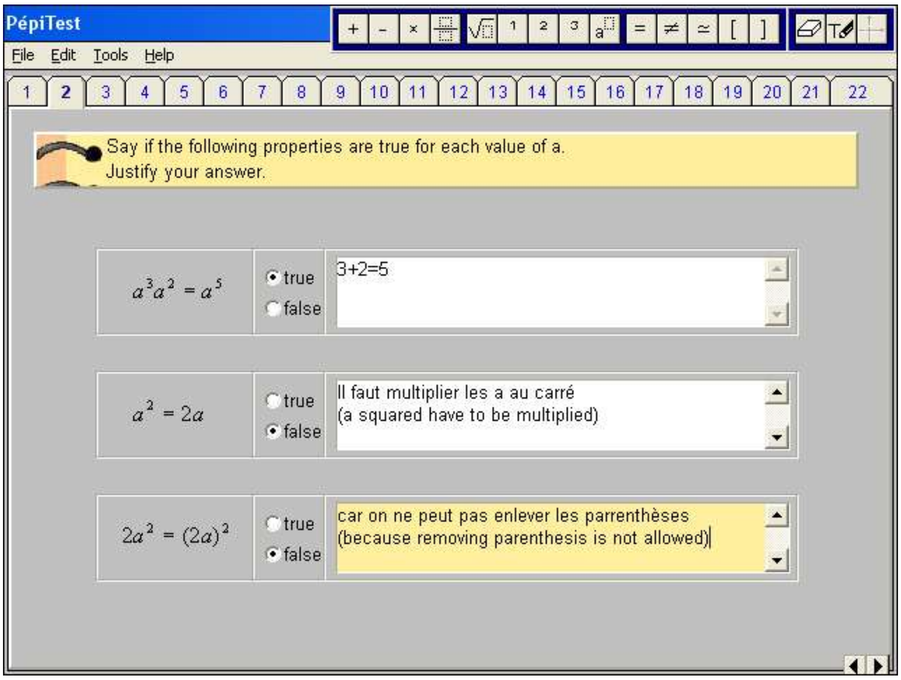

Figure 1: Juliette's answers to exercise 2 in Pépite

Table 1 shows the example of what the educational researchers in our team diagnosed in students' justifications [3, 8]. The diagnosis is based on a classification of justifications like in other research work [1, 10, 14]. Pépite implements this analysis and first diagnoses whether the justification is algebraic, numerical or expressed in mathural language. Then it assesses whether numerical or algebraic answers are correct. For "mathural" answers it only detects automatically that students rely on "school authority" by using markers like "il faut" (it is necessary), "on doit" (you have to), "on ne peut pas" (it is not allowed). In other words, for these students mathematics consist in respecting formal rules without having to understand them.

Workshop and classrooms experiments with teachers showed that, except in very special occasions, they need a fully automated diagnosis to get a quick and accurate overview of the student's competencies [6]. Thus, one of our research targets is to enhance the diagnosis software by analyzing answers expressed in "mathural" language in a more efficient way. We also noticed that our first classification (Cf. Table 1) was too specific to a high school level and that teachers were more tolerant than Pépite toward mathural justifications. For instance for the following answer "the product of two identical numbers with different exponents is this same number but with both exponents added, thus a to the power 2+3", Pépite does not consider it as an algebraic proof when human assessors do.

Table 1. Current Pepite's analysis of some students' justifications when stating that "$a^2 * a^3 = a^5$" is true for every a.

Proof by …	Type of justification	Representation mode	Examples of students' justifications
algebra	To give correct rule or definition	algebraic	$a^3 a^2 = a \times a \times a \times a \times a = a^5$ $a^2 a^3 = a^{2+3}$ $a^n a^p = a^{n\times p}$
numerical example	To try with one or several numbers	numerical	$2^3 \times 2^2 = 32$
explanation	To give explanations	mathural language	"it is true because both exponents are added" "in multiplications with powers, exponents are added
school authority	To give rule relying on authority	mathural language	"We must never multiply exponents" "It's a fundamental law"

