Publié dans Travaux de logique 7, 91-102, 1992,
source qui doit être utilisée pour toute référence à ce travail

REMARQUES SUR L'INTERPRETATION DU THEOREME D'INDECIDABILITE DE GÖDEL

Henri Lauener

Ce n'est pas seulement dans des cercles de mathématiciens que le théorème d'incomplétude de Gödel a été salué comme un résultat ouvrant de larges perspectives, mais aussi par des philosophes. Cependant l'interprétation philosophique du célèbre théorème ne me semble pas aussi irréprochable que sa démonstration purement formelle. On attribue entre autres à Gödel le mérite d'avoir ébranlé le programme de Hilbert en apportant la preuve que dans les mathématiques - à la différence de la logique - la classe des thèses (axiomes et théorèmes) ne coïncide pas avec la classe des propositions vraies et que, pour cette raison, la première ne peut être ramenée à la seconde. On affirme généralement que le résultat de ses travaux démontre l'incomplétude de tout système formel de l'arithmétique qui soit suffisamment fort et exempt de contradictions. Mes difficultés à comprendre cette affirmation proviennent du fait que, dans la partie formelle du célèbre article de 1931, on trouve des propositions mathématiques que l'on ne peut pas, à mon avis, transformer sans autre en des évidences philosophiques. Comment parvenons-nous des expressions purement arithmétiques aux interprétations, que l'auteur lui-même place en tête de son article? Pour les justifier, il faudrait préalablement expliciter les raisons pour lesquelles une sémantique formelle doit passer pour la (seule) interprétation correcte.

Qu'un philosophe soit prêt à accepter l'interprétation habituelle ou non, cela dépend de sa conception fondamentale des mathématiques. Gödel était un platonicien déclaré. Pour lui les objets mathématiques existent indépendamment du fait que nous autres, êtres humains, devons utiliser, pour en parler, une langue

spécifiquement élaborée dans ce but. Dans «What is Cantor's continuum problem?» (1947, il postule, à l'instar de Kant, une forme d'intuition pure, à l'aide de laquelle nous percevons des objets abstraits. Le fait que les axiomes s'imposent à nous comme vrais constitue pour lui une preuve de l'existence de ces objets. L'interprétation traditionnelle du théorème ne me semble en effet compréhensible qu'en faisant une présupposition de ce genre. Pour Gödel la propriété d'être une thèse était visiblement dépendante de la langue, mais pas celle d'être vrai; ainsi sa conception débouche en ultime analyse sur la thèse kantienne affirmant que certaines vérités *a priori* sont synthétiques. Celui qui ne partage pas cette conviction n'accordera pas la même signification philosophique aux résultats de Gödel. L'enthousiasme initial avec lequel certains membres du Cercle de Vienne ont accueilli le premier théorème reposait sur un malentendu, car celui-ci, en fait, entraînait moins la fin de la métaphysique que la restauration du platonisme.

Si l'on admet des vérités absolues, indépendantes de la langue et des théories, l'affirmation qu'il existe dans l'arithmétique une proposition vraie, mais non démontrable, peut paraître intuitivement plausible. Par contre, si l'on considère la question du point de vue de ma philosophie transcendantale pragmatiquement relativisée, qui exige une relativisation systématique de toute vérité à un contexte (théorique), cela n'est pas le cas. En effet, selon ce point de vue, on produit les objets mathématiques en utilisant un système linguistique fixé par des règles. En adoptant des conventions qui déterminent les conditions d'utilisation des termes logiques et mathématiques, nous *créons* par des définitions les vérités de la logique et des mathématiques («truth by *fiat*»). Par le choix du cadre conceptuel, nous fixons aussi une ontologie qui, dans le cas des mathématiques, consiste en objets abstraits, c'est-à-dire en objets produits par la seule utilisation d'un système linguistique; ceci a pour conséquence qu'en mathématiques, contrairement à ce qui se passe dans les sciences naturelles, il faut considérer les affirmations d'existence comme analytiques. Or, si la signification d'une expression mathématique est déterminée par l'ensemble des règles qui régissent son usage dans le cadre de la théorie en

question et s'il faut limiter le domaine des valeurs des variables en conséquence, il ne peut y avoir aucune vérité *externe* dans le sens du réalisme platonicien. La méthode de la relativisation en fonction des contextes aboutit à un concept purement interne, relativisé - conformément à la définition de Tarski - à un langage particulier.

