Educ Stud Math (2009) 71:19–41
DOI 10.1007/s10649-008-9157-x

From arithmetical thought to algebraic thought: The role of the "variable"

Elsa Malisani · Filippo Spagnolo

Published online: 15 October 2008

Abstract The introduction of the concept of the variable represents a critical point in the arithmetic–algebraic transition. This concept is complex because it is used with different meanings in different situations. Its management depends on the particular way of using it in problem-solving. The aim of this paper was to analyse whether the notion of "unknown" interferes with the interpretation of the variable "in functional relation" and the kinds of languages used by the students in problem-solving. We also wanted to study the concept of the variable in the process of translation from algebraic language into natural language. We present two experimental studies. In the first one, we administered a questionnaire to 111 students aged 16–19 years. Drawing on the conclusions of this research we carried out the second study with two pairs of students aged 16–17 years.

Keywords Arithmetic–algebra transition · Concept of variable · Unknown · Variable in functional relation · Translation from algebraic language into natural language

1 Introduction

Since the early 1980s, all over the world, numerous researchers have studied the obstacles that students meet when they begin to study algebra. The results of many of these studies have already become part of the culture of researchers in mathematics education. We will refer here only to those results that are significant to our work.

At school, arithmetic does not operate at the same level of abstraction as algebra does, although they both involve written symbols and an understanding of operations (Herscovics

This research was carried out in the different semiotic contexts of Algebra and Analytic Geometry

E. Malisani · F. Spagnolo (✉)
G.R.I.M. (Group of Research in Mathematics Education), Department of Mathematics,
University of Palermo, Palermo, Italy
e-mail: spagnolo@math.unipa.it
URL: http://www.math.unipa.it/~grim

& Linchevski 1994). Arithmetic is habitually associated with numbers and numerical computations (Sfard & Linchevski 1994). Solving an arithmetic problem means to carrry out one or more operations with specific data for reaching a solution (almost always unique), proceeding in a sequential way and using fundamentally natural language that semantic control facilitates.

Elementary algebra, on the other hand, involves making explicit the relationships between the unknown and the data of a problem, and then moves on to a relatively automatic manipulation of these, in order to reach a solution (Sadovsky & Sessa 2005). Algebra focuses on relationships among quantities, using symbolic language that hampers the control, showing a certain semantic weakness.

Certainly, the use of a suitable symbolism facilitates algebraic thinking, but in the phase of arithmetic–algebra transition it is possible to think algebraically without letters (Linchevski 1995). Radford (2003, 2006) claims not only that: "Using letters does not amount to doing algebra", but also that it is possible to distinguish algebraic thinking in terms of layers of generality (factual, contextual and symbolic).

However, one of the fundamental problems of learning algebra remains *making sense* of letters. A letter is a sign, something that is designated to mean something else. The letters "x", "y", "n" designate particular objects, namely variables. A variable is not a number in an arithmetic sense, e.g. the number 8 does not vary. A variable is an algebraic object that can be replaced by a number (Bardini, Radford, & Sabena 2005). But the notion of a variable has a plurality of conceptions: *generalised number* (an indeterminate number that appears in generalisations and in general methods); *unknown* (a specific number but unidentified, its value could be calculated with consideration of the restrictions of the problem); "in functional relation" (relation of the variation to other variables); *totally arbitrary sign* (it appears in the study of algebraic structures); and *register of memory* (in informatics; Usiskin 1988). The variable in functional relation is simply named "variable" by Küchemann (1981), who defines it as follows: "the letter is seen as representing a range of unspecified values and a systematic relationship is seen to exist between two such sets of values". We shall define a variable in functional relation as a "thing that varies". This last expression chiefly points out that the letter can represent a range of unspecified values, even if the systematic relation is not rendered explicitly yet. From a historical point of view this notion precedes the conception of variable in functional relation[1].

Denotation represents a difficulty for the students: the same symbols are used for indicating different conceptions of the variable, e.g. x can represent an unknown, a generalised number or an independent variable in a function. Different symbols can be used for representing the same conception of the variable, e.g. x, y or z can denote generalised numbers (Matz 1982). This contributes to dispersing the differences among the various conceptions and to hiding the conditions that determine where and how the value of the variable can vary.

[1] According to Collins English Dictionary (1991), *variable* is: "an expression that can be assigned any of set of values. (As modifier) able to take any of a range of values: a variable sum. A symbol, such as x, y or z, representing an unspecified member of a class of objects, numbers etc. Variables may be used either existentially or universally: in elementary algebra variables occur in *conditional* equations representing unknown quantities of which the values are to be found….". *Unknown* is: "a variable, or the quantity it represents, whose value is to be discovered by solving an equation; a variable in a conditional equation". Freudenthal (1983) also groups under the name *variable*: the unknown, the variable in functional relation and the "polyvalent names", that is, those that can take a set of different values, but that do not vary according to a functional dependence.

Some researchers (Matz 1982; Wagner 1981, 1983) consider that the concept of variable represents a critical point of the arithmetic–algebra transition. This concept is complex because it is used with different meanings in different situations. Its management depends on the particular way of using it in problem-solving. An appropriate approach should allow the student to individualise the useful conceptions to solve a problem and to manipulate these different notions in the solving procedure, to find how and in which set the value of the variable can vary. Then the students should be able to interpret in different ways the symbols that represent the variable and to be able to pass from one interpretation to another with flexibility, in relation to the requirements of the problem to solve.

Bardini *et al.* (2005), however, demonstrate that some students meet difficulty in using the concept of the variable and they consider it to be a *temporally indeterminate number* whose fate is to be determined at a certain point. Then, for these students, the letter is an *index* (Radford 2003), a sign that indicates the place that an actual number will occupy in a *process* (Sfard 1991) temporarily in abeyance.

Other studies show that the students use different conceptions to interpret the letters that represent the variables of a problem (Küchemann 1980). We observed that students (aged 11–12 and 14–15 years) spontaneously consider the letters as: numerical values, constants, unknowns, “things that vary”, even if they do not master the symbolic standard algebraic language (Malisani & Spagnolo 2005).

The results of several research studies suggest, however, that the different conceptions of the variable have different degrees of difficulty for the students. The “letter as an unknown” is simpler (Küchemann 1980), and even 8- and 9-year-old students can attain a certain understanding of the algebraic unknown (Carraher, Schliemann, & Brizuela 2001). The conceptions of the generalised number and the variable in functional relation, instead, seem to result in greater difficulties. Küchemann (1981) showed that most students between 13 and 15 years of age treat the letters in expressions or in equations as specific unknowns before they treat them as generalised numbers or variables in functional relation. Trigueros, Reyes, Ursini, and Quintero (1996) demonstrated that students beginning university have a fairly poor conception of the variable in functional relation and the generalised number. The students chiefly have difficulty in understanding the variation in a dynamic form, that is, the relation of the variation to other variables. The obstacles are greater when the resolution of the questions does not take place by manipulation, but through interpretation and symbolisation.

Some studies point out that the conception of the unknown could interfere with the variable in functional relation (Janvier, Charbonneau, & Cotret 1989). Panizza, Sadovsky, and Sessa (1999) showed that the notion of the unknown would not be effective in interpreting the role of the letters in a linear equation with two variables, because this equation is not recognised by the students as an object that defines a set of pairs of infinite numbers. Instead, if the student uses the concept of function, he/she can calculate different solutions more easily.

