ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

E. W. BETH

Observations concernant la théorie de la définition

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 7, série *Mathématiques*, nº 1 (1962), p. 83-87

<http://www.numdam.org/item?id=ASCFM_1962__7_1_83_0>

OBSERVATIONS CONCERNANT LA THÉORIE DE LA DÉFINITION

E.W. BETH
Amsterdam

Hommage à D.K. de Jongh

I - INTRODUCTION

Dans son admirable opuscule *De l'esprit géométrique et de l'art de persuader* [1] Pascal fait observer qu'on "ne reconnaît en géométrie que les seules définitions de nom". Ce procédé permet "de substituer... la définition à la place du défini". Et "rien n'éloigne plus promptement et plus puissamment les surprises captieuses des sophistes que cette méthode, qu'il faut avoir toujours présente, et qui suffit seule pour bannir toutes sortes de difficultés et d'équivoques".

Il convient de décrire la situation à laquelle Pascal fait allusion au moyen de la terminologie de la méthodologie contemporaine. Soit T_0 une théorie déductive qui fait intervenir certaines notions primitives A, B, C, ... et soit K_0 un système d'axiomes pour T_0. Soit F un terme qui n'intervient pas dans T_0 ; pour fixer les idées, il convient de supposer que F est un paramètre de prédicat binaire. Une formule close :

$$(x)\ (y)\ [F(x, y) \longleftrightarrow U(x, y)] \tag{F}$$

où $U(x, y)$ ne contient que les variables libres x et y et ne fait intervenir que les notions primitives A, B, C, ... de T_0 sera appelée une *définition nominale* du terme F.

Si (F) est adjointe à K_0, alors T_0 est remplacée par une théorie T qui fait intervenir les notions primitives A, B, C, ... et F et qui contient tout théorème X de T_0. Cependant, une formule X qui ne fait intervenir que les notions A, B, C, ... ne peut se présenter dans T que si elle se présente déjà dans T_0. Notamment, T ne peut être contradictoire que si T_0 est déjà contradictoire ; par conséquence, la non-contradiction de T_0 implique la non-contradiction de T.

II - DEFINISSABILITE DEMONTRABLE -

D'autre part, soit donnée une théorie déductive T qui fait intervenir les notions primitives A, B, C, ..., et F. Supposons qu'il est possible d'obtenir T en partant d'une théorie T_0 qui ne contient que les notions A, B, C, ... par l'adjonction d'une définition nominale (F) appropriée. Alors il faudra qu'une formule (F) appropriée se présente parmi les théorèmes de T et que T_0 se constitue exactement de tous les théorèmes de T qui ne contiennent que les notions A, B, C, ... Et si T contient comme théorème une formule (F) appropriée, alors l'adjonction de (F) à la théorie T_0 constituée par tous les théorèmes de T qui ne contiennent que les notions A, B, C, ... produira exactement la théorie T.

Si cette situation se présente, nous dirons que la notion F est *démontrablement définissable* par rapport à T. Soit < $\underline{S}$, $\underline{A}$, $\underline{B}$, $\underline{C}$, ... > un *modèle* quelconque de la théorie T_0, où $\underline{S}$ est le domaine des individus du modèle tandis que $\underline{A}$, $\underline{B}$, $\underline{C}$, ... sont les interprétations respectives des notions A, B, C, ... Alors la définition nominale (F) détermine de façon univoque une interprétation $\underline{F}$ du terme F telle que < S, A, B, C, ..., F > sera un modèle de la théorie T.

III - METHODE DE PADOA - DEFINISSABILITE IMPLICITE -

Supposons que la théorie T_0 possède un modèle $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots >$ pour lequel on peut trouver deux interprétations distinctes $\underline{F}$ et $\underline{F}'$ du terme F telles que $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots, \underline{F} >$ et $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots, \underline{F}' >$ constituent l'un et l'autre des modèles de la théorie T. Alors le terme F ne peut pas être démontrablement définissable par rapport à T. Or, la méthode de Padoa [2] dans la théorie de la définition consiste précisément à construire deux modèles de ce genre ; il en résulte que, de notre point de vue, l'application de cette méthode est justifiée.