Le fait qu'on peut conclure à la vérité d'énoncés mathématiques au moyen de méthodes purement analytiques exclut certes une éventuelle réfutation empirique, mais il ne signifie pas pour autant que toute révision soit exclue. Si, par exemple, quelqu'un se décide à utiliser la logique ou les mathématiques intuitionnistes au lieu de leurs versions classiques, il procède à une altération de tout l'appareil conceptuel en redéfinissant les termes en question. Comme aucune preuve objective de la vérité externe du système, considéré comme un tout, n'est possible, les raisons qui le poussent à faire ce choix doivent être de nature pragmatique. Dans la mesure où les axiomes de la géométrie se bornent à stipuler comment nous devons parler des objets spaciaux et de leurs relations, il ne faut pas considérer les différents systèmes géométriques comme des théories rivales qui rendraient compte d'un même domaine de réalité donné de façon neutre, mais bien plutôt comme des systèmes conceptuels différents. Ces deux exemples illustrent la relativité de la notion d'analycité telle que je la conçois; j'insiste sur le fait que celle-ci ne repose nullement sur la présupposition de significations immuables, comme entités platoniciennes; elle met bien au contraire un accent particulier sur le fait que des expressions homonymes peuvent avoir des significations différentes selon leur usage dans des contextes différents. Ainsi, il peut se produire que la même phrase soit utilisée dans une théorie scientifique comme définition (analytique) et, dans une autre théorie, comme loi naturelle (synthétique).

Ces indications générales sur ma conception des mathématiques devraient déjà faire apparaître les raisons qui suscitent en moi un doute quant à la portée philosophique du théorème de Gödel. Dans ce qui suit, j'aimerais mettre en évidence les points particuliers, où résident à mon avis les difficultés. Afin d'éviter certaines erreurs d'interprétation que les mathématiciens ont re-

prochées aux philosophes, il est important de distinguer les trois systèmes (linguistiques) que Gödel introduit. D'une part, nous avons un système formel, *P*, qui représente pour l'essentiel une version (non interprétée) de l'axiomatique de Peano exprimée dans la notation de la logique des prédicats (avec des nombres comme individus et le symbole de fonction pour la relation de succession, en tant que terme primitif); les formules de ce système sont déductibles ou non déductibles - par contraste avec les formules vraies ou fausses d'un système interprété. Le but visé est de prouver que ce système contient des propositions non-décidables. On utilise la langue naturelle comme métalangue *M*, dans laquelle on discourt sur les expressions de *P* et dont les énoncés sont vrais ou faux. Enfin, nous avons encore besoin du système non formalisé de l'arithmétique récursive *A*, dans lequel nous formulons des propositions (vraies ou fausses) sur les nombres naturels et leurs relations. Or la célèbre trouvaille de Gödel, qu'on a baptisée la numérotation de Gödel, a consisté en la découverte d'un procédé permettant d'attribuer à chaque expression de *P* un nombre naturel dans la métalangue *M*. On affirme parfois - en donnant au terme un sens quelque peu mystérieux pour moi - que l'on obtient ainsi une *interprétation* mathématique de *P*. La prudence nous commande pourtant de souligner qu'il ne s'agit pas d'une réduction ontologique - on ne déclare pas que les nombres naturels et les expressions métalinguistiques sont identiques! - mais seulement d'une correspondance biunivoque, qui doit servir à démontrer l'identité de la structure des deux domaines. La correspondance se limite au fait qu'à chaque symbole on associe un nombre naturel, à chaque formule une suite de nombres naturels et à chaque preuve une suite de suites de nombres naturels. Il s'agit donc d'une interprétation dans un sens purement formel. Pour démontrer l'incomplétude de l'arithmétique, il nous faut d'abord montrer que toutes les expressions de l'arithmétique récursive primitive *A* peuvent être corrélées à des expressions dans *P* de telle façon qu'une formule (respectivement, la négation d'une formule) soit démontrable dans *P*, si la proposition correspondante dans *A* est vraie (respectivement, fausse). Il nous faut de plus prouver que le système *P*, pour autant qu'il soit exempt de contradiction, doit