1.1 Historical-epistemological analysis

Several experimental studies (Harper 1987; Sfard 1992; Sierpinska 1994) seem to confirm that some of the students’ difficulties can be grouped around some obstacles that have been met in history. The historical studies carried out by Malisani (1996, 1999, 2006) show that for many centuries algebra lagged behind in comparison with geometry, with its symbolism experiencing difficulty in gaining acceptance (see Chevallard 1989). These studies highlight the principal conceptions, the precursory procedures, the passage from one concept to the other and, particularly, the passage through the linguistic levels of the different phases: rhetoric, syncopated and symbolic[2]. These conclusions can be functional in the research of the epistemological obstacles (Sierpinska 1994) to the arithmetic–algebra transition. For the implementation of our experimental work we consider that the following conclusions are very important:

- The construction of symbolic language was very slow and difficult: passing from certain *names* denoting the unknown and certain relations, through the *abbreviations of these words*, through the *intermediary codes* between rhetorical language and syncopated language, and finally to the *symbols*. In the process of the elaboration of a syntactically correct and operationally efficient algebraic symbolism, the progressive abandonment of natural language as a mediator of expression is observed.
- The absence of an adequate algebraic language conditioned the evolution of the solving procedures, because mathematicians had to have recourse to other languages—natural, arithmetic or geometric—that were semantically richer than the algebraic language. These languages allowed a problem to be interpreted adequately, its solution to be obtained, the steps to be justified algebraically and the rules to be formulated. With the elaboration of a more adequate algebraic language, the support languages were gradually abandoned. That is, from the historical viewpoint it is possible to observe the passage from "elementary mathematics" to "advanced mathematics". In the former, perceptive aspects and a strong use of natural language prevail; in the latter, instead, formal aspects and demonstrative activity prevail (see Harel and Tall 1991).
- In the construction of algebraic language, two levels of conception of the generality of a method exist: one regarding the feasibility of applying it to a plurality of specific cases, and the other one concerning the possibility of expressing it through the language of symbolic algebra.
- The visual representative registers are present in the different solving procedures that use geometric language (Euclid, al-Khowârizmî, al-Khayyam etc.), but also in some arithmetic strategy (double false position) and analytic method (al-Tusi).
- The historical study (Malisani 2006) also highlights the differences between the notions of the unknown and the variable as a thing that varies. These two conceptions have a totally different genesis and evolution. Radford (1996) considers that their processes of conceptualisations seem to be entirely different, even if both notions deal with numbers.
- The notion of the unknown has its origin in the resolution of problems that ask for the calculation of one or more quantities. It was introduced by Diophantus with the name "*arithmos*", that is, the number of the problem. The preponderance of this notion in the solving procedures is notable up to 1600, although some attempts to consider the dependence among variables are recorded in Diophantus and al-Tusi (Malisani 2006).

[2] Nesselman (1842) individualises three periods in the evolution of algebraic symbolism: the **rhetorical phase**—anterior to Diophantus in Alexandria (250 AD), in which natural language was used exclusively, without resorting to any signs; the **syncopated phase**—from Diophantus up to the end of the 16th century, in which some abbreviations for the unknown and the relations in more frequent use were introduced, but the calculations were performed in natural language; and the **symbolic phase**—introduced by Viète (1540–1603), in which letters are used for all the quantities and signs to represent the operations, symbolic language is utilised not only to solve equations but also to demonstrate general rules.

- The notion of the variable as a thing that varies is very ancient, but it is difficult to establish exactly the origin of this conception in the history of algebra. Its evolution is very slow: from the relationship among the numbers contained in the tables (Babylonians, Tolomeo, Indians); through the dynamic, but discrete, quantities reported by the concept of formula (Diophantus); through the variable connected with continuous quantities in the study of the physical laws (Oresme, Galileo); through the curves described in kinematical terms (17th century); through the description of the relation among variables (Gregory, Newton, Leibniz, Bernoulli, Cauchy) that leads precisely to the concept of function (Malisani 2006).

1.2 Some considerations of algebra

The historical-epistemological survey provides evidence of the variety of problems that have been historically tackled through a diversity of methods that we now group together under the label of "algebra". Many contemporary mathematics educators still maintain that the conceptions of the *unknown* and the variable as a thing that varies are the same, mainly because they represent unknown numbers and—on the level of symbols—they can be manipulated in the same way. However, the historical-epistemological analysis makes it possible to show, in a decisive way, that they belong to two different conceptual formations (see Radford 1996), unified under the name "variable". It is interesting to note that the historical-epistemological survey allows a subtle interplay of different semiotic systems to be shown, to give us an idea of the level of semantic reduction that the contemporary algebraic symbolism imposes on its users.

From a historical viewpoint it is possible to observe that the origins of algebraic thinking are individualised in an effort to apply a computational process to more specific cases (first level of generalisation). At this level, algebraic symbolism does not represent the only tool that carries out processes of generalisation. According to this vision, algebraic thinking starts before the symbolism (see Arzarello *et al.* 1994; Linchevski 1995). Radford (2006) claims that algebra not only uses letters, it rather uses letters or other signs (spoken words, gestures etc.) to think in a certain way.

Therefore, we think that the idea of considering "algebra as generalised arithmetic" is too reductive, because it is intended to ignore the influence that other languages have had on the development of algebra. Radford (1996) considers that algebra is much more than generalised arithmetic and that the algebra we know owes a lot to geometry too (to witness, the term *square* root).

2 Purpose of the research

The mathematical concept of the variable can have various aspects. Its study constitutes a very wide field of research. Therefore, it is necessary to circumscribe the dominion of the survey.

The aim of this research was to analyse whether the notion of unknown interferes with the interpretation of the variable in functional relation and the kinds of language used by students in problem-solving. We also wanted to study the concept of the variable in the process of translation from algebraic language into natural language.

The semantics of algebraic language is less rich than the semantics of natural, arithmetic or geometric languages. Therefore, in the syncopated phase it is necessary to use these languages to operate algebraically. Hence, this analysis will take place in the semiotic contexts of Algebra and Analytic Geometry[3].

2.1 Research questions

Beginning from the review of literature and the reflections on the epistemological and historical-epistemological representations carried out by Malisani (1996, 1999, 2006), we planned two experimental studies. We formulated four research questions:

- **Group of questions number 1 (Q_1):** which conception of the variable do the students use in the context of a situation/problem[4]? Can they evoke more conceptions? Does the notion of the unknown interfere with that of variable in functional relation? Is the evoked conception dependent on the situation/problem? Does it depend on the context? From a historical viewpoint, the evolution of symbolic language was very slow and difficult. In the absence of an adequate algebraic language, the mathematicians must have recourse to other languages, arithmetic, natural or geometric, that are semantically richer.
- **Question number 2 (Q_2):** if the students do not master the symbolic standard algebraic language, which language do they use for problem-solving?
  We think that the formulation of a problem could reveal certain difficulties the students have of interpreting the concept of the variable that is not observable through the solution.
- **Group of questions number 3 (Q_3):** do the students succeed in finding a context giving meaning to the equation? Do they use the conception of the unknown or the variable in functional relation? Do they succeed in identifying the variables? How do they interpret the terms of a linear equation?
  Historically, the notion of the variable as a thing that varies appears with the use of tables (astronomic, trigonometric), the graphic representation of some physical laws of motion, the curves described in kinematical terms (trajectory of a mobile point) etc. From a historical viewpoint this notion is closely connected to visual representative registers.
- **Question number 4 (Q_4):** from the student's viewpoint, does this connection exist? That is, would the presence of visual registers allow the conception of the variable in functional relation to be evoked and/or used more easily in problem-solving?

2.2 Problems

To carry out this research we chose the linear equation in two variables for two reasons: first, because it represents a nodal point from which the students derive the conceptions of the letters as unknowns or things that vary. Second, this kind of equation is well known by

[3] The digital technologies (computer algebra systems, graphic symbolic calculators, microworlds etc.) play a very important role in students' development of at least one kind of use of algebraic language in general and the idea of the variable in particular (Hoyles, Noss, & Adamson 2002; Lagrange and Chiappini 2007 etc.). A lot of research confirms this viewpoint, but in this paper we chose not to include the dimension of digital technologies. It will be the aim of future research.

[4] By situation/problem we mean a learning situation in which: there are initial data that specify the context and that are useful for solving the problem; there is an aim to pursue that makes sense of the mobilisation and the organisation of the things learned; there are some obligations where the obstacles to be overcome require a reorganisation of a student's knowledge and that bring him/her to finding other means and therefore to new learning; the procedure and the solution are not evident—the student must do some research to know how to proceed.

the students, because they have studied it from different viewpoints: linear function, equation of a straight line and as a component of linear systems.

The questionnaire presents four problems. The **first** of them is a situation/problem of "money and bets". We also asked the students to think over the solution set. The aim of this problem is to find answers to $\mathbf{Q_1}$ and $\mathbf{Q_2}$.