Inversement, supposons qu'un modèle quelconque $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots >$ de T_0 ne produise jamais qu'un seul modèle $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots, \underline{F} >$ de T ; alors on dira que le terme F est *implicitement définissable* par rapport à T. Il va sans dire que, sous des conditions fort générales, la définissabilité démontrable implique la définissabilité implicite.

Quant à l'implication inverse, on connaît en ce moment plusieurs conditions suffisantes pour sa validité. Il convient toutefois de préciser quelque peu les notions. Soit K un système d'axiomes pour la théorie T et soit K' le système d'axiomes obtenu en remplaçant dans K le paramètre F par un paramètre F'. Soit (FF') la formule :

$$(x)\ (y)\ [F\ (x\ ,\ y) \longleftrightarrow F'(x\ ,\ y)].$$

1/ Si T est une théorie formalisée au moyen d'un système logique d'ordre supérieur, alors la définissabilité implicite n'implique pas nécessairement la démontrabilité de la formule (FF') en partant de l'union des ensembles K et K'. D'autre part, la démontrabilité de la formule (FF') en partant de l'union des ensembles K et K' implique la définissabilité démontrable du terme F par rapport à T (A. Tarski 1935).

2/ Si T est une théorie formalisée au moyen de la logique élémentaire, alors en vertu du théorème de complétude (K. Godel 1930) la définissabilité implicite de F par rapport à T implique la démontrabilité de la formule (FF') en partant de l'union des ensembles K et K'. Et ensuite la démontrabilité de la formule (FF') en partant de l'union de K et K' implique la définissabilité démontrable de F par rapport à T (E.W. Beth 1953). La démonstration originale de ce dernier résultat fut assez pénible ; le lemme d'interpolation de W. Craig (1957) permet en ce moment d'en donner une démonstration plus élégante.

3/ Tout récemment K. Schütte a donné une démonstration du lemme d'interpolation pour la logique élémentaire intuitioniste (communication de G. Kreisel [3]). Par conséquent, si T est une théorie formalisée au moyen de ce système logique, la démontrabilité de la formule (FF') en partant de l'union de K et K' implique la définissabilité démontrable du terme F par rapport à T. L'application de la notion de définissabilité implicite en logique intuitioniste soulève malheureusement certaines difficultés.

IV - UNE VARIANTE DE LA METHODE DE PADOA - DEFINISSABILITE FAIBLE -

Il existe une variante de la méthode de Padoa qui mérite également d'être discutée dans le présent contexte [4]. Il peut arriver que pour un certain modèle $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots >$ de la théorie T_0 il n'existe aucune interprétation $\underline{F}$ du terme F telle que $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots, \underline{F} >$ constitue un modèle de la théorie T. Alors il résulte de notre discussion que le terme F n'est pas démontrablement définissable par rapport à T.

K.L. de Bouvère (1959) [5] a fait l'observation que si l'on fait une supposition supplémentaire on peut tirer une conclusion plus forte et particulièrement intéressante.

Si F n'est pas démontrablement définissable par rapport à T, il peut arriver qu'une certaine formule (F) soit *compatible* avec T ; on dira alors que F est *faiblement définissable* par rapport à T. Soit T' la théorie qui résulte de l'adjonction de (F) à T et soit T_0' la théorie constituée par les théorèmes de T' qui ne contiennent que les notions A, B, C, ... Il va sans dire qu'on aura $T_0 \subseteq T_0'$.

Etant donné que le terme F est démontrablement définissable par rapport à T', il résulte de notre discussion au par. 2 qu'un modèle quelconque $< \underline{S}', \underline{A}', \underline{B}', \underline{C}', \ldots >$ de T_0' produit un modèle $< \underline{S}', \underline{A}', \underline{B}', \underline{C}', \ldots, \underline{F}' >$ de T' qui sera forcément un modèle de T. Or, cette conclusion ne contredit pas l'existence du modèle $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots >$ mentionné tout à l'heure ; il peut y avoir, en effet, des modèles de T_0 qui ne sont pas des modèles de T_0'.

