DIANE, a diagnosis system for arithmetical problem solving

Khider Hakem, Emmanuel Sander, Jean-Marc Labat, Jean-François Richard

DIANE, a diagnosis system for arithmetical problem solving

Khider Hakem*, Emmanuel Sander*, Jean-Marc Labat**, Jean-François Richard*
** Cognition & Usages, 2 Rue de la Liberté, 93526 Saint Denis Cedex 02France*
khider.hakem@cognition-usages.org, sander@univ-paris8.fr, richard@univ-paris8.fr
*** UTES , Université Pierre et Marie Curie, 12, rue Cuvier 75270 Paris cedex 05, France*
jean-marc.labat@upmc.fr

Abstract. We hereby describe DIANE an environment that aims at performing an automatic diagnosis on arithmetic problems depending on the productions of the learners. This work relies on results from cognitive psychology studies that insist on the fact that problem solving depends to a great extent on the construction of an adequate representation of the problem, which is highly constrained. DIANE allows large-scale experimentations and has the specificity of providing diagnosis at a very detailed level of precision, whether it concerns adequate or erroneous strategies, allowing one to analyze cognitive mechanisms involved in the solving process. The quality of the diagnosis module has been assessed and, concerning non verbal cues, 93.4% of the protocols were diagnosed in the same way as with manual analysis.
Key Words: cognitive diagnosis, arithmetical problem solving, models of learners.

Introduction

DIANE (French acronym for Computerized Diagnosis on Arithmetic at Elementary School) is part of a project named « conceptualization and semantic properties of situations in arithmetical problem solving » [12]; it is articulated around the idea that traditional approaches in terms of typologies, schemas or situation models, the relevance of which remains undisputable, do not account for some of the determinants of problem difficulties: transverse semantic dimensions, which rely on the nature of the variables or the entities involved independently of an actual problem schema, influence problem interpretation, and consequently, influence also solving strategies, learning and transfer between problems. The identification of these dimensions relies on studying isomorphic problems as well as on an accurate analysis of the strategies used by the pupils, whether they lead to a correct result or not. We believe that fundamental insight in understanding learning processes and modeling learners may be gained through studying a “relevant” micro domain in a detailed manner. Thus, even if our target is to enlarge in the long run the scope of exercises treated by DIANE, the range covered is not so crucial for us compared to the choice of the micro domain and the precision of the analysis. We consider as well that a data analysis at a procedural level is a prerequisite to more epistemic analyses: the automatic generation of a protocol analysis is a level of diagnostic that seems crucial to us and which is the one implemented in DIANE right now. It makes possible to test at a fine level hypotheses regarding problem solving and learning mechanisms with straightforward educational implications. Having introduced our theoretical background that stresses the importance of interpretive aspects and transverse semantic dimensions in arithmetical problem solving, we will then present the kind of problems we are working with, describe DIANE in more details and provide some results of experiments of cognitive psychology that we conducted.

1. Toward a semantic account of arithmetical problem solving

1.1 From schemas to mental models

The 80's were the golden age for the experimental works and the theories concerning arithmetical problem solving. The previously prevalent conception was that solving a story problem consisted mainly in identifying the accurate procedure and applying it to the accurate data from the problem. This conception evolved towards stressing the importance of the conceptual dimensions involved. Riley, Greeno, & Heller [10] established a typology of one-step additive problems, differentiating combination problems, comparison problems and transformation problems. Kinstch & Greeno [7] have developed a formal model for solving transformation problems relying on problem schemas. Later on, the emphasis on interpretive aspects in problem solving has led to the notion of the mental model of the problem introduced by Reusser [9], which is an intermediate step between reading the text of the problem and searching for a solution. This view made it possible to explain the role of some semantic aspects which were out of the scope of Kinstch & Greeno's [7] model; for instance, Hudson [6] showed that in a comparison problem, where a set of birds and a set of worms are presented together, the question *How many birds will not get a worm ?* is easier to answer than the more traditional form *How many more birds are there than worms ?,* and many studies have shown that a lot of mistakes are due to misinterpretations [4]. Thus, these researches emphasized the importance of two aspects: conceptual structure and interpretive aspects, which have to be described more precisely. Informative results come from works on analogical transfer.