Charles and Lucy won the total sum of € 300 in the lottery. We know that Charles won the triple of the betted money, while Lucy won the quadruple of her own.

a) Calculate[5] the sums of money that Charles and Lucy betted. Comment on the procedure that you have followed.
b) How many possible solutions are there? Give reasons for your answer.

The **second query** asks for the formulation of a problem. This must be solved by means of a given equation; namely, the student must translate from algebraic language into natural one. The aim of this query is to find answers to $\mathbf{Q_3}$.

Invent a possible situation problem that could be solved using the following relation of equality: $6x-3y=18$. Comment on the procedure that you have followed.

The **third query** is a checking question. We want to compare the models and the contexts evoked by the students, from the two relations of equality listed below, with those activated through the particular equations of problems 1 and 4. The aim of this query is to find answers to $\mathbf{Q_1}$ (second part).

What is it? Interpret with a "short answer" the following expressions:

a) $ax+by+c=0$
b) $y=mx+q$

The **fourth question** is a situation/problem of "monthly telephone expenses". We also asked the students to think over the solution set.

To use the telephone of another person, a man arranges to pay a monthly fee of € 5 and in addition € 2 per hour for the phone calls he actually makes.
Let x be the number of monthly hours of phone calls made and let y be the total sum that he pays monthly.

a) Establish which type of relation intervenes between x and y and represent it graphically in the Cartesian plane.
b) Calculate[5] the total sum that he pays monthly and the number of hours monthly of phone calls made. Give reasons for your answer. How many possible solutions are there? Give reasons for your answer.

While in the first problem the student was free to choose the solving context, in this one, instead, the context is given (analytic geometry). We want to compare the results of problems 1 and 4 for analysing the influence of the context. Namely, the aim of this problem is to find answers to $\mathbf{Q_4}$.

[5] In Italian we used the term "*determinare*", in the sense of "determine through the calculation" or "calculate".

3 First experimental study

3.1 Methodology of research

One hundred and eleven students between the ages of 16 and 18 of the Experimental High School of the city of Ribera, Italy participated in this research. The students worked individually, and we did not allow them to consult books or notes. They were given 60 min.

Qualitative and quantitative analyses are necessary for educational research. Each of them is useful for integrating with the other. The methodology that we introduce here uses the quantitative tool because the number of variables to be checked is very large. Thus, the quantitative tool can demonstrate variables or groups of variables that might be impossible to examine using only qualitative analysis. To carry out the quantitative analysis, it is necessary to do an a priori analysis[6] on the four questions of the questionnaire.

The methodological framework relative to the a priori analysis is the Theory of Didactic Situations of G. Brousseau (1997). As regards the statistical methods adopted, we make reference to the works of Gras (1997, 2000), Gras *et al.* (1996, 2001, 2005) and Spagnolo, Gras, Suzuki, and Guillet[7] (2008) with regard to the implicative and factorial analysis.

The aim of the a priori analysis was to individualise all the possible strategies that the students could use, the calculation of the solution set (number of solutions, bounds imposed by the context), the kinds of language utilised, any errors that the students might make etc. (see Appendix 1).

We filled in the table with a double input “students/experimental variables”. For every student, we indicated with the value “1” the experimental variables (or strategy) that he/she used in the answer; with the value “0” for the experimental variables that he/she did not apply.

The data were analysed in a quantitative way, using the software of inferential statistics CHIC 2000 (Classification Hiérarchique Implicative et Cohésitive) and the factorial statistical survey SPSS (Statistical Package for Social Sciences). We introduced five supplementary variables in the “students” component[8].

[6] The “a priori analysis” is the analysis of the “Epistemological Representations”, “Historical-epistemological Representations” and “Supposed Behaviours”, correct and incorrect, to solve a given didactic situation. The epistemological representations are the representations of the possible cognitive paths regarding a particular concept. Such representations can be prepared by a beginner subject or by a scientific community in a specific historical period. Historical-epistemological representations are the representations of the possible cognitive paths regarding the syntactic, semantic and pragmatic reconstruction of a specific concept. The supposed behaviours of students in facing the situation/problem are all the possible strategies of the solution, both correct and incorrect. Among the erroneous strategies, those that can become correct strategies will be taken into account (Spagnolo 2006).

[7] In particular, Spagnolo *et al.* (2008) demonstrate the theoretical and methodological questions in didactic research. They also introduce experimental works that show the effectiveness of some of the new methodologies used in this article (supplementary variable etc.).

[8] The supplementary variables represent different cognitive styles and they allow the experimental data to be analysed in a better way. The introduction of supplementary variables as ideal individuals was used in numerous experimental works of the GRIM: Spagnolo (1998, 2006, 2008). These works led us to validate this method both experimentally and theoretically. In this context, with the high number of variables in play, such an investigative method allows better highlighting of the fundamental characteristics of the a priori analysis.

3.1.1 Example of classification of the students' answers

The student “3A14” answers Problem 1 in this way:

> “300/2 = € 150 (A)
> 150. 1/3 = € 50 (B)
> 150. 1/4 = € 37.50 (C)
> I have divided the money equally between Charles and Lucy (A).
> I have multiplied by 1/3 in the case of Charles (B) and by 1/4 in the case of Lucy (C) and I have reached the conclusion that Charles bets € 50, while Lucy bets € 37.50. The number of possible solutions, in my opinion, is 1 (one) from the calculations reported above.”

From the a priori analysis, we find the following experimental variables in this answer: AL1 (the student answers), AL2 (he shows a procedure in natural language), AL4 (he adds a datum), AL4.1 (he considers the winnings to be divided in half), ALb1: (he calculates the solution set) and ALb2 (he shows one particular solution to the equation). In the table “students/experimental variables”, we assign the “value 1” to these experimental variables, and the “value 0” to all the others.

3.2 Statistical methodology and analyses of the results

3.2.1 Implicative graph of the first problem

The implicative graph was created with the software CHIC 2000. This shows three well-defined groups of the experimental variables (with statistical percentages of 95% and 99%). They are indicated by the dots cloud, the vertical lines cloud and the net cloud (the horizontal lines cloud indicates the intersection between the vertical lines cloud and the net cloud). A different kind of strategy used by the students corresponds to every group (Fig. 1):

- ***Procedure in natural language*** (dots cloud): If the students apply this procedure (AL2), they add a datum (AL4) considering the winnings to be equal (AL4.1 or AL4.2) or that the bets are equal (AL4.3). Thus, the student transforms the question into a typical arithmetical problem. He/she solves it and calculates the solution set (ALb1), finding only one particular solution to the equation (ALb2). This procedure is the one most frequently used by the students[9] and it leads to the single solution. So the predominant conception of a variable is that of an *unknown*. Precisely, “adding a datum” is equivalent to introducing a new equation and thus, with the equation of the problem, to form a system of two linear equations. The solution of this system is a “particular solution of the equation of the problem”[10].
- ***Method by trial and error in natural language or in half-formalised language*** (vertical lines cloud): the student generally assigns several values to a variable (e.g.

[9] The chief experimental variables AL2, AL4, AL4.1 and ALb.2, which define this procedure, have greater frequencies than the variables of the other two procedures (see the table of frequencies in Appendix 2).

[10] The experimental variable **AL4** “he/she adds a datum” considers two possibilities: equal winnings (AL4.1 and AL4.2) or equal bets (AL4.3). This is equivalent to introducing a new equation and to forming a system of two linear equations: “The winnings of € 300 are divided in half” (AL4.1)—equivalent to the system $3x + 4y = 300$ and $3x = 4y = 150$; “The winnings are equal to € 300 for both the teenagers” (AL4.2)—corresponds to the system $3x = 300$ and $4y = 300$; “The bets are equal” (AL4.3)—is equivalent to the system $3x + 4y = 300$ and $x = y$.