être incomplet en ce sens qu'il doit contenir au moins une formule qui, de même que sa négation, n'est pas démontrable. On peut ensuite prouver que l'incomplétude de *P* est liée à l'incomplétude de *A*.

Grâce au procédé de la numérotation de Gödel, nous sommes en mesure de produire un pendant arithmétique du concept méta-mathématique de la démonstration. Au prédicat de la métalangue 'v_1 est une démonstration dans *P* pour v_2' ['v_1' désignant une variable dont le domaine est constitué de suites de suites d'expressions et 'v_2' une variable dont le domaine est constitué de suites d'expressions] correspond dans *A* la fonction propositionnelle 'xBy', que l'on transforme en une proposition vraie en substituant à 'y' le nombre de Gödel d'une formule et à 'x' le nombre de Gödel de sa preuve. Nous pouvons alors définir pour *P*

$$Dem\, x \;\text{<->}\; (\mathrm{E}y)\; yBx\,,$$

de telle façon que '*Dem x*' soit vraie précisément lorsque 'x' a la valeur du nombre de Gödel d'une formule démontrable dans *P*.

Pour formuler la proposition-clef de la preuve, c'est-à-dire V, qui exprime la correspondance, il nous faut définir quelques autres fonctions récursives primitives. Le pendant arithmétique du concept métalinguistique de substitution est l'expression

$$Sb\,(x\,{}^{y}_{z})\,,$$

qui désigne le nombre de Gödel de la formule qui résulte de la substitution, dans la formule dont le nombre de Gödel est *x*, de la variable libre dont le nombre de Gödel est *y* par le terme dont le nombre de Gödel est *z*. Posons en outre que

$$Z\,(x)$$

désigne le nombre de Gödel du symbole représentant x dans P[1] et que

$$Neg\ (x)$$

désigne le nom du nombre de Gödel de la négation de la formule dont le nombre de Gödel est x. Enfin, nous aurons encore besoin de l'opération de la généralisation universelle:

$$x\ Gen\ y$$

désigne le nombre de Gödel de la formule qui résulte de ce que nous plaçons le quantificateur universel '(v)' devant la formule dont le nombre de Gödel est y (où x est le nombre de Gödel de la variable 'v'). Nous obtenons ainsi l'outillage nécessaire pour exprimer l'affirmation suivante, qui est à mon avis difficile à saisir intuitivement:

> «Le fait intuitif que toute relation récursive est définissable dans le système P (si l'on donne leur signification habituelle aux formules de ce système) se trouve exprimé précisément et *sans* qu'il soit fait référence à aucune interprétation des formules de P, dans le théorème suivant:
>
> Théorème V. *Pour toute relation récursive $R(x_1, ..., x_n)$, il existe un SIGNE DE RELATION r à n places (avec les VARIABLES LIBRES $u_1, u_2, ..., u_n$), tel que pour tous les n-uples de nombres $(x_1,..., x_n)$, on ait:*
>
> $$R\ (x_1...x_n) \text{ -> Dem } [Sb\ (r\ {}^{u_1\ \ ...u_n}_{Z(x_1)...Z(x_n)})]$$
>
> $$\overline{R}\ (x_1, ..., x_n) \text{ -> Dem } [\text{Neg } Sb\ (r\ {}^{u_1\ \ ...u_n}_{Z(x_1)...Z(x_n)})]»$$
>
> (Traduction française 1989: 125-126; édition allemande 1931: 186).

