Les théorèmes d'incomplétude de Gödel

Alexandre Miquel

On se propose de présenter la démonstration des deux théorèmes d'incomplétude dûs au logicien Kurt Gödel [3, 5], dont les énoncés sont les suivants :

> **Premier théorème d'incomplétude.** — *Si $\mathscr{T}$ est une théorie du premier ordre cohérente, récursivement axiomatisable et contenant l'arithmétique de Robinson (*PA^-*), alors $\mathscr{T}$ est incomplète, en ce sens qu'il existe une formule close G dans le langage de $\mathscr{T}$ telle qu'aucune des formules G et $\neg G$ n'est conséquence des axiomes de $\mathscr{T}$.*

> **Second théorème d'incomplétude.** — *Si $\mathscr{T}$ est une théorie du premier ordre cohérente, récursivement axiomatisable et contenant l'arithmétique de Peano (*PA*), alors la formule « Cons$\mathscr{T}$ » (qui dans le langage de $\mathscr{T}$ exprime la cohérence de la théorie $\mathscr{T}$) n'est pas une conséquence des axiomes de $\mathscr{T}$.*

Les termes utilisés dans ces énoncés seront précisés au fil des pages qui viennent.

Survol de la démonstration

La démonstration des théorèmes d'incomplétude de Gödel ne présente pas de difficulté majeure sur le plan conceptuel, mais elle repose sur des codages et des lemmes techniques dont la lourdeur (malheureusement inévitable dans un tel cadre) peuvent rendre la preuve complètement inintelligible. Pour cette raison, nous allons d'abord effectuer un survol de la démonstration, en mettant de côté les détails techniques sur lesquels on reviendra plus loin.

La première mise en abyme : la numérisation

Le premier théorème d'incomplétude de Gödel repose sur l'observation que toutes les structures syntaxiques de l'arithmétique de Peano (PA) — les termes, les formules et les démonstrations — sont des structures de données finies qui peuvent être représentées par des entiers naturels à l'aide d'un codage approprié. Cette observation, qui n'était déjà pas nouvelle à l'époque de Gödel, nous est aujourd'hui familière car l'informatique nous a habitués au fait que toutes les données que nous manipulons sur nos ordinateurs (les textes, les programmes, les bases de données, et même les images et les sons) sont représentées dans les entrailles de la machine par des suites finies de 0 et de 1. Et qu'est-ce qu'une suite finie de 0 et de 1 sinon un entier naturel écrit en base 2 ?

L'intérêt d'une telle *numérisation* des structures syntaxiques de l'arithmétique (i.e. des termes, des formules et des démonstrations) réside dans le fait qu'elle nous permet

d'exprimer les propriétés formelles de ces structures (une fois qu'on les a représentées par des entiers naturels) dans le langage même de l'arithmétique, et de raisonner sur ces structures dans l'arithmétique formelle. La démonstration du théorème de Gödel procède donc d'une véritable mise en abyme, dans laquelle on construit des formules (syntaxiques) qui parlent des formules (numériques) et des démonstrations (syntaxiques) qui établissent les propriétés des démonstrations (numériques).

Les fonctions récursives et leur représentation

Formellement, on *numérise* ainsi l'arithmétique en définissant un codage $e \mapsto \ulcorner e \urcorner$ (injectif) qui à chaque expression syntaxique e de l'arithmétique (un terme, une formule ou une démonstration) associe un entier naturel noté $\ulcorner e \urcorner$. La difficulté ne réside pas dans la définition du codage lui-même, mais dans le fait que chaque opération sur la syntaxe doit ensuite être représentée par une formule dans le langage (rudimentaire) de l'arithmétique. Par exemple, on devra représenter la substitution $A\{x := t\}$ par une formule arithmétique $\mathit{Subst}(a, v, b, a')$ exprimant que

> « a est le code d'une formule A, v est le code d'une variable x, b est le code d'un terme t, et a' est le code de la formule $A\{x := t\}$ »

(Et ainsi de suite pour toutes les autres opérations.)

Pour cela, on procède en deux temps.

D'abord, on définit le codage de telle sorte que toutes les opérations de construction et de destruction des expressions syntaxiques correspondent (à travers le codage) à des fonctions récursives. C'est pour cette raison qu'on ne peut pas se contenter d'une énumération arbitraire des expressions, et qu'il nous faut un véritable codage. Une fois le codage défini, il est facile de vérifier que toutes les autres opérations sur la syntaxe (comme par exemple la substitution) peuvent être elles-aussi définies (du côté des entiers) à l'aide de fonctions récursives. De même, on doit s'assurer que la fonction qui teste si un entier est un code de démonstration est également une fonction récursive, et c'est la raison pour laquelle on doit se restreindre aux théories récursives, c'est-à-dire aux théories $\mathscr{T}$ telles que l'ensemble des codes des axiomes de $\mathscr{T}$ est récursif.

Ensuite, on démontre un *théorème de représentation*, qui exprime que le graphe de chaque fonction récursive peut être représenté par une formule du langage de l'arithmétique (en un sens qui devra être précisé). De la sorte, on obtient toutes les formules qui permettent d'exprimer les opérations désirées.

Le paradoxe du menteur

Le cœur de la démonstration du premier théorème d'incomplétude réside dans la construction (par un procédé de diagonalisation assez standard) d'une formule close G telle que

$$G \;\equiv\; \neg\exists z\; \mathit{Dem}(z, \ulcorner G \urcorner)\,,$$

où $\mathit{Dem}(d, a)$ est la formule arithmétique qui exprime que d est le code d'une démonstration de la formule représentée par le code a, et où $\ulcorner G \urcorner$ désigne le code de la formule G elle-même. En substance, la formule G nous dit donc : « je ne suis pas démontrable ». On retrouve là l'idée du paradoxe du menteur (« je mens ») transposée à la logique, et il est fort possible que Gödel ait construit la formule G dans l'espoir de dériver une contradiction dans l'arithmétique.

Cependant, la formule G ne conduit pas (du moins pas directement) à une contradiction, car elle ne parle pas de *vérité* (« je ne suis pas vraie ») mais de *prouvabilité*

(« je ne suis pas démontrable »). Cette limitation vient du fait que, contrairement à la prouvabilité, la notion de vérité n'est pas définissable à l'aide d'une formule de l'arithmétique. (C'est un résultat classique dû à Tarski, et dont la démonstration est en fait calquée sur celle du théorème de Gödel.) Grâce à cette subtile différence, on évite la contradiction, mais on vérifie sans trop de difficulté que si l'arithmétique est cohérente, alors la formule G ne peut pas être démontrée dans la théorie, de même que sa négation. (Dans le second cas, on a besoin d'une hypothèse un peu plus forte que la cohérence, sur laquelle on reviendra par la suite.) Voilà pour le premier théorème.

Précisons qu'il n'est pas nécessaire de chercher à comprendre le sens de la fameuse formule G, car cette formule a précisément été construite (de manière très artificielle) pour n'avoir aucun sens, du moins pas tel qu'on l'entend ordinairement en mathématiques. Intuitivement, la formule G ne parle plus vraiment des entiers, elle ne parle que d'elle-même. Si cette formule n'est pas décidable dans le système formel de l'arithmétique, c'est sans doute parce qu'elle n'énonce pas grand chose de plus que du bruit [1].

La portée du premier théorème

Si la formule G en elle-même n'a pas grand intérêt, son existence est en revanche très instructive, à la fois philosophiquement et techniquement.

Sur le plan philosophique, l'existence de la formule G montre que même un système formel défini de la manière la plus rigoureuse qui soit est tout-à-fait capable d'engendrer du bruit, c'est-à-dire des formules qui ne parlent plus vraiment des objets que le système formel est censé décrire. (D'où leur indécidabilité.) Autrement dit, la rigueur et le formalisme de la méthode axiomatique ne nous protègent pas totalement contre la production de non-sens.

Sur le plan technique, on peut démontrer que la formule G est impliquée par certaines formules arithmétiques qui, pour le coup, sont très intéressantes, et même prouvables dans des théories strictement plus fortes que l'arithmétique. L'archétype d'une telle formule est la formule « $Cons$PA » qui exprime (dans le langage de l'arithmétique, et à travers le codage des formules et des démonstrations dans les entiers) que l'arithmétique de Peano est cohérente. (La preuve de l'implication $Cons\text{PA} \Rightarrow G$ constitue en effet le cœur de la démonstration du second théorème d'incomplétude.) Or il est clair que si une implication de la forme $A \Rightarrow G$ est démontrable dans l'arithmétique (où G est la formule indécidable de Gödel), alors la formule A ne peut pas non plus être démontrée dans l'arithmétique.

Cette méthode nous permet donc d'entrevoir les *limites* de l'arithmétique formelle, en nous donnant des exemples de formules arithmétiques qu'on sait démontrer dans des systèmes plus puissants (par exemple la théorie des ensembles), mais qu'on ne peut pas démontrer dans PA. (Par exemple, la formule $Cons\text{PA}$ est démontrable dans la théorie des ensembles de Zermelo-Fraenkel, et même dans des systèmes formels beaucoup moins puissants.)

La seconde mise en abyme : l'internalisation

Le premier théorème d'incomplétude de Gödel établit donc que :

Si l'arithmétique est cohérente, alors G n'est pas démontrable.

1. La production de bruit est une conséquence fréquente de l'abus du procédé de mise en abyme. Il suffit de penser à la *webcam* qui filme le moniteur branché à la *webcam*, ou au micro qui enregistre le haut-parleur branché au micro, qui est à l'origine du fameux *effet Larsen*. De fait, on peut légitimement considérer la formule G comme une forme d'effet Larsen en logique.

L'examen approfondi de cette démonstration (effectuée en dehors de l'arithmétique) montre qu'elle n'utilise aucun principe de raisonnement qui ne soit arithmétisable. On peut donc procéder à une seconde mise en abyme, en reconstruisant la démonstration du premier théorème d'incomplétude *à l'intérieur* du système formel de l'arithmétique. Évidemment, ce processus de reconstruction nécessite de remplacer chaque énoncé par sa version numérisée : l'énoncé « l'arithmétique est cohérente » est remplacé par la formule « $Cons$PA » tandis que l'énoncé « G n'est pas démontrable » est remplacé par la formule « $\neg\exists z\, Dem(z, \ulcorner G \urcorner)$ ». On obtient ainsi une démonstration formelle, dans l'arithmétique de Peano, de la formule

$$Cons\text{PA} \Rightarrow \neg\exists z\, Dem(z, \ulcorner G \urcorner)$$

c'est-à-dire précisément une démonstration de l'implication

$$Cons\text{PA} \Rightarrow G\,.$$

Comme on sait que la formule G n'est pas démontrable dans l'arithmétique (sous l'hypothèse que l'arithmétique est cohérente), la formule $Cons$PA ne peut donc pas non plus être démontrée dans l'arithmétique. Autrement dit, l'arithmétique ne peut pas démontrer sa propre cohérence... sauf si elle incohérente, auquel cas elle peut évidemment démontrer tout et n'importe quoi !

Les hypothèses des deux théorèmes d'incomplétude

Une analyse de la preuve du premier théorème d'incomplétude montre qu'elle n'utilise en réalité qu'un tout petit fragment de l'arithmétique de Peano, essentiellement celui qui est nécessaire pour établir le théorème de représentation. Ce fragment, qu'on appelle l'*arithmétique de Robinson* (PA$^-$), est le fragment qui permet d'exprimer toutes les propriétés calculatoires des entiers *standard*. Formellement, l'arithmétique de Robinson est définie en supprimant toutes les instances du schéma de récurrence sauf une, à savoir celle qui permet de démontrer que tout entier est soit nul, soit le successeur d'un autre entier. Dans l'arithmétique de Robinson, on peut encore démontrer que toutes les additions commutent ($2 + 3 = 3 + 2$, $42 + 18 = 18 + 42$, etc.) mais pas que l'addition est commutative en général ($\forall x\, \forall y\, (x + y = y + x)$).

Grâce à cette remarque, il est possible d'établir le premier théorème d'incomplétude pour l'arithmétique de Robinson (PA$^-$), et plus généralement pour n'importe quelle théorie $\mathscr{T}$ qui est une extension récursive de l'arithmétique de Robinson. (Ce qui est le cas de l'arithmétique de Peano.) On notera que l'hypothèse selon laquelle la théorie $\mathscr{T}$ est récursive est essentielle, car c'est elle qui permet d'appliquer le théorème de représentation à la fonction testant si un entier est le code d'une démonstration correcte dans $\mathscr{T}$. Par ailleurs, l'arithmétique (de Robinson ou de Peano) a des extensions *non récursives* qui sont complètes, ainsi que le montre la preuve du théorème de complétude de la logique du premier ordre — un autre résultat fameux dû à Gödel.

Si la preuve du premier théorème d'incomplétude n'utilise pas le schéma de récurrence interne (i.e. de l'arithmétique formelle, numérisée), elle utilise en revanche à tour de bras les récurrences externes : inductions structurelles sur les termes, les formules et les démonstrations, ne serait-ce que pour définir le codage et montrer ses propriétés. Lorsque la preuve du premier théorème d'incomplétude est à son tour internalisée dans l'arithmétique (lors de la seconde mise en abyme), ces récurrences se retrouvent donc au niveau interne, et c'est pourquoi le second théorème d'incomplétude ne vaut que pour les théories récursives contenant toute l'arithmétique de Peano (PA).

1 L'arithmétique de Peano (PA)

Dans ce qui suit, on ne s'intéresse qu'aux théories du premier ordre définies sur le langage de l'arithmétique (cf section 1.1).

1.1 Le langage de l'arithmétique

Les termes (notation : t, u, etc.) et les formules (notation : A, B, C, etc.) du langage de l'arithmétique (noté $\mathcal{L}$) sont définis par la grammaire suivante :

$$\begin{array}{lrcl} \textbf{Termes} & t, u & ::= & x \mid 0 \mid s(t) \mid t + u \mid t \times u \\ \textbf{Formules} & A, B & ::= & t = u \mid \bot \mid A \Rightarrow B \\ & & & \mid A \wedge B \mid A \vee B \mid \forall x\, A \mid \exists x\, A \end{array}$$

On adopte ici le point de vue suivant lequel les termes et les formules sont des arbres finis dont les nœuds sont étiquetés par des symboles qui comprennent :

- les symboles de variables (notés x, y, z, etc.) que l'on suppose donnés en nombre infini dénombrable ;
- le symbole de constante 0 (« zéro »), le symbole de fonction unaire s (« successeur »), et les symboles de fonction binaires $+$ (« plus ») et $\times$ (« fois ») ;
- le symbole de prédicat binaire $=$ (« égale ») ;
- la constante propositionnelle $\bot$ (« absurde ») et les connecteurs binaires $\Rightarrow$ (« implique »), $\wedge$ (« et ») et $\vee$ (« ou ») ;
- les quantificateurs $\forall$ (« pour tout ») et $\exists$ (« il existe »).

Cette convention nous dispense en particulier d'introduire des symboles supplémentaires « (» et «) » à des fins de parenthésage. On continuera cependant d'écrire les termes et les formules de manière linéaire, avec toutes les parenthèses nécessaires pour prévenir les ambiguïtés de lecture.

Opérations sur les termes L'ensemble des variables (libres) d'un terme t est noté $FV(t)$. (On rappelle que dans un terme, toute occurrence d'une variable est nécessairement une occurrence libre.) On dit que le terme t est *clos* lorsque $FV(t) = \varnothing$; sinon on dit que t est *ouvert*.

Si t, u sont des termes, et si x est une variable, on note $t\{x := u\}$ le terme obtenu en substituant le terme u à chaque occurrence (libre) de la variable x dans le terme t. (La définition de cette forme de substitution ne présente aucune difficulté car elle ne nécessite aucun renommage.)

Enfin, on note $1 = s(0)$, $2 = s(1)$, $3 = s(2)$, etc. et plus généralement, on associe à tout entier naturel n le terme clos $\overline{n}$ défini par

$$\overline{n} \;\equiv\; s^n(0) \;\equiv\; \underbrace{s(\cdots s}_{n}(0)\cdots)\,.$$

Opérations sur les formules La manipulation des variables est plus délicate dans les formules que dans les termes, car celles-ci peuvent contenir des quantifications ($\forall x$, $\exists x$) qui rendent muette la variable x sur laquelle elles portent. On doit donc en permanence distinguer dans une formule les occurrences de variables qui sont liées (i.e. qui figurent dans le scope d'une quantification portant sur la même variable) de celles qui sont libres (i.e. qui ne figurent dans le scope d'aucune quantification portant sur la

même variable), en faisant attention au fait qu'une même variable peut avoir à la fois des occurrences libres et des occurrences liées dans une même formule.

L'ensemble des variables libres d'une formule A est noté $FV(A)$, tandis que l'ensemble de ses variables liées est noté $BV(A)$. On dit que la formule A est *close* lorsque $FV(A) = \varnothing$; sinon on dit que A est *ouverte*.

Étant données deux formules A et A', on écrit $A \equiv_\alpha A'$ lorsque les formules A et A' sont α-*équivalentes*, c'est-à-dire lorsque ces deux formules se déduisent l'une de l'autre par un renommage de certaines de leurs variables liées. On ne donnera pas la définition formelle de la relation d'α-équivalence ni de la notion de renommage (ces deux notions sont en effet assez délicates à définir formellement), et le lecteur intéressé pourra se reporter à [1] pour une présentation détaillée.

Enfin, si A est une formule, x une variable et u un terme, on note $A\{x := u\}$ la formule obtenue en substituant le terme u à chaque occurrence libre de la variable x dans la formule A. On notera que cette opération de substitution nécessite de procéder à des renommages de variables liées dans la formule A afin d'éviter qu'une variable libre du terme u ne soit capturée par une quantification sur le même nom de variable dans la formule A. On trouvera dans [1] une définition formelle de cette opération.