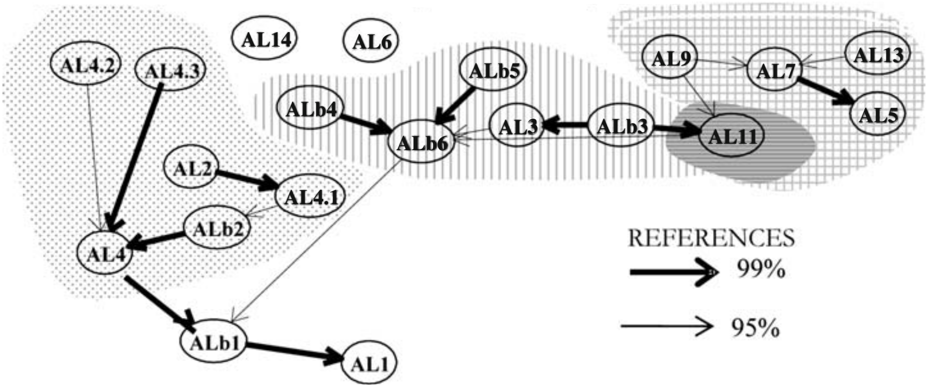


Fig. 1 Implicative graph of the first problem[11]

Charles' bet) and he/she finds the corresponding values in the other variable (Lucy's bet; AL3). Thus, the student shows some solutions of the equation (ALb3) and he/she considers the variables of the problem in functional relation in an implicit way (AL11). This method leads to many solutions (ALb4, ALb5, ALb6), and allows the dependence between the variables to be evoked.

- ***Pseudo-algebraic***[12] ***strategy*** (net cloud): the student translates the text of the problem into an equation of the first degree with two unknowns ($4x-3y = 300$; AL5). He/she applies an incorrect method where he/she writes one variable as a function of the other. He/she replaces this variable in the original equation and thus he/she obtains an identity (AL7). Since the student does not succeed in interpreting the identity, either he/she changes his/her solving procedure, abandoning the pseudo-algebraic one (AL9) or he/she resumes the resolution of the equation and he/she makes some errors trying to find only one solution (AL13). It is interesting to observe that, if the student abandons this strategy (AL9) then he/she considers the variable conception to be the thing that varies (AL11 connects the procedure by trial and error and the pseudo-algebraic strategy, indicated by the horizontal lines cloud). However, the pseudo-algebraic strategy is rarely used and it leads to the correct solution only in some cases, because it involves a purely syntactic manipulation of the equation as an interplay of signs without sense.

3.2.2 Students' profiles

A certain correspondence between the solving strategies used by the students and the conceptions of the variable as an unknown and a thing that varies emerges from the

[11] With this statistical methodology, e.g. the implication AL2→AL4.1 of 99% is interpreted in this way: if there is AL2 in the student's procedure, then with 99% probability the procedure will also include AL4.1.

[12] The purely syntactic manipulation of the equation does not lead to the correct solution of the problem; therefore, we use the term "pseudo-algebraic".

Table 1 Definition of the supplementary variables

Supplementary variables	Strategies
NAT	In natural language
FUNZ	By trial and error in natural language and/or half-formalised language
PALG1	Pseudo-algebraic+ resolution of the equation with errors of syntactic kind
PALG2	Pseudo-algebraic + other strategy
ALG	Algebraic

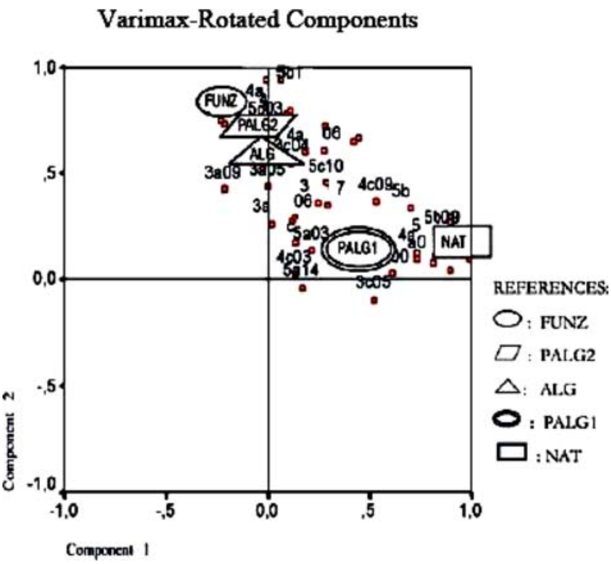
previous analysis. To examine carefully these results, we introduce five supplementary variables in the "students" component[13] (Table 1; Malisani, Scimone, & Spagnolo 2008).

3.3 Correspondence factor analysis of the first problem

We observe that the supplementary variables NAT and PALG1 assume a determining role. They strongly characterise the horizontal component. These near profiles lead to the single solution (conception of the *unknown*; Fig. 2).

The supplementary variables FUNZ, PALG2 and ALG form a cloud that strongly characterises the vertical component. The profile PALG2 is very near to FUNZ, because the student who abandons the pseudo-algebraic procedure generally adopts the method described

Fig. 2 Correspondence factor analysis



[13] The supplementary variables represent different cognitive styles and they allow the experimental data to be analysed in a better way. The introduction of supplementary variables as ideal individuals was used in numerous experimental works of the GRIM: Spagnolo (1998, 2006, 2008). These works led us to validate this method both experimentally and theoretically. In this context, with the high number of variables in play, such an investigative method allows better highlighting of the fundamental characteristics of the a priori analysis.

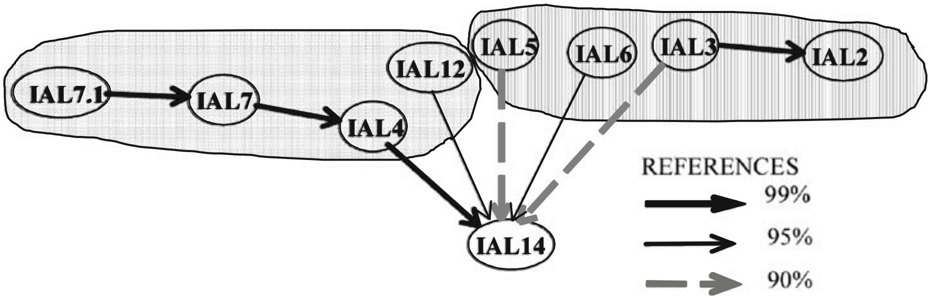


Fig. 3 Implicative graph of the second problem

in FUNZ. The winning strategies are precisely those described in the profiles FUNZ, PALG2 and ALG, which lead to multiple solutions[14] (variable *in functional relation*).

This corresponds strongly with the different conceptions of the variable. Thus, the horizontal axis represents the unknown, the vertical axis, instead, reproduces the variable as the thing that varies.

3.3.1 Answers to Q_1 and Q_2

The implicative graph shows two disjoint groups of experimental variables: the dots cloud related to the procedure in natural language (conception of the *unknown*) and the vertical lines cloud associated with the method of trial and error (variable in *functional relation*). The factorial analysis highlights the correspondence between the horizontal component and the conception of the *unknown*, the vertical axis and the variable as a thing that varies. Then, these results allow the first question to be answered, that is, in the context of a situation/problem, the students chiefly use the notion of the unknown. If they apply this conception, they do not evoke the variable in functional relation.

Moreover, we observe that the students predominantly use natural language as an expressive means to solve the first problem. They also use, in an explicit or implicit way, arithmetic language in a not purely algebraic context. Twenty-four percent of the students translate the problem into a first degree equation with two unknowns, but most of them (17%) do not use symbolic language to solve it. We observe a purely syntactic manipulation of the equation that leads to leaving the pseudo-algebraic procedure or to making errors trying to find only one solution. Then the students use natural and/or arithmetic language in a situation/problem, if they do not master the symbolic standard algebraic language (**Q_2**; see also Malisani 2006, pp. 98–100).

3.3.2 Implicative graph of the second problem and answer to Q_3

The formulation of a problem that must be solved by means of a given equation ($6x–3y = 18$) is a difficult exercise for the students, because only 60% answered the query: 7% correctly and 53% incorrectly.

[14] In this study we prefer to use the term "multiple solutions" rather than "infinite solutions", because we have not considered the possible connotations of the word "infinite". However, we defined the experimental variable ALb4 to take into account cases in which the student explicitly considers the existence of infinite solutions.