We assumed that a linguistic study of our corpus might give important insights to improve the classification as a first step to automatically analyze the quality of the justifications in mathural language. Our preliminary study aimed to point out how linguistic structures used by students could be connected with their algebra thinking. Hence we adopted a dual point of view: linguistic and didactical. The study was made up of five steps: (1) an empirical analysis from a purely linguistic point of view in order to provide new ideas ; (2) a categorization of justifications by cross fertilizing the first and second authors' linguistic and didactical points of view; (3) a review of this categorization in a workshop with teachers, educational researchers, psychological ergonomists, Pépite designers and a linguist (the first author); (4) a final categorization refined by the four authors and presented here; (5) a validation of the categorization set up by the Pépite team. In the following sections we present our methodology, the final categorization and the data analysis. The paper ends with a discussion of the results and with perspectives: first to confirm these early results with other data and then to use these results to build systems that understand some open answers uttered in "mathural" language in a more efficient way.

2. Methodology

This study is based on an empirical analysis of linguistic productions given by 168 students (aged 15 – 16) from French secondary schools while solving algebraic tasks using 'PépiTest'. We focused on a specific task (exercise 2, Figure 1) where students were asked to say if some algebraic properties are true or false and to justify their statements in their own words. The exercise is made up of three *questions* and for each question the student's *answer* is composed of a *choice* (between true or false) and a *justification* (the arguments students give to justify their choice between true or false). These justifications are various: no justification, only algebraic or numerical expressions, "mathural language" statements. In the following sections we focus only on students who gave at least one justification expressed in "mathural language" (52 students).

The statements written by students were studied as speech acts [16] performed by students in the context of the task. We aimed to connect what students said (locutionary act), what they meant (illocutionary act) and what they performed (perlocutionary act) [4]. Speech acts performed by students were conditioned by context (here to justify their choices). At a pragmatic level, we assessed the illocutionary strength of students' statements in relation with the objective of the utterances: the task they were asked to perform ("validate or invalidate an algebraic equality and justify this choice"). Our approach is situated within the Integrated Pragmatic Framework [5]. We pointed out different formal linguistic markers expressed by students. We interpreted these markers as providing a specific orientation to the statement. In our opinion this orientation characterizes a discourse mode. Discourse produced in an assessment context is a very specific written dialog between one student and an unknown reader who will judge him/her. The contract for the writer is very different from a conversational dialog as studied in [11] or from Socratic Dialog as studied in ITS community [1, 2, 7, 10]. But as in some of these studies [13, 14] we are looking for a classification of the quality of students' justifications and criteria to classify these justifications.

Table 2 Classifications of the students' justifications

Correctness	**Discursive mode**	**Level of development in algebra thinking**
CC	*Argumentative:* Students use connections between their arguments to articulate their justifications (consequence, restriction, opposition)	*Conceptual:* Students handle concepts
CC or CP	*Descriptive:* Students describe some elements from the context set by the given equality	*Contextual:* Students select some elements that make sense in the context
CP, II	*Explanatory:* Students require causality often with wrong arguments.	*Formal or school authority:* Students apply or mention formal rules or malrules
CP, II	*Legal:* Students base their justifications on legal or authoritative arguments.	

We first distinguished two groups of students' answers according to the correctness of their choices: Group 1: Students who gave right choices « true/false » to the three questions (24 students), Group 2: Students who gave, at least, one wrong choice to one of the three questions (28 students). Then, for each question we codified: correct choice / correct justification (CC), correct choice / partial justification (CP), correct choice / incorrect justification (CI) ; incorrect choice / incorrect justification (II). Secondly, for each question, we started by highlighting the features of the equality from a mathematical point of view. Then, for each category of coded answers, we pointed out specific linguistic forms used by students and we proposed a typology of justifications from a discursive point of view. So we obtained a quantitative analysis of the corpus that linked students' performance level on the task (correctness) to four discursive modes: argumentative, descriptive, explanatory and legal. Thirdly, from a didactic point of view, we *a priori* hypothesized that these different discursive modes were closely linked with different levels of development in the students' algebra thinking that we qualify as: conceptual, contextual and formal (or school authority). Table 2 summarizes the categorization and the next section describes its application to the corpus we studied.