Abréviations La négation ($\neg A$) et l'équivalence logique ($A \Leftrightarrow B$) sont définies dans le langage de l'arithmétique à travers les abréviations

$$\begin{aligned} \neg A &\equiv A \Rightarrow \bot \\ A \Leftrightarrow B &\equiv (A \Rightarrow B) \wedge (B \Rightarrow A) \end{aligned}$$

de même que la formule $\exists! x\, A$ (« il existe un unique x tel que A ») :

$$\exists! x\, A \equiv \exists x\, (A \wedge \forall z\, (A\{x := z\} \Rightarrow z = x)) \qquad (z \notin FV(A))$$

La relation d'ordre $x \leq y$ (au sens large) et la relation d'ordre strict $x < y$ sont quant à elles définies par

$$\begin{aligned} t \leq u &\equiv \exists z\, (t + z = u) && (z \notin FV(t) \cup FV(u)) \\ t < u &\equiv s(t) \leq u \equiv \exists z\, (s(t) + z = u) && (z \notin FV(t) \cup FV(u)) \end{aligned}$$

Enfin, on utilisera les abréviations

$$t \neq u \equiv \neg(t = u), \qquad t \not\leq u \equiv \neg(t \leq u), \qquad t \not< u \equiv \neg(t < u)\,.$$

1.2 Le système de déduction

Il existe de nombreux systèmes de déduction capturant la notion de conséquence logique (au sens du calcul des prédicats classique), dont les plus connus sont le système de déduction de Hilbert, la déduction naturelle classique et le calcul des séquents de Gentzen. On choisit ici d'utiliser la déduction naturelle classique (avec contextes explicites), qui est un bon compromis entre la commodité d'utilisation (pour écrire des démonstrations formelles) et la facilité de codage (cf section 4.6).

Formellement, on appelle un *contexte* toute liste (éventuellement vide) de formules notée $\Gamma \equiv A_1, \ldots, A_n$. Les notations $FV(\Gamma)$ et $\Gamma\{x := u\}$ sont étendues aux contextes en posant :

$$\begin{aligned} FV(A_1, \ldots, A_n) &\equiv FV(A_1) \cup \cdots \cup FV(A_n) \\ (A_1, \ldots, A_n)\{x := u\} &\equiv A_1\{x := u\}, \ldots, A_n\{x := u\}\,. \end{aligned}$$

Un *séquent* est un couple (Γ, A) noté $\Gamma \vdash A$, où Γ est un contexte (i.e. le *subséquent*) et où A est une formule (i.e le *conséquent*). Intuitivement, le séquent $\Gamma \vdash A$ exprime que la formule A est conséquence logique des hypothèses Γ.

Les séquents se déduisent les uns des autres à l'aide des 14 règles d'inférence données dans la Fig. 1. Chacune de ces règles est de la forme

$$\frac{S_1 \quad \cdots \quad S_n}{S}$$

où le séquent S est la *conclusion* de la règle et où les séquents $S_1, \ldots, S_n$ sont ses *prémisses*. Certaines règles (la règle Axiome, la règle d'introduction du $\forall$ et la règle d'élimination du $\exists$) comportent une condition de bord (figurant à droite de la règle) qui en restreint la portée. Par exemple, la règle Axiome n'est utilisable qui si la formule A apparaît dans la liste Γ des hypothèses, tandis que la règle d'introduction du quantificateur universel n'est utilisable que si la variable x n'a pas d'occurrence libre dans le contexte Γ (intuitivement : x désigne un objet quelconque).

(Axiome, $\bot$) $\quad \dfrac{}{\Gamma \vdash A}\ {\scriptstyle (A \in \Gamma)} \qquad \dfrac{\Gamma, \neg A \vdash \bot}{\Gamma \vdash A}$

($\Rightarrow$) $\quad \dfrac{\Gamma, A \vdash B}{\Gamma \vdash A \Rightarrow B} \qquad \dfrac{\Gamma \vdash A \Rightarrow B \quad \Gamma \vdash A}{\Gamma \vdash B}$

($\wedge$) $\quad \dfrac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B} \qquad \dfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \qquad \dfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B}$

($\vee$) $\quad \dfrac{\Gamma \vdash A}{\Gamma \vdash A \vee B} \qquad \dfrac{\Gamma \vdash B}{\Gamma \vdash A \vee B} \qquad \dfrac{\Gamma \vdash A \vee B \quad \Gamma, A \vdash C \quad \Gamma, B \vdash C}{\Gamma \vdash C}$

($\forall$) $\quad \dfrac{\Gamma \vdash A}{\Gamma \vdash \forall x\, A}\ {\scriptstyle (x \notin FV(\Gamma))} \qquad \dfrac{\Gamma \vdash \forall x\, A}{\Gamma \vdash A\{x := t\}}$

($\exists$) $\quad \dfrac{\Gamma \vdash A\{x := t\}}{\Gamma \vdash \exists x\, A} \qquad \dfrac{\Gamma \vdash \exists x, A \quad \Gamma, A \vdash B}{\Gamma \vdash B}\ {\scriptstyle (x \notin FV(\Gamma, B))}$

FIGURE 1 – Les règles d'inférence de la déduction naturelle classique

Les enchaînements d'inférences sont alors disposés sous forme d'arbres de dérivation, dont la racine est traditionnellement placée en bas et les feuilles en haut. Formellement, on appelle un *arbre de dérivation* (ou une *dérivation*) tout arbre fini étiqueté par des séquents tel qu'à chaque nœud de cet arbre, si S désigne l'étiquette de ce nœud et si $P_1, \ldots, P_n$ désignent respectivement les étiquettes des n fils de ce nœud, alors les séquents S (la conclusion) et $P_1, \ldots, P_n$ (les prémisses) sont reliés entre eux par l'une des 14 règles d'inférence de la Fig. 1 (en respectant la condition de bord s'il y a lieu).

On dit alors qu'un séquent $\Gamma \vdash A$ est *dérivable* s'il existe un arbre de dérivation dont ce séquent est la racine.

1.3 La notion de théorie

On ne s'intéresse ici qu'aux théories du premier ordre construites sur le langage de l'arithmétique. Dans ce cadre, une *théorie* $\mathscr{T}$ est essentiellement un ensemble (fini ou infini) de formules closes, qu'on appelle les *axiomes* de $\mathscr{T}$.

Une théorie $\mathscr{T}$ permet de déduire un certain nombre de théorèmes par voie de conséquence logique. Formellement, on appelle une *démonstration* de la formule A dans la théorie $\mathscr{T}$ toute dérivation D dont la conclusion est de la forme $\Gamma \vdash A$, où Γ est un ensemble (fini) d'axiomes de $\mathscr{T}$. Lorsqu'une telle démonstration existe, on dit que A est un *théorème* de $\mathscr{T}$, ce que l'on note $\mathscr{T} \vdash A$.

On dit que la théorie $\mathscr{T}$ est *incohérente* lorsque la formule $\bot$ est un théorème de $\mathscr{T}$ ou, ce qui revient au même, lorsque toute formule A est un théorème de $\mathscr{T}$. Dans le cas contraire, on dit que la théorie $\mathscr{T}$ est *cohérente*.

1.4 La théorie de l'égalité ($\mathscr{E}$)

On appelle la *théorie de l'égalité* (sur le langage de l'arithmétique) et on note $\mathscr{E}$ la théorie dont les 7 axiomes sont les suivants :

(E1) $\quad \forall x\ (x = x)$
(E2) $\quad \forall x\ \forall y\ \forall z\ (x = y \wedge x = z \Rightarrow y = z)$
(E3) $\quad \forall x\ \forall y\ (x = y \Rightarrow s(x) = s(y))$
(E4) $\quad \forall x\ \forall y\ \forall z\ (x = y \Rightarrow x + z = y + z)$
(E5) $\quad \forall x\ \forall y\ \forall z\ (x = y \Rightarrow z + x = z + y)$
(E6) $\quad \forall x\ \forall y\ \forall z\ (x = y \Rightarrow x \times z = y \times z)$
(E7) $\quad \forall x\ \forall y\ \forall z\ (x = y \Rightarrow z \times x = z \times y)$

L'axiome (E1) exprime que l'égalité est réflexive, et la conjonction de (E1) et de (E2) permet de dériver que l'égalité est symétrique et transitive :

$$\mathscr{E} \vdash x = y \Rightarrow y = x$$
$$\mathscr{E} \vdash x = y \wedge y = z \Rightarrow x = z$$

(Il s'agit donc d'une relation d'équivalence dans la théorie $\mathscr{E}$.)

Par ailleurs, l'axiome (E3) exprime que la fonction successeur est compatible avec la relation d'égalité, tandis que les axiomes (E4)–(E7) expriment que l'addition et la multiplication sont toutes les deux compatibles (à gauche et à droite) avec l'égalité. De ces axiomes on peut tirer le *principe de remplacement de Leibniz*, à savoir le schéma de théorèmes

$$\mathscr{E} \vdash x = y \ \Rightarrow\ (A\{z := x\} \Leftrightarrow A\{z := y\})$$

où A désigne n'importe quelle formule du langage. (La démonstration de ce principe est construite par récurrence sur la structure de la formule A.)

1.5 L'arithmétique de Peano (PA)

L'*arithmétique de Peano*, notée PA est la théorie obtenue en ajoutant à la théorie de l'égalité ($\mathscr{E}$) les six axiomes ci-dessous

(PA1) $\forall y\ (0 + y = y)$
(PA2) $\forall x\ \forall y\ (s(x) + y = s(x + y))$
(PA3) $\forall y\ (0 \times y = 0)$
(PA4) $\forall x\ \forall y\ (s(x) \times y = (x \times y) + y)$
(PA5) $\forall x\ \forall y\ (s(x) = s(y) \Rightarrow x = y)$
(PA6) $\forall x\ (s(x) \neq 0)$

ainsi que les *axiomes de récurrence*

(PA7) $$\forall z_1 \cdots z_n\ (A\{x := 0\} \wedge \forall x\,(A \Rightarrow A\{x := s(x)\}) \Rightarrow \forall x\, A)$$

pour chaque formule A de variables libres $x, z_1, \ldots, z_n$.

Expressivité de l'arithmétique de Peano Bien que son langage soit rudimentaire, l'arithmétique de Peano est une théorie très expressive qui permet de démontrer les propriétés élémentaires de l'addition et de la multiplication (associativité, commutativité, éléments neutres, distributivité, etc.), de l'ordre (réflexivité, transitivité, antisymétrie et totalité) et de l'ordre strict. On notera en particulier que le principe de récurrence forte

$$\text{PA} \ \vdash\ \forall x\,(\forall y\,(y < x \Rightarrow A(y)) \Rightarrow A(x)) \ \Rightarrow\ \forall x\, A(x)$$

et la propriété de bon ordre

$$\text{PA} \ \vdash\ \exists x\, A(x) \ \Rightarrow\ \exists x\,(A(x) \wedge \forall y\,(A(y) \Rightarrow x \leq y))$$

sont tous les deux prouvables dans PA (pour toute formule $A(x)$).

Les théorèmes de l'arithmétique de Peano ne se limitent pas aux propriétés élémentaires mentionnées ci-dessus, mais concernent également les propriétés de la divisibilité et de l'arithmétique modulaire (division euclidienne, théorème de Bézout, théorème des restes Chinois, etc.) de même que les propriétés des nombres premiers, en utilisant l'abréviation

$$Prim(x) \ \equiv\ x \neq 1 \wedge \forall y\, \forall z\,(x = y \times z \Rightarrow y = 1 \vee z = 1)\,.$$

On notera cependant que la formalisation de ces résultats dans PA nécessite un certain travail d'adaptation, tant au niveau de l'énoncé que de la démonstration. Ainsi on énoncera le théorème d'Euclide (« il existe une infinité de nombres premiers ») sous sa forme originelle

$$\forall x\, \exists y\,(y > x \wedge Prim(y))\,,$$

et on construira la démonstration de ce théorème en vérifiant successivement que les formules suivantes sont démontrables dans PA :

1. $\forall x\,(x \geq 2 \Rightarrow \exists y\,(Prim(y) \wedge y|x))$
2. $\forall x\, \exists y\,(y \neq 0 \wedge \forall x'\,(1 \leq x' \wedge x' \leq x \Rightarrow x'|y))$
3. $\forall x\, \forall y\,\big(y \neq 0 \wedge \forall x'\,(1 \leq x' \wedge x' \leq x \Rightarrow x'|y)$
   $\Rightarrow\ \forall z\,(Prim(z) \wedge z|s(y) \Rightarrow x < z)\big)$

(On notera que par rapport à la démonstration traditionnelle, le produit des nombres premiers inférieurs ou égaux à x est remplacé par un nombre y divisible par tous les entiers de 1 à x, qui fait tout aussi bien l'affaire.)

1.6 Le modèle standard

On appelle le *modèle standard* de l'arithmétique la $\mathcal{L}$-structure (où $\mathcal{L}$ désigne le langage de l'arithmétique) dont le domaine est l'ensemble $\mathbb{N}$ des entiers naturels, et dans laquelle chaque symbole de $\mathcal{L}$ est interprété de la manière évidente (le symbole 0 par l'entier 0, le symbole s par la fonction $n \mapsto n+1$, etc.) Par abus de langage, cette $\mathcal{L}$-structure formée sur l'ensemble $\mathbb{N}$ est encore notée $\mathbb{N}$.

Formellement, on associe à chaque terme clos t du langage $\mathcal{L}$ un entier naturel noté $\mathrm{Val}(t)$ et appelé la *valeur* du terme t. Cette valeur est définie par récurrence sur la structure de t à l'aide des équations :

$$\begin{array}{rclcrcl} \mathrm{Val}(0) & = & 0 & \qquad & \mathrm{Val}(t+u) & = & \mathrm{Val}(t)+\mathrm{Val}(u) \\ \mathrm{Val}(s(t)) & = & \mathrm{Val}(t)+1 & & \mathrm{Val}(t\times u) & = & \mathrm{Val}(t)\mathrm{Val}(u) \end{array}$$

On notera en particulier que $\mathrm{Val}(\overline{n}) = n$ pour tout $n \in \mathbb{N}$.

On définit ensuite la relation de satisfaction $\mathbb{N} \vDash A$ (où A est une formule close), qui exprime que la formule A est vraie dans le modèle standard :

Définition 1 (Satisfaction d'une formule dans le modèle standard) — La relation de satisfaction $\mathbb{N} \vDash A$ (où A est une formule close) est définie par récurrence sur le nombre de quantificateurs et de connecteurs de la formule A en posant :

1. $\mathbb{N} \vDash t = u$ ssi $\mathrm{Val}(t) = \mathrm{Val}(u)$;
2. $\mathbb{N} \nvDash \bot$;
3. $\mathbb{N} \vDash A \Rightarrow B$ ssi $\mathbb{N} \nvDash A$ ou $\mathbb{N} \vDash B$;
4. $\mathbb{N} \vDash A \wedge B$ ssi $\mathbb{N} \vDash A$ et $\mathbb{N} \vDash B$;
5. $\mathbb{N} \vDash A \vee B$ ssi $\mathbb{N} \vDash A$ ou $\mathbb{N} \vDash B$;
6. $\mathbb{N} \vDash \forall x\, A$ ssi pour tout $n \in \mathbb{N}$ on a $\mathbb{N} \vDash A\{x := \overline{n}\}$;
7. $\mathbb{N} \vDash \exists x\, A$ ssi il existe $n \in \mathbb{N}$ tel que $\mathbb{N} \vDash A\{x := \overline{n}\}$.

On vérifie d'abord que chaque axiome de Peano est vrai dans le modèle standard :

Proposition 1 — *Si A est un axiome de* PA*, alors $\mathbb{N} \vDash A$.*

D'où il s'ensuit que toute formule prouvable dans l'arithmétique de Peano est vraie dans le modèle standard :

Proposition 2 — *Si A est une formule close telle que* $\mathrm{PA} \vdash A$*, alors $\mathbb{N} \vDash A$.*

Démonstration. On commence par démontrer (par récurrence sur la structure de la dérivation) que si le séquent $A_1, \ldots, A_p \vdash B$ de variables libres $x_1, \ldots, x_k$ est dérivable, alors pour tous $n_1, \ldots, n_k \in \mathbb{N}$ tels que $\mathbb{N} \vDash A_i\{x_1 := \overline{n}_1; \ldots; x_k := \overline{n}_k\}$ (pour $i \in [1..p]$), on a $\mathbb{N} \vDash B\{x_1 := \overline{n}_1; \ldots; x_k := \overline{n}_k\}$. La proposition découle de ce résultat d'après la Prop. 1 et la définition de la prouvabilité dans PA. □

Une définition infinitaire Par construction, la relation de satisfaction $\mathbb{N} \vDash A$ (dans le modèle standard) effectue une véritable traduction de la formule close A dans le langage de la « théorie ambiante » [2], en donnant aux symboles qui figurent dans A leur signification usuelle dans l'ensemble des entiers naturels. Par exemple, l'énoncé

$$\mathbb{N} \;\vDash\; \forall x\, \exists y_1 \exists y_2\, (Prim(y_1) \wedge Prim(y_2) \wedge y_1 + y_2 = 2 \times x + 4)$$

2. C'est à dire la théorie (au sens intuitif) dans laquelle on se place pour raisonner sur les termes, les formules, les démonstrations, etc. de l'arithmétique formelle, par opposition à la théorie *formelle* de l'arithmétique proprement dite, sur laquelle porte notre raisonnement. La théorie ambiante (il s'agit évidemment d'une notion intuitive, non définie) est également appelée la *méta-théorie*, ou encore la *théorie externe*.

est par définition équivalent (dans la théorie ambiante) à l'énoncé

Tout nombre pair supérieur à 2 est la somme de deux nombres premiers.

Ainsi pour déterminer si la formule

$$\forall x\, \exists y_1 \exists y_2\, (Prim(y_1) \wedge Prim(y_2) \wedge y_1 + y_2 = 2 \times x + 4)$$

est vraie dans le modèle standard, on devra résoudre (positivement ou négativement) la conjecture de Goldbach dans la théorie ambiante, ni plus ni moins [3]. On voit donc avec cet exemple que la définition de la relation de satisfaction $\mathbb{N} \vDash A$ ne relève pas d'un simple calcul (contrairement à toutes les notions que nous avons introduites auparavant), mais constitue de manière authentique une définition infinitaire.

Techniquement, la définition de la relation de satisfaction $\mathbb{N} \vDash A$ (Déf. 1) est une définition récursive qui repose implicitement sur un ordre bien fondé sur l'ensemble des formules closes. Suivant cette définition, on ne peut définir la satisfaction de la formule $\forall x\, A(x)$ qu'après avoir défini la satisfaction de toutes les formules $A(\overline{n})$, où n parcourt l'ensemble des entiers naturels. Une telle définition n'est possible en pratique que si la théorie ambiante dispose de mécanismes permettant de construire des objets infinitaires (par exemple des ensembles infinis) et de raisonner sur ces objets.