1.2 Influence of semantic dimensions

More recently, work on analogical transfer showed that semantic features have a major role in problem solving process. Positive spontaneous transfer is usually observed when both semantic and structural features are common [1]. When the problems are similar in their surface features but dissimilar in their structure, the transfer is equally high but negative [11], [8]. Some studies have explicitly studied the role of semantic aspects and attributed the differences between some isomorphic problem solving strategies to the way the situations are encoded [2]. Several possibilities exist for coding the objects of the situation and a source of error is the use of an inappropriate coding, partially compatible with the relevant one [13].

Within the framework of arithmetic problems, our claim is that the variables involved in the problem are an essential factor that is transverse to problem schemas or problem types. We propose that the different types of quantities used in arithmetic problems do not behave in a similar way. Certain variables call for some specific operations. Quantities such as weights, prices, and numbers of elements may be easily added, because we are used to situations where these quantities are accumulated to give a unique quantity. In this kind of situations, the salient dimension of these variables is the cardinal one. Conversely, dates, ages, durations are not so easy to add: although a given value of age may be added to a duration to provide a new value of age; in this case, the quantities which are added are not of the same type. On the other hand, temporal or spatial quantities are more suited to comparison and call for the operation of subtraction, which measures the difference in a comparison. In this kind of situations, the salient dimension of these variables is the ordinal one.

We want to describe in a more precise way the semantic differences between isomorphic problems by characterizing their influence. For this purpose, it seems necessary to study problem solving mechanism at a detailed level which makes it possible to identify not only the performance but the solving process itself and to characterize the effect of the interpretive aspects induced by the semantic dimensions. Thus, we constructed a structure of problems from which we manipulated the semantic features.

2. A set of structured exercises and their solving models

Several constraints were applied in order to choose the exercises. (i) Concerning the conceptual structure, the part-whole dimension is a fundamental issue in additive problem solving; it appears as being a prerequisite in order for children to solve additive word problems efficiently [14]; thus our problems are focused on a part-whole structure. (ii) We looked for problems that could be described in an isomorphic manner through a change of some semantic dimensions. We decided to manipulate the variables involved. (iii) We looked for a variety of problems, more precisely problems that would allow the measure of the influence of the variable on the combination/comparison dimension. Hence, we built combination problems as well as comparison problems (iv) In order to focus on the role of transverse semantic dimensions, we looked for problems that did not involve either procedural or calculation difficulties. Therefore, we chose problems involving small numbers. (v) We looked for problems allowing several ways to reach the solution so as to study not only the rate of success but the mechanisms involved in the choice of a strategy, whether it is adequate or not and to assess the quality of DIANE's diagnosis in non trivial situations. As a result, we built problems that might require several steps to solve.

The following problems illustrate how those constraints were embedded:

John bought a 8-Euro pen and an exercise book. He paid 14 Euros. Followed by one of these four wordings:

- *Paul bought an exercise book and 5-Euro scissors. How much did he pay?*
- *Paul bought an exercise book and scissors that costs 3 Euros less than the exercise book. How much did he pay?*
- *Paul bought an exercise book and scissors. He paid 10 Euros. How much are the scissors?*
- *Paul bought an exercise book and scissors. He paid 3 Euros less than John. How much are the scissors?*

Those problems have the following structure: all problems involve two wholes (Whole1 and Whole2) and three parts (Part1, Part2, Part3); Part2 is common to Whole1 and Whole2. The values of a part (Part1) and of a whole (Whole1) are given first (John bought a 8 Euros pen and an exercise book. He paid 14 Euros). Then, a new set is introduced, sharing the second part (Part2) with the first set. In the condition in which the final question concerns the second whole (Whole2) a piece of information is stated concerning the non common part (Part3), this information being either explicit (combination problems: Paul bought an exercise book and 5-Euro pair of scissors) either defined by comparison with Part1 (comparison problems: Paul bought *an exercise book* and scissors that cost 3 Euros less than the exercise book). In the condition in which the final question concerns the third part (Part3) a piece of information is stated concerning the second whole (Whole2), this information being either explicit (combination problems: Paul bought *an exercise book* and scissors. He paid 10 Euros) either defined by comparison with Whole1 (comparison problems: Paul bought an exercise book and scissors. He paid 3 Euros less than John). Then a question concerns the missing entity: Part 3 (How much are the scissors?) or Whole2 (How much did Paul pay?).