Different implicative paths are drawn in Fig. 3, but all have the same consequent: IAL14 “the student translates algebraic language with difficulty”. We notice two groups of experimental variables that correspond to two different strategies for solving the question. The first group (the vertical lines cloud) comprises the variables IAL2, IAL3, IAL5 and IAL6, which characterise the activity of the purely syntactic manipulation of the formula (write the equation into its explicit form, show several solutions, add another equation and form a system etc.). The second group (the dots cloud) contains the variable IAL12 and the route IAL7.1→ILA7→IAL4, which corresponds to the production of the text of a problem that is not meaningful for the given relation (Appendix 2).

Therefore, some students do not succeed in finding a context giving meaning to the equation (first group). Other students find a context, but either they do not succeed in identifying the variables (the text of a classical arithmetic problem with specific numerical values: IAL7) or they do not succeed in translating exactly the equation (IAL12). For example, the student “5B08” does not succeed in linking the coefficients with the variables, translating the equation $x - y = 18$:

> “Mark listened only to 6 songs on a CD; he listened only to 3 songs on another. If the difference in the songs recorded on the two CDs is 18, how many songs are recorded on the first and on the second CD?”

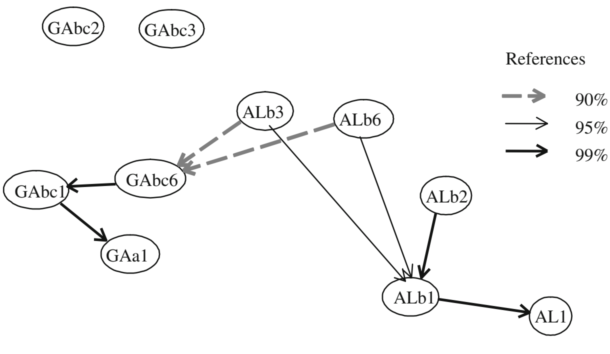
3.3.3 Results of the third query

In the third query, 76% of the students interpret the expression $ax + by + c = 0$ within the framework of analytic geometry (equation of a straight line 49%, a circle or a parabola 26% and a bundle of rays 1%), while 26% recall the algebraic context (linear equation with two variables 21% and polynomial 5%). For the expression $y = mx + q$, instead, all of the students refer to analytic geometry (equation of a straight line 68%, a bundle of rays 30% and a parabola 1%).

3.3.4 Comparison between the first and the fourth problem

In the first problem, a very important implicative link is ALb2→ALb1 (99%): the students, who calculate the solution set, consider that the equation has only one solution. The other links, ALb3→ALb1 and ALb6→ALb1 (95%), the students calculate more solutions or multiple solutions, instead, are less significant. This corresponds clearly to the percentages of the experimental variables: ALb2, ALb3 and ALb6 (Fig. 4; Appendix 2).

Fig. 4 Implicative graph—first and fourth problems



In the fourth problem, the student, who answers on the possible solutions, basically considers that the problem has a plurality of solutions (implicative link GAbc6→ GAbc1 of 99%). The experimental variables GAbc2 and GAbc3 ("he/she calculates one or several solutions to the equation"), instead, are not connected to the variable GAbc1 ("he/she calculates the solution set"). It is interesting to notice two implications that link the first problem to the fourth one: ALb3→GAbc6 and ALb6→GAbc6 (of 90%). In other words, if the student considers that the equation in the first question is verified by multiple solutions, then he/she also allows for multiple solutions in the fourth problem.

3.3.5 Answer to Q_4

We consider that the student understands the conception of variable in functional relation more easily, when he/she is able to allow for multiple solutions that satisfy the linear equation. This can be verified especially in the fourth problem in the presence of visual representative registers and when the students do not master the symbolic standard algebraic language. Thus, from the student's viewpoint the conception of the variable in functional relation is connected with visual representative registers.

3.4 Discussion

From the study with the implicative and the factorial analysis, we observe that the students chiefly use the conception of the unknown. If they apply this notion, they do not evoke the variable in functional relation ($\mathbf{Q_1}$). These results coincide with those found in Harel and Sowder (2005). We also deduce that the solving procedures are supported predominantly by natural or arithmetic language, when the students do not master the symbolic standard algebraic language ($\mathbf{Q_2}$).

It is interesting to observe that many students consider that the first situation/problem has only one solution (conception of the *unknown*). The fourth problem presents a concrete situation similar to the first one, but in the context of analytic geometry. For this query the students calculate the solution set directly considering multiple solutions (variable in *functional relation*). Thus, the evoked conception depends on the context ($\mathbf{Q_1}$).

Therefore, the students with insufficient mastery of symbolic standard language were able to consider multiple solutions more easily, in the presence of visual representative registers, by evoking the mental model of the equation of the straight line. Thus, the relationship between variables becomes "perceivable" through the graph and the student can more easily "visualise" multiple solutions ($\mathbf{Q_4}$).

The student is more inclined to consider the conception of the unknown (searching for the single solution of the linear equation) in the absence of representative graphic registers. In a few cases, although the conception of the unknown prevails, we observe the passage from the single solution to multiple solutions. Therefore, we were able to affirm that there is a certain interference of the notion of the unknown with the conception of the variable in functional relation ($\mathbf{Q_1}$). However, we believe that the matter must still be deepened analysing, in detail, the solving strategies used. We should investigate how the conceptions of the unknown and the thing that varies are activated in the process of problem-solving.

For the third question almost all the students interpreted the expressions $ax + by + c = 0$ and $y = mx + q$ within the framework of analytic geometry, but in the first problem, the model of the straight line has not been recalled with the linear equation and the graphic representation is totally absent in the solving process. This behaviour, called "avoidance of visualisation", has already been found in didactic research (see Eisenberg & Dreyfus 1991; Vinner 1989; Furinghetti & Somaglia 1994).

In this situation we think that the "avoidance of visualisation" is linked to a matter of didactic contract. Usually, problems with equations given at school are solved in an algebraic context using the conception of the unknown. Situations/problems generally are never solved within analytic geometry, recalling visual representative registers. The problems of analytic geometry given at school are different.

The students translate algebraic language with difficulty. Some of them carry out a purely syntactic manipulation of the formula. Other students, instead, are able to produce the text of a problem, which does not have a meaningful result for the given relation; at least they find a context, but they succeed in neither identifying the variables nor in translating exactly the equation (**Q_3**). Only 8 out of all 111 students, succeed in formulating a problem correctly and 6 of those chose a different context from the "money and bets" of the first problem.

4 Second experimental study

A series of questions emerged from the first experimental study for which we had not found any answer. These questions draw on those that we formulated at the beginning (see section 2.1):

Q5: How are the conception of the unknown and the variable in functional relation activated and used? Is the passage from one conception to the other possible? If so, how does it happen? The answers would allow us to understand if the notion of the unknown interferes with the conception of the variable in functional relation.

Q6: Which languages are used? Is symbolic language present? If so, with which function?

Q7: How does the process of translation from algebraic language into natural language come about? What difficulty do the students find in interpreting the concept of the variable in the process of translation?

To answer these questions we proposed a second experimental study.

4.1 Methodology of the research

Four students of aged 16–17 years of the Scientific Experimental High School of Ribera (Italy) participated in the second experiment. They had not participated in the first investigation.

The students worked in pairs: Serena with Graziella and Vita with Alessandra. They had to reach an agreement in their discussion before they could write. In order to answer the above-listed questions, it was sufficient to analyse the resolution of the first two queries of the questionnaire.

The team that conducted the interview was composed of two teachers: an interviewer and an observer. The first one had the assignment of explaining the problems and conducting the interview, the second one of taking note of all the elements that he thought important. The interviewer tried to stimulate the students only when they were in difficulty and in a neutral way. The whole interview was audio-recorded and afterwards it was transcribed.

4.2 Analyses of the results

4.2.1 First query

In the first protocol natural language prevails, the symbolic standard algebraic language is completely absent. The students use the conception of the unknown. They calculate two particular solutions by adding a datum (equal bets or equal winnings). This is equivalent to solving two linear systems of equations:

1. $3x + 4y = 300$ and $x = y$
2. $3x + 4y = 300$ and $3x = 4y$

Then Graziella says: "…it does not depend on how many parts they win or from the sum that they have bet…; at least, we do not know how they have divided the money, or what sum they have bet…" (Line 8); at least, the resolution of the problem is independent of the assumptions that they can make on the winnings or the bets (from another relation between x and y). She concludes: "the possible solutions are infinite" (Line 11). The impossibility of finding another criterion for dividing the money is equivalent to the impossibility of forming a unique system adding another equation. Moreover, they do not calculate the solution set.