Par exemple en théorie des ensembles (prise comme théorie ambiante), on peut définir la relation de satisfaction dans le modèle standard de la manière suivante :

1. Pour chaque formule $|A|$, on note $|A|$ la taille de la formule A, c'est-à-dire le nombre de connecteurs et de quantificateurs de A, et pour chaque entier $k \in \mathbb{N}$ on note $\mathcal{L}_k$ l'ensemble des formules closes de taille k.
2. Pour chaque entier naturel k, on définit ensuite un ensemble $\mathscr{V}_k \subseteq \mathcal{L}_k$ dont les éléments sont les formules closes de taille k qui sont satisfaites dans le modèle standard. La définition se fait par récurrence sur k en posant

$$\begin{array}{rcll}
\mathscr{V}_0 & = & \{t = u \in \mathcal{L}_0 \,:\, \mathrm{Val}(t) = \mathrm{Val}(u)\} & \\
\mathscr{V}_k & = & \{A \Rightarrow B \in \mathcal{L}_k \,:\, A \notin \mathscr{V}_{|A|} \text{ ou } B \in \mathscr{V}_{|B|}\} & \cup \\
 & & \{A \wedge B \in \mathcal{L}_k \,:\, A \in \mathscr{V}_{|A|} \text{ et } B \in \mathscr{V}_{|B|}\} & \cup \\
 & & \{A \vee B \in \mathcal{L}_k \,:\, A \in \mathscr{V}_{|A|} \text{ ou } B \in \mathscr{V}_{|B|}\} & \cup \\
 & & \{\forall x\, A \in \mathcal{L}_k \,:\, \text{pour tout } n \in \mathbb{N},\ A\{x := \overline{n}\} \in \mathscr{V}_{|A|}\} & \cup \\
 & & \{\exists x\, A \in \mathcal{L}_k \,:\, \text{il existe } n \in \mathbb{N} \text{ t.q. } A\{x := \overline{n}\} \in \mathscr{V}_{|A|}\} & (k \geq 1)
\end{array}$$

   On notera que cette définition est bien formée, car chaque ensemble $\mathscr{V}_k$ est défini uniquement à partir des ensembles $\mathscr{V}_0, \ldots, \mathscr{V}_{k-1}$.
3. Enfin, on pose $\mathbb{N} \vDash A$ ssi $A \in \mathscr{V}_{|A|}$.

On pourra remarquer que :

1. La construction par récurrence de la suite d'ensembles *infinis* $\mathscr{V}_k \subseteq \mathscr{F}_k$ est la seule partie essentiellement infinitaire de la définition ci-dessus.
2. La même construction peut être effectuée dans des formalismes beaucoup plus faibles que la théorie des ensembles, comme par exemple l'arithmétique du second ordre ou même la théorie des types de Martin-Löf (avec univers).

Quoi qu'il en soit, dès que la théorie ambiante est suffisamment expressive pour permettre la construction du modèle standard (c'est-à-dire, en fait, la définition de la relation de satisfaction $\mathbb{N} \vDash A$), il est facile de vérifier que :

3. Autrement dit, la notion de satisfaction dans le modèle standard ne fait rien de plus que déplacer un problème exprimé à l'origine dans la théorie formelle (par une formule) vers la théorie ambiante.

Proposition 3 (Cohérence de PA) — *L'arithmétique de Peano est cohérente.*

Démonstration. Puisque la formule $\bot$ est fausse dans le modèle standard (par définition), celle-ci n'est pas démontrable dans PA d'après la Prop. 2. □

Raisonnements dans un cadre finitaire Évidemment, la preuve de cohérence ci-dessus ne fonctionne que si la théorie ambiante permet de construire des objets infinitaires et de raisonner sur ces objets. Dans un cadre finitaire en revanche — c'est-à-dire dans une théorie ambiante où l'on ne peut raisonner que sur des objets finis [4] — il est toujours possible de raisonner sur les entiers, les termes ou les formules, et même de définir les opérations syntaxiques telles que la substitution dans les formules ou la valeur d'un terme clos. (Le lecteur pourra facilement se convaincre que toutes ces opérations sont finitaires.) Cependant, il n'est plus possible de définir le modèle standard de l'arithmétique, et la question de la cohérence de PA reste ouverte a priori.

Le lecteur pourra vérifier que la plupart des raisonnements que nous effectuerons dans les pages qui suivent sont des raisonnement purement finitaires, et que ces raisonnements peuvent être entièrement écrits (par exemple) dans la théorie des ensembles héréditairement finis. On ne supposera donc pas (en règle générale) que l'arithmétique de Peano est une théorie cohérente, sauf bien entendu dans les cas où on utilisera explicitement le modèle standard.

1.7 Formules à quantifications bornées

Nous venons de voir que la relation de satisfaction $\mathbb{N} \vDash A$ (dans le modèle standard $\mathbb{N}$) est une notion infinitaire. On peut cependant définir cette relation de manière finitaire dans le cas où la formule A a une complexité logique suffisamment faible, et notamment lorsque la formule A est à quantifications bornées.

Pour cela, on introduit les abréviations

$$\begin{array}{rcl} \forall x \leq t\ A & \equiv & \forall x\,(x \leq t \Rightarrow A) \\ \forall x < t\ A & \equiv & \forall x\,(x < t \Rightarrow A) \\ \exists x \leq t\ A & \equiv & \exists x\,(x \leq t \wedge A) \\ \exists x < t\ A & \equiv & \exists x\,(x < t \wedge A) \end{array}$$

où A et t sont quelconques, et où x est telle que $x \notin FV(t)$.

Définition 2 (Formules à quantifications bornées) — On dit qu'une formule est *à quantifications bornées* si elle est construite uniquement à partir des règles suivantes :

1. Si t et u sont des termes, alors les formules $\bot$, $t = u$ et $t \leq u$ sont des formules à quantifications bornées.
2. Si A et B sont des formules à quantifications bornées, alors les formules $A \Rightarrow B$, $A \wedge B$ et $A \vee B$ sont des formules à quantifications bornées.

4. Ce qui n'implique évidemment pas que l'*univers du discours* (formé par tous les objets sur lesquels on s'autorise de raisonner) soit fini. L'arithmétique de Peano est l'archétype des théories finitaires, mais il existe des théories finitaires beaucoup plus faibles, comme par exemple l'arithmétique de Robinson que nous étudierons à la section 2. Un autre exemple de théorie finitaire (en réalité équivalente à l'arithmétique de Peano) est la *théorie des ensembles héréditairement finis*, obtenue en remplaçant (dans la théorie des ensembles) l'axiome de l'infini par sa négation, c'est-à-dire par un axiome exprimant que tous les objets de l'univers sont des ensembles finis (et donc héréditairement finis). C'est plus ou moins le cadre dans lequel on travaillera (informellement) dans ce qui suit.

3. Si x est une variable, t un terme tel que $x \notin FV(t)$, et si A est une formule à quantifications bornées, alors les formules $\forall x < t\ A$ et $\exists x < t\ A$ sont des formules à quantifications bornées [5].

Dans le cas où A est une formule close à quantifications bornées, on peut définir de manière finitaire la relation de satisfaction $\mathbb{N} \vDash A$ (dans le modèle standard) en ajoutant à la Déf. 1 la nouvelle clause

1a. $\mathbb{N} \vDash t \leq u$ ssi $\mathrm{Val}(t) \leq \mathrm{Val}(u)$;

et en remplaçant les clauses (infinitaires) 6. et 7. par les clauses (finitaires) ci-dessous :

6a. $\mathbb{N} \vDash \forall x < t\ A$ ssi pour tout $n < \mathrm{Val}(t)$ on a $\mathbb{N} \vDash A\{x := \overline{n}\}$;

7a. $\mathbb{N} \vDash \exists x < t\ A$ ssi il existe $n < \mathrm{Val}(t)$ tel que $\mathbb{N} \vDash A\{x := \overline{n}\}$.

On peut formaliser (de manière non récursive !) cette définition en théorie des ensembles héréditairement finis en introduisant la notion suivante :

Définition 3 (Certificat) — On appelle un *certificat* tout couple $(\mathscr{V}, \mathscr{F})$ constitué de deux ensembles finis de formules closes à quantifications bornées $\mathscr{V}$ (« formules acceptées ») et $\mathscr{F}$ (« formules rejetées ») tels que :

1. $\bot \notin \mathscr{V}$
2. si $t = u \in \mathscr{V}$, alors $\mathrm{Val}(t) = \mathrm{Val}(u)$;
3. si $t = u \in \mathscr{F}$, alors $\mathrm{Val}(t) \neq \mathrm{Val}(u)$;
4. si $t \leq u \in \mathscr{V}$, alors $\mathrm{Val}(t) \leq \mathrm{Val}(u)$;
5. si $t \leq u \in \mathscr{F}$, alors $\mathrm{Val}(t) > \mathrm{Val}(u)$;
6. si $A \Rightarrow B \in \mathscr{V}$, alors $A \in \mathscr{F}$ ou $B \in \mathscr{V}$;
7. si $A \Rightarrow B \in \mathscr{F}$, alors $A \in \mathscr{V}$ et $B \in \mathscr{F}$;
8. si $A \wedge B \in \mathscr{V}$, alors $A \in \mathscr{V}$ et $B \in \mathscr{V}$;
9. si $A \wedge B \in \mathscr{F}$, alors $A \in \mathscr{F}$ ou $B \in \mathscr{F}$;
10. si $A \vee B \in \mathscr{V}$, alors $A \in \mathscr{V}$ ou $B \in \mathscr{V}$;
11. si $A \vee B \in \mathscr{F}$, alors $A \in \mathscr{F}$ et $B \in \mathscr{F}$;
12. si $\forall x < t\ A \in \mathscr{V}$, alors pour tout $n < \mathrm{Val}(t)$, $A\{x := \overline{n}\} \in \mathscr{V}$;
13. si $\forall x < t\ A \in \mathscr{F}$, alors il existe $n < \mathrm{Val}(t)$ tel que $A\{x := \overline{n}\} \in \mathscr{F}$;
14. si $\exists x < t\ A \in \mathscr{V}$, alors il existe $n < \mathrm{Val}(t)$ tel que $A\{x := \overline{n}\} \in \mathscr{V}$;
15. si $\exists x < t\ A \in \mathscr{F}$, alors pour tout $n < \mathrm{Val}(t)$, $A\{x := \overline{n}\} \in \mathscr{F}$.

Intuitivement, un certificat $(\mathscr{V}, \mathscr{F})$ représente la trace d'un calcul permettant de tester si une formule close à quantifications bornées est satisfaite ou pas dans le modèle standard : les éléments de $\mathscr{V}$ représentent les formules examinées et reconnues comme vraies au cours du calcul, tandis que les éléments de $\mathscr{F}$ représentent les formules examinées et reconnues comme fausses au cours du calcul. On vérifie aisément que :

Lemme 1 — *Pour toute formule close A à quantifications bornées :*
- *soit il existe un certificat $(\mathscr{V}, \mathscr{F})$ tel que $A \in \mathscr{V}$,*
- *soit il existe un certificat $(\mathscr{V}, \mathscr{F})$ tel que $A \in \mathscr{F}$;*

ces deux conditions s'excluant mutuellement.

Démonstration. Ce résultat se démontre par récurrence sur le nombre de connecteurs et de quantificateurs dans la formule A. □

5. Il n'est pas nécessaire de considérer les formules $\forall x \leq t\ A$ et $\exists x \leq t\ A$ dans la définition, car celles-ci sont prouvablement équivalentes aux formules $\forall x < s(t)\ A$ et $\exists x < s(t)\ A$ dans l'arithmétique de Robinson (cf section 2), qui est une forme très affaiblie de l'arithmétique de Peano.

Ce qui justifie la définition :

Définition 4 (Satisfaction d'une formule close à quantifications bornées) — On dit qu'une formule close A à quantifications bornées est satisfaite dans le modèle standard et on note $\mathbb{N} \vDash A$ s'il existe un certificat $(\mathscr{V}, \mathscr{F})$ tel que $A \in \mathscr{V}$.

On vérifie aisément que la définition ci-dessus satisfait les équivalences 1. à 5. de la Déf. 1 ainsi que les équivalences 1a., 6a. et 7a. introduites plus haut. Par ailleurs, si on se place dans un cadre infinitaire (où l'on dispose du modèle standard), on vérifie que pour toute formule close A à quantification bornées, $N \vDash A$ au sens de la Déf. 4 si et seulement si $N \vDash A$ au sens de la Déf. 1.

Le caractère finitaire de la Déf. 4 suggère que les formules closes à quantifications bornées sont en réalité décidables dans PA, en ce sens que pour toute formule close A à quantifications bornées, on a ou bien $\mathrm{PA} \vdash A$ ou bien $\mathrm{PA} \vdash \neg A$ (ces deux conditions ne s'excluant que si PA est cohérente). Nous allons voir que c'est effectivement le cas, et qu'il est même possible de décider les formules closes à quantifications bornées dans une théorie bien plus faible que l'arithmétique de Peano, à savoir l'arithmétique de Peano dans laquelle on a supprimé le principe de récurrence.

2 L'arithmétique de Robinson (PA^-)

2.1 Les axiomes de l'arithmétique de Robinson

L'*arithmétique de Robinson*, notée PA^-, est la théorie du premier ordre obtenue à partir de l'arithmétique de Peano en supprimant tous les axiomes de récurrence (PA7) et en les remplaçant par un unique axiome de « filtrage »

$$(\mathrm{PA7}^-) \qquad \forall x\,(x = 0 \vee \exists y\,(x = s(y)))$$

qui exprime que tout entier naturel est soit nul, soit le successeur d'un autre entier. (Le fait qu'il s'agisse d'un « ou » exclusif découle de (PA6).)

Formellement, l'arithmétique de Robinson est donc la théorie dont les 14 axiomes sont les axiomes (E1)–(E7) définis à la section 1.4, les axiomes (PA1)–(PA6) définis à la section 1.5 ainsi que l'axiome ($\mathrm{PA7}^-$) introduit ci-dessus. (On notera que contrairement à PA, la théorie PA^- est définie à partir d'un ensemble *fini* d'axiomes.)

Il est utile de remarquer que l'axiome de filtrage ($\mathrm{PA7}^-$) est logiquement équivalent à l'axiome de récurrence

$$A(0) \wedge \forall x\,(A(x) \Rightarrow A(s(x))) \Rightarrow \forall x\,A(x)$$

pour la formule $A(x) \equiv x = 0 \vee \exists y\,(x = s(y))$. L'axiome ($\mathrm{PA7}^-$) est donc implicitement contenu dans les axiomes de récurrence (PA7) de l'arithmétique de Peano, ce qui explique pourquoi on ne l'a pas inclus dans la liste d'axiomes donnée à la section 1.5. D'après cette remarque, il est clair que toute formule prouvable dans PA^- est également prouvable dans PA, ce qui nous permet de considérer l'arithmétique de Robinson (PA^-) comme un sous-système de l'arithmétique de Peano (PA).

Expressivité de l'arithmétique de Robinson L'arithmétique de Robinson est évidemment beaucoup moins expressive que l'arithmétique de Peano, car il n'est même pas possible d'y démontrer que l'addition est associative ou commutative :

$$\mathrm{PA}^- \nvdash (x + y) + z = x + (y + z), \qquad \mathrm{PA}^- \nvdash x + y = y + x\,.$$

(Pour s'en convaincre, il suffit de construire un modèle de PA^- — forcément non standard — dans lequel l'addition n'est ni associative, ni commutative.)

Cependant, nous allons voir que le système PA^- reste suffisamment expressif pour qu'il soit possible d'y décider (c'est-à-dire : démontrer ou réfuter) toutes les formules closes à quantifications bornées.

2.2 Décision des égalités et des inégalités closes

On commence par vérifier que :

Lemme 2 — *Pour tous entiers naturels n et m on a :*

1. $\text{PA}^- \vdash \overline{n} + \overline{m} = \overline{n+m}$
2. $\text{PA}^- \vdash \overline{n} \times \overline{m} = \overline{nm}$

Démonstration. Item 1 : La démonstration de l'égalité $\overline{n} + \overline{m} = \overline{n+m}$ (dans PA^-) est construite par récurrence [6] sur l'entier n. Le cas de base correspond à une instance de (PA1), et le cas de récurrence utilise (PA2) combiné avec les axiomes d'égalité.
Item 2 : Même schéma de construction de la démonstration, en utilisant (PA3) pour le cas de base et (PA4) combiné avec l'item 1 pour le cas de récurrence. □

De ce lemme on tire immédiatement que :

Lemme 3 — *Pour tout terme clos t :* $\text{PA}^- \vdash t = \overline{\text{Val}(t)}$.

Démonstration. Par récurrence sur t en utilisant le Lemme 2. □

Par ailleurs :

Lemme 4 — *Pour tous entiers naturels n, m tels que $n \neq m$, on a :* $\text{PA}^- \vdash \overline{n} \neq \overline{m}$.

Démonstration. La démonstration (dans PA^-) de la formule $\overline{n} \neq \overline{m}$ est construite par récurrence sur le plus petit des deux entiers n et m, en utilisant (PA6) pour le cas de base et (PA5) pour le cas de récurrence. □

On peut alors démontrer que :

Proposition 4 (Décision des égalités closes) — *Pour tous termes clos t et u :*

1. *Si* $\mathbb{N} \vDash t = u$, *alors* $\text{PA}^- \vdash t = u$.
2. *Si* $\mathbb{N} \nvDash t = u$, *alors* $\text{PA}^- \vdash t \neq u$.

Démonstration. Soient $n = \text{Val}(t)$ et $m = \text{Val}(u)$. D'après le Lemme 3, il vient $\text{PA}^- \vdash t = \overline{n}$ et $\text{PA}^- \vdash u = \overline{m}$. Dans le cas où $\mathbb{N} \vDash t = u$, c'est-à-dire $n = m$, on conclut que $\text{PA}^- \vdash t = u$ par symétrie et transitivité de l'égalité. Dans le cas où $\mathbb{N} \nvDash t = u$, c'est-à-dire $n \neq m$, on a alors $\text{PA}^- \vdash t = u \Rightarrow \overline{n} = \overline{m}$, d'où l'on tire que $\text{PA}^- \vdash t = u \Rightarrow \bot$ à l'aide du Lemme 4. □

6. Il s'agit bien entendu d'une récurrence au sens de la théorie ambiante, par opposition à la théorie formelle qu'on est en train d'étudier — qui d'ailleurs ne comporte pas d'axiome de récurrence. L'utilisation de cette récurrence « externe » sert ici à construire une démonstration dont la structure et même la taille dépend de l'entier n. L'idée qui est sous-jacente dans toutes les constructions présentées dans cette section est que l'on peut toujours se passer du principe (interne) de récurrence tant qu'on travaille avec des termes clos et avec des formules closes suffisamment simples. Le principe de récurrence est indispensable pour démontrer que $x + y = y + x$, mais pas pour démontrer que $5 + 7 = 7 + 5$ puisque dans ce dernier cas il suffit de faire le calcul. On notera que cette idée est similaire à l'idée (bien connue en compilation) selon laquelle on peut toujours remplacer une boucle $\textbf{for } i = 1 \textbf{ to } n \textbf{ do } C \textbf{ done}$ dont les bornes sont connues à l'avance par le morceau de code $C\{i := 1\}; C\{i := 2\}; \cdots; C\{i := n\}$ dans lequel la boucle a disparu. Le lecteur pourra vérifier que les techniques de déroulage de l'ordre et des quantifications que nous présenterons au fil des pages qui viennent suivent le même esprit que la technique de déroulage des boucles.

