From usage analysis to automatic diagnosis: The case of the learning of algebra

Emmanuel Sander, Jean-François Nicaud, Hamid Chaachoua, Marie-Caroline Croset

From usage analysis to automatic diagnosis: The case of the learning of algebra.

Emmanuel SANDER*, Jean-François NICAUD**, Hamid CHACHOUA***, Marie-Caroline CROSET**

** Université Paris 8, 2 Rue de la Liberté 93526, Saint Denis, France sander@univ-paris8.fr*
*** Université Joseph Fourier, Leibniz- MeTAH, 46 Ave Viallet, 38031 Grenoble France Jean-Francois.Nicaud@imag.fr & mkcroset@wanadoo.fr*
**** IUFM-Leibniz-MeTAH, 46 Ave Viallet, 38031 Grenoble, France, Hamid.Chaachoua@imag.fr*

Summary: User modelling is a significant part of usage analysis. We present a case study in the field of the learning of algebra that aims at producing automatic diagnosis rules, based on the analysis of tracks of students solving algebra exercises within the Aplusix learning environment. We present two experiments that were conducted among 8th and 9th grade students. Manual analyses performed on the data made it possible to contribute to the construction of a library of rules aiming at modelling students by hand or automatically. The automatic diagnosis, based on the use of a library of correct and incorrect rules, and on a heuristic search algorithm, reveals a high performance on some of the algebra fields and will be extended to other fields through iterative comparisons with the results of the manual diagnosis.

1. Problematic

Within the scope of a project aiming at automatic student's modelling in algebra [9], we conducted tracks analyses with a variety of students from 8 and 9 grades, in order to identify the systematic errors they commit when solving algebra exercises, and to use the identified incorrect rules [5, 7] for automatic diagnosis of the students' transformations in term of rule applications. In our view, this requires: (i) Designing relevant tasks, i.e., relevant algebra exercises; (ii) Identifying, a set of correct and incorrect rules; (iii) Designing an automatic diagnosis algorithm; (iv) Assessing the quality of this diagnosis. For gathering the data, we have used the Aplusix learning environment [3] that allows students to freely make calculation steps and records all the students' actions (Figure 1). This part of our research work is described in this paper. The diagnoses obtained are next used to model students in term of conceptions[1] [4]. In a longer term, we plan to insert our global process in the Aplusix system to be used in ecological situations where the students will learn algebra skills with Aplusix in usual school situations and Aplusix will calculate on-line the students' conceptions and inform the teacher.

Students rely on conceptions, inadequate in some contexts, that are likely to subsist despite learning [1, 2, 8]. One target of this study, due to the lack of converging and exhaustive results on conceptions in algebra, is to build a panel of exercises and to analyse the errors observed. In other words, we aim at identifying a "map" of incorrect rules and of conceptions. This project is quite ambitious since, to our knowledge, this is the first time

[1] An example of part of a conception is the following: *When a sub-expression is moved from one side to the other in an (in)equation, its sign is always changed*. Note that this is sometimes correct (e.g., $x-3 = 5 \rightarrow x = 5+3$) and sometimes incorrect (e.g., $-3x = 5 \rightarrow x = 5/3$).

that a work aims at achieving exhaustiveness in the field of elementary algebra[2] with a data driven approach. More precisely, we aim, in a first step, at providing a methodology that could be applied at a larger scale in order to provide the expected exhaustiveness. For this reason, the experimental work focuses only on semi-beginners in algebra (8th and 9th grades) and part of the study is achieved only on a subfield, namely solving linear equations.