In the second protocol, the students show a long resolutive procedure, and they use predominantly natural language enriched with arithmetic language. Symbolic language appears only in the final part of the resolution. But they immediately point out the necessity of giving the variables x and y a meaning in relation to the context of the problem: "Because x and y represent… the money that was bet…" (Lines 354 and 451). This pair use the symbols in the verbal description as a way of communicating: "maybe, because x and y were different…" (Line 352), but not of solving the problem.

The students use the conception of the variable in functional relation. They discuss animatedly what criteria to adopt to determine the bets: "There are no data on the bets…; we must provide one…" (Line 22) and they continue formulating some hypotheses: "…We admit that Charles has bet € 10…" (Lines 13 and 31), "Let's suppose, if Charles bets € 50…" (Line 30), "If it (the bet) is € 30…"(Line 37). The students realise that the problem has got infinite solutions and as an example they fix one of the bets and they determine the other using inverse arithmetic operations:

Alessandra	Vita
39. If it is € 30, the triple … should win 90….	40. …the total sum…, from 90 to 300…
41. Then from 90 to 300, correct? There are 210…	42. Yes
43. …then to win 210, Lucy …, correct?	44. Yes
45. …to win € 210 …	46. Lucy is the quadruple of that…, though it is divided by 4, isn't it?

The idea of linear dependence between the two bets appears implicit in this discussion. Alessandra believes that it is necessary to calculate the solution set, "…for me, there is a limit, there are some solutions that go…" (Line 63), "in my opinion… the possible solutions… go from "tot" to "tot"…, but we must see…" (Line 67). Then she considers the bounds of the problem in an explicit way: "…If we are speaking of bets, this means that it cannot be a negative number" (Line 86). The students determine the minimal bet equal to 0 and the maximal bet calculating 300÷3 and 300÷4. But they succeed only in expressing in writing the solution set of x: $0 \leq x \leq 100$, because they consider that the solution set of y ($0 \leq x \leq 75$) is included in the first one. Namely, the students designate the same letter "x" to both of the variables.

4.2.2 Second query

The two pairs of students begin the second question with a purely syntactic manipulation of the equation *(6x–3y = 18)* to find some solutions. In the first protocol Graziella says: "…I instinctively search for some numbers…" (Line 21). They write the solution: "*x = 3, y = 0*" and Graziella comments: "…we found some numbers that make the equality true…, then we can also build a problem on these two numbers…" (Line 27). At a certain moment Graziella asks: "Find the solution to a problem, but what does it mean to find the problem?" (Line 39). After the interviewer has explained, the students invent a problem similar to the first one, using the equation *3x–0y = 18*; namely, they mistook the coefficients for the solution pair. Then Graziella corrects herself: "It seems to me that we must do an inverse procedure, and put these two numbers…" (Line 51) and she points out the coefficients 6 and 3. Finally, the students re-phrase the following text:

60. Graziella: "There are two persons who play these two different sums of money. The first one wins six times the money that he/she bet, the second wins three times the money that he/she bet, the difference…"
61. Serena: "… between the winnings…"
62. Graziella: "…between the winnings is equal to 18. Find how much they have bet".

In the second protocol the students tackle the query asking: "What does it mean to invent a possible situation/problem? Should we invent a problem?" (Line 193). They syntactically manipulate the equation with serious errors as if it were a play of signs that do not make sense. The discussion becomes animated because it seems that these students do not understand what it means "to invent a possible situation/problem". After some trials the students calculate a solution pair and they formulate the text of a classical arithmetic problem with specific numerical values (the coefficients of the equation):

> "… there are 18 apples at the market, …Mark has taken 6 of them. How many…, (then… you must put… another datum,…eh?). There were 3 in the cupboard, how many are left altogether?" (Line 308).

In an attempt to improve the statement, they succeed only in inserting a variable:

> "…a merchant buys 18 apples, he puts out 6 of them so, he gives the others 3 …, he hands…, he puts inside some part…, and then, the merchant takes '**a certain number**' of apples. How many are there between the ones left and those…? (Line 316).

Finally, Vita and Alessandra abandon this context. They link the second query with the first problem and they try to formulate a similar problem. However, they felt the need to interpret the minus sign[15] and to bring out the variables in the text of the problem: "This *x* is the money that was bet. *x* and *y* are the unknowns of what they have bet…" (Line 450). Their final formulation is the consequence of a gradual elaboration.

[15] According to Radford (2002), sometimes the students' signs (in this case the minus sign) constitute short *scripts* recounting salient parts of the original story problem (told in natural language). This author considers that, for some students, the minus sign in the expression *x-2* does not perform a subtraction from the unknown *x*, but it is an orienting mark of a short script about the story problem. Then Vita and Alessandra had the necessity to interpret the minus sign, because this exigency is linked to the possibility of conferring to the expression the correct algebraic meaning (it is different from the scripts' meaning).

4.3 Discussion

From the analysis of the protocols of the first problem we notice that the solving procedures are based on natural language enriched by arithmetic language. In the first protocol, the symbolic language of algebra is completely absent. In the second protocol it appears in the final part of the resolution, but the students manifest the need to connect the variables with the "original story of the problem" (word problem). In this protocol, the students use symbols in a superficial way, only in the verbal description to communicate, but not in solving the problem (**Q_6**). Therefore, what is missing is the control that a formula can operate of the flow of the verbal reasoning.

In the first protocol the students use chiefly the conception of the unknown to solve the first problem. The passage from single solutions to infinite solutions is produced through systems of equations. Thus, it is not possible to observe the passage from the conception of the unknown to that of the variable in functional relation. In other words, for this pair of students, the infinite solutions constitute a set of single solutions coming from the resolution of different linear systems that contain the given equation. Accordingly, they do not consider the bounds imposed by the context in which the expression is considered. Then, the notion of the unknown interferes with the conception of the variable in functional relation (**Q_5**).

In the second protocol, instead, the conception of the variable in functional relation prevails. Thus, the infinite solutions constitute a set of value pairs that are obtained by varying one of them and calculating the other, beginning from the linear dependence between the variables. The students consider the bounds to the numerical universe that the contextual sense of the equation imposes (**Q_5**).

The students carry out a purely syntactic manipulation of the equation of the second query *($6x–3y = 18$)* to find some solutions. From the protocols, we deduce that the students actually confuse the activity of solving an equation with that of inventing a problem that originates from an equation. We think that this difficulty is due to a matter of a didactic contract: at school the students usually solve problems, they do not invent problems.

The formulation of a problem from an equation implicates fundamentally three activities:

1. Choosing an adequate context to give meaning to the equation
2. Identifying the quantities of the context that represent the variables
3. Individualising the properties of the quantities that are underlined by the relation expressed in the equation

We believe that the critical stage is precisely: "individualising the quantities of the context to be associated with the variables". In the second protocol we observe an attempt to choose a context of "market and apples", but the students do not succeed in identifying x and y with the quantities of apples of two different subject-objects: two shopkeepers, two different varieties, two different crates etc. Finally, the pairs of students solve the query by producing a text similar to the first problem. This means having at their disposal the context "money and bets" and the elements "two persons who bet". They have only to adapt the properties of the quantities to the new relation that the equation expresses (**Q_7**).

In the two protocols, we clearly observe an important gap between symbolic language and the possibility of finding a different context from "money and bets", to give meaning to the equation. We think that this is not the consequence of the lack of a certain amount of creativity, but the result of insufficient control of the symbols (using the same letter to designate two different variables, mistaking the coefficients for the solution pair). This is revealed in the impossibility of associating the variables with some of the quantities of the context. This difficulty could point out an inadequate representation of the relation between semantics and syntax inside the algebraic code (**Q_7**).