Proposition 5 (Décision des inégalités closes) — *Pour tous termes clos t et u :*

1. *Si* $\mathbb{N} \vDash t \leq u$, *alors* $\mathrm{PA}^- \vdash t \leq u$;
2. *Si* $\mathbb{N} \nvDash t \leq u$, *alors* $\mathrm{PA}^- \vdash t \not\leq u$.
3. *Si* $\mathbb{N} \vDash t < u$, *alors* $\mathrm{PA}^- \vdash t < u$;
4. *Si* $\mathbb{N} \nvDash t < u$, *alors* $\mathrm{PA}^- \vdash t \not< u$;

Démonstration. Il suffit de traiter les inégalités larges, puisque les inégalités strictes constituent un cas particulier (car $t < u \equiv s(t) \leq u$). Soient $n = \mathrm{Val}(t)$ et $m = \mathrm{Val}(u)$. On distingue deux cas :

1. $\mathbb{N} \vDash t \leq u$, c'est-à-dire : $n \leq m$. Dans ce cas on a :
   - $\mathrm{PA}^- \vdash t + \overline{m-n} = u$ (Prop. 4)
   - $\mathrm{PA}^- \vdash \exists z\,(t + z = u)$ ($\exists$-intro)
2. $\mathbb{N} \nvDash t \leq u$, c'est-à-dire : $n > m$. Dans ce cas on a :
   - $\mathrm{PA}^- \vdash t \leq u \Rightarrow \exists z\,(\overline{n} + z = \overline{m})$ (Lemme 3)
   - $\mathrm{PA}^- \vdash t \leq u \Rightarrow \exists z\,(s^n(z) = \overline{m})$ (PA1), (PA2)
   - $\mathrm{PA}^- \vdash t \leq u \Rightarrow \exists z\,(s^{n-m}(z) = 0)$ (PA5)
   - $\mathrm{PA}^- \vdash t \leq u \Rightarrow \bot$ (PA6) □

2.3 Déroulage des quantifications et de l'ordre

L'axiome (PA7$^-$) permet de démontrer dans PA^- toute propriété universelle ou existentielle en séparant le cas zéro du cas successeur :

Lemme 5 — *Pour toute formule $A(x)$:*

1. $\mathrm{PA}^- \vdash \exists x\, A(x) \Leftrightarrow A(0) \vee \exists x\, A(s(x))$
2. $\mathrm{PA}^- \vdash \forall x\, A(x) \Leftrightarrow A(0) \wedge \forall x\, A(s(x))$

Démonstration. Pour la quantification existentielle, on vérifie successivement que :
- $\mathrm{PA}^- \vdash A(x) \Leftrightarrow (x = 0 \vee \exists y\,(x = s(y))) \wedge A(x)$ (PA7$^-$)
- $\mathrm{PA}^- \vdash A(x) \Leftrightarrow (x = 0 \wedge A(x)) \vee \exists y\,(x = s(y) \wedge A(x))$
- $\mathrm{PA}^- \vdash A(x) \Leftrightarrow (x = 0 \wedge A(0)) \vee \exists y\,(x = s(y) \wedge A(s(y)))$
- $\mathrm{PA}^- \vdash \exists x\, A(x) \Leftrightarrow \exists x\,(x = 0 \wedge A(0)) \vee \exists x\, \exists y\,(x = s(y) \wedge A(s(y)))$
- $\mathrm{PA}^- \vdash \exists x\, A(x) \Leftrightarrow A(0) \vee \exists y\, A(s(y))$

Le cas de la quantification universelle se traite de manière similaire. □

Grâce à ce lemme, il est possible de dérouler toutes les quantifications existentielles et universelles de la manière suivante :

Proposition 6 — *Pour toute formule $A(x)$ et pour tout entier naturel n :*

1. $\mathrm{PA}^- \vdash \exists x\, A(x) \Leftrightarrow A(0) \vee \cdots \vee A(\overline{n-1}) \vee \exists x\, A(s^n(x))$
2. $\mathrm{PA}^- \vdash \forall x\, A(x) \Leftrightarrow A(0) \wedge \cdots \wedge A(\overline{n-1}) \wedge \forall x\, A(s^n(x))$

Lemme 6 — *Dans l'arithmétique de Robinson :*

1. $s(x) \leq s(y) \Leftrightarrow x \leq y$
2. $s(x) < s(y) \Leftrightarrow x < y$
3. $x \not< 0$
4. $x \leq 0 \Leftrightarrow x = 0$

Démonstration. Item 1 :

- $\text{PA}^- \vdash s(x) \leq s(y) \Leftrightarrow \exists z\ (s(x) + z = s(y))$
- $\text{PA}^- \vdash s(x) \leq s(y) \Leftrightarrow \exists z\ (s(x + z) = s(y))$ (PA2)
- $\text{PA}^- \vdash s(x) \leq s(y) \Leftrightarrow \exists z\ (x + z = y)$ (PA5)

L'item 2 est une conséquence immédiate de l'item 1. Item 3 :

- $\text{PA}^- \vdash x < 0 \Leftrightarrow \exists z\,(s(x) + z = 0)$
- $\text{PA}^- \vdash x < 0 \Leftrightarrow \exists z\,(s(x + z) = 0)$ (PA2)
- $\text{PA}^- \vdash x < 0 \Rightarrow \bot$ (PA6)

Item 4 :

- $\text{PA}^- \vdash x \leq 0 \Leftrightarrow \exists y\,(y = x \wedge y \leq 0)$
- $\text{PA}^- \vdash x \leq 0 \Leftrightarrow (x = 0 \wedge 0 \leq 0) \vee \exists y\,(x = s(y) \wedge s(y) \leq 0)$
- $\text{PA}^- \vdash x \leq 0 \Leftrightarrow x = 0 \vee \bot$ (item 3)
- $\text{PA}^- \vdash x \leq 0 \Leftrightarrow x = 0$ □

Proposition 7 (Déroulage de l'ordre) — *Pour tout entier naturel n :*

1. $\text{PA}^- \vdash x \leq \overline{n} \ \Leftrightarrow \bigvee_{m \leq n} x = \overline{m}$
2. $\text{PA}^- \vdash x < \overline{n} \ \Leftrightarrow \bigvee_{m < n} x = \overline{m}$

Démonstration. On commence par traiter l'ordre strict :

- $\text{PA}^- \vdash x < \overline{n} \Leftrightarrow \exists y\,(x = y \wedge y < \overline{n})$
- $\text{PA}^- \vdash x < \overline{n} \Leftrightarrow \bigvee_{m<n} (x = \overline{m} \wedge \overline{m} < \overline{n}) \ \vee \ \exists y\,(x = s^n(y) \wedge s^n(y) < \overline{n})$
- $\text{PA}^- \vdash x < \overline{n} \Leftrightarrow \bigvee_{m<n} x = \overline{m} \ \vee \ \exists y\,(x = s^n(y) \wedge s^n(y) < \overline{n})$
- $\text{PA}^- \vdash x < \overline{n} \Leftrightarrow \bigvee_{m<n} x = \overline{m} \ \vee \ \exists y\,(x = s^n(y) \wedge y < 0)$ (Lemme 6, 2.)
- $\text{PA}^- \vdash x < \overline{n} \Leftrightarrow \bigvee_{m<n} x = \overline{m}$ (Lemme 6, 3.)

Pour l'ordre large, on remarque que :

- $\text{PA}^- \vdash x \leq \overline{n} \Leftrightarrow x < \overline{n+1}$ (Lemme 6, 1.)
- $\text{PA}^- \vdash x \leq \overline{n} \Leftrightarrow \bigvee_{m \leq n} x = \overline{m}$ □

Sans le principe de récurrence, il n'est pas possible de démontrer dans l'arithmétique de Robinson que la relation d'ordre $x \leq y$ est transitive ou antisymétrique, et encore moins qu'elle est totale. Cependant, si on impose qu'un des termes soit un entier standard, alors ces propriétés deviennent démontrables dans PA^- [7] :

Proposition 8 — *Pour tout entier naturel n :*

1. $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \overline{n} \leq y$ *(transitivité restreinte)*
2. $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq \overline{n} \Rightarrow x = \overline{n}$ *(antisymétrie restreinte)*
3. $\text{PA}^- \vdash \overline{n} \leq x \vee x < \overline{n}$ *(totalité restreinte 1)*
4. $\text{PA}^- \vdash \overline{n} < x \vee x \leq \overline{n}$ *(totalité restreinte 2)*
5. $\text{PA}^- \vdash \overline{n} \leq x \Leftrightarrow x \nless \overline{n}$
6. $\text{PA}^- \vdash \overline{n} < x \Leftrightarrow x \nleq \overline{n}$

Démonstration. Item 1 :

- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \exists z\ \exists z'\,(\overline{n} + z = x \wedge x + z' = y)$

7. Ce résultat est très utilisé dans la démonstration du théorème de représentation.

- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \exists z\, \exists z' \, (s^n(z) = x \wedge x + z' = y)$
- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \exists z\, \exists z' \, (s^n(z) + z' = y)$
- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \exists z\, \exists z' \, (s^n(z + z') = y)$
- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \exists z\, \exists z' \, (\overline{n} + (z + z') = y)$
- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq y \Rightarrow \overline{n} \leq y$

Item 2 :

- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq \overline{n} \Rightarrow \overline{n} \leq x \wedge \bigvee_{m \leq n} x = \overline{m}$
- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq \overline{n} \Rightarrow \bigvee_{m \leq n} (x = \overline{m} \wedge \overline{n} \leq \overline{m})$
- $\text{PA}^- \vdash \overline{n} \leq x \wedge x \leq \overline{n} \Rightarrow x = \overline{n}$

Item 3 :

- $\text{PA}^- \vdash \exists y \, (x = y)$
- $\text{PA}^- \vdash \bigvee_{m < n} x = \overline{m} \ \vee \ \exists y \, (x = s^n(y))$
- $\text{PA}^- \vdash x < \overline{n} \ \vee \ \exists y \, (\overline{n} + y = x)$
- $\text{PA}^- \vdash x < \overline{n} \vee \overline{n} \leq x$

Les items 4, 5 et 6 se déduisent facilement des items 1, 2 et 3. □

2.4 Décision des formules closes à quantifications bornées

Proposition 9 (Déroulage des quantifications bornées) — *Pour toute formule $A(x)$ et pour tout entier naturel n on a :*

1. $\text{PA}^- \ \vdash \ \forall x < \overline{n} \, A(x) \ \Leftrightarrow \ \bigwedge_{m < n} A(\overline{m})$
2. $\text{PA}^- \ \vdash \ \exists x < \overline{n} \, A(x) \ \Leftrightarrow \ \bigvee_{m < n} A(\overline{m})$

Démonstration. Pour la quantification universelle :

- $\text{PA}^- \vdash \forall x < \overline{n} \, A(x) \Leftrightarrow \bigwedge_{m < n} (\overline{m} < \overline{m} \Rightarrow A(\overline{m})) \wedge \forall x \, (s^n(x) < \overline{n} \Rightarrow A(s^n(x)))$
- $\text{PA}^- \vdash \forall x < \overline{n} \, A(x) \Leftrightarrow \bigwedge_{m < n} A(\overline{m}) \wedge \forall x \, (x < 0 \Rightarrow A(s^n(x)))$
- $\text{PA}^- \vdash \forall x < \overline{n} \, A(x) \Leftrightarrow \bigwedge_{m < n} A(\overline{m})$

Le cas de la quantification existentielle se traite de la même manière. □

Proposition 10 (Décision des formules closes à quantifications bornées) — *Soit A une formule close à quantifications bornées :*

1. *Si* $\mathbb{N} \vDash A$, *alors* $\text{PA}^- \vdash A$.
2. *Si* $\mathbb{N} \nvDash A$, *alors* $\text{PA}^- \vdash \neg A$.

Démonstration. La propriété se démontre par récurrence sur le nombre de connecteurs et de quantificateurs de la formule A. On distingue les cas suivants :

- $A \equiv \bot$. Immédiat, car $\mathbb{N} \nvDash \bot$ et $\text{PA}^- \vdash \neg\bot$.
- $A \equiv t = u$. Ce cas a déjà été traité à la Prop. 4.
- $A \equiv t \leq u$. Ce cas a déjà été traité à la Prop. 5.
- $A \equiv B \Rightarrow C$. On distingue deux cas :
  - Soit $\mathbb{N} \vDash B \Rightarrow C$. On distingue deux sous-cas :
    - Ou bien $\mathbb{N} \nvDash B$. Dans ce cas on a $\text{PA}^- \vdash \neg B$ (par hypothèse d'induction), d'où $\text{PA}^- \vdash B \Rightarrow C$ en utilisant la tautologie $\neg B \Rightarrow (B \Rightarrow C)$.

  - Ou bien $\mathbb{N} \models C$. Dans ce cas on a $\text{PA}^- \vdash C$ (par hypothèse d'induction), d'où $\text{PA}^- \vdash B \Rightarrow C$ en utilisant la tautologie $C \Rightarrow (B \Rightarrow C)$.
  - Soit $\mathbb{N} \not\models B \Rightarrow C$. Dans ce cas, on a $\mathbb{N} \models B$ et $\mathbb{N} \not\models C$. Par hypothèse d'induction, on a $\text{PA}^- \vdash B$ et $\text{PA}^- \vdash \neg C$, d'où l'on tire que $\text{PA}^- \vdash \neg(B \Rightarrow C)$ en utilisant la tautologie $B \wedge \neg C \Rightarrow \neg(B \Rightarrow C)$.
- $A \equiv B \wedge C$ ou $A \equiv B \vee C$. Ces deux cas sont similaires au précédent.
- $A \equiv \forall x{<}t\ B(x)$. On distingue deux cas :
  - Soit $\mathbb{N} \models \forall x{<}t\ B(x)$. Dans ce cas on a $\mathbb{N} \models B(\overline{n})$ pour tout $n < \text{Val}(t)$. Par hypothèse d'induction, il vient
    - $\text{PA}^- \vdash \bigwedge_{n<\text{Val}(t)} B(\overline{n})$
    - $\text{PA}^- \vdash \forall x{<}t\ B(x)$ (Prop. 9 et Lemme 3)
  - Soit $\mathbb{N} \not\models \forall x{<}t\ B(x)$. Dans ce cas il existe $n < \text{Val}(t)$ tel que $\mathbb{N} \not\models B(\overline{n})$. Par hypothèse d'induction, il vient
    - $\text{PA}^- \vdash \neg B(\overline{n})$
    - $\text{PA}^- \vdash \neg \bigwedge_{m<\text{Val}(t)} B(\overline{m})$
    - $\text{PA}^- \vdash \neg\forall x{<}t\ B(x)$ (Prop. 9 et Lemme 3)
- $A \equiv \exists x{<}t\ B(x)$. Ce cas se traite de la même manière. □

3 Le théorème de représentation

Dans cette section, on se propose de démontrer que l'arithmétique de Robinson (PA^-) est suffisamment expressive pour représenter toutes les fonctions récursives.

3.1 Fonctions représentables

Définition 5 (Fonction représentable) — Soient $\mathscr{T}$ une théorie et f une fonction partielle de $\mathbb{N}^k$ dans $\mathbb{N}$ ($k \geq 1$). On dit que la fonction f est *représentable dans la théorie* $\mathscr{T}$ s'il existe une formule $R(x_1, \ldots, x_k, y)$ n'ayant pas d'autres variables libres que les variables $x_1, \ldots, x_k, y$, et telle que

$$\mathscr{T} \ \vdash\ R(\overline{n}_1, \ldots, \overline{n}_k, y) \Leftrightarrow y = \overline{f(n_1, \ldots, n_k)}$$

pour tout $(n_1, \ldots, n_k) \in \text{dom}(f)$. Une telle formule, lorsqu'elle existe, est alors appelée une *représentation* de la fonction f dans la théorie $\mathscr{T}$.

De manière équivalente, une formule $R(x_1, \ldots, x_k, y)$ représente une fonction partielle $f : \mathbb{N}^k \rightharpoonup \mathbb{N}$ dans la théorie $\mathscr{T}$ lorsque

(i) $\mathscr{T} \ \vdash\ R(\overline{n}_1, \ldots, \overline{n}_k, \overline{f(n_1, \ldots, n_k)})$

(ii) $\mathscr{T} \ \vdash\ R(\overline{n}_1, \ldots, \overline{n}_k, y) \Rightarrow y = \overline{f(n_1, \ldots, n_k)}$

pour tout $(n_1, \ldots, n_k) \in \text{dom}(f)$, ce qui entraîne en particulier que :

(iii) $\mathscr{T} \ \vdash\ \exists! y\ R(\overline{n}_1, \ldots, \overline{n}_k, y)$

On notera que cette définition ne nous dit rien sur les propriétés de la relation $R(\overline{n}_1, \ldots, \overline{n}_k, y)$ dans le cas où la fonction f n'est pas définie au point $(n_1, \ldots, n_k)$. En dehors du domaine de f, la relation $R(\overline{n}_1, \ldots, \overline{n}_k, y)$ peut donc être définie de manière arbitraire. Cependant, on utilisera le plus souvent cette définition dans le cas où la fonction f est totale (i.e. $\text{dom}(f) = \mathbb{N}^k$), le cas général n'étant vraiment utile que dans la démonstration du théorème de représentation.