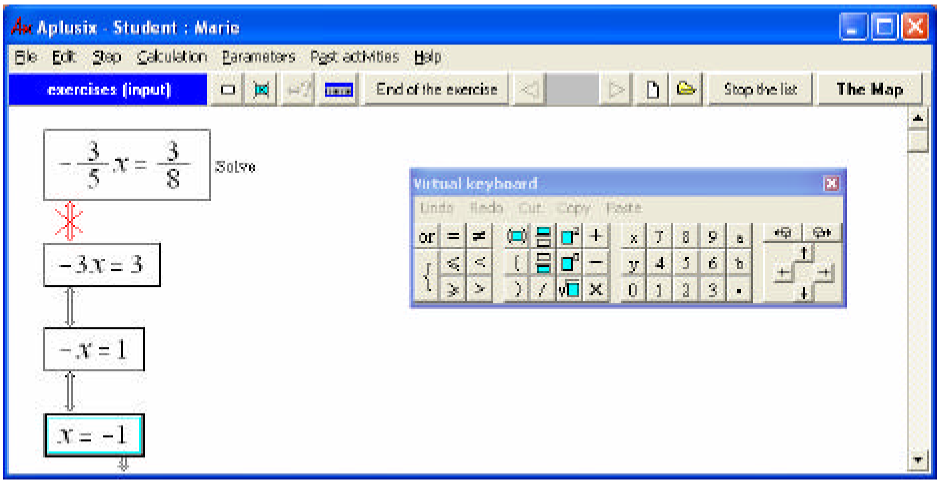

Figure 1. Aplusix problem solving interface

Beyond the incorrect rules identified, the manual analyses performed hereby provide contributions in at least two directions that address issues relevant to many cases of usage analysis involving user modelling: (i) Reaching a conclusion about the stability of the behaviours of the students in the use of the incorrect rules, measured as the repetition of the same behaviour within the same context. This stability is crucial since it conditions the relevance of the diagnosis. (ii) Evaluating the quality of the automatic diagnosis by providing a basis of comparison.

2. Methodology

2.1 Choice of the relevant exercises

We designed two experiments, with complementary purposes. The first one, designed for grade 9 students, intends to make possible gathering a large set of data covering the whole range of rules that might be applied at this level of the curriculum. It is composed of 31 exercises listed below (Table 1). This set of exercises makes possible to observe a large range of incorrect rules in order to build a library of the rules involved. However, this experimental setting reveals some limitations since the importance of the range covered is not compatible with an accurate assessment of the context of application of the rules and of their stability; an apparent lack of stability in the behaviour of a user might reveal that s/he categorizes the situation as different from the previous one despite that they are equivalent from an expert point of view. For this purpose, we designed a second experiment that investigates a subfield in a more systematic manner, namely solving linear equations.

This second experiment was designed in order to focus on a subfield and to check out the possibility of increasing the assessment of the stability of the behaviours identified through a more accurate description of the context of use of the rules. Thus, we built 15

[2] We term elementary algebra, the calculations made on polynomial and rational expressions, polynomial and rational equation and inequation up to highschool. Most of the research devoted to student's modelling in algebra concerns very beginners.

linear equation exercises listed below (Table 2) and we manipulated in quite a systematic manner the parameters that could lead to the identification of the context of application of incorrect rules. More precisely, in an equation of the form ax+b=cx+d we varied the nature of a, b, c, and d and the relations between these values along the dimensions that might be relevant from a naïve point of view; for instance, whether there value is zero or not, and if not whether they are positive or negative numbers, integers or fractions. This methodology is intended to be reproducible in other subfields, in order to encounter the whole range of exercises in algebra as they are categorized by the learners.

2.2 Gathering data and the tool for visualization and construction of numerical indicators

Each action of the student is recorded and can be transcribed by a "video tape recorder" integrated into the software. As a first step of the diagnosis and before its automation, manual analyses were performed. For each student the detail of the resolution of each exercise was looked at with the video tape recorder of APLUSIX (Figure 2) and the rules which made it possible to explain the transformation of an expression into another were identified by the analyst.