5 Conclusions

In the two experiments, we observe that the solving procedures predominantly use natural language. They follow the pace of spoken thought in which the semantic control of the situation is developed and takes place. The students also use arithmetic language in a not purely algebraic context (**Q_2**, **Q_6**). From this viewpoint, the individual development introduces some characteristic lines of the historical development.

We also notice that, in the context of a situation/problem, the students do not evoke the variable in functional relation if they use the notion of the unknown (**Q_1**, **Q_5**). The students, with insufficient mastery of the symbolic standard algebraic language, can consider the variable in functional relation more easily, in the presence of visual representative registers, by evoking the mental model of a straight line (**Q_4**). A certain convergence with the historical point of view is manifested. The notion of the unknown appears in the resolution of problems that require the calculation of one or more quantities. This concept exerted a strong predominance in the solving procedures up until 1600. Historically, the variable in functional relation often appears in the presence of visual registers: tables, curves, descriptions of a motion, graphs etc.

It is interesting to observe that many students add an extra condition ($3x = 4y$ or $x = y$) in order to obtain one solution of the linear equation ($3x + 4y = 300$) in the absence of representative graphic registers. Thus, the predominant conception of the variable is that of the unknown. In the few cases in which we notice the passage from the single solution to multiple solutions, it takes place through systems of equations. This behaviour may hide a deep cognitive problem: “the students need to add extra data in order to transform the statement of the problem into an interpretable model within their cognitive scope”. That is, they cannot interpret the problem in functional relation terms, because this conception of the variable is beyond their grasp. Thus, the students transform the original problem into another problem that they can solve with the knowledge of the unknown that they possess (**Q_1**, **Q_5**).

Our experimental studies show that the students meet serious difficulties in translating from algebraic language ($6x–3y = 18$) into natural language. Only 6 students (out of all 115 participants in the first and second experiments) correctly formulate a problem in a different context from the “money and bets” of the first question. We observe an important rupture between symbolic language and the possibility of finding a context to give meaning to the equation. We think that this behaviour is the result of insufficient control of the symbols. But it may also represent another problem: the impossibility of interpreting a linear equation in functional relation terms is translated into the impossibility of identifying a function in a situation/problem (**Q_3**, **Q_7**).

To study these conclusions in depth, it would be interesting to analyse the existing relation between the variables of an equation and the quantities of the context that represents them, from a semiotic perspective of the discourse. It would be important to study how the construction of the sense of a linear equation takes place. Other central matters emerging from this research that would be interesting to study in depth are:

- What would the role of symbolism, natural language, drawings, visual registers etc. be in the creation of a deeper understanding of the variable in functional relation by the student?
- How would the semiotic context influence the conceptions of the variable from the student’s viewpoint? At least, it would be interesting to study the interaction of other contexts—natural language, geometric language, perceptive schemes etc.—with the students operating in a strictly algebraic context.

- Is it possible to establish different layers of generality in the students' conceptions of the variable in functional relation? If yes, can these layers correspond to the layers determined by Radford (2006) for the generalisation of patterns?
- In the solving procedures, the students follow the pace of spoken thought. When algebraic symbolism is used, that is, when an extremely encoded syntax replaces oral reasoning, the pace of thought is disrupted and a new configuration guides the students' pace of thought. It would be interesting to know what the qualities and the particularities of this new cognitive configuration are.

Acknowledgements The authors wish to thank Prof. Luis Radford for his opportune and useful remarks on the doctoral thesis: *The concept of variable in the passage from arithmetic language to algebraic language in different semiotic contexts*. This paper is inspired by the experimental studies carried out in this thesis.

Appendix 1

We determined the principal experimental variables from an a priori analysis (the complete table is in Malisani 2006, pp. 114–115). They were the following for the **first query**:

AL1: The student answers the question.
AL2: He/she shows a procedure in natural language.
AL3: He/she shows a procedure by trial and error in natural language and/or in a half formalised language.
AL4: He/she adds a datum.
AL4.1: He/she adds a datum, but he/she considers that the winnings are divided in half.
AL4.2: He/she adds a datum, but he/she considers that the winnings of the two teenagers are equal to € 300.
AL4.3: He/she adds a datum, but he/she considers that the bets are equal.
AL5: He/she translates the problem into a first degree equation with two unknowns.
AL6: He/she explicitly considers the bounds of the problem.
AL7: He/she translates the problem into a first degree equation with two unknowns and he/she applies an incorrect method where he/she writes one variable in function of the other. Then he/she replaces this variable in the original equation and thus he/she obtains an identity. In short, the student applies the method of substitution used to solve systems of equations to a single equation.
AL9: He/she abandons the pseudo-algebraic procedure and he/she tries another method.
AL11: He/she considers, in an explicit or implicit way, the variables of the problem in a functional relation.
AL13: He/she makes some errors in the resolution of the equation and he/she finds (or he/she tries to find) only one solution.
AL14: He/she considers that a relation of proportionality exists between x and y.
ALb1: The student calculates the solution set.
ALb2: He/she shows a particular solution of the equation.
ALb3: He/she shows several solutions of the equation.
ALb4: He/she considers infinite solutions expressly.
ALb5: He/she explicitly considers that the data are insufficient to determine only one solution.
ALb6: He/she considers multiple solutions (includes ALb4 and ALb5).

Second query::

IAL1: The student answers the question.
IAL2: He/she transforms the equation into its explicit form.
IAL3: He/she solves the equation by applying an incorrect method where he/she writes one variable in function of the other.
IAL4: He/she shows a particular solution of the equation.
IAL5: He/she shows several solutions of the equation.
IAL6: He/she adds another equation and forms a system.
IAL7: He/she produces a text considering only constants.
IAL7.1: The question refers to the second member of the equation, that is, to 18.
IAL11: He/she answers correctly.
IAL12: He/she produces a text considering two variables, but it does not translate the given equation exactly.
IAL14: He/she translates algebraic language with difficulty.

Fourth query:

GAa1: The student answers the question.
GAbc1: The student calculates the solution set.
GAbc2: He/she shows a particular solution of the equation.
GAbc3: He/she shows several solutions of the equation.
GAbc4: He/she considers multiple solutions.

Appendix 2

Table 2 shows the frequencies of the variables.

Table 2 Frequencies

Variable	Absolute frequency	Percentage	Variable	Absolute frequency	Percentage
AL1	106	95	ALb5	13	12
AL2	44	40	ALb6	36	32
AL3	34	31	IAL1	67	60
AL4	71	64	IAL2	8	7
AL4.1	53	48	IAL3	5	5
AL4.2	11	10	IAL4	27	24
AL4.3	13	12	IAL5	5	5
AL5	27	24	IAL6	7	6
AL6	4	4	IAL7	20	18
AL7	13	12	IAL7.1	18	16
AL9	9	8	IAL11	8	7
AL11	34	31	IAL12	11	10
AL13	8	7	IAL14	59	53
AL14	3	3	GAa1	84	76
ALb1	99	89	GAbc1	61	55
ALb2	63	57	GAbc2	3	3
ALb3	33	30	GAbc3	4	4
ALb4	25	23	GAbc6	57	51

For the complete table of frequencies, see Malisani (2006, pp. 111–113).

References

Arzarello, F., Bazzini, L., & Chiappini, G. (1994). *L'Algebra come strumento di pensiero. Analisi teorica e considerazioni didattiche*. Progetto Strategico del C.N.R. Pavia: Dipartimento di Matematica: Università di Pavia, Quaderno n. 6.

Bardini, C., Radford, L., & Sabena, C. (2005). Struggling with variables, parameters, and indeterminate objects or how to go insane in mathematics. *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp.129–136). Melbourne, Australia.

Brousseau, G. (1997). *Theory of didactical situations in mathematics.* N. Balacheff, M. Cooper, R. Sutherland & V. Warfield, Eds. and Trans.). Dordrecht: Kluwer.

Carraher, D., Schliemann, A., & Brizuela, B. (2001). Can young students operate on unknowns? *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 130–140). Utrecht, The Netherlands.

Chevallard, Y. (1989). *Arithmetique, algebre, modelisation*. Aix, France: IREM.

Collins English Dictionary. (1991). Bologna, Italy: Zanichelli.