Si $f : \mathbb{N}^k \to \mathbb{N}$ est une fonction totale représentée dans la théorie $\mathscr{T}$ par une relation $R(x_1, \ldots, x_k, y)$, on dit que la relation $R(x_1, \ldots, x_k, y)$ est :

– *prouvablement totale* (dans $\mathscr{T}$) lorsque

$$\mathscr{T} \vdash \exists y\, R(x_1, \ldots, x_k, y)\,;$$

– *prouvablement fonctionnelle* (dans $\mathscr{T}$) lorsque

$$\mathscr{T} \vdash \forall y\, \forall y' \left(R(x_1, \ldots, x_k, y) \wedge R(x_1, \ldots, x_k, y') \Rightarrow y = y'\right).$$

Les conditions (i) et (ii) n'entraînent pas (en général) que la relation $R(x_1, \ldots, x_k, y)$ est prouvablement totale ni même qu'elle est prouvablement fonctionnelle dans $\mathscr{T}$. Lorsque ces deux propriétés supplémentaires sont réunies, on dit que la fonction f est *fortement représentable* dans la théorie $\mathscr{T}$.

Enfin, on notera que toute fonction (partielle ou totale) qui est représentable dans une théorie $\mathscr{T}$ reste représentable dans toute extension de la théorie $\mathscr{T}$. (En particulier, toutes les fonctions sont représentables dans une théorie incohérente.)

3.2 Ensembles représentables

Définition 6 (Ensemble représentable) — Soient $\mathscr{T}$ une théorie et $E \subseteq \mathbb{N}^k$ un sous-ensemble de $\mathbb{N}^k$. On dit que le sous-ensemble E est *représentable dans la théorie* $\mathscr{T}$ s'il existe une formule $R(x_1, \ldots, x_k)$ n'ayant pas d'autres variables libres que les variables $x_1, \ldots, x_k$, et telle que pour tout $(n_1, \ldots, n_k) \in \mathbb{N}^k$ on ait :

$$\begin{array}{lll} (i) & \mathscr{T} \vdash R(\overline{n}_1, \ldots, \overline{n}_k) & \text{si } (n_1, \ldots, n_k) \in E \\ (ii) & \mathscr{T} \vdash \neg R(\overline{n}_1, \ldots, \overline{n}_k) & \text{si } (n_1, \ldots, n_k) \notin E \end{array}$$

Une telle formule $R(x_1, \ldots, x_k)$, lorsqu'elle existe, est alors appelée une *représentation* de l'ensemble E dans la théorie $\mathscr{T}$.

On vérifie aisément que :

Lemme 7 — *Soient $\mathscr{T}$ une théorie et $E \subseteq \mathbb{N}^k$ un sous-ensemble de $\mathbb{N}^k$.*

1. *Si $R(x_1, \ldots, x_k, y)$ est une représentation dans la théorie $\mathscr{T}$ de la fonction caractéristique $\mathbf{1}_E : \mathbb{N}^k \to \mathbb{N}$ de l'ensemble E (au sens de la Déf. 5), alors la relation*

$$R'(x_1, \ldots, x_k) \equiv R(x_1, \ldots, x_k, 0)$$

   est une représentation de l'ensemble E dans $\mathscr{T}$ (au sens de la Déf. 6).

2. *Réciproquement, si $R'(x_1, \ldots, x_k)$ est une représentation de l'ensemble E dans la théorie $\mathscr{T}$ (au sens de la Déf. 6), alors la relation*

$$R(x_1, \ldots, x_k, y) \equiv (y = 1 \wedge R'(x_1, \ldots, x_k)) \vee (y = 0 \wedge \neg R'(x_1, \ldots, x_k))$$

   est une représentation de la fonction caractéristique de E dans $\mathscr{T}$ (Déf. 5), et même une représentation forte de cette fonction.

(La démonstration est laissée en exercice au lecteur.)
Par conséquent :

Proposition 11 — *Soient $\mathscr{T}$ une théorie et $E \subseteq \mathbb{N}^k$ un sous-ensemble. Les trois assertions suivantes sont équivalentes :*

1. *L'ensemble E est représentable dans $\mathscr{T}$;*

2. *La fonction caractéristique* $\mathbf{1}_E : \mathbb{N}^k \to \mathbb{N}$ *est représentable dans* $\mathscr{T}$ *;*
3. *La fonction caractéristique* $\mathbf{1}_E : \mathbb{N}^k \to \mathbb{N}$ *est fortement représentable dans* $\mathscr{T}$*.*

La suite de la section 3 est consacrée à la démonstration du *théorème de représentation*, qui énonce que toutes les fonctions récursives (partielles et totales) sont représentables dans l'arithmétique de Robinson (PA^-).

3.3 Les fonctions de base

Les fonctions récursives sont construites à partir des fonctions de base suivantes, qui sont toutes des fonctions totales :

Les fonctions nulles $\mathsf{zero}_k : \mathbb{N}^k \to \mathbb{N}$ (pour $k \geq 1$), définies par

$$\mathsf{zero}_k(n_1, \ldots, n_k) = 0 \qquad ((n_1, \ldots, n_k) \in \mathbb{N}^k)$$

La fonction successeur $\mathsf{succ} : \mathbb{N} \to \mathbb{N}$, définie par

$$\mathsf{succ}(n) = n + 1 \qquad (n \in \mathbb{N})$$

Les fonctions de projection $\mathsf{proj}_{k,i} : \mathbb{N}_k \to \mathbb{N}$ (pour $k \geq i \geq 1$), définies par

$$\mathsf{proj}_{k,i}(n_1, \ldots, n_k) = n_i \qquad ((n_1, \ldots, n_k) \in \mathbb{N}^k)$$

Il est immédiat que :

Proposition 12 (Représentation des fonctions de base) — *Les fonctions de base* zero_k*,* succ *et* $\mathsf{proj}_{k,i}$ *(*$k \geq i \geq 1$*) sont fortement représentées dans l'arithmétique de Robinson par les relations :*

$$\begin{array}{rcl} R_{\mathsf{zero}_k}(x_1, \ldots, x_k, y) & \equiv & y = 0 \\ R_{\mathsf{succ}}(x, y) & \equiv & y = s(x) \\ R_{\mathsf{proj}_{k,i}}(x_1, \ldots, x_k, y) & \equiv & y = x_i \end{array}$$

Il s'agit à présent de démontrer que la classe des fonctions partielles représentables dans PA^- est close par les schémas de composition, de récursion et de minimisation.

3.4 Le schéma de composition

Soient $f_1, \ldots, f_p : \mathbb{N}^k \rightharpoonup \mathbb{N}$ et $g : \mathbb{N}^p \rightharpoonup \mathbb{N}$ des fonctions partielles. La composée des fonctions $f_1, \ldots, f_p$ par la fonction g est la fonction $h : \mathbb{N}^k \rightharpoonup \mathbb{N}$ définie par

$$h(n_1, \ldots, n_k) = g(f_1(n_1, \ldots, n_k), \ldots, f_p(n_1, \ldots, n_k))$$

pour tout $(n_1, \ldots, n_k) \in \mathbb{N}^k$ où le membre droit de l'égalité ci-dessus est défini.

Par construction, la fonction h n'est pas définie au point $(n_1, \ldots, n_k) \in \mathbb{N}^k$ dans les cas suivants :

1. si $f_i(n_1, \ldots, n_k)$ n'est pas défini pour au moins un entier $i \in [1..p]$;
2. si $f_i(n_1, \ldots, n_k)$ est défini pour tout entier $i \in [1..p]$, mais $g(f_1(n_1, \ldots, n_k), \ldots, f_p(n_1, \ldots, n_k))$ n'est pas défini.

Là encore, on vérifie sans difficulté que :

Proposition 13 (Représentation du schéma de composition) — *Si les fonctions f_i ($i \in [1..p]$) et g sont représentées par des relations $R_{f_i}(x_1, \ldots, x_k, y)$ ($i \in [1..p]$) et $R_g(x_1, \ldots, x_p, y)$ dans l'arithmétique de Robinson, alors la fonction h est elle-même représentée dans cette même théorie par la relation $R_h(x_1, \ldots, x_k, y)$ définie par*

$$R_h(x_1, \ldots, x_k, y) \equiv \exists z_1 \cdots \exists z_p \left(R_{f_1}(x_1, \ldots, x_k, z_1) \wedge \vdots \wedge R_{f_p}(x_1, \ldots, x_k, z_p) \wedge R_g(z_1, \ldots, z_p, y) \right).$$

(La démonstration est laissée en exercice au lecteur.)

Contrairement à ce qu'on pourrait penser, les fonctions définies par le schéma de minimisation sont bien plus faciles à représenter que celles qui sont définies par le schéma de récursion, et c'est donc par elles que nous allons commencer.

3.5 Le schéma de minimisation

Soit $f : \mathbb{N}^{k+1} \rightharpoonup \mathbb{N}$ une fonction partielle. La minimisée de la fonction f (par rapport à son premier argument) est la fonction partielle $h : \mathbb{N}^k \rightharpoonup \mathbb{N}$ définie par

$$h(m_1, \ldots, m_k) = \text{l'entier } n \text{ défini par : } f(n, m_1, \ldots, m_k) = 0 \text{ et } f(n', m_1, \ldots, m_k) \neq 0 \text{ pour tout } n' < n$$

pour tout $(m_1, \ldots, m_k) \in \mathbb{N}^k$ pour lequel l'entier n défini ci-dessus existe.

Par construction, la fonction h n'est pas définie au point $(m_1, \ldots, m_k) \in \mathbb{N}^k$ dans les cas suivants :

1. s'il existe $n \in \mathbb{N}$ pour lequel $f(n, m_1, \ldots, m_k)$ n'est pas défini, et tel que pour tout $n' < n$, $f(n', m_1, \ldots, m_k)$ est défini et $f(n', m_1, \ldots, m_k) \neq 0$;
2. si pour tout $n \in \mathbb{N}$, $f(n, m_1, \ldots, m_k)$ est défini et $f(n, m_1, \ldots, m_k) \neq 0$.

Proposition 14 (Représentation du schéma de minimisation) — *Si la fonction f est représentée dans l'arithmétique de Robinson par une relation $R_f(x, x_1, \ldots, x_k, y)$, alors la fonction h est elle aussi représentée dans cette même théorie par la relation $R_h(x_1, \ldots, x_k, y)$ définie par*

$$R_h(x_1, \ldots, x_k, y) \equiv R_f(y, x_1, \ldots, x_k, 0) \wedge \forall y' < y \, \neg R_f(y', x_1, \ldots, x_k, 0).$$

Démonstration. Soient $(m_1, \ldots, m_k) \in \mathrm{dom}(h)$ et $r = h(m_1, \ldots, m_k)$. On a donc $f(r, m_1, \ldots, m_k) = 0$ et $f(r', m_1, \ldots, m_k) \neq 0$ pour tout $r' < r$. On considère les abréviations $A_f(x, y) \equiv R_f(x, \overline{m}_1, \ldots, \overline{m}_k, y)$ et $A_h(y) \equiv R_h(\overline{m}_1, \ldots, \overline{m}_k, y) \equiv A_f(y, 0) \wedge \forall y < y' \, \neg A_f(y', 0)$. On vérifie que :

1. $\mathrm{PA}^- \vdash A_f(\overline{r}, 0)$ (par représentation)
2. $\mathrm{PA}^- \vdash \neg A_f(\overline{r'}, 0)$ pour tout $r' < r$ (par représentation)
3. $\mathrm{PA}^- \vdash A_f(\overline{r}, 0) \wedge \bigwedge_{r' < r} \neg A_f(\overline{r'}, 0)$ (1. et 2.)
4. $\mathrm{PA}^- \vdash A_h(\overline{r})$ (3. et Prop. 9)
5. $\mathrm{PA}^- \vdash y < \overline{r} \Rightarrow \neg A_f(y, 0)$ (4.)
6. $\mathrm{PA}^- \vdash A_f(y, 0) \Rightarrow y \not< \overline{r}$ (5.)
7. $\mathrm{PA}^- \vdash A_f(y, 0) \Rightarrow \overline{r} \leq y$ (6. et Prop. 8)

8. $\mathrm{PA}^- \vdash A_h(y) \Rightarrow (\overline{r} < y \Rightarrow \neg A_f(\overline{r}, 0))$ (Déf. de A_h)
9. $\mathrm{PA}^- \vdash A_h(y) \Rightarrow \overline{r} \not< y$ (1. et 8.)
10. $\mathrm{PA}^- \vdash A_h(y) \Rightarrow y \leq \overline{r}$ (9. et Prop. 8)
11. $\mathrm{PA}^- \vdash A_h(y) \Rightarrow y = \overline{r}$ (7., 10. et Prop.8) □

3.6 Restes Chinois et fonction β de Gödel

La représentation d'une fonction f définie par récursion est délicate, car pour décrire la valeur retournée par f au point n (à l'aide d'une formule non récursive !), il est nécessaire d'introduire la suite (finie) des valeurs retournées par les appels récursifs successifs de f jusqu'à l'entier n. Mais pour pouvoir parler de cette suite dans le langage de l'arithmétique, il nous faut d'abord un codage des suites finies d'entiers dans les entiers qui soit facilement exprimable dans le langage.

L'astuce de Gödel consiste à utiliser le lemme des restes Chinois pour représenter chaque suite finie $r_1, \ldots, r_k$ par un entier naturel n tel que

$$r_1 = n \bmod p_1, \quad r_2 = n \bmod p_2, \quad \cdots, \quad r_k = n \bmod p_k\,,$$

où $n \bmod p$ désigne le reste de la division euclidienne de n par p, et où $p_1, \ldots, p_k$ est une suite finie judicieusement choisie. (Nous verrons un peu plus loin comment.)

Le lemme des restes Chinois s'énonce ainsi :

Lemme 8 (Restes Chinois) — *Soient $p_1, \ldots, p_k$ ($k \geq 1$) des entiers strictement positifs deux à deux premiers entre eux. Alors pour toute suite de k entiers $r_1, \ldots, r_k$, il existe un entier naturel n tel que $n \equiv r_i \pmod{p_i}$ pour tout $i \in [1..k]$.*

Démonstration. Il suffit de démontrer l'existence d'un entier relatif $n \in \mathbb{Z}$ tel que $n \equiv r_i \pmod{p_i}$ pour tout $i \in [1..k]$; pour obtenir un entier naturel ayant la même propriété, on prendra $n + qp_1 \cdots p_k$ avec q suffisamment grand.

On traite d'abord le cas binaire ($k = 2$). Pour cela, on suppose $p_1, p_2 > 0$ premiers entre eux, et on considère deux entiers $x_1, x_2 \in \mathbb{Z}$ tels que $x_1 p_1 + x_2 p_2 = 1$ (lemme de Bézout). On vérifie alors que pour tous $r_1, r_2 \in \mathbb{Z}$, l'entier $n = r_2 x_1 p_1 + r_1 x_2 p_2$ est tel que $n \equiv r_1 \pmod{p_1}$ et $n \equiv r_2 \pmod{p_2}$.

Le lemme se démontre ensuite par récurrence sur $k \geq 1$.

- $k = 1$. Immédiat.
- $k \geq 2$. On remarque que si les entiers $p_1, \ldots, p_k$ sont deux à deux premiers entre eux, alors les deux entiers $p_1 \cdots p_{k-1}$ et p_k sont premiers entre eux. On construit d'abord un entier m tel que $m \equiv r_i \pmod{p_i}$ pour tout $i \in [1..k-1]$ (hypothèse de récurrence), puis un entier n tel que $n \equiv m \pmod{p_1 \cdots p_{k-1}}$ et $n \equiv r_k \pmod{p_k}$ en utilisant le cas binaire. □

En cours d'Algèbre, on énonce plutôt ce résultat de la manière suivante :

Corollaire 1 (Restes Chinois, forme algébrique) — *Pour toute suite finie $p_1, \ldots, p_k$ ($k \geq 1$) d'entiers non nuls deux à deux premiers entre eux, on a l'isomorphisme : $\mathbb{Z}/p_1 \cdots p_k\mathbb{Z} \cong (\mathbb{Z}/p_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k\mathbb{Z})$.*

Démonstration. On considère l'homomorphisme de groupes

$$f : \mathbb{Z}/p_1 \cdots p_k\mathbb{Z} \to (\mathbb{Z}/p_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k\mathbb{Z})$$

défini par $f(n \bmod p_1 \cdots p_k) = (n \bmod p_1, \ldots, n \bmod p_k)$ pour tout $n \in \mathbb{N}$. D'après le lemme 8, cet homomorphisme est surjectif, et comme les ensembles de départ et d'arrivée ont le même cardinal, il s'agit d'un isomorphisme. □

Étant donnée une suite finie d'entiers naturels $r_1, \ldots, r_k$, il s'agit à présent de construire des entiers strictement positifs $p_1, \ldots, p_k$ deux à deux premiers entre eux tels que $r_i < p_i$ pour tout $i \in [1..k]$. Pour cela on s'appuie sur le lemme suivant :

Lemme 9 — *Pour tout entier naturel k et pour tout entier strictement positif p qui est divisible par tous les entiers de 1 à k, les $k+1$ nombres*

$$p+1, \quad 2p+1, \quad \ldots, \quad (k+1)p+1$$

sont deux à deux premiers entre eux.

Démonstration. Sous les hypothèses du lemme, on considère deux indices i, j tels que $1 \leq i < j \leq k+1$ et on suppose qu'il existe un nombre premier q tel que $q|ip+1$ et $q|jp+1$. Par soustraction, il vient $q|(j-i)p$. On distingue deux cas :

- $q|p$. Dans ce cas on a $q|(ip+1-ip)$ donc $q|1$, ce qui est impossible.
- $q|(j-i)$. Dans ce cas on a $q|(j-i)|p$, car p est divisible par tous les entiers de 1 à k, et $(j-i) \in [1..k]$. Ce qui nous ramène au cas précédent.

Dans les deux cas, l'hypothèse selon laquelle $q|ip+1$ et $q|jp+1$ est absurde, ce qui montre que les deux nombres $ip+1$ et $jp+1$ sont premiers entre eux. □

La fonction β de Gödel On appelle la *fonction β de Gödel* la fonction récursive primitive $\beta : \mathbb{N}^3 \to \mathbb{N}$ définie pour tout $(n,p,i) \in \mathbb{N}^3$ par

$$\beta(n,p,i) \;=\; n \bmod ((i+1)p+1)$$

où $n \bmod q$ désigne le reste de la division euclidienne de n par q (avec $q > 0$).