Cependant nous introduisons maintenant la supposition supplémentaire que T_0 est *complète*. Puisque $T_0 \subseteq T_0'$ tandis que T_0' est non-contradictoire, on a forcément $T_0 = T_0'$, en sorte que nous retombons sur la contradiction que nous pouvions éviter tout à l'heure. Il en résulte : si T_0 est complète et si T_0 a un modèle qui ne produit pas un modèle de T, alors F est *essentiellement non définissable* par rapport à T.

F est appelé essentiellement non définissable par rapport à T s'il n'est pas faiblement définissable par rapport à T. Cette terminologie s'explique du fait que si F n'est pas faiblement définissable par rapport à T, alors F ne sera démontrablement définissable ni par rapport à T ni par rapport à n'importe quelle extension non-contradictoire T' de T.

Le problème des conditions suffisantes pour l'existence d'un modèle de T_0 qui ne produit pas un modèle de T reste ouvert. De Bouvère a montré par un contre-exemple qu'il ne suffit pas de supposer que F est essentiellement non définissable par rapport à T et que T_0 est complète.

Pour terminer, je voudrais toutefois faire l'observation suivante : s'il existe un modèle de T_0 qui ne produit aucun modèle de T, alors il existe une extension T' de T par rapport auquel F est essentiellement non définissable. En effet, le modèle donné détermine une extension complète T_0' de T_0 qui détermine à son tour une extension T' de T qui se prête à l'application du résultat de de Bouvère.

V - DEFINITION NON EXPLICITE -

Il convient d'appeler *définition explicite*, toute introduction d'un terme F qui peut être remplacée par l'adjonction d'une formule (F) appropriée. Il va sans dire que toute définition explicite remplit les conditions énoncées par Pascal.

Toutefois on a bien souvent recours à d'autres moyens lorsqu'il s'agit d'introduire de nouvelles notions. Soit, par exemple, T_0 la théorie élémentaire de l'addition pour les nombres naturels ; il est bien connu que cette théorie est complète (et décidable) [6]. Alors il est possible d'introduire la multiplication par l'adjonction des deux formules :

$$(x)\ [x\ .\ 0 = 0], \tag{M1}$$

$$(x)\ (y)\ [x\ .\ (y + 1) = (x\ .\ y) + x] \tag{M2}$$

à T_0 ; soit T la théorie qui résulte.

Selon la terminologie courante, les formules (M1) et (M2) constituent une *définition récursive* de la multiplication. Cette terminologie s'explique du fait que la théorie T permet de démontrer toute formule numérique :

$$\underline{m}\ .\ \underline{n} = \underline{p},$$

où $\underline{m}$, $\underline{n}$ et $\underline{p}$ sont respectivement les notations, dans T, de trois nombres naturels m, n et mn.

Toutefois les formules (M1) et (M2) ne constituent pas une définition explicite de la multiplication, étant donné qu'elles ne peuvent pas être remplacées par une formule (F). En effet, il n'est pas difficile de construire un modèle $< \underline{N}, +, 0, 1 >$ de T_0 qui ne produit aucun modèle de T. Puisque T_0 est complète, nous pouvons appliquer le résultat de de Bouvère ; par conséquent, la multiplication est essentiellement non définissable par rapport à T.

Cette discussion montre, me semble-t-il, les deux aspects complémentaires du recours à des définitions non explicites. D'un côté, la définition non explicite constitue pour ainsi dire un pis-aller qui s'impose lorsqu'un terme F se montre essentiellement non définissable par rapport à une théorie T où elle intervient. D'autre part, certains types de définitions non explicites présentent parfois des avantages considérables sur les définitions explicites. Notamment, les définitions récursives ont l'avantage particulier de mettre en évidence le caractère effectif de certaines notions.