In fact, three factors were manipulated in a systematic manner for constructing the problems presented hereby:

- The nature of the variable involved.
- The kind of problem (2 modalities: complementation or comparison): if the problem can be solved by a double complementation, we call it a complementation problem; if it can be solved by a complementation followed by a comparison, we call it a comparison problem.
- The nature of the question (2 modalities: part or whole): If the question concerns Whole2, we call it a whole problem and if the question concerns Part3, we call it a part problem.

The two last factors define four families of problems that share some structural dimensions (two wholes, a common part and the explicit statement of Whole1 and Part1) but differ in others (the 2x2 previous modalities). Among each family, we built isomorphic problems through the use of several variables that we will describe more precisely later on.

One major interest of those problems is that they can all be solved by two alternative strategies that we named *step by step* strategy and *difference* strategy. The *step by step* strategy requires to calculate Part2 before determining whether Part3 or Whole2 (calculating that the price of the *exercise book* is 6 Euros in the previous example). The *difference* strategy does not require to calculate the common part and is based on the fact that if two sets share a common part, then their wholes differ by the same value as do the specific parts (the price of the pen and the price of the scissors differ by the same value as the total prices paid). It has to be noted that, if in complementation problems both strategies are in two steps, in the case of the comparison problem, the *step by step* strategy require three steps whereas the *difference* strategy requires only one. There exists as well a *mixed* strategy, that leads to the correct result even though it involves a non useful calculation; it starts with the calculation of Part 2 and ends with the *difference* strategy.

The solving model used for DIANE is composed of the following triple RM=(T, S, H). T refers to the problem Type and depends on the three parameters defined above (kind of problem, nature of the question, nature of the variable). S refers to the Strategy at hand (*step by step, difference* or *mixed* strategy). H refers to the Heuristics used and is mostly used to model the erroneous resolution; for instance applying an arithmetic operator to the last data of the problem and the result of the intermediate calculation.

3. Description of DIANE

DIANE is a web based application relying on open source technologies. DIANE is composed of an administrator interface dedicated to the researcher or the teacher and of a problem solving interface dedicated to the pupil. The administrator interface allows the user to add problems, according to the factors defined above, to create series of exercises, to look for the protocol of a student, or to download the results of a diagnosis. The role of the problem solving interface is to enable the pupil to solve a series of problems that will be analyzed later on and will be the basis for the diagnosis. This interface (Figure 1) provides some functions aimed at facilitating the calculation and writing parts of the process in order to let the pupil concentrate on the problem solving. The use of the keyboard is optional: all the problems can be solved by using the mouse only. The answers of the pupils are a mix of algebraic expressions and natural language. All the words which are necessary to write an answer are present in the text; the words were made clickable for this purpose. Using only the words of the problem for writing the solution helps to work with a restrained lexicon and avoids typing and spelling mistakes; it allows us to analyze a constrained natural language.