3. Data Analysis

In this section, we present how we characterized each question from a mathematical point of view and how we classified students answer according to (i) performance on the task, (ii) discursive mode, (iii) level of development in algebra thinking.

Question 1: $a^3 a^2 = a^5$

From a mathematical point of view, this equality has three main features. First, this equality is true. Second, it is very similar to an algebraic rule ($a^m a^n = a^{m+n}$) that is found in every textbook and teacher' courses as part of the curriculum. Third, the both members of the equality can be developed ($a^3 a^2 = (a\times a\times a) \times a\times a$ et $a^5 = a\times a\times a\times a\times a$). For this questions we determined five categories..

CC (Correct choice/correct justification), argumentative mode (consequence, restriction), conceptual level :3 students from Group 1
Students use coordinating conjunctions such as: « donc » (thus) , « mais » (but) to establish relationship such as *consequence* or *restriction*. Thus, we assumed that their discourse was *argumentative* and that through their arguments their algebra thinking was situated on a *conceptual level*. For instance:
« Le produit de deux nombres identiques à exposants différents est ce même nombre mais avec leurs exposants ajoutés tous deux, donc a puissance 2+3 » (the product of two identical numbers with different exponents is this same number but with both exponents added, thus a to the power 2+3).

CC (Correct choice/correct justification) descriptive mode, contextual level: 5 students , 1 from Group 1, 4 from Group 2
Students use a complex sentence, including a main clause and a situating subordinate clause: these clauses are juxtaposed or embedded. The main clause defines the action (« on ajoute, on additionne »: one adds, one adds up). The second clause indicates the *context* which is defined by students and which is necessary for the action (« lors d'une multiplication », « dans une multiplication »: when multiplying, in a multiplication). Their discourses are *descriptive* and their arguments reflect a *contextual* level. Specific linguistic forms are used such as: « lorsque » (when), « quand » (when), « dans » (in). For instance:
« quand on multiplie des mêmes nombres avec des puissances, on adition les puissances et le nombre reste inchangé » (when you multiply numbers with powers, you add the power and the number remains unchanged)

CP (Correct choice/partial justification), descriptive mode and contextual level: 15 *students, 12 from Group 1, 3 from Group 2*

Students use a complex sentence similar to the previous one. Nevertheless, the justification is coded as partial because students do not mention every condition required for the application of the general mathematical rule. In fact, students often forget the following condition: variable a, which is exponentiated, has to be the same. Students focus on the components of the equality that change from the first member to the second member: exponents 3, 2 and 5. They overlook a, the stable component. So we classified these justifications as situated on the *contextual* level. Specific linguistic forms are used such as: « lorsque » (when), « quand » (when) « dans » (in). For instance: (i) « Dans les multiplication à puissances, on additionne les exposants » (in multiplications with powers, exponents are added), (ii) « quand on multiplie des nombres avec des puissances il faut additionner les puissances » (When numbers with powers are multiplied, it is necessary to add up the powers).

CP (Correct choice/partial justification), explanatory and legal mode, school authority level: 6 students from Group 2

Like in the previous type of answer, students focus only on changing features from left member to right member of the equality. But instead of setting a context, students require *causality*, beginning an explanation with connectors such as « car » (it's true because, as). Some of them use *modal verbs* expressing feasibility, possibility or obligation such as « il faut » (it is necessary, you have to). Through the usage of such linguistic forms, we qualify these discourses as *explanatory*. Moreover, as they formulate only partially the rule without mentioning its context of validity we assume that students in that case give a *legal dimension* to their explanations. In other words, they feel this equality respects « formal laws » in algebraic calculus. Thus we classified their algebra thinking in a "*school authority*" category. For example:
(i) « car il faut additionner les puissants » (because it is necessary to add the powers),
(ii) « c'est vrai car on additionne les 2exposent » (it is true because both exponents are added).