Les lemmes 8 (restes Chinois) et 9 entraînent que

Lemme 10 — *Pour tout entier $k \geq 0$ et pour toute suite finie $r_0, \ldots, r_k \in \mathbb{N}$, il existe deux entiers n et p tels que $\beta(n,p,i) = r_i$ pour tout $i \in [0..k]$.*

Démonstration. On prend un entier $p > 0$ tel que $i|p$ pour tout entier $i \in [1..k]$ et tel que $r_i \leq p$ pour tout $i \in [0..k]$. D'après le Lemme 9, les entiers naturels de la forme $(i+1)p+1$ (pour $i \in [0..k]$) sont deux à deux premiers entre eux. Il suffit alors de prendre un entier naturel n tel que $n \bmod ((i+1)p+1) = r_i$ pour tout $i \in [0..k]$, sachant qu'un tel entier existe d'après le lemme des restes Chinois (Lemme 8). □

On vérifie sans difficulté que :

Proposition 15 (Représentation de la fonction β) — *La fonction β de Gödel est représentée par la relation $R_\beta(u,v,x,y)$ définie par :*

$$R_\beta(u,v,x,y) \;\equiv\; \exists z\, \big(u = s(s(x) \times v) \times z + y \wedge y < s(s(x) \times v)\big)\,.$$

Dans ce qui suit, on aura besoin d'utiliser une autre représentation de la fonction β qui a de meilleures propriétés, à savoir la relation :

$$S_\beta(u,v,x,y) \;\equiv\; R_\beta(u,v,x,y) \wedge \forall y' < y\, \neg R_\beta(u,v,x,y')\,.$$

On vérifie que :

Proposition 16 — *La fonction β est représentée par la relation $S_\beta(u,v,x,y)$, et pour tout entier naturel r on a :* $\mathrm{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow y = \overline{r}$.

Démonstration. Supposons que $r = \beta(n,p,i)$. On a alors

- $\text{PA}^- \vdash R_\beta(\overline{n},\overline{p},\overline{i},\overline{r})$ (par représentation)
- $\text{PA}^- \vdash \neg R_\beta(\overline{n},\overline{p},\overline{i},\overline{r}')$ pour tout $r' < r$ (idem)
- $\text{PA}^- \vdash R_\beta(\overline{n},\overline{p},\overline{i},\overline{r}) \wedge \forall y' < \overline{r}\ \neg R_\beta(\overline{n},\overline{p},\overline{i},y')$ (Prop. 9)
- $\text{PA}^- \vdash S_\beta(\overline{n},\overline{p},\overline{i},\overline{r})$

Par ailleurs :

- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow R_\beta(u,v,x,\overline{r})$
- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow (\overline{r} < y \Rightarrow \neg R_\beta(u,v,x,\overline{r}))$
- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow y \leq \overline{r}$
- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow R_\beta(u,v,x,y)$
- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow (y < \overline{r} \Rightarrow \neg R_\beta(u,v,x,y))$
- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow \overline{r} \leq y$
- $\text{PA}^- \vdash S_\beta(u,v,x,\overline{r}) \wedge S_\beta(u,v,x,y) \Rightarrow y = \overline{r}$ □

3.7 Le schéma de récursion

Soient des fonctions partielles $f : \mathbb{N}^k \rightharpoonup \mathbb{N}$ et $g : \mathbb{N}^{k+2} \rightharpoonup \mathbb{N}$. On appelle la *fonction définie par récursion à partir de f et g* la fonction $h : \mathbb{N}^{k+1} \rightharpoonup \mathbb{N}$ définie par

$$\begin{aligned} h(0,m_1,\ldots,m_k) &= f(m_1,\ldots,m_k) \\ h(n+1,m_1,\ldots,m_k) &= g(n,h(n,m_1,\ldots,m_k),m_1,\ldots,m_k) \end{aligned}$$

pour tout $(n,m_1,\ldots,m_k) \in \mathbb{N}^{k+1}$ pour lequel cette définition a un sens.

Par construction, la fonction h n'est pas définie au point $(n,m_1,\ldots,m_k) \in \mathbb{N}^{k+1}$ dans les cas suivants :

1. si $n = 0$ et $f(m_1,\ldots,m_k)$ n'est pas défini ;
2. si $n > 0$ et $h(n-1,m_1,\ldots,m_k)$ n'est pas défini ;
3. si $n > 0$ et $h(n-1,m_1,\ldots,m_k)$ est défini, mais $g(n-1,h(n-1,m_1,\ldots,m_k),m_1,\ldots,m_k)$ n'est pas défini.

Proposition 17 (Représentation du schéma de récursion) — *Si les fonctions f et g sont représentées dans l'arithmétique de Robinson par des relations $R_f(x_1,\ldots,x_k,y)$ et $R_g(x,z,x_1,\ldots,x_k,y)$, alors la fonction h est elle-même représentée dans cette théorie par la relation $R_h(x,x_1,\ldots,x_k,y)$ définie par*

$$\begin{aligned} R_h(x,x_1,\ldots,x_k,y) \ \equiv\ & \\ & \exists u\ \exists v\ \Big(S_\beta(u,v,x,y) \qquad\qquad\qquad\qquad \wedge \\ & \qquad \exists z\ \big(R_f(x_1,\ldots,x_k,z) \wedge S_\beta(u,v,0,z) \big) \qquad \wedge \\ & \qquad \forall x' < x\ \exists z\ \exists z'\ \big(R_g(x',z,x_1,\ldots,x_k,z') \wedge \\ & \qquad\qquad S_\beta(u,v,x',z) \wedge S_\beta(u,v,s(x'),z') \big) \Big). \end{aligned}$$

La démonstration (fort technique) est laissée en exercice au lecteur. On notera que le choix de la représentation $S_\beta(u,v,x,y)$ (Prop. 16) joue un rôle essentiel.

En combinant les Prop. 12, 13, 14 et 17, on obtient ainsi le :

Théorème 1 (Représentation) — *Toute fonction récursive partielle est représentable dans l'arithmétique de Robinson (au sens de la Déf. 5).*

4 Numérisation de la syntaxe : les codages

Nous allons maintenant définir un codage des expressions dans les entiers, en associant à chaque expression formelle e (une variable, un terme, une formule, un contexte ou une démonstration) un entier naturel noté $\ulcorner e \urcorner$. On notera que ce codage s'applique à toutes les expressions du langage, y compris aux expressions ouvertes (i.e. dépendant de certaines variables libres), bien que les codes qu'on associera à de telles expressions — des entiers naturels — soient par nature des objets clos.

4.1 Codage des variables

Les expressions du langage de l'arithmétique sont construites à partir d'un jeu de variables qui est dénombrable, ce qui suppose que l'ensemble des variables vient avec une bijection sur l'ensemble des entiers naturels.

Dans ce qui suit, on note $x \mapsto \tilde{x}$ cette bijection qui à chaque variable x associe un entier naturel $\tilde{x}$, qu'on appelle le *numéro* de la variable x.

4.2 Codage des couples

On considère la bijection $(n, m) \mapsto \langle n, m \rangle$ (de $\mathbb{N} \times \mathbb{N}$ sur $\mathbb{N}$) qui à chaque couple d'entiers naturels n et m associe l'entier naturel $\langle n, m \rangle$ défini par

$$\langle n, m \rangle = (n + m)(n + m + 1)/2 + n\,.$$

Il s'agit là de l'énumération diagonale bien connue des couples d'entiers :

4	14	⋯			
3	9	13	⋯		
2	5	8	12	⋯	
1	2	4	7	11	⋯
0	0	1	3	6	10
n/m	0	1	2	3	4

La fonction $(n, m) \mapsto \langle n, m \rangle$ est une fonction récursive primitive, de même que les projections correspondantes $\mathsf{fst} : \mathbb{N}^2 \to \mathbb{N}$ et $\mathsf{snd} : \mathbb{N}^2 \to \mathbb{N}$, définies par

$$\mathsf{fst}(\langle n, m \rangle) = n \qquad \text{et} \qquad \mathsf{snd}(\langle n, m \rangle) = m\,.$$

Dans ce qui suit, on associe la notation $\langle n, m \rangle$ à droite pour définir les triplets, les quadruplets, etc. en posant :

$$\langle n_1, n_2, n_3 \rangle = \langle n_1, \langle n_2, n_3 \rangle \rangle, \quad \langle n_1, n_2, n_3, n_4 \rangle = \langle n_1, \langle n_2, \langle n_3, n_4 \rangle \rangle \rangle, \quad \text{etc.}$$

4.3 Codage des termes

À chaque terme t du langage de l'arithmétique on associe un entier naturel noté $\ulcorner t \urcorner$ et défini par récurrence sur la structure de t par les équations :

$$\begin{aligned}
\ulcorner 0 \urcorner &= \langle 0, 0 \rangle \\
\ulcorner x \urcorner &= \langle 1, \tilde{x} \rangle \\
\ulcorner s(t) \urcorner &= \langle 2, \ulcorner t \urcorner \rangle \\
\ulcorner t + u \urcorner &= \langle 3, \ulcorner t \urcorner, \ulcorner u \urcorner \rangle \\
\ulcorner t \times u \urcorner &= \langle 4, \ulcorner t \urcorner, \ulcorner u \urcorner \rangle
\end{aligned}$$

On vérifie aisément que :

1. La fonction $t \mapsto \ulcorner t \urcorner$ est injective.
2. La fonction $\mathsf{term} : \mathbb{N} \to \mathbb{N}$ définie pour tout $n \in \mathbb{N}$ par

$$\mathsf{term}(n) \;=\; \begin{cases} 1 & \text{si } n = \ulcorner t \urcorner \text{ pour un certain terme } t \\ 0 & \text{sinon} \end{cases}$$

est une fonction récursive primitive.

On vérifie plus généralement que les fonctions usuelles de manipulation des termes (test d'occurrence d'une variable, substitution, etc.) correspondent à travers ce codage à des fonctions récursives primitives. Dans ce qui suit, on aura besoin notamment de la fonction $\mathsf{num} : \mathbb{N} \to \mathbb{N}$ qui à chaque $n \in \mathbb{N}$ associe le code du terme $\overline{n} = s^n 0$, et qui est définie par récursion primitive à l'aide des équations

$$\begin{aligned} \mathsf{num}(0) &= \langle 0, 0 \rangle \\ \mathsf{num}(n+1) &= \langle 2, \mathsf{num}(n) \rangle \end{aligned}$$

4.4 Codage des formules

À chaque formule A du langage de l'arithmétique on associe un entier naturel noté $\ulcorner A \urcorner$ et défini par récurrence sur la structure de A par les équations :

$$\begin{aligned} \ulcorner \bot \urcorner &= \langle 0, 0 \rangle \\ \ulcorner t = u \urcorner &= \langle 1, \ulcorner t \urcorner, \ulcorner u \urcorner \rangle \\ \ulcorner A \Rightarrow B \urcorner &= \langle 2, \ulcorner A \urcorner, \ulcorner B \urcorner \rangle \\ \ulcorner A \wedge B \urcorner &= \langle 3, \ulcorner A \urcorner, \ulcorner B \urcorner \rangle \\ \ulcorner A \vee B \urcorner &= \langle 4, \ulcorner A \urcorner, \ulcorner B \urcorner \rangle \\ \ulcorner \forall x\, A \urcorner &= \langle 5, \tilde{x}, \ulcorner A \urcorner \rangle \\ \ulcorner \exists x\, A \urcorner &= \langle 6, \tilde{x}, \ulcorner A \urcorner \rangle \end{aligned}$$

On notera qu'avec cette définition, l'ensemble des codes de formules n'est pas disjoint de l'ensemble des codes de termes (par exemple, la formule $\bot$ a le même code que le terme 0). Ce recouvrement ne posera pas de problème en pratique, car les termes et les formules sont utilisés dans des contextes différents.

Là encore :

1. La fonction $A \mapsto \ulcorner A \urcorner$ est injective.
2. La fonction $\mathsf{form} : \mathbb{N} \to \mathbb{N}$ définie par

$$\mathsf{form}(n) \;=\; \begin{cases} 1 & \text{si } n = \ulcorner A \urcorner \text{ pour une certaine formule } A \\ 0 & \text{sinon} \end{cases}$$

est une fonction récursive primitive.

Les fonctions usuelles de manipulation des formules (test d'occurrence libre/liée d'une variable, substitution, test d'α-conversion, etc.) correspondent elles-aussi à des fonctions récursives primitives à travers ce codage. On utilisera notamment la fonction récursive primitive $\mathsf{subst} : \mathbb{N}^3 \to \mathbb{N}$ définie par

$$\mathsf{subst}(\ulcorner A \urcorner, \tilde{x}, \ulcorner t \urcorner) \;=\; \ulcorner A\{x := t\} \urcorner$$

(et complétée par l'équation $\mathsf{subst}(n, v, m) = 0$ dans le cas où n n'est pas un code de formule ou dans le cas où m n'est pas un code de terme), de même que la fonction $\mathsf{substn} : \mathbb{N}^3 \to \mathbb{N}$ définie par

$$\mathsf{substn}(n, v, m) \;=\; \mathsf{subst}(n, v, \mathsf{num}(n))$$

pour tous $(n, v, m) \in \mathbb{N}^3$, et qui satisfait par construction l'équation

$$\mathsf{substn}(\ulcorner A \urcorner, \tilde{x}, n) \;=\; \ulcorner A\{x := \overline{n}\} \urcorner \,.$$

4.5 Codage des contextes

À chaque contexte Γ on associe un entier naturel noté $\ulcorner \Gamma \urcorner$ et défini par récurrence sur la longueur de Γ par :

$$\begin{array}{rcl} \ulcorner \varnothing \urcorner & = & 0 \\ \ulcorner \Gamma, A \urcorner & = & 1 + \langle \ulcorner \Gamma \urcorner, \ulcorner A \urcorner \rangle \end{array}$$

Là encore, il est clair que les fonctions usuelles de manipulation des contextes (calcul de la longueur, test d'occurrence d'une variable libre, substitution, etc.) sont représentées par des fonctions récursives primitives.

4.6 Codage des démonstrations

Il s'agit à présent de définir le codage $D \mapsto \ulcorner D \urcorner$ des démonstrations.

Pour cela, on commence par définir un premier codage $D \mapsto D^*$ des arbres de dérivation dans lequel chaque arbre de la forme

$$D = \left\{ \begin{array}{c} \begin{array}{ccc} \vdots & & \vdots \\ D_1 & & D_n \\ \vdots & & \vdots \\ S_1 & \cdots & S_n \end{array} \\ \hline \Gamma \vdash A \end{array} \right.$$

est représenté par un entier de la forme

$$D^* \;=\; \langle k, \ulcorner A \urcorner, D_1^*, \ldots, D_n^* \rangle \,,$$

où k désigne le numéro de la règle appliquée (en considérant les règles suivant leur ordre d'apparition dans la Fig. 1). On notera qu'un tel codage ne prend pas en compte les contextes, car ceux-ci peuvent être reconstitués dans toute la dérivation à partir du contexte situé à la racine de la dérivation. (Autrement dit, l'entier D^* est un codage de la dérivation D au sens de la déduction naturelle avec contextes implicites.) Formellement, la fonction $D \mapsto D^*$ est définie par les équations données dans la Fig. 2.

Le code d'une dérivation D est alors défini par $\ulcorner D \urcorner = \langle \ulcorner \Gamma \urcorner, D^* \rangle$, où Γ est le contexte figurant dans la conclusion de la dérivation D. Comme précédemment, on vérifie que la fonction $\mathsf{deriv} : \mathbb{N} \to \mathbb{N}$ définie par

$$\mathsf{deriv}(n) \;=\; \begin{cases} 1 & \text{si } n = \ulcorner D \urcorner \text{ pour une certaine dérivation } D \\ 0 & \text{sinon} \end{cases}$$

est une fonction récursive primitive.

4.7 Théories récursives

Définition 7 (Théories récursives) — On dit qu'une théorie $\mathscr{T}$ (définie sur le langage de l'arithmétique) est *récursive* si l'ensemble

$$\mathbf{Ax}_{\mathscr{T}} = \{ \ulcorner A \urcorner \;:\; A \text{ axiome de } \mathscr{T} \}$$

constitué des codes des axiomes de $\mathscr{T}$ est un ensemble récursif.

$$\left(\dfrac{}{\Gamma \vdash A}\ {\scriptstyle (A \in \Gamma)} \right)^* = \langle 0, \ulcorner A \urcorner, 0 \rangle \qquad \left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma, \neg A \vdash \bot\end{array}}{\Gamma \vdash A} \right)^* = \langle 1, \ulcorner A \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma, A \vdash B\end{array}}{\Gamma \vdash A \Rightarrow B} \right)^* = \langle 2, \ulcorner A \Rightarrow B \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{cc}\vdots\ D_1 & \vdots\ D_2 \\ \Gamma \vdash A \Rightarrow B & \Gamma \vdash A\end{array}}{\Gamma \vdash B} \right)^* = \langle 3, \ulcorner B \urcorner, D_1^*, D_2^* \rangle$$

$$\left(\dfrac{\begin{array}{cc}\vdots\ D_1 & \vdots\ D_2 \\ \Gamma \vdash A & \Gamma \vdash B\end{array}}{\Gamma \vdash A \wedge B} \right)^* = \langle 4, \ulcorner A \wedge B \urcorner, D_1^*, D_2^* \rangle$$

$$\left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash A \wedge B\end{array}}{\Gamma \vdash A} \right)^* = \langle 5, \ulcorner A \urcorner, D_1^* \rangle \qquad \left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash A \wedge B\end{array}}{\Gamma \vdash B} \right)^* = \langle 6, \ulcorner B \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash A\end{array}}{\Gamma \vdash A \vee B} \right)^* = \langle 7, \ulcorner A \vee B \urcorner, D_1^* \rangle \qquad \left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash B\end{array}}{\Gamma \vdash A \vee B} \right)^* = \langle 8, \ulcorner A \vee B \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{ccc}\vdots\ D_1 & \vdots\ D_2 & \vdots\ D_3 \\ \Gamma \vdash A \vee B & \Gamma, A \vdash C & \Gamma, B \vdash C\end{array}}{\Gamma \vdash C} \right)^* = \langle 9, \ulcorner C \urcorner, D_1^*, D_2^*, D_3^* \rangle$$

$$\left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash A\end{array}}{\Gamma \vdash \forall x\, A}\ {\scriptstyle (x \notin FV(\Gamma))} \right)^* = \langle 10, \ulcorner \forall x\, A \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash A\end{array}}{\Gamma \vdash A\{x := t\}} \right)^* = \langle 11, \ulcorner A\{x := t\} \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{c}\vdots\ D_1 \\ \Gamma \vdash A\{x := t\}\end{array}}{\Gamma \vdash \exists x\, A} \right)^* = \langle 12, \ulcorner \exists x\, A \urcorner, D_1^* \rangle$$

$$\left(\dfrac{\begin{array}{cc}\vdots\ D_1 & \vdots\ D_2 \\ \Gamma \vdash \exists x\, A & \Gamma, A \vdash B\end{array}}{\Gamma \vdash B}\ {\scriptstyle (x \notin FV(\Gamma, B))} \right)^* = \langle 13, \ulcorner B \urcorner, D_1^* \rangle$$

FIGURE 2 – Codage des dérivations $D \mapsto D^*$ (sans contexte)

On remarque alors que :

1. Si $\mathscr{T}$ a un nombre fini d'axiomes, alors $\mathscr{T}$ est une théorie récursive. (C'est immédiat car tout ensemble fini est récursif.) En particulier, l'arithmétique de Robinson (PA^-) est une théorie récursive.
2. L'arithmétique de Peano (PA) est également une théorie récursive. Pour s'en convaincre, il suffit de remarquer que l'ensemble

$$\mathbf{Rec} \;=\; \{\ulcorner A \urcorner \;:\; A \text{ formule (close) de récurrence}\}$$

constitué par tous les codes des axiomes de récurrence est un ensemble récursif. Il est alors clair que l'ensemble des codes des axiomes de PA, donné par

$$\mathbf{Ax}_{\mathrm{PA}} \;=\; \mathbf{Rec} \cup \{\ulcorner(\mathrm{E1})\urcorner ; \ldots ; \ulcorner(\mathrm{E7})\urcorner ; \ulcorner(\mathrm{PA1})\urcorner ; \ldots ; \ulcorner(\mathrm{PA6})\urcorner\},$$

est lui-même un ensemble récursif.