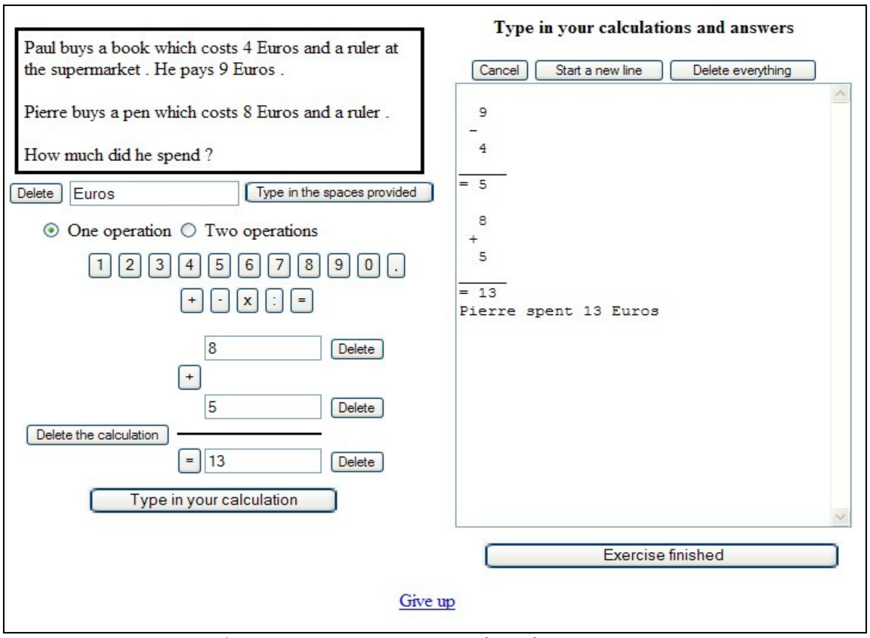

Figure 1. *The pupil interface*

4. Diagnosis with DIANE

Diagnosis with DIANE is a tool for analyzing and understanding the behavior of the learners at a detailed level when they solve arithmetic problems. The diagnosis is generic in that it might be applied to all the classes of problems that are defined and is not influenced by the surface features of the exercises. Diagnosis concerns not only success or failure or the different kinds of successful strategies, but erroneous results are coded at the same detailed level as the successful strategies. As we have already mentioned, our main rationale is that understanding the influence of representation on problem solving requires the analysis of behavior at a very detailed level. Note that more than half of the modalities of the table of analysis are used for encoding errors.

Diagnosis is reported in a 18 column table. Depending on the strategies and the nature of the problem up to 14 columns are effectively used for analyzing one particular resolution. The first column encodes the strategy. It is followed by several groups of four columns. The first column of each group encodes the procedure (addition, subtraction, etc), the second one indicates whether the data are relevant, the third one indicates whether the result is correct and the fourth one indicates whether a sentence is formulated and evaluates the sentence (this column is not yet encoded automatically). Another column, the 14^{th} is used to identify the nature of what is calculated in the last step of the resolution (a part, a whole, the result of a comparison, an operation involving the intermediary result and the last item of data, etc.)

The answer of the pupil, a string of characters, is treated following the pattern of regular expressions. This treatment turns the answer of the pupil into four tables, which are used for the analysis. The first table contains all the numbers included in the answer, the second one contains all the operations, the third one all numbers that are not operands and the fourth one contains all the words separated by spaces.

The data extracted or inferred from the problem (Whole1, Part1, Part3 …) are stored in a database. The automatic diagnosis is based on comparisons between the data extracted and inferred from the text and the tables, through using heuristics derived from the table of analysis.

The following table (Table 1) provides two examples of diagnosis for the problem:

John bought a 8-Euro pen and an exercise book. He paid 14 Euros. Paul bought an exercise book and scissors. He paid 3 Euros less than John. How much are the scissors?

Pupil 1		Pupil 2	
Response	Diagnosis by DIANE	Response	Diagnosis by DIANE
14 - 8 = 7 14 - 3 = 11 11 - 7 = 4 The scissors cost 4 Euros	Col 1: *step by step* strategy Col 2-4: subtraction, relevant data, calculation error Col 6-8: subtraction, relevant data, exact result Col 14: calculation of a part Col 15-17: subtraction, relevant data (the calculation error is taken into account), exact result	14 - 8 = 6 14 - 3 = 11 The scissors cost 11 Euros	Col 1: Erroneous *comparison* strategy Col 2-4: subtraction, relevant data, exact result Col 14: calculation of comparison Col 15-17: subtraction, data correct for the comparison but not for the solution, exact result

Table 1: An example of Diagnosis with DIANE

DIANE provides a fine grained diagnosis that identifies the errors made by the pupils. For instance, pupil 1 (Table 1) made a calculation mistake when calculating Part 2 (14-8=7), which implies an erroneous value for the solution (11-7=4). DIANE indicates that an item of data is incorrect in the last calculation due to a calculation error at the first step. The same holds true for erroneous strategies. Pupil 2 (Table 1), after having performed a correct first step ends his/her resolution with the calculation of the comparison (14-3=11). In this situation, DIANE diagnosis indicates that the pupil used an erroneous strategy that provided a result which is correct for the calculation of the comparison but not for the solution. This situation is a case of use of the heuristic previously described (using the last data and the result of the intermediate calculation).