II (Incorrect choice/Incorrect justification), legal mode, school authority level: 4 students from Group 2

Student use *modal verbs*, such as « falloir » (« il faut», it is necessary) or « devoir » (to have to) to justify their wrong choice. In our opinion by using such verbs they situate their discourse in a *legal dimension*. Here the formal law is implicit or sometimes explicit malrule (such as $a^m a^n = a^{mn}$). For example: (i) « on doit faire une soustraction entre les deux chiffres du haut » (a subtraction between the two upper digits has to be made), (ii) « il ne faut pas additionné les puissances mais les multiplier » (we are not allowed to add up the powers but we have to multiply them).

Question 2: $a^2 = 2a$

The given equality ($a^2 = 2a$) is false. Furthermore, as it is not similar to any classical rule given in algebra courses, students cannot evoke such a rule.[2] Each algebraic

[2] In that, it is different of other false equalities: as $a^3 a^2 = a^6$ which is similar to the form of the rule $a^m a^n = a^{m+n}$.

expression of this equality can be developed - in (a×a) and (a + a) or (2×a)-. We defined three categories of students' answers.

CC, argumentative mode (opposition), conceptual level: 11 students, 9 from Group 1, 2 from Group 2

Students use a complex sentence, including a main clause and a subordinate clause: linked by a *conjunctive locution* marking an *opposition* between the two members of the equality: « tandis que » or « alors que » (while/whereas), « et non pas » (and not). Their discourse is *argumentative*, and reflects a *conceptual level*. For example: (i) « a^2 signifie a×a alors que 2a signifie a×2 » (a squared means a×a while 2a means a×2), (ii) « L'expression a^2 équivaut à a×a, et non pas à 2×a » (the expression a^2 is equivalent to a×a, and not 2×a) .

CC, argumentative mode (coordination), conceptual level: 9 students, 5 from Group 1, 4 from Group 2

As previously, students use a complex sentence but main and subordinate clauses are linked by a *coordinating conjunction*: « et » (and). The link between the two clauses is established, but not specified, contrary to the previous case where students expressed an opposition. For such justifications the conjunction « et » (and) is used. For example: (i) « car le premier ça fait a fois a et le deuxième ça fait 2 fois a » (because the first results in a times a and the second results in 2 times a), (ii) « a^2 = a×a et 2a = 2×a » (a^2 = a×a and 2a = 2×a).

CP, descriptive mode, contextual level: 5 students, 3 from Group 1, 2 from Group 2

In this category, the connection with the second member has become implicit: only one member of the equality is considered. Students describe some algebraic expressions equivalent to this member and introduced their justification by « c'est » « ça fait » (it is, that results in). Their discourse is *descriptive* and the level *contextual*. For example: (i) « ça fait a×a. » (it results in a×a), (ii) « c'est « a+a » qui est égal à 2a. » (it is « a+a » who is equal to 2a).

II, explanatory mode, school authority level: 6 students from Group 2

Students require causality, beginning their justification with connectives such as « car » (because, as) or « c'est vrai car » (it's true because). Their discourse is explanatory using wrong arguments. For example: (i) « car le a au carré vaut bien deux fois a» (because the value of a squared is actually twice a), (ii) « c'est vrai car la lettre a qui est élevé au carré donne 2a (a×a = 2a). » (it is true because the squared letter a results in 2a (a×a = 2a).

Question 3: $2a^2 = (2a)^2$

The given equality ($2a^2 = (2a)^2$) is false. Like the previous equality, it is not similar to any classical rules given in algebra courses. Each member can be developed ($2a^2$ = 2×a×a, $(2a)^2$ = 2×a×2×a). The right part of the equality contains parentheses:

mathematics teachers often underline the role of parentheses in numerical and algebraic calculus. For this question we have obtained five categories.