---

1 On interprète Z comme une fonction qui représente des nombres par le nombre de Gödel de leur représentation dans P.

Avec cette proposition, nous avons atteint le point d'articulation qui est déterminant pour la conception selon laquelle la notion syntaxique de la démontrabilité dans *P* doit être un pendant du concept sémantique de la vérité dans l'arithmétique récursive. La démonstration que *P*, pour autant qu'il est exempt de contradiction, contient une formule qui n'est ni démontrable, ni réfutable, s'appuie essentiellement sur cette proposition. Si nous considérons le cas particulier:

$$\overline{x\, B\, (Sb\, (y^{19}_{Z(y)}))} \;\text{->}\; Dem\, (Sb\, (q^{17}_{Z(x)}\, {}^{19}_{Z(y)}))$$

et si nous admettons qu'est vrai l'antécédent de toutes les conditionnelles qui résultent de la généralisation universelle des variables '*x*', nous obtenons alors la proposition arithmétique:

G_A: $(x)\, (x\, B\, (Sb(p^{19}_{Z(p)}))\,)$

[*p* est la généralisation de *q*: *p* = 17 *Gen* q]

Si G_A est vraie, la formule qui a le nombre de Gödel

G: $Sb\, (p^{19}_{Z(p)})$

n'est pas démontrable dans *P*, et par conséquent la proposition métalinguistique suivante:

G_M: G n'est pas démontrable dans *P*

doit être également vraie. Nous arrivons donc à la conclusion que 'G_A est vraie', 'G_M est vraie' et 'G n'est pas démontrable dans P' sont équivalentes. Gödel a prouvé que l'affirmation que G est démontrable et l'affirmation que la négation de G est démontrable conduisent toutes deux à une contradiction.

J'ai déjà laissé entendre que pour moi la difficulté réside en mon incapacité à comprendre la relation exacte entre l'incomplétude de *P* et l'incomplétude de *A*. Mes doutes ne por-

tent pas sur la preuve formelle démontrant que la proposition 'G_A doit être vrai, si *P* est exempt de contradiction', mais bien sur l'argument selon lequel G devrait signifier la même chose que G_A, parce que nous interprétons *P* arithmétiquement. Que doit-on entendre ici par 'signifier la même chose'? On ne saurait penser à une relation extensionnelle, puisque en ce cas les objets seraient identiques, de sorte qu'il ne nous serait pas possible d'établir une correspondance biunivoque entre deux domaines *différents*. Que faisons-nous exactement lorsque nous *interprétons* des énoncés métalinguistiques portant sur des expressions comme des énoncés arithmétiques portant sur des nombres? La confusion est créée par les affirmations que G_M parle de G, que G_A parle des nombres naturels, que G, en tant que formule d'un système non interprété, ne parle de rien du tout et enfin que G est interprété comme G_A, affirmations qui sont incompatibles à mon avis. Car si elles étaient toutes vraies, G_M serait ambigu, dans la mesure où l'énoncé devrait à la fois traiter de rien et des nombres naturels. Nous aurions donc d'une certaine façon deux versions du théorème - une syntaxique et une sémantique -, qui n'ont rien à voir l'une avec l'autre. La seconde version recèle d'ailleurs en elle le risque d'un paradoxe du fait de son caractère auto-référentiel. Je ne pense pas que l'on puisse conjurer ce risque de manière convaincante en arguant qu'il ne s'agit pas d'une identification, mais bien d'une correspondance biunivoque.

Ma tentative d'interpréter intuitivement le premier théorème de Gödel échoue finalement sur le doute que j'ai quant à la nature mathématique des énoncés censés exprimer la correspondance. Je n'arrive pas à me persuader que l'antécédent et le conséquent de ces conditionnelles sont de même nature. L'interprétation d'après laquelle la formule G exprime sa propre indémontrabilité ne pourrait acquérir son éventuelle crédibilité intuitive qu'en présupposant l'interprétation standard de l'arithmétique et en concédant que *B* exprime numériquement la relation de preuve '*Dem*'. Si toutefois on n'exige pour *B* rien de plus que la satisfaction des conditionnelles exprimant la correspondance, une telle utilisation du mot 'preuve' paraîtra arbitraire du fait qu'on ne présente aucune raison pour la justifier. Il me

semble insuffisant de rétorquer que des systèmes formels pour lesquels la proposition V n'est pas valable ne pourraient pas raisonnablement être interprétés comme des systèmes mathématiques. D'ailleurs, comme le modèle standard n'est pas le seul, on se demandera s'il ne pourrait pas y avoir d'autres systèmes, dans lesquels cette condition serait satisfaite sans qu'ils ne contiennent de formule indécidable.