5. Results from experimental psychology

Experimentation has been conducted on a large scale [12]; 402 pupils (168 5^{th} graders, 234 6^{th} graders) from 15 schools in Paris and the Toulouse area participating. The experimental design was the following: each child solved, within two sessions, complementation and comparison problems for three kinds of variables and the two kinds of questions, that is twelve problems. Even if the experimental results are not the main scope of this paper, let us mention that the main hypotheses were confirmed (for each of the four families of problems, we found a main effect of the kind of variable on the score of success ($17,79<F(2, 401)<51,12$; $p<0.0001$ for all the analyses). As predicted, we also found that cardinal variables made combination problems easier and ordinal variables made comparison problems easier. Furthermore, similar results were observed concerning the strategies at hand: strategies were highly dependent on the variable involves. For instance, in a comparison problem in which the variable was an age, 64% of the pupils used a strategy that did not require to calculate the intermediate part. Conversely, for the isomorphic problem in which the variable was a price, only 4% did so. We were also able to generalize our results to a larger scale of variables [5]. The table of analysis, on which DIANE's diagnosis is based was tested manually on those protocols. Except that human coding requiring a long training period for the coder, was slow and difficult, results were very satisfactory: (i) between judge agreement was always more than 95% for well trained coders for all the samples that we tested, and (ii) the detailed level of description made it possible to distinguish between and to embrace a large variety of behaviors.

6. Assessment of the quality of DIANE's diagnosis

In order to assess the quality of the automatic diagnosis, we carried out two experiments.

For the first one, we typed the protocols issued from a pen and pencil experiment in a 5th grade class [12] with 29 pupils. Each protocol included 12 problems, thus we analyzed 308 productions. In the second one, the experimentation was conducted directly with the interface and concerned 46 pupils from one 5th grade class and one 6th grade class. Each of the children solved 6 problems in this situation [3] and we analyzed 276 productions. For this second situation we might note that no difficulty due to the use of the interface was identified neither by the children nor by the experimenter; the interface was very easily used and well accepted by the children. The main experimental measures provided no significant results concerning the success rate or the strategy used between the two experiments [3]. However, the question of the difference of behavior between the pen and pencil situation and the interface situation will be looked at more deeply in forthcoming studies.