CC, argumentative mode(opposition), conceptual level: 14 students, 12 from Group 1, 2 from Group 2

As for previously, students use complex sentence including a main clause and a subordinate clause. These clauses are linked by a *conjunctive locution* which marks the *opposition* between the two members of the equality (focusing on the role of parentheses): « tandis que » or « alors que » (while/whereas), « et non pas » (and not). Their discourse is argumentative and their argument conceptual. For example: (i) « Dans la première partie de l'équation, seul a est au carré alors que dans la deuxième, le produit de 2a est au carré » (In the first part of the equation, only a is squared while in the second part, the product of 2a is squared), (ii) « $2a^2 = 2 \times a^2$ et non pas $(2a)^2$ car ce serait égal à $4a^2$. » (« $2a^2 = 2 \times a^2$ and not $(2a)^2$ because it would be equal to $4a^2$).

CC, argumentative mode (coordination), conceptual level: 5 students, 3 from Group 1, 2 from Group 2

Students use a complex sentence, similar to the previous one but the main and subordinate clauses are linked by a *coordinating conjunction*: « et » (and). Some juxtapose two main clauses, considering each member separately. Students do not mark explicit opposition or explicit links between the clauses. For example: (i) « car $2a^2$, c'est a qui est au carré. Et $(2a)^2$, c'est 2a qui est au carré. » (because $2a^2$, it is a that is squared. And $(2a)^2$, it is 2a that is squared), (ii) «$2a^2 = 2 \times a^2$: il n'y a que le a qui est au carré. $(2a)^2 = 4a^2$: le tout est au carré. » ($2a^2 = 2 \times a^2$: only a is squared. $(2a)^2 = 4a^2$: the whole is squared)

CP, descriptive mode (restriction), contextual level: 4 students, 2 from Group 1, 2 from Group 2

The connection with the second member is implicit. Only one member (the right one) of the equality is considered by students. They focus on the right member, introducing their description by « c'est » (that is), thus underlining the restrictive function of the square which concerns only variable a (because of the absence of parentheses) by « juste », « seulement » (only). Their discourse is descriptive and the level contextual. For example: (i) « c'est juste le a qui est au carré. » (only a is squared), (ii) « comme il n'y a pas de parenthèses, c'est seulement la valeur « a » que l'on multiplie par elle-même.» (as there is no parentheses, only value a is to be multiplied by itself).

II, legal mode, school authority level: 2 students from Group 2
Students frequently use modal verbs such as « pouvoir » (can) or « avoir le droit » (be allowed to) to justify their wrong choice, focusing on the importance of parentheses. By using such verbs, they situate their discourse on a legal dimension. For instance: (i) « on a le droit de mettre des parenthèses à un chiffre » (we are allowed to put parentheses to a digit), (ii) « on peut mettre une parenthese, cela ne change rien sauf lors d'un calcul, quand il y a des prioritées. » (we can put a parenthesis, it does not change anything except when you have priorities in a calculation).

II, explanatory mode, school authority level: 2 students from Group 2
Students use causality beginning their justification with *connectives* such as « car » (because, as). Their discourse is an explanation using wrong arguments.. For instance: (i) « car on multiplie de gauche à droite » (because we multiply from left to right), (ii) « car les deux résultats sont égaux. » (because both results are equal).

4. Results and Perspectives

This study is exploratory but offers some significant results and promising perspectives. We a priori hypothesized links between the discursive modes and the level of development in students' algebra thinking. This empirical study allowed us to define a classification of the students' answers based on these links. Applying it systematically to our data did not invalidate our a priori hypothesis. So this study takes an important step in our project to improve the automatic assessment of students' "mathural" answers.
Our first perspective is to validate this hypothesized correlation in the two following ways. First it remains to be confirmed by systematically triangulating performance (correctness), level in algebra thinking (classification with linguistic markers) and students' profile (built by PepiTest with the whole test), this for every single student in the corpus we studied here. We began to testing our categorization, on some students. We compared their level of development in algebra thinking (as described in this paper by classifying their answer to this specific exercise) with their cognitive profile established by Pépite (by analyzing their answers along the whole test). We noticed that, even in group 1 (correct choices for the three questions), the distinction between school authority/contextual/conceptual levels we derived from linguistic markers is relevant from a cognitive point of view. As suggested by Grugeon [8], students situated in school authority level have difficulties in other exercises to interpret algebraic expressions and often invoke malrules when they make algebraic calculations. Moreover, students adopting argumentative discourse at a conceptual level obtain good results to the whole test. Concerning the contextual category, the interpretation of data seems to be more complex. In particular we hypothesize that the mathematical features of the equality may influence the discourse mode and we will have to investigate that. Second, we will test our typology on other corpora to assess its robustness. We have built a new set of questions based on the same task (to validate or invalidate the equality of two algebraic expressions) but modulating the variables pointed out in this study (true or false equality, features of the expression. We expect to shade light on the nature of partial justifications and contextual level.