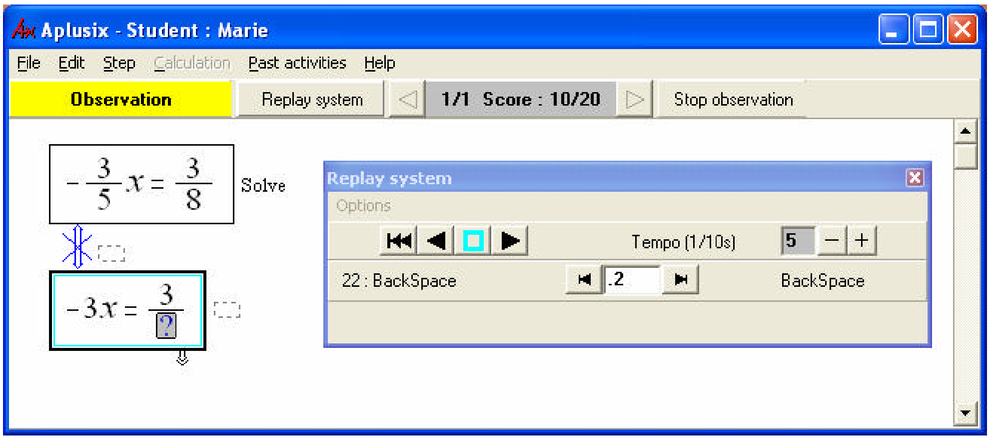

Figure 2. Aplusix's videotape recorder

We built several numerical indicators. Because of the length of these exhaustive manual analyses (several hours for one hour of one student problem solving session), we applied them only to a small part of the whole population that we tested in other studies [9]; namely 3 classes of grade 8 students and 3 classes of grade 9 students.

We measured the rate of occurrence of each incorrect rule among our population. First, this indicator makes possible to distinguish the marginal rules, known as orphan, that seldom occur and that do not deserve designing specific remediation strategies, from the dominant ones that are present in a significant part of the population and deserve to be taken into account seriously. Second, the orphan rules shall not be implemented in the automatic diagnosis in order to avoid combinatorial explosion, whether the most frequent ones have to be implemented in order to avoid diagnosis failures or psychologically implausible automatic diagnoses.

We measured the predictive character of the identified rules with a binary indicator. A rule is regarded as predictive when its use alone leads to the result given by the student. We took a restrictive criterion since we considered that the use of several rules altogether implies that none of the rules are predictive, just as calculation or copy errors interfering with an incorrect rule. This is a strict criterion that might be considered as gathering the cases for which the automation of the diagnosis appears more easily accessible.

We measured the stability of each rule for each student. One of our concerns was to test the robustness of the analyses carried out by the construction of indices on the systematic character of the use of the identified rules, being quite obvious that the diagnoses related to the protocols have interest if one observes a certain stability of the behaviours. The stability is calculated by a ratio between the frequency of use of the rule and the total frequency of cases in which it could be used.

3 Manual diagnoses

3.1. Results with 9th graders

We analysed in a systematic manner the protocols of 73 students (3 classes from a Parisian school) solving the 31 exercises of the first experiment. We organized experimental settings in small groups (from 8 to 10 students) and with as many sessions as necessary in order to solve the whole range of exercises. The list of the exercises, as well as the frequency of success, are provided Table 1.We then performed manual protocol analyses: all in all, 104 rules were identified, and gathered within a typology: (i) Power rules (P1 to P10); (ii) Priority of operators rules (O1 to O11); (iii) Factorization rules (F1 to F10); (iv) Distributivity rules (D1 to D18); (v) Sign rules (S1 to S12); (vi) Elimination of the coefficient of the unknown rules (Eu1 to Eu18); (vii) Elimination of the fixed value rules (Ef1 to Ef16); (viii) Calculation rules (C1 to C9)

Table 1. List of exercises and frequency of success for the first experiment

	Exercises		Class1/21	Class2/25	Class3/27	Sum /73
1	$5x^2+3x-7-3x2+2x+8$	Simplify and order	18	20	20	58
2	(-3-6)*(6-8)	Calculate	14	19	23	56
3	x+2=-3	Solve	20	24	24	68
4	7x+(2x-8)-(-3x+12)	Expand, simplify and order	13	12	17	42
5	(-2)*(-5)*(+3)+(-2)*(-4)	Calculate	20	24	24	68
6	9-x=12	Solve	17	17	16	50
7	2-3(-5x-5)+5(4x+8)	Expand, simplify and order	17	13	12	42
8	4x=16	Solve	20	18	26	64
9	8a+8b	Factor	21	22	23	66
10	5x=9	Solve	20	19	22	61
11	7x(3x+5)	Expand, simplify and order	18	17	17	52
12	8a+40	Factor	21	22	22	65
13	(9x-5)(-6x+2)	Expand, simplify and order	14	8	11	33
14	$12x^2-7x$	Factor	19	21	23	63
15	8x-4=3x-2	Solve	14	19	15	48
16	$10x+1-6x^2+5-3x^2+6x-6$	Simplify and order	19	15	19	53
17	-9*(-2)-7*(-6+2)	Calculate	17	13	15	45
18	10+x=-8	Solve	21	18	24	63
19	9x-(-4+5x)-(5x+10)	Expand, simplify and order	16	7	12	35
20	5x=25	Solve	19	23	26	68
21	4/3+7/6	Calculate	20	23	24	67
22	4x(-1-7x)	Expand, simplify and order	20	11	16	47
23	x/3=-7	Solve	20	16	23	59
24	2/5-1/7	Calculate	17	17	25	59
25	$10(-4x-1)-2(4x^2-6)$	Expand, simplify and order	18	12	16	46
26	-8=-7x+5	Solve	12	14	16	42
27	-10/9*-6/-5	Calculate	16	14	17	47
28	(1+5x)(2x-3)	Expand, simplify and order	13	11	17	41
29	-2x+8=3+2x	Solve	12	11	10	33
30	2-5*5-7*3	Calculate	17	12	10	39
31	7x=4/5	Solve	15	17	10	42