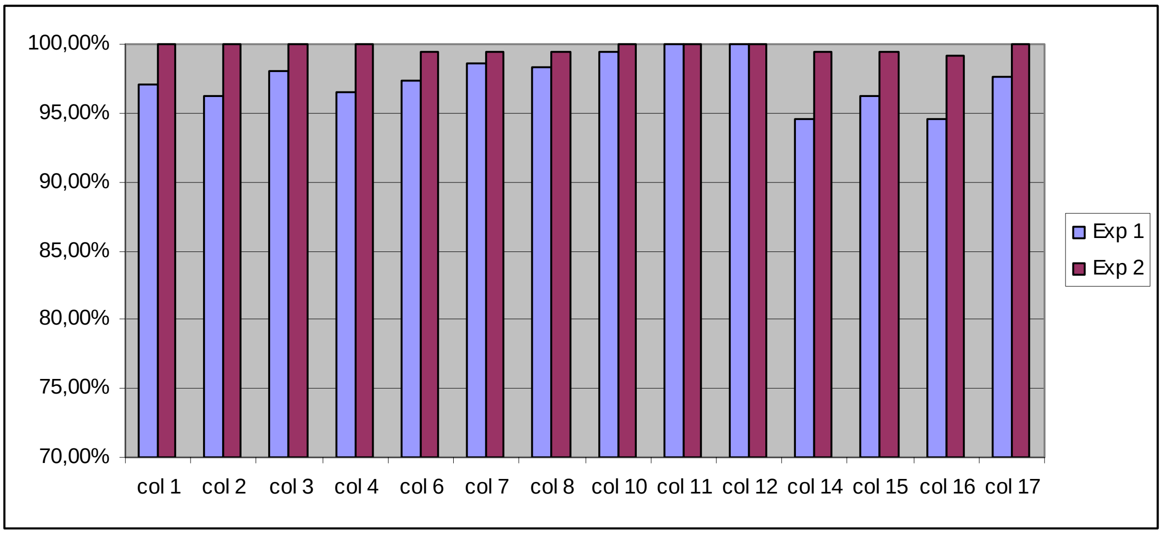

Figure 2. Distribution of rates of equality between automatic and manual coding

Figure 2 illustrates that for all the columns, and for each of the two experiments, the rate of equality between the manual encoding and the automatic one was between 94.5 and 100%. Furthermore, for 93.4% of the problems, automatic encoding is equal to manual encoding for all the columns encoded. Thus, these two experiments confirmed that DIANE is actually able to make a diagnosis of a quality close to the manual one.

7. Conclusion and perspectives

In this paper, we introduced DIANE, an environment aimed at diagnosing at a detailed level arithmetical word problems and currently specialized in four families of two-steps additive combination and comparison problems. These problems were designed in order to make it possible to test hypotheses on cognitive mechanisms involved in arithmetical problem solving that have direct educational implications. We are now working in three directions with DIANE. (i) We want to build a fine-grained typology of the strategies based on DIANE's diagnosis that will serve as a basis for the remediation module. We already constructed [12] a manual typology that includes all the successful strategies, and nearly 80% of the erroneous ones. (ii) We want to produce diagnoses that are straightforwardly understandable by teachers. The diagnosis produced by DIANE provides information on the solving process that teachers found very informative. We are now working on a module that will produce the diagnosis in natural language. (iii) We plan to enlarge the range of problems considered by DIANE: all the one-step additive word problems can be nearly readily integrated, and this diagnosis will be involved in the remediation module.

References

[1] Susan M. BARNETT; Stephen J. CECI (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological bulletin, vol. 128, no 4, pp. 612 - 637

[2] Bassok, M. Wu, L. & Olseth, L.K. (1995), Judging a book by its cover: Interpretative effects of content on problem solving transfer. Memory & Cognition, 23, 354-367.

[3] Calestroupat, J., Catégorisation d'interprétations conduisant à des erreurs dans la résolution de problèmes arithmétiques, Mémoire de DEA, Université de Paris 8, 2004.

[4] Cummins, D.D., Kintsch, W., Reusser, K. & Weimer, R. «The role of understanding in solving words problems». *Cognitive Psychology*, vol. 20, p. 405-438.

[5] Gamo, S. Aspects sémantiques et rôle de l'amorçage dans la résolution de Problèmes additifs à étapes. Mémoire de DEA. Université de Paris 8

[6] Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. Child Development, 54, 84-90.

[7] Kintsch, W. & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92,1, 109-129.

[8] Novick, L. (1988), Analogical transfer, problem similariry, and expertise. Journal of Experimental Psychology: Learning, Memory, & Cognition, 14, 510-520.

[9] Reusser, K., (1989). From text to situation to equation Cognitive simulation of understanding and solving mathematical word problems. European Journal in an International Context, 2, 2, 477-498.

[10] Riley, M.S., Greeno, J.G. &Heller, J.I. (1983). Development of problem solving ability in arithmetic. In H.P. Ginsberg(Ed.)

[11] Ross, B.H. (1989), Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. Journal of Experimental Psychology: Learning, Memory, & Cognition, 15, 456-468.

[12] Sander, E., Levrat, B., Brissiaud, R., Porcheron, P., Richard, R (2003). Conceptualisation et propriétés sémantiques des situations dans la résolution de problèmes arithmétiques : rapport d'étape. Ministère de la Recherche ; appel d'offre 2002 Ecole et Sciences Cognitives : les apprentissages et leurs dysfonctionnements.

[13] Sander, E., & Richard, J.-F. (1998). Analogy making as a categorization and an abstraction process. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.) Advances in analogy research: Integration of theory and data from the cognitive, computational and neural sciences (pp. 381-389). Sofia: NBU Series in Cognitive Sciences.

[14] Sophian, C, & McCorgray, P. (1994) Part-whole knowledge and early arithmetic problem solving. Cognition and Instruction, 12, 1, 3-33