On vérifie sans difficulté que :

Proposition 18 — *Si $\mathscr{T}$ est une théorie récursive, alors l'ensemble*

$$\mathbf{Dem}_{\mathscr{T}} \;=\; \{(\ulcorner D \urcorner, \ulcorner A \urcorner) \;:\; D \textit{ démonstration de } A\} \;\subseteq\; \mathbb{N} \times \mathbb{N}$$

est un ensemble récursif.

Démonstration. Il s'agit essentiellement de construire une fonction récursive à deux arguments d et a qui teste si :

1. a est le code d'une formule A,
2. d est le code d'une dérivation d'un séquent de la forme $\Gamma \vdash A$, et
3. toutes les formules figurant dans Γ sont des axiomes de $\mathscr{T}$;

et qui échoue sinon. (L'hypothèse de récursivité de $\mathscr{T}$ est cruciale dans l'étape 3.) □

5 Le premier théorème d'incomplétude

5.1 Le lemme du point fixe

On commence par établir une propriété d'existence de point fixe (à travers le codage des formules) qui nous sera très utile dans la suite :

Lemme 11 (Point fixe) — *Si $\mathscr{T}$ est une théorie dans laquelle toutes les fonctions récursives sont représentables, alors pour tout formule $A(y)$ n'ayant pas d'autre variable libre que y, il existe une formule close F telle que : $\mathscr{T} \vdash F \Leftrightarrow A(\overline{\ulcorner F \urcorner})$.*

Démonstration. On considère une fonction récursive $\mathsf{substn} : \mathbb{N}^3 \to \mathbb{N}$ (voir section 4.4) telle que $\mathsf{substn}(\ulcorner A \urcorner, \widetilde{x}, n) = \ulcorner A\{x := \overline{n}\} \urcorner$ pour toute formule A, pour toute variable x et pour tout entier naturel n. D'après notre hypothèse sur la théorie $\mathscr{T}$, la fonction substn est représentée dans $\mathscr{T}$ par une formule $SubstN(x_1, x_2, x_3, y)$ qui satisfait donc la propriété

$$(*) \qquad \mathscr{T} \;\vdash\; SubstN(\overline{\ulcorner A \urcorner}, \overline{\widetilde{x}}, \overline{n}, y) \Leftrightarrow y = \overline{\ulcorner A\{x := \overline{n}\} \urcorner}$$

(où A, x et n sont quelconques). Supposons à présent que $A(y)$ est une formule n'ayant pas d'autre variable libre que y. On se fixe une variable x_0 distincte de y et on pose

$$\begin{array}{rcl} D & \equiv & \exists y\,(SubstN(x_0, \overline{\widetilde{x_0}}, x_0, y) \wedge A(y)) \\ F & \equiv & D\{x_0 := \overline{\ulcorner D \urcorner}\} \end{array}$$

Par construction, la formule D (la « diagonale ») ne dépend que de la variable x_0, et la formule F (le « point fixe ») est close. On vérifie alors que :

- $F \equiv \exists y\,(SubstN(\overline{\ulcorner D \urcorner}, \overline{\widetilde{x_0}}, \overline{\ulcorner D \urcorner}, y) \wedge A(y))$ (définition de F)
- $\mathscr{T} \vdash SubstN(\overline{\ulcorner D \urcorner}, \overline{\widetilde{x_0}}, \overline{\ulcorner D \urcorner}, y) \Leftrightarrow y = \overline{\ulcorner D\{x_0 := \overline{\ulcorner D \urcorner}\} \urcorner}$ (d'après $(*)$)
- $\mathscr{T} \vdash SubstN(\overline{\ulcorner D \urcorner}, \overline{\widetilde{x_0}}, \overline{\ulcorner D \urcorner}, y) \Leftrightarrow y = \overline{\ulcorner F \urcorner}$ (définition de F)
- $\mathscr{T} \vdash \exists y\,(SubstN(\overline{\ulcorner D \urcorner}, \overline{\widetilde{x_0}}, \overline{\ulcorner D \urcorner}, y) \wedge A(y)) \Leftrightarrow \exists y\,(y = \overline{\ulcorner F \urcorner} \wedge A(y))$
- $\mathscr{T} \vdash \exists y\,(SubstN(\overline{\ulcorner D \urcorner}, \overline{\widetilde{x_0}}, \overline{\ulcorner D \urcorner}, y) \wedge A(y)) \Leftrightarrow A(\overline{\ulcorner F \urcorner})$

c'est-à-dire précisément : $\mathscr{T} \vdash F \Leftrightarrow A(\overline{\ulcorner F \urcorner})$. □

5.2 Construction d'une formule G non démontrable (mais vraie)

On suppose donnée une théorie récursive $\mathscr{T}$ dans laquelle toutes les fonctions récursives sont représentables. (Cette dernière hypothèse est automatiquement satisfaite si $\mathscr{T}$ contient l'arithmétique de Robinson d'après le Théorème 1.)

D'après ces hypothèses, l'ensemble

$$\mathbf{Dem}_{\mathscr{T}} = \{(\ulcorner D \urcorner, \ulcorner A \urcorner) \,:\, D \text{ démonstration de } A\}$$

est récursif (Prop. 18), et il existe une formule $Dem_{\mathscr{T}}(x_1, x_2)$ telle que

1. $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{d}, \overline{\ulcorner A \urcorner})$ si d est le code d'une démonstration de A
2. $\mathscr{T} \vdash \neg Dem_{\mathscr{T}}(\overline{d}, \overline{\ulcorner A \urcorner})$ si d n'est pas le code d'une démonstration de A

pour tout formule A et pour tout entier naturel d. On pose

$$Th(x) \equiv \exists z\; Dem_{\mathscr{T}}(z, x) \qquad \text{« } x \text{ est le code d'un théorème de } \mathscr{T} \text{ »}$$

et on considère une formule close G telle que

$$\mathscr{T} \vdash G \Leftrightarrow \neg\, Th(\ulcorner G \urcorner)\,,$$

dont l'existence découle du lemme du point fixe (Lemme 11). On vérifie alors que :

Proposition 19 — *Si la théorie $\mathscr{T}$ est cohérente, alors la formule G n'est pas démontrable dans $\mathscr{T}$.*

Démonstration. On raisonne par contraposition en supposant que $\mathscr{T} \vdash G$, c'est-à-dire que G admet une démonstration D dans $\mathscr{T}$. On a alors

- $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner G \urcorner})$ (d'après la définition de $Dem_{\mathscr{T}}(x_1, x_2)$)
- $\mathscr{T} \vdash \exists z\, Dem_{\mathscr{T}}(z, \overline{\ulcorner G \urcorner})$ (règle $\exists$-intro)
- $\mathscr{T} \vdash Th(\overline{\ulcorner G \urcorner})$ (définition de la formule $Th(x)$)
- $\mathscr{T} \vdash \neg G$ (car $\mathscr{T} \vdash G \Leftrightarrow \neg\, Th(\overline{\ulcorner G \urcorner})$)

d'où il ressort que la théorie $\mathscr{T}$ est incohérente. □

On notera que la démonstration ci-dessus est une démonstration purement syntaxique qui ne fait aucune hypothèse sur la vérité de la formule G ou des axiomes de la théorie $\mathscr{T}$ dans le modèle standard. On vérifie cependant que :

Proposition 20 — *Si tous les axiomes de $\mathscr{T}$ sont vrais dans le modèle standard, alors la formule G est vraie et non démontrable dans $\mathscr{T}$:* $\mathbb{N} \vDash G$ *et* $\mathscr{T} \nvdash G$.

Démonstration. Si tous les axiomes de $\mathscr{T}$ sont vrais dans le modèle standard, alors tous les théorèmes (clos) de $\mathscr{T}$ sont également vrais dans le modèle standard (même raisonnement qu'à la Prop. 2) et la théorie $\mathscr{T}$ est cohérente (même raisonnement qu'à la Prop. 3). D'après la Prop. 19, la formule G n'est donc pas démontrable dans la théorie $\mathscr{T}$. Par conséquent, aucun entier naturel n n'est le code d'une démonstration de la formule G dans la théorie $\mathscr{T}$. On a donc

- $(n, \ulcorner G \urcorner) \notin \mathbf{Dem}_{\mathscr{T}}$ pour tout $n \in \mathbb{N}$
- $\mathscr{T} \vdash \neg Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner G \urcorner})$ pour tout $n \in \mathbb{N}$ (déf. de la formule $Dem_{\mathscr{T}}(x_1, x_2)$)
- $\mathbb{N} \vDash \neg Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner G \urcorner})$ pour tout $n \in \mathbb{N}$ (car $\mathbb{N} \vDash \mathscr{T}$)
- $\mathbb{N} \vDash \forall z \neg Dem_{\mathscr{T}}(z, \overline{\ulcorner G \urcorner})$
- $\mathbb{N} \vDash \neg \exists z\, Dem_{\mathscr{T}}(z, \overline{\ulcorner G \urcorner})$
- $\mathbb{N} \vDash \neg Th(\overline{\ulcorner G \urcorner})$ (car $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z, x)$)

d'où il ressort que $\mathbb{N} \vDash G$ (car $\mathbb{N} \vDash \mathscr{T} \vdash G \Leftrightarrow \neg Th(\overline{\ulcorner G \urcorner})$). □

Nous obtenons ainsi un premier résultat d'incomplétude, qui exprime qu'une théorie récursive ne peut pas capturer la notion de vérité au sens du modèle standard. Ce résultat s'applique non seulement à l'arithmétique de Peano, mais aussi à toute extension récursive de PA, pourvu que cette extension ne contienne que des axiomes vrais dans le modèle standard. Cette forme d'incomplétude n'est donc pas due au fait qu'on a « oublié » des axiomes en définissant l'arithmétique formelle, mais elle traduit un phénomène plus profond (et bien connu en philosophie) qui est que la vérité n'est pas réductible à la prouvabilité.

On notera cependant que l'hypothèse de récursivité sur la théorie $\mathscr{T}$ est essentielle dans la démonstration de la Prop. 20. En effet, si on considère la théorie $\mathscr{T}$ dont les axiomes sont précisément toutes les formules closes qui sont vraies dans le modèle standard, il est évident que la prouvabilité dans la théorie $\mathscr{T}$ coïncide (par définition !) avec la vérité dans le modèle standard. Mais le prix à payer est très élevé, car dans une telle théorie, c'est la notion même de démonstration qui perd son caractère effectif. (Nous reviendrons sur ce problème à la section 5.4.)

De la Prop. 20 on peut également déduire que :

Proposition 21 — *Si tous les axiomes de $\mathscr{T}$ sont vrais dans le modèle standard, alors la formule $\neg G$ n'est pas démontrable dans $\mathscr{T}$:* $\mathscr{T} \nvdash \neg G$.

Démonstration. Sous ces hypothèses, on a en effet $\mathbb{N} \vDash G$ d'après la Prop. 20, d'où $\mathbb{N} \nvDash \neg G$ et par conséquent $\mathscr{T} \nvdash \neg G$. □

Contrairement à la Prop. 19, ce deuxième résultat de non-prouvabilité repose sur des hypothèses sémantiques très fortes, à savoir :

(1) qu'il existe un modèle standard (cf la discussion de la section 1.6) ; et

(2) que tous les axiomes de $\mathscr{T}$ sont vrais dans le modèle standard.

On peut néanmoins établir que la formule $\neg G$ n'est pas démontrable dans la théorie $\mathscr{T}$ sous des hypothèses beaucoup plus faibles, et notamment sous l'hypothèse que la théorie $\mathscr{T}$ est 1-cohérente.

5.3 L'hypothèse de 1-cohérence

Définition 8 (Théorie 1-cohérente) — Soit $\mathscr{T}$ une théorie dans laquelle tous les ensembles récursifs sont représentables. On dit que $\mathscr{T}$ est 1-*cohérente* lorsque pour tout ensemble récursif $E \subseteq \mathbb{N}$ représenté par la formule $R_E(x)$ (au sens de la Déf. 6) :

$$\mathscr{T} \vdash \exists x\, R_E(x) \quad \text{entraîne} \quad E \neq \varnothing\,.$$

Intuitivement, une théorie 1-cohérente est une théorie $\mathscr{T}$ dans laquelle tout ensemble récursif $E \subseteq \mathbb{N}$ qui est *prouvablement non vide* (dans la théorie $\mathscr{T}$) est effectivement non vide. La 1-cohérence de $\mathscr{T}$ entraîne la cohérence de $\mathscr{T}$, et nous verrons un peu plus loin que la réciproque est fausse :

Lemme 12 — *Si $\mathscr{T}$ est 1-cohérente, alors $\mathscr{T}$ est cohérente.*

Démonstration. On raisonne par contraposition en supposant que la théorie $\mathscr{T}$ est incohérente. Dans ce cas, on a $\mathscr{T} \vdash \exists x\, R_E(x)$ pour tout ensemble récursif E, même dans le cas où E est vide. Donc la théorie $\mathscr{T}$ n'est pas 1-cohérente. □

On suppose à présent (comme dans la section 5.2) que $\mathscr{T}$ est une théorie récursive dans laquelle toutes les fonctions récursives sont représentables, et on reprend la formule G (construite par point fixe) telle que $\mathscr{T} \vdash G \Leftrightarrow \neg\, Th(\overline{\ulcorner G \urcorner})$.

Nous pouvons alors vérifier que :

Proposition 22 — *Si la théorie $\mathscr{T}$ est 1-cohérente, alors la formule $\neg G$ n'est pas démontrable dans $\mathscr{T}$.*

Démonstration. On suppose que $\mathscr{T}$ est 1-cohérente, et on considère l'ensemble récursif $E = \{n \in \mathbb{N} : (n, \ulcorner G \urcorner) \in \mathbf{Dem}_{\mathscr{T}}\}$ qui est représenté dans la théorie $\mathscr{T}$ par la formule $R_E(x) \equiv Dem_{\mathscr{T}}(x, \overline{\ulcorner G \urcorner})$. Comme la formule G n'est pas démontrable dans $\mathscr{T}$ (Prop. 19), l'ensemble E est vide. D'après l'hypothèse de 1-cohérence, la formule $\exists z\, Dem_{\mathscr{T}}(z, \ulcorner G \urcorner)$ n'est donc pas démontrable dans $\mathscr{T}$. Mais comme $\mathscr{T} \vdash \neg G \Leftrightarrow \exists z\, Dem_{\mathscr{T}}(z, \ulcorner G \urcorner)$, la formule $\neg G$ n'est pas démontrable dans $\mathscr{T}$. □

En résumé, nous avons démontré que dans une théorie récursive $\mathscr{T}$ dans laquelle toutes les fonctions récursives sont représentables :

1. Si $\mathscr{T}$ est cohérente, alors $\mathscr{T} \nvdash G$ (Prop. 19).
2. Si $\mathscr{T}$ est 1-cohérente, alors $\mathscr{T} \nvdash \neg G$ (Prop. 22).

Ces deux résultats nous permettent de donner une démonstration d'une forme affaiblie[8] du théorème de Gödel qui est la suivante :

Proposition 23 — *Si $\mathscr{T}$ est une théorie 1-cohérente, récursive et contenant l'arithmétique calculatoire (PA$^-$), alors $\mathscr{T}$ est incomplète en ce sens qu'il existe une formule close G telle que $\mathscr{T} \nvdash G$ et $\mathscr{T} \nvdash \neg G$.*

Nous verrons à la section 5.5 comment nous passer de l'hypothèse de 1-cohérence pour démontrer le théorème dans toute sa généralité.

8. La démonstration originelle de Gödel [3] repose sur une hypothèse encore plus forte que la 1-cohérence, à savoir la propriété de ω-consistance de $\mathscr{T}$, qui exprime que si $\mathscr{T} \vdash \exists x\, P(x)$, alors il existe au moins un entier $n \in \mathbb{N}$ tel que $\mathscr{T} \nvdash \neg P(n)$ (où $P(x)$ est une formule arbitraire). Le lecteur pourra vérifier que la ω-consistance de $\mathscr{T}$ entraîne la 1-cohérence de $\mathscr{T}$.