Thus, we achieved our goal of identifying a large set of incorrect rules that are used by students solving algebra exercises. In this sense, the usage analysis appeared to be very informative. However, if this first experimentation made it possible to identify a library of rules, the analyses of the associated numerical indicators described above, which is not detailed hereby, revealed that the stability was often quite low. We reached the conclusion that because of the variety of the exercises, the identification of the context of application of the rules is made dubious, and some indices of non stability of the rules might in fact reveal different contexts of application: conditions that seemed strictly equivalent from the expert point of view were not for the students. So it seemed necessary to carry out a more systematic analysis on a subfield, as discussed above. In order to avoid ceilings effects related to a too high level of the students, this experimentation was carried out with grade 8 students, after they already studied the resolution of linear equations (ax+b=cx+d type).

3.2 Results with the 8th grades

Table 2 indicates the list of exercises as well with the frequency of success (the numbers are not always integer numbers since some intermediate 0,5 mark were attributed in some specific cases).

Table 2: List of exercises and frequency of success for the second experiment

	Exercise	Type	Class1/30	Class2/30	Class3/30	Sum/90
1	-1/4 x=6	Solve	4,5	1,5	14	20
2	$\frac{x}{7}=3$	Solve	17	19	26	62
3	7=28x	Solve	8	6	15	29
4	-4x=-27	Solve	12	6	21	39
5	12-6x=-15x-3	Solve	11	6	16	33
6	8x-11=7+10x	Solve	12	6	10	28
7	2=-x+15	Solve	20	15	18	53
8	-9=x-7	Solve	9	9	14	32
9	11-x=-12	Solve	13	11	11	35
10	-x+2=7+x	Solve	12	4	13	29
11	-3+2x=-2x-2	Solve	15	10	16	41
12	$-\frac{7}{2}x=5$	Solve	8,5	1,5	8	18
13	$\frac{3}{8}x=4$	Solve	10	3,5	15	28,5
14	$9x=\frac{27}{2}$	Solve	9	4,5	13	26,5
15	$-11x=-\frac{22}{2}$	Solve	8	4	16	28

We used the same typology of rules than with the previous experiment. Due to the systematic manipulation of the factors that might influence the choice of the rules, we were able to identify, for each student, the context of application of the rule and it's stability.
Table 3 and Table 4 are extracted for the tables gathering the data from the students. Table 3 indicates, for a sample of the participants, the numerical indicators that were performed: percentage of rules identified relative to the number of exercises that the student got wrong, predictability relative to the total number of failures, and predictability of the identified rules (2nd column divided by first column). Table 4 indicates, for a sample of the participants, the total number of incorrect rules identified for a given student and the degree of stability of these rules.

Table 3. Numerical indicators regarding rate of incorrect rules and rules predictibility

Student	Rules/Failures	Predictibility/Failures	Predictibility
Beck	62,50%	37,50%	60,00%
Bert	40,00%	40,00%	100,00%
Beri	42,86%	42,86%	100,00%
Bong	50,00%	36,36%	72,73%
Dasi	65,22%	52,17%	80,00%

Among the failures, we were able to identify an incorrect rule for 58% of the failures in average. The cases in which no rules were identified concern mostly calculation errors and unachieved exercises. The average rate of predictability among the rules identified is 67%.

Table 4. Stability of the identified rules.