Mon impuissance à comprendre le contenu philosophique du théorème de Gödel provient de ma conception normative de la logique et des mathématiques, qui se distance fondamentalement de la vision platonicienne et descriptive de Gödel, d'après laquelle nous découvrons et décrivons de prétendues vérités (au moyen d'une intuition intellectuelle). Accepter que des formules dépourvues de sens puissent être équivalentes à des énoncés vrais de la métalangue me semble déraisonnable; ceci reviendrait à affirmer qu'il existe des vérités inexprimables, qui se trouvent au-delà d'un système des mathématiques constitué par des règles. Pour pouvoir utiliser le prédicat sémantique 'vrai' dans le sens de Tarski, il nous faut déterminer un domaine fixant les valeurs des variables qui ne peut être donné que de façon interne, c'est-à-dire dans le cadre d'une théorie établie et axiomatiquement construite. Pour cette raison, il est impossible de définir un concept *externe* de la vérité, autant pour les mathématiques que pour les sciences naturelles; ceci vaut d'ailleurs également pour le concept de la démontrabilité, qu'on ne peut introduire que relativement à un système donné. Selon ma conception, la signification des énoncés mathématiques est fixée par des règles intra-linguistiques, qui ne *dé*crivent pas une concordance existante entre des expressions et une «réalité» platonicienne ou empirique, mais qui, en tant que normes, *pre*scrivent l'usage des termes dans le système en question. De mon point de vue, il faut insister sur le fait que toute connaissance dépend de nos moyens d'expression et que toute vérité *a priori* est analytique, dans le sens qu'elle est produite par l'adoption d'un système linguistique. La perspective platonicienne pèche précisément par le fait qu'elle présuppose des vérités synthétiques *a priori*.

En vue du problème que pose l'interprétation philosophique des théorèmes de Gödel, il est indispensable d'élucider exactement ce que leur auteur obtient par son procédé. On nous assure que l'arithmétisation nous permet d'«interpréter» comme des expressions portant sur des relations arithmétiques les expressions de la métalangue dans laquelle nous parlons du calcul. Dans la mesure où l'on n'entend pas par là opérer une simple traduction dans une autre langue ou introduire une nouvelle notation, on prétend convertir des propositions de la métalangue en des propositions (nécessairement vraies) exprimant des relations purement mathématiques. Comme d'autre part on nous assure qu'il ne s'agit pas d'une réduction ontologique, il ne me reste plus qu'à jeter l'éponge. En effet, je ne vois pas comment s'opère la transition de la simple correspondance biunivoque à une interprétation dans le sens propre du terme, qui confère un contenu aux symboles. Si l'on transforme, par arithmétisation, en une proposition mathématique vraie une règle disant qu'une suite de formules représente la preuve d'une formule déterminée si et seulement si le nombre de Gödel de la démonstration se trouve dans la relation «*Dem (x,y)*» avec le nombre de Gödel de la conclusion, alors on commet - me semble-t-il - une erreur analogue à celle connue sous le nom de «naturalistic fallacy». Je suspecte précisément cette transmutation subreptice d'être à l'origine de mon incompréhension de l'interprétation que l'on donne du théorème.