Our second perspective is to study how using those linguistic patterns can improve the diagnosis system of Pépite. The current diagnosis system assesses students' choices. Then it distinguishes whether the justification is numerical, algebraic or "mathural". It can both analyze most algebraic or numerical expressions and detect some modal auxiliaries to diagnose a "school authority" level. But so far it has been unable to assess the correctness of justifications in "mathural" language. Once our categorization is validated we will be able to implement a system that links linguistic markers and a level in algebra thinking. The correctness of justification cannot be always automatically derived but (i) an argumentative level is likely to be linked to a correct justification, (ii) a contextual level to a correct or partial (iii) a legal level to a partial or incorrect. Moreover, we will investigate whether the level assigned by this study can be useful to implement an adaptive testing system.

References

1. V. Aleven, K. Koedinger, O. Popescu, A Tutorial Dialog System to Support Self-explanation: Evaluation and Open Questions, Artificial Intelligence in Education (2003) 39-46.
2. V. Aleven, O. Popescu, A. Ogan, K. Koedinger, A Formative Classroom Evaluation of a Tutorial Dialog System that Supports Self-explanation: Workshop on Tutorial Dialog Systems, Supplementary Proceedings of Artificial Intelligence in Education (2003) 303-312.
3. M. Artigue, T. Assude, B. Grugeon, A. Lenfant, Teaching and Learning Algebra : approaching complexity through complementary perspectives, Proceedings of 12 th ICMI Study Conference, Melbourne, December 9-14 (2001) 21-32.
4. J. L. Austin, How to do the things with words. Cambridge, Cambridge University Press (1962)
5. O. Ducrot, Le dire et le dit, Paris, Minuit (1984)
6. É. Delozanne, D. Prévit, B. Grugeon, P. Jacoboni, Supporting teachers when diagnosing their students in algebra, Workshop Advanced Technologies for Mathematics Education, Supplementary Proceedings of Artificial Intelligence in Education (2003) 461-470.
7. A. Graesser, K. Moreno, J. Marineau, A. Adcock, A. Olney, N. Person, Auto-Tutor Improves Deep Learning of Computer Literacy: Is it the Dialog or the Talking Head ? Artificial Intelligence in Education (2003) 47-54.
8. B. Grugeon, 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 (1995).
9. S. Jean, E. Delozanne, P. Jacoboni, B. Grugeon, A diagnostic based on a qualitative model of competence in elementary algebra, Artificial Intelligence in Education (1999) 491-498
10. P. Jordan, S. Siler, Student Initiative and Questioning Strategies in Computer-Mediated Human Tutoring Strategies, on Empirical Methods for Tutorial Dialog Systems, International Conference on Intelligent Tutoring Systems, 2002.
11. M. M. Louwerse, H. H. Mitchell, Towards a Taxonomy of a Set of Discourse Markers in Dialog: A Theoritical and Computational Linguistic Account, *Discourse Processes, 35*(3) (2004) 199-239.
13. C. P. Rosé, A. Roque, D. Bhembe, K. VanLehn, 2003, A Hybrid Text Classification Approach for Analysis of Student Essays, Proceedings of the HLT-NAACL 03 Workshop on Educational Applications of NLP (2003).
14. C. P. Rosé, A. Roque, D. Bhembe, K. VanLehn, Overcoming the Knowledge Engineering Bottleneck for Understanding Student Language Input, Artificial Intelligence in Education, (2003) 315-322.

15. J. R. Searle, Speech Acts, An essay in the Philosophy of Language, Cambridge, CUP (1969)