Student	Stable	Intermediate	Non stable	Total
Beck	1	2	3	6
Bert			1	1
Beri	1	1		2
Bong	2	1	1	4
Dasi		4	1	5
……	……	……	……	……
total	46	162	152	360

If the rate of full stability (the rule is used in 100% of the cases) appears to be quite low (13% of the total), partly due to calculation errors, 45% show intermediate stability (the rule is used in at least 50% of the cases), which might show competition between the correct rules and the ones taught in school, as well as the use of opportunistic strategies that we identified with grade 9 students as well and which consist in using in an ad hoc manner a rule that make the problem simpler; the opportunistic rule appear to have a low stability because their context of application is not captured by the manipulated factors.

4. Local automatic diagnosis

The purpose of the local diagnosis is to automatically find a sequence of rules (correct or incorrect) that explains a transformation made by a student (e.g., $7=28x \rightarrow x=28-7$). The term "local" is used because we consider only one transformation at this point. Such diagnosis is achieved to be followed by other automatic treatments: (1) Calculation of the frequencies of incorrect rules used by a student or a class; (2) Attribution of conceptions to students, conceptions being more global representations of the students' knowledge, see details in [9]. We only develop the local diagnosis work in the rest of this paper.

We have implemented formal rules in a computer language and we have implemented an algorithm for providing diagnoses. At the present time, our focus is on the rules that apply to linear (in)equations and the rules for performing expansions and reduction (for other fields, like factorization and fractions, just a few rules have been implemented). As the goal is to automatically diagnose a lot of students' transformations (a class working during 2 hours with Aplusix produces about 1000 transformations to be diagnosed and we have more that 100 classes to study), we did not implement the rules that are very specific and very rare.

We have combined the above cognitive study of the students' productions and epistemic study of (in)equations to produce the rules to be implemented. This led to consider two sets of rule. First, we consider the correct fundamental operations on both sides of the (in)equations (addition, subtraction, multiplication, division) like $A = B \rightarrow A+C = B+C$. Second we consider the correct *movement rules* that are compiled form of these rules when they are combined with reduction; there are *additive movement rules* like: $A+C = B \rightarrow A = B–C$ and *multiplicative movement rules* like $AC = B \rightarrow A = B/C$ (C?0); in these rules, C is said to be moved from one side to the other; and we consider incorrect movement rules obtained by the following processes: (1) Incorrect (un)change of the sign of the moved expression, like in $4x+5=7 \rightarrow 4x=7+5$; (2) Incorrect (un)change of the orientation of the inequality sign, like in $-4x<7 \rightarrow x<-7/4$; (3) Incorrect operator linking the moved expression to the global expression (e.g., move C from a multiplicative to an additive position like in $28x=7 \rightarrow x=28+7$). The combination of these processes and of the sort of relation (=, <, etc.) and of the orientation (left to right, right to left) lead to more that 1000 rules. We did not implement 1000 specific rules but a general rule with features that correspond to the above "processes". For example, the transformation $4x<7 \rightarrow x>7-4$ is diagnosed as the application of the movement rule with the features: <, LefToRight, NumeratorToNumerator, InitAdditive, FinAdditive, SignChanged, OrientationChanged.

The implemented rules are based on the above cognitive study but some of the rules were slightly generalized and the very specific and very rare rules were abandoned. We have not yet implemented a complete set of rules for the calculation of fractions.

4.1. The algorithm of the diagnosis

The diagnosis algorithm that we have implemented is a heuristic search algorithm of the "best first" type [6]. Such an algorithm manipulates objects or states (algebraic expressions in our case) based on the use of operators (rewriting rules in our case) and uses a heuristic in order to constrain the search; the heuristic being a function that provides a proximity measure between two objects. Initial data are composed of two objects, the algebraic expressions A and B in our case, B being the result of the transformation of A by the student. The algorithm searches a list of operators (correct or incorrect rewriting rules in our case) allowing the transformation of A into B. For achieving this purpose, it builds a search tree, A being included in the root, and develops successive nodes. Developing a node N consists of applying all the rules that are applicable to the object that it contains, and to generate a successor of N each time a new object is generated. Algebra is a difficult domain for this kind of search because of: (1) an important branch factor (number of successors of a node) coming from a large number of applicable rules; (2) the presence of cycles in the application of rules; (3) the absence of a good distance to evaluate the proximity of a produce expression with the target. For these reasons, we had to adapt the general algorithm, in particular, some rules take into account the goal: they are not applied when they are applicable if some conditions regarding the goal are not verified. This is the case of the above movement rule that can be applied for 8 expressions in $2x+3+4x+5=6x+7+3x+2$ and each time with a lot of features, and can be applied as many times in the produced nodes. Such an algorithm sometimes fails. When the algorithm does not reach the target after a chosen number of developed nodes (we often chose 30 nodes), it fails. When the target is reached, the obtained diagnosis can be considered as inappropriate because it makes a bizarre combination of rules when the analyst has a better diagnosis.