Eisenberg, T., & Dreyfus, T. (1991). On the reluctance to visualize in mathematics. In W. Zimmermann, & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics* (pp. 25–37). Washington, DC: Mathematical Association of America.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Furinghetti, F., & Somaglia, A. (1994). Functions in algebraic and graphical environments. *Proceedings of 46th Commission Internationale pour l'Étude et l'Amélioration de l'Enseignement des Mathématiques.* Toulouse, France.

Gras, R. (1997). Metodologia di analisi di indagine. *Quaderni di Ricerca in Didattica (G.R.I.M.)*, *7*, 99–109.

Gras, R. (2000). Les fondements de l'analyse statistique implicative. *Quaderni di Ricerca in Didattica (G.R. I.M.)*, 9, 189–209 Available from: http://dipmat.math.unipa.it/grim/quaderno9.

Gras, R., et al. (1996). *L'implication statistique (Nouvelle méthode exploratoire de données*). Grenoble, France: La Pensée Sauvage.

Gras, R., Couturier, R., Guillet, F., & Spagnolo, F. (2005). Extraction de règles en incertain par la méthode statistique implicative. *Comptes rendus des 12èmes Rencontres de la Société Francophone de Classification*, (pp. 148–151). Montréal, Canada: UQAM.

Gras, R., Diday, E., Kuntz, P., & Couturier, R. (2001). Variables sur intervalles et variables-intervalles en analyse implicative. *Actes du 8ème Congrès de la Société Francophone de Classification de Pointe à Pitre* (pp. 166–173).

Harel, G., & Sowder, L. (2005). Advanced mathematical-thinking at any age: its nature and its development. *Mathematical Thinking and Learning*, *7*(1), 27–50. doi:10.1207/s15327833mtl0701_3.

Harel, G., & Tall, D. (1991). The general, the abstract, and the generic in advanced mathematics. *For the Learning of Mathematics*, *11*, 38–42.

Harper, E. (1987). Ghosts of Diophantus. *Educational Studies in Mathematics*, *18*(1), 75–90. doi:10.1007/BF00367915.

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics*, *27*(1), 59–78. doi:10.1007/BF01284528.

Hoyles, C., Noss, R., & Adamson, R. (2002). Rethinking the microworld idea. *Journal of Educational Computing Research*, *27*(1–2), 29–53. doi:10.2190/U6X9–0M6H-MU1Q-V36X.

Janvier, C., Charbonneau, L., & Cotret, S. (1989). Obstacles épistémologiques à la notion de variable: Perspectives historiques. *Construction des savoirs, obstacles et conflits*, (pp. 64–75). Montréal, Canada: CIRADE-Université de Québec.

Küchemann, D. (1980). The understanding of generalised arithmetic by secondary school children. (Unpublished doctoral dissertation, Chelsea College, University of London).

Kücherman, D. (1981). Algebra. In K. M. Hart (Ed.), *Children's understanding of mathematics: 11–16* (pp. 102–119). London: John Murray.

Lagrange, J. P., & Chiappini, G. (2007). Integrating the learning of algebra with technology at the European level: two examples in the remath project. *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 903–912). Larnaca, Cyprus.

Linchevski, L. (1995). Algebra with numbers and arithmetic with letters: A definition of pre-algebra. *The Journal of Mathematical Behavior*, *14*, 113–120. doi:10.1016/0732–3123(95)90026–8.

Malisani, E. (1996). Storia del pensiero algebrico fino al cinquecento. Costruzione del simbolismo e risoluzione di equazioni. *Quaderni di Ricerca in Didattica (G.R.I.M.)*, *6*, 26–77. Available from: http://dipmat math.unipa.it/~grim/quaderno6.htm.

Malisani, E. (1999). Los obstáculos epistemológicos en el desarrollo del pensamiento algebraico. Visión histórica. *Revista IRICE*, *13*, 105–132. Available from: http://dipmat math.unipa.it/ ~grim/quaderno6.htm.

Malisani, E. (2006). The concept of variable in the passage from the arithmetic language to the algebraic language in different semiotic contexts (Doctoral Dissertation, Comenius University of Bratislava, 2006). *Quaderni di Ricerca in Didattica (G.R.I.M.), 16*(1). Available from: http://math.unipa.it/ ~grim/quad_16_suppl_1.htm).

Malisani, E., Scimone, A., & Spagnolo, F. (2008). Modelling by statistic in research of mathematics education. In F. Spagnolo, R. Gras, E. Suzuki, & F. Guillet (Eds.), *Statistical implicative analysis: theory and applications* (pp. 247–276). Berlin: Springer.

Malisani, E., & Spagnolo, F. (2005). Difficulty and obstacles with the concept of variable. *Proceedings of 57th Commission Internationale pour l'Étude et l'Amélioration de l'Enseignement des Mathématiques* (pp. 226–231). Piazza Armerina, Italy.

Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman, & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 25–50). London: Academic Press.

Nesselman, G. H. F. (1842). Die algebra der Griechen (pp. 302–306). Berlin.

Panizza, M., Sadovsky, P., & Sessa, C. (1999). La ecuación lineal con dos variables: entre la unicidad y el infinito. *Ensenañza de las Ciencias*, *17*(3), 453–461.

Radford, L. (1996). The roles of geometry and arithmetic in the development of algebra: historical remarks form a didactic perspective. In N. Bednarz, C. Kieran, & L. Lee (Eds.), *Approaches to algebra. Perspectives for research and teaching* (pp. 39–53). Dordrecht, The Netherlands: Kluwer.

Radford, L. (2002). On heroes and the collapse of narratives: a contribution to the study of symbolic thinking. *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 81–88). Norwich, UK.

Radford, L. (2003). Gestures, speech and the sprouting of signs. *Mathematical Thinking and Learning*, *5*(1), 37–70. doi:10.1207/S15327833MTL0501_02.

Radford, L. (2006). Algebraic thinking and the generalization of patterns: a semiotic perspective. *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 2–21). Prague, Czech Republic.

Sadovsky, P., & Sessa, C. (2005). The adidactic interaction with the procedures of peers in the transition from arithmetic to algebra: a milieu for the emergence of new questions. *Educational Studies in Mathematics*, *59*, 85–112. doi:10.1007/s10649–005–5886–2.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1–36. doi:10.1007/BF00302715.

Sfard, A. (1992). The development of algebra. Confronting historical and psychological perspectives. *Algebra Working Group of Seventh International Congress on Mathematical Education.* Quebec, Canada.

Sfard, A., & Linchevski, L. (1994). The gains and the pitfalls of reification—the case of algebra. *Educational Studies in Mathematics*, *26*, 191–228. doi:10.1007/BF01273663.

Sierpinska, A. (1994). *Understanding in mathematics*. London: Falmer Press.

Spagnolo, F. (1998). *Insegnare le Matematiche nella scuola secondaria*. Florence: La Nuova Italia.

Spagnolo, F. (2006). La modélisation dans la recherche en didactiques des mathématiques: les obstacles épistémologiques. *Recherches en Didactiques des Mathématiques*, *26*(3), 337–379.

Spagnolo, F., Gras, R., Suzuki, E., & Guillet, F. (Eds.). (2008). *Statistical implicative analysis: Theory and applications*. Berlin: Springer.

Trigueros, M., Reyes, A., Ursini, S., & Quintero, R. (1996). Diseño di un cuestionario de diágnostico acerca del manejo del concepto de variable en el álgebra. *Ensenañza de las Ciencias*, *14*(3), 351–363.

Usiskin, Z. (1988). Conceptions of school algebra and uses of variables. In A. F. Coxford, & A. P. Shulte (Eds.), *The ideas of algebra, K-12* (pp. 8–19). Reston, VA: NCTM.

Vinner, S. (1989). Avoidance of visual considerations in calculus students. *FOCUS on Learning Problems in Mathematics*, *11*(2), 149–156.

Wagner, S. (1981). An analytical framework for mathematical variables. *Proceedings of the 5th Conference of the International Group for the Psychology of Mathematics Education* (pp. 165–170). Grenoble, France.

Wagner, S. (1983). What are these things called variables. *Mathematics Teacher*, *76*(7), 474–479.