Dans les remarques informelles qu'il place au début de son article, Gödel constate que du fait que '*R (q);q*' affirme sa propre indémontrabilité découle immédiatement sa vérité, puisque, en effet, '*R (q);q*' n'est pas démontrable[2]. Mais comment la construction d'une proposition arithmétique qui «représente» l'affirmation métalinguistique de sa propre indémontrabilité est-elle possible? On prétend échapper au paradoxe sémantique parce que l'énoncé dans *P* dont le nombre de Gödel est *i* n'est pas auto-référentiel, malgré le fait que son prédicat soit corrélé avec la notion métalinguistique de la démontrabilité. Delong,

2 Cf. *op. cit.*, p. 176. Il me semble caractéristique de l'utilisation vague du mot 'vrai' avant la célèbre définition de Tarski que Gödel lui-même emploie l'expression 'correct' ('richtig').

par exemple, tout en admettant qu'il a en effet le nombre de Gödel *i*, avance qu'il n'exprime pas lui-même ce fait[3]. Cet argument me semble pourtant spécieux, car je ne vois pas comment l'énoncé en question pourrait ne pas être auto-référentiel, s'il doit passer pour une proposition authentiquement mathématique.

Bien que je ne prétende pas connaître l'«essence» des mathématiques, je crois qu'il y a de bonnes raisons de repousser le platonisme. On peut résumer la conception relativiste que je lui oppose en insistant sur le fait qu'en pratiquant les mathématiques nous ne décrivons pas des états de choses mathématiques, mais que nous les créons nous-mêmes par l'utilisation de certains systèmes linguistiques. Dans la mesure où j'attribue un rôle constitutif à la langue dans l'ensemble des activités scientifiques et que je considère comme inéluctable la différentiation entre la langue-objet et la métalangue, il m'est difficile de trouver un sens à «l'interprétation» d'énoncés méta-mathématiques par des énoncés arithmétiques. Comme les similitudes de structure mises en lumière par Gödel me semblent être de nature externe, je doute que ses théorèmes ouvrent réellement toutes les perspectives philosophiques qu'on leur attribue. En effet, ma conviction profonde que le sens des expressions linguistiques est établi de façon strictement interne, c'est-à-dire à l'aide de règles d'utilisation fixées dans un contexte théorique, m'empêche de voir un rapport significatif entre *M* et *A*. Je pense que les énoncés mathématiques ne peuvent être compris que relativement à un système donné ou une théorie établie. Considéré de ce point de vue 'vrai, mais indémontrable (à l'intérieur d'un système!)' ne peut que donner l'impression d'une contradiction.

Mes remarques critiques ne concernent pas la construction formelle de la démonstration, qui est - autant que je peux en juger - irréprochable. Il me semble, en revanche, que les interprétations philosophiques qu'on y a ajoutées ne sont pas vraiment

---

3 Cf. entre autres Howard Delong, *A Profile of Mathematical Logic*. Reading (Mass.): Addison-Wesley Pub. Co 1971, p. 176 sq. et S.G. Shanker, Wittgenstein's Remarks on the Significance of Gödel's Theorem, in: *Gödel's Theorem in Focus*. London/New York: Routledge, 1988, p. 223.

satisfaisantes si on les juge d'après les critères de ma propre philosophie.[4]

Séminaire de Philosophie
Université de Berne
Falkenplatz 16, CH - 3000 Berne

Références bibliographiques

GÖDEL, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, *Monatshefte für Mathematik,* 38, Vienne. (Traduction française in: E. Nagel *et al.* (eds), *Le théorème de Gödel.* Paris: Seuil, 1989, 105-143.

GÖDEL, K. (1947). What Cantor's continuum problem? *American Mathematical Monthly*, 54.

4 Je remercie Markus Sprenger et Jürg Freudiger, qui m'ont rendu attentif à une erreur et à des formulations inexactes dans la première version de ce texte. Dans une communication personnelle, le premier souligne l'importance du théorème de Gödel pour les mathématiques, en signalant qu'en informatique il permet en particulier d'exclure *a priori* de nombreux problèmes comme insolubles pour l'ordinateur. Je remercie également Alain Perrinjaquet qui m'a grandement facilité la tâche en établissant une excellente traduction de mon texte écrit en allemand. Je me suis permis de modifier un certain nombre de formulations qui me semblaient déformer ma pensée.