4.2. Results

Here is an example of diagnosis of the transformation of $2x–6 = 7x-8$ into $–5x = –14$. It is diagnosed with 4 rules: (1) Incorrect additive move of 6, leading to $2x = 7x–8–6$; (2) Correct additive move of 7x, leading to $2x-7x = –8–6$; (3) Correct additive reduction, leading to $–5x = –8–6$; (4) Correct additive reduction, leading to $–5x = –14$. The current application of the automatic diagnosis on recorded data of a grade 8 class (29 students) and a grade 9 (21 students) provides a good ratio of "success" presented in table 6. The "appropriateness" ratios of these diagnoses are presented in table 7 for the incorrect transformations of the two families studied (expansions, reductions and transformations on equations).

Table 6. Success of the automatic diagnosis (i.e., when it did not fail).

Class	Type	Number	Success	Ratio
Grade 8	Correct transformations	1070	1005	94%
Grade 8	Incorrect transformations	434	351	81%
Grade 9	Correct transformations	1071	985	92%
Grade 9	Incorrect transformations	155	121	78%

Table 7. Appropriateness of the automatic diagnosis for incorrect expansions, reductions and transformations on equations (i.e., when it is judged appropriate by the analyst).

Class	Type	Number	Appropriate	Ratio
Grade 8	Incorrect transformations on equations	78	76	97%
Grade 8	Incorrect expansions and reductions	126	103	82%
Grade 9	Incorrect transformations on equations	33	28	85%
Grade 9	Incorrect expansions and reductions	50	46	92%

5. Perspectives

We are now working on the process of improving the quality of the automatic diagnosis through the implementation of other fields (e.g., fractions, factorization) and iterative comparisons with the results of the manual diagnosis. The diagnoses produced by this process are used in another part of our project devoted to the production of conceptions, conceptions being more global representation of the knowledge of the student. This work is presented in another communication of the workshop.

Besides providing libraries of incorrect rules and of conceptions to the scientific community, our results will be used later for two purposes: (1) to calculate students' conceptions in the Aplusix system and present them to the teacher; (2) to build artificial tutor devoted to remediation of inadequate conceptions.

References

[1] Brousseau G. (1997). *Theory of didactical situations in mathematics*. Kluwer Academic Publishers, Dordrecht.

[2] Fischbein, E., (1989). Tacit models and mathematical reasoning, *For the Learning of Mathematics, 9*, 9-14

[3] Nicaud, J.F., Bouhineau, D., and Chaachoua H. (2004). Mixing microworld and CAS features in building computer systems that help students to learn algebra, in *Int. Journal of Computers for Mathematical Learning*, Volume 9 (2), Springer-Verlag.

[4] Nicaud J.F., Chachoua H., Bittar M., Bouhineau D.. (2005) Student's modelling with a lattice of conceptions in the domain of linear equations and inequations. AIED Workshop on Usage Analysis in Learning Systems. Ibid.

[5] Payne S. and Squibb H. (1994) Algebra Mal-Rules and Cognitive Accounts of Error. Cognitive Science, 14.

[6] Pearl J. (1984). *Heuristics*. Addison-Wesley, London.

[7] Sleeman D.H. (1983). Inferring student model for intelligent aided instruction. In Michalski, Carbonell and Mitchell (Eds) Machine Learning: An Artificial Intelligence Approach. Morgan Kaufman.

[8] Tirosh, D., & Graeber, A.O. (1991). The influence of problem type and primitive models on preservice elementary tearchers' about division. *School Science and Mathematics, 91*, 157-163.

[9] Modélisation cognitive d'élèves en algèbre et construction de stratégies d'enseignement dans un contexte technologique. Project report of the "Ecole et sciences cognitives" research programme. Cahier #123 of the Leibniz laboratory, http://www-leibniz.imag.fr