VI - PSEUDO-DEFINITIONS -

Supposons que dans une certaine théorie T_0 on puisse démontrer un théorème :

$$(x)\ (y)\ (E\, z)\ V\ (x\ ,\ y\ ,\ z) \tag{f°}$$

et soit T la théorie qui résulte de l'adjonction de l'axiome :

$$(x)\ (y)\ V\ (x\ ,\ y\ ,\ f(x\ ,y)) \tag{f}$$

à T_0. Alors T ne peut contenir aucun théorème X qui ne soit pas contenu dans T_0 et qui ne fasse pas intervenir le foncteur f ; en particulier, la non-contradiction de T_0 implique la non-contradiction de T. [Si le théorème (f°) présente la forme particulière :

$$(x)\ (y)\ (Ez)\ (t)\ [U\ (x\ ,\ y\ ,\ t) \longleftrightarrow z\ =\ t]$$

alors cette version faible du *théorème d'élimination* produit la version courante de ce théorème [7]]. Il convient d'appeler l'axiome (f) une *pseudo-définition* du foncteur f par rapport à la théorie T.

Si f est pseudo-définissable par rapport à T, c'est-à-dire, si T résulte de la théorie T_0 correspondante par l'adjonction d'un axiome (f) approprié, alors un modèle quelconque $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots >$ de T_0 produira tout au moins un modèle $< \underline{S}, \underline{A}, \underline{B}, \underline{C}, \ldots, \underline{f} >$ de T. Par conséquent, si T_0 a un modèle qui ne produit aucun modèle de T, alors f ne peut pas être pseudo-définissable par rapport à T. Les théories T_0 et T que nous avons discutées au par. 5 fournissent un exemple d'une situation de ce genre.

Je veux faire quelques remarques supplémentaires sur le cas général. Supposons que f soit pseudo-définissable par rapport à T et que T_0 contienne le théorème :

$$(x)\ (y)\ (z)\ (t)\ [\{V\ (x\ ,\ y\ ,\ z)\ \ \&\ \ V(y\ ,\ x\ ,\ t)\} \longrightarrow z\ =\ t].$$

Alors il est clair que T contiendra les théorèmes :

$$(x)\ (y)\ (z)\ [f(x\ ,\ y) = z \longleftrightarrow V\ (x\ ,\ y\ ,\ z)] \tag{1}$$

et :

$$(x)\ (y)\ [f(x\ ,\ y) = f(y\ ,x)]. \tag{2}$$

D'autre part, supposons que T_0 contienne le théorème :

$$(x)\ (y)\ (Ez)\ [V\ (x\ ,\ y\ ,z)\ \ \&\ \ V\ (y\ ,\ x\ ,\ z)];$$

alors la formule (2) sera compatible avec la théorie T.

Notons, pour terminer, que les pseudo-définitions jouent un rôle important dans la logique d'ordre supérieur. Notamment, les *définitions par abstraction* constituent un exemple classique de leur application. Soit S une classe et soit R une relation réflexive, symétrique et transitive définie dans S. Alors la définition par abstraction qui se rapporte à la relation R consiste dans l'introduction des classes :

$$C\ (x) = E_z\ [z\ \varepsilon\ S\ \ \&\ \ R\ (x\ ,\ z)].$$

Or, l'existence de toutes ces classes est garantie par le théorème :

$$(x)\ (EY)\ (z)\ [z\ \varepsilon\ Y \longleftrightarrow \{z\ \varepsilon\ S\ \ \&\ \ R\ (x\ ,\ z)\}]$$

qui à son tour est fondé sur un principe qui est couramment appelé *axiome de réductibilité* ou *axiome des pseudo-définitions*.

BIBLIOGRAPHIE

[1] Voir par exemple : L'oeuvre de Pascal, texte établi et annoté par J. Chevalier, Bibliothèque de la Pléiade, Paris 1936, pp. 358-386.

[2] A. PADOA - Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une théorie déductive quelconque, Biblioth. du Congrès Int. de Philos., vol. III, Paris 1901.

[3] Lettre du 10 avril 1962.

[4] E.W. BETH - L'existence en mathématiques, Paris-Louvain 1956.

[5] K.L. DE BOUVERE - A Method in Proofs of Undefinability, Amsterdam 1959.

[6] E.W. BETH, l.c., K.L. DE BOUVERE, l.c.

[7] S.C. KLEENE - Introduction to Metamathematics, Amsterdam-Groningen 1952.