Théorie cohérentes et non 1-cohérentes Grâce à ce qui précède, il est facile de construire artificiellement des théories cohérentes (et même récursives !) qui ne satisfont pas la propriété de 1-cohérence :

Proposition 24 — *Si $\mathscr{T}$ est une théorie cohérente et récursive dans laquelle toutes les fonctions récursives sont représentables, et si G désigne la formule de Gödel associée à $\mathscr{T}$ (i.e. telle que $\mathscr{T} \vdash G \Leftrightarrow \neg Th(\overline{\ulcorner G \urcorner})$), alors la théorie $\mathscr{T} + \neg G$ est cohérente et récursive, mais n'est pas 1-cohérente.*

Démonstration. Comme d'après la Prop. 19 la formule G n'est pas démontrable dans la théorie $\mathscr{T}$, la théorie $\mathscr{T} + \neg G$ est évidemment cohérente. La théorie $\mathscr{T} + \neg G$ est également récursive, puisque $\mathbf{Ax}_{\mathscr{T}+\neg G} = \mathbf{Ax}_{\mathscr{T}} \cup \{\ulcorner \neg G \urcorner\}$. L'ensemble récursif $E = \{n \in \mathbb{N} : (n, \ulcorner G \urcorner) \in Dem_{\mathscr{T}}\}$ est représenté dans la théorie $\mathscr{T}$ — et donc aussi dans la théorie $\mathscr{T} + \neg G$ — par la formule $Dem_{\mathscr{T}}(x, \overline{\ulcorner G \urcorner})$. Par construction de la théorie $\mathscr{T} + \neg G$, on a $\mathscr{T} + \neg G \vdash \exists x\, Dem_{\mathscr{T}}(x, \overline{\ulcorner G \urcorner})$. Mais comme G n'est pas démontrable dans $\mathscr{T}$, l'ensemble E est vide. Par conséquent, $\mathscr{T} + \neg G$ n'est pas 1-cohérente. □

5.4 Vérité, prouvabilité et récursivité

La Prop. 20 entraîne immédiatement le résultat suivant :

Proposition 25 — *Aucun des deux ensembles V et F définis par*

$$\begin{array}{rcll} V & = & \{\ulcorner A \urcorner : A \textit{ formule close et } \mathbb{N} \vDash A\} & \textit{(formules vraies)} \\ F & = & \{\ulcorner A \urcorner : A \textit{ formule close et } \mathbb{N} \nvDash A\} & \textit{(formules fausses)} \end{array}$$

n'est semi-récursif[9]*, de même que leurs complémentaires $\complement V$ et $\complement F$.*

Démonstration. Soit E l'ensemble des codes des formules closes, qui est un ensemble récursif (cf section 4.4). On vérifie successivement que :

- Si V est semi-récursif, alors $\complement F$ est semi-récursif (car $\complement F = V \cup \complement E$).
- Si $\complement F$ est semi-récursif, alors V est semi-récursif (car $V = E \cap \complement F$).
- Si F est semi-récursif, alors $\complement V$ est semi-récursif (car $\complement V = F \cup \complement E$).
- Si $\complement V$ est semi-récursif, alors F est semi-récursif (car $F = E \cap \complement V$).
- Si V est semi-récursif, alors F est semi-récursif (car $F = \mathsf{not}^{-1}(V)$)
- Si F est semi-récursif, alors V est semi-récursif (car $V = \mathsf{not}^{-1}(F)$)

(où $\mathsf{not} : \mathbb{N} \to \mathbb{N}$ est la fonction récursive définie par $\mathsf{not}(\ulcorner A \urcorner) = \ulcorner \neg A \urcorner$). Par conséquent, si l'un des quatre ensembles V, F, $\complement V$ ou $\complement F$ est semi-récursif, alors les trois autres le sont aussi. Faisons l'hypothèse que tous ces ensembles sont semi-récursifs, et donc récursifs (puisque leurs complémentaires sont semi-récursifs également). Dans ce cas, la théorie $\mathscr{T} = \{A : \ulcorner A \urcorner \in V\}$ est récursive. Cette théorie contient par ailleurs l'arithmétique de Peano (Prop. 2) et permet donc de représenter toutes les fonctions récursives (Théorème 1). D'après la Prop. 20, il existe une formule close G qui est vraie et non démontrable dans $\mathscr{T}$, ce qui est absurde puisque G, qui est vraie, est un axiome de $\mathscr{T}$. Donc l'hypothèse faite sur les ensembles V, F, $\complement V$ et $\complement F$ est absurde, et aucun de ceux-ci n'est un ensemble semi-récursif. □

9. C'est-à-dire : récursivement énumérable.

En revanche, l'ensemble des (codes des) formules *prouvables* (resp. *réfutables*) dans une théorie récursive donnée est bien évidemment un ensemble semi-récursif :

Proposition 26 — *Si $\mathscr{T}$ est une théorie récursive, alors les ensembles*

$$\begin{array}{rcll} P & = & \{\ulcorner A\urcorner \,:\, A \textit{ formule close et } \mathscr{T} \vdash A\} & \textit{(formules prouvables)} \\ R & = & \{\ulcorner A\urcorner \,:\, A \textit{ formule close et } \mathscr{T} \vdash \neg A\} & \textit{(formules réfutables)} \end{array}$$

sont tous les deux semi-récursifs.

Démonstration. On construit une énumération des codes des formules closes prouvables (ou réfutables) dans $\mathscr{T}$ à partir d'une énumération des codes des démonstrations, en ne conservant que la formule de conclusion (ou sa négation). □

On insistera sur le fait que la proposition ci-dessus énonce que les ensembles P (codes des formules prouvables) et R (codes des formules réfutables) sont tous les deux semi-récursifs (c'est-à-dire : récursivement énumérables), ce qui n'implique en aucun cas qu'ils sont récursifs. En effet, on peut démontrer que :

Théorème 2 (Indécidabilité de la prouvabilité) — *Si $\mathscr{T}$ est une théorie récursive cohérente dans laquelle toutes les fonctions récursives sont représentables, alors aucun des deux ensembles*

$$\begin{array}{rcll} P & = & \{\ulcorner A\urcorner \,:\, A \textit{ formule close et } \mathscr{T} \vdash A\} & \textit{(formules prouvables)} \\ R & = & \{\ulcorner A\urcorner \,:\, A \textit{ formule close et } \mathscr{T} \vdash \neg A\} & \textit{(formules réfutables)} \end{array}$$

n'est récursif.

Démonstration. On commence par remarquer que l'un des deux ensembles P ou R est récursif si et seulement si l'autre l'est (car $\ulcorner A\urcorner \in R$ ssi $\mathsf{not}(\ulcorner A\urcorner) \in P$, et vice-versa). Soit x une variable fixée, et E l'ensemble (récursif) des codes des formules n'ayant pas d'autre variable libre que x, c'est-à-dire : $E = \{\ulcorner A\urcorner : FV(A) \subseteq \{x\}\}$. On considère à présent le sous-ensemble $\Omega \subseteq E$ défini par

$$\begin{array}{rcl} \Omega & = & \{n = \ulcorner A(x)\urcorner \in E \,:\, \mathscr{T} \nvdash A(\overline{n})\} \\ & = & \{n = \ulcorner A(x)\urcorner \in E \,:\, \ulcorner A(\overline{n})\urcorner \notin P\} \end{array}$$

Si l'on suppose que P est récursif, alors il est clair que l'ensemble Ω est lui-même récursif. On considère alors une formule $B(x)$ qui représente cet ensemble et on pose $b = \ulcorner B(x)\urcorner \in E$. On distingue deux cas :

- Soit $b \in \Omega$. Dans ce cas on a (par représentation) $\mathscr{T} \vdash B(\overline{b})$, d'où $b \notin \Omega$.
- Soit $b \notin \Omega$. Dans ce cas on a (par représentation) $\mathscr{T} \vdash \neg B(\overline{b})$, d'où $\mathscr{T} \nvdash B(\overline{b})$ (d'après l'hypothèse de cohérence) et finalement $b \in \Omega$.

Les deux cas conduisent à une contradiction, ce qui montre que l'hypothèse de départ (i.e. P est récursif) est absurde. □

(On notera que l'hypothèse de cohérence est essentielle dans la démonstration ci-dessus, car dans une théorie incohérente, les ensembles P et R sont égaux à l'ensemble des codes des formules closes, qui est évidemment récursif.)

Le lecteur aura sans doute remarqué que les démonstrations de la Prop. 26 et du Théorème 2 sont toutes les deux indépendantes du premier théorème d'incomplétude de Gödel. Nous l'invitons à vérifier (en exercice) que ces deux résultats impliquent directement le premier théorème d'incomplétude de Gödel, mais de manière non constructive (i.e. sans exhiber la formule indécidable).

5.5 La variante de Rosser

La démonstration de la Prop. 23 (théorème d'incomplétude faible) repose sur l'hypothèse que la théorie $\mathscr{T}$ est 1-cohérente. On peut se passer de cette hypothèse en utilisant une astuce proposée par Rosser, qui consiste à remplacer la relation $Dem_{\mathscr{T}}(x,y)$ par la relation $Dem'_{\mathscr{T}}(x,y)$ définie par :

$$\begin{array}{rcl} Dem'_{\mathscr{T}}(x,y) & \equiv & Dem_{\mathscr{T}}(x,y) \ \wedge \\ & & \exists y' \, (Not(y,y') \ \wedge \ \forall x' < x \ \neg Dem_{\mathscr{T}}(x',y')) \, . \end{array}$$

(où $Not(x,y)$ est la formule qui représente la fonction récursive $\mathsf{not} : \mathbb{N} \to \mathbb{N}$ telle que $\mathsf{not}(\ulcorner A \urcorner) = \ulcorner \neg A \urcorner$ pour toute formule A).

On se place alors dans une théorie récursive $\mathscr{T}$ contenant l'arithmétique de Robinson (PA^-), et on construit par point fixe (Lemme 11) une formule G' telle que

$$\begin{array}{rcl} \mathscr{T} \ \vdash \ G' & \Leftrightarrow & \neg \exists z \, Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner}) \\ & \Leftrightarrow & \neg \exists z \, \big(Dem_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner}) \ \wedge \ \forall z' < z \ \neg Dem_{\mathscr{T}}(z, \overline{\ulcorner \neg G' \urcorner}) \big) \end{array}$$

Proposition 27 — *Si la théorie $\mathscr{T}$ est cohérente, alors la formule G' n'est pas démontrable dans $\mathscr{T}$.*

Démonstration. Supposons que $\mathscr{T}$ est cohérente et que la formule G' admet une démonstration D dans la théorie $\mathscr{T}$. On a alors

- $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner G' \urcorner})$ (représentation)
- $\mathscr{T} \vdash \neg Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner \neg G' \urcorner})$ pour tout entier n (repr. et cohérence de $\mathscr{T}$)
- $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner G' \urcorner}) \wedge \bigwedge_{n < \ulcorner D \urcorner} \neg Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner \neg G' \urcorner})$
- $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner G' \urcorner}) \wedge \forall z' < \overline{\ulcorner D \urcorner} \ \neg Dem_{\mathscr{T}}(z', \overline{\ulcorner \neg G' \urcorner})$ (Prop. 9)
- $\mathscr{T} \vdash \exists z \, Dem'_{\mathscr{T}}(z, \overline{\ulcorner G \urcorner})$ ($\exists$-intro)
- $\mathscr{T} \vdash \neg G$ (car $\mathscr{T} \vdash G \Leftrightarrow \neg \exists z \, Dem'_{\mathscr{T}}(z, \overline{\ulcorner G \urcorner})$)

ce qui est absurde puisque $\mathscr{T}$ est cohérente. □

Proposition 28 — *Si la théorie $\mathscr{T}$ est cohérente, alors la formule $\neg G'$ n'est pas démontrable dans $\mathscr{T}$.*

Démonstration. Supposons que $\mathscr{T}$ est cohérente et que la formule $\neg G'$ admet une démonstration D dans la théorie $\mathscr{T}$.

1. $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner \neg G \urcorner})$ (représentation)
2. $\mathscr{T} \vdash Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner}) \Rightarrow \big(\overline{\ulcorner D \urcorner} < z \Rightarrow \neg Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner \neg G \urcorner}) \big)$
3. $\mathscr{T} \vdash Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner}) \Rightarrow \overline{\ulcorner D \urcorner} \not< z$ (d'après 1. et 2.)
4. $\mathscr{T} \vdash Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner}) \Rightarrow z \leq \overline{\ulcorner D \urcorner}$ (Prop. 8)
5. $\mathscr{T} \vdash \neg G'$ (par hypothèse)
6. $\mathscr{T} \vdash \exists z \, Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner})$ (car $\mathscr{T} \vdash G' \Leftrightarrow \neg \exists z \, Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner})$)
7. $\mathscr{T} \vdash \exists z \leq \overline{\ulcorner D \urcorner} \ Dem'_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner})$ (d'après 4. et 6.)
8. $\mathscr{T} \vdash \exists z \leq \overline{\ulcorner D \urcorner} \ Dem_{\mathscr{T}}(z, \overline{\ulcorner G' \urcorner})$

9. $\mathscr{T} \vdash \bigvee_{n \leq \ulcorner D' \urcorner} Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner G' \urcorner})$ (Prop. 9)

10. $\mathscr{T} \vdash \neg Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner G' \urcorner})$ pour tout entier n (repr. et cohérence de $\mathscr{T}$)

11. $\mathscr{T} \vdash \bigwedge_{n \leq \ulcorner D' \urcorner} \neg Dem_{\mathscr{T}}(\overline{n}, \overline{\ulcorner G' \urcorner})$ (d'après 10.)

12. $\mathscr{T} \vdash \bot$ (d'après 9. et 11.)

ce qui est absurde. □

Nous venons ainsi de terminer la démonstration du

Théorème 3 (Premier théorème d'incomplétude) — *Si $\mathscr{T}$ est une théorie cohérente, récursive et contenant* PA$^-$*, alors $\mathscr{T}$ est incomplète en ce sens qu'il existe une formule close G' telle que $\mathscr{T} \nvdash G'$ et $\mathscr{T} \nvdash \neg G'$.*

6 Le second théorème d'incomplétude

6.1 L'argument de Gödel

Soit $\mathscr{T}$ une théorie récursive contenant l'arithmétique de Robinson (PA$^-$).

Nous avons vu que la démonstration du premier théorème d'incomplétude repose sur la construction d'une formule G telle que $\mathscr{T} \vdash G \Leftrightarrow \neg Th(\overline{\ulcorner G \urcorner})$, formule dont nous avons démontré qu'elle n'est pas prouvable dans la théorie $\mathscr{T}$ si l'on suppose que $\mathscr{T}$ est cohérente. Pour démontrer le second théorème d'incomplétude, Gödel remarque alors que la démonstration (dans la théorie ambiante) de l'implication

$$\textit{Si } \mathscr{T} \textit{ est cohérente, alors } \mathscr{T} \nvdash G$$

n'utilise en réalité que des principes de raisonnements finitaires. Grâce à cette observation, on peut donc reformuler l'implication ci-dessus dans le langage de l'arithmétique (en utilisant le codage des formules et des démonstrations dans les entiers), et la démontrer formellement dans PA. Les énoncés « $\mathscr{T}$ est cohérente » et « G n'est pas démontrable » sont remplacés par les formules $Cons\mathscr{T} \equiv \neg Th(\overline{\ulcorner \bot \urcorner})$ et $\neg Th(\overline{\ulcorner G \urcorner})$, et l'énoncé du premier théorème d'incomplétude devient alors

$$Cons\mathscr{T} \Rightarrow \neg Th(\overline{\ulcorner G \urcorner})$$

où $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z, x)$. Gödel construit ensuite une démonstration formelle de la formule $Cons\mathscr{T} \Rightarrow \neg Th(\overline{\ulcorner G \urcorner})$ dans l'arithmétique de Peano, en *internalisant* (dans PA) la démonstration du premier théorème d'incomplétude. (On reviendra plus loin sur les difficultés posées par une telle internalisation.)

Si l'on suppose en outre que la théorie $\mathscr{T}$ contient elle-même l'arithmétique de Peano (et pas seulement l'arithmétique de Robinson), on déduit de ce qui précède que $\mathscr{T} \vdash Cons\mathscr{T} \Rightarrow \neg Th(\overline{\ulcorner G \urcorner})$ (car PA $\subseteq \mathscr{T}$), et finalement que

$$\mathscr{T} \vdash Cons\mathscr{T} \Rightarrow G\,,$$

puisque $\mathscr{T} \vdash G \Leftrightarrow \neg Th(\overline{\ulcorner G \urcorner})$ d'après la définition de la formule G.

Mais comme $\mathscr{T} \nvdash G$ (sous l'hypothèse que $\mathscr{T}$ est cohérente), il est immédiat que $\mathscr{T} \nvdash Cons\mathscr{T}$. La théorie $\mathscr{T}$ ne peut donc pas démontrer la formule $Cons\mathscr{T}$ qui exprime la cohérence de $\mathscr{T}$ à travers le codage dans les entiers.

Internalisation de la démonstration du premier théorème Évidemment, la partie la plus technique de la démonstration ci-dessus (qui autrement ne présente aucune difficulté conceptuelle) réside dans la construction effective de la démonstration de la formule $Cons\mathscr{T} \Rightarrow \neg Th(\overline{\ulcorner G \urcorner})$ dans l'arithmétique de Peano, en internalisant la démonstration du premier théorème d'incomplétude.

On peut tout de suite remarquer qu'il n'est pas nécessaire d'internaliser le théorème de représentation (Théorème 1), car celui-ci n'est nécessaire que pour construire certaines formules cruciales dans l'énoncé du premier théorème d'incomplétude. On pourra donc reprendre les formules telles quelles, en s'assurant qu'elles vérifient les bonnes propriétés. Par exemple, on devra démontrer que la substitution est compatible avec la relation d'α-équivalence, c'est-à-dire, dans PA :

$$\begin{array}{l} Form(x) \wedge Form(x') \wedge Term(y) \wedge \\ Subst(x,v,y,z) \wedge Subst(x',v,y,z') \wedge Equiv(x,x') \;\Rightarrow\; Equiv(z,z') \end{array}$$

(où $Form(x)$ exprime que x est le code d'une formule, $Term(y)$ que y est le code d'un terme, $Equiv(x,x')$ que les formules de code x et x' sont α-convertibles, etc.) On notera que c'est pour démontrer des lemmes administratifs comme celui-ci qu'on aura besoin du principe de récurrence dans PA.

Malgré cela, la démonstration formelle de l'implication $Cons\mathscr{T} \Rightarrow \neg Th(\overline{\ulcorner G \urcorner})$ *dans* PA reste un exercice techniquement difficile, et c'est pourquoi il peut être utile de considérer le problème sous une forme un peu plus abstraite afin de dégager les ingrédients cruciaux de cette démonstration. Ce qui nous conduit à nous pencher sur les conditions de dérivabilité introduites par Hilbert et Bernays.

6.2 Les conditions de dérivabilité de Hilbert-Bernays

Définition 9 (Conditions de dérivabilité de Hilbert-Bernays) — Soit $\mathscr{T}$ une théorie munie d'une formule $Th(x)$ à une seule variable libre x. On dit que $\mathscr{T}$ satisfait les *conditions de dérivabilité de Hilbert-Bernays* si pour toutes formules closes A et B les trois propriétés suivantes sont satisfaites :

(HB1) Si $\mathscr{T} \vdash A$, alors $\mathscr{T} \vdash Th(\overline{\ulcorner A \urcorner})$;

(HB2) $\mathscr{T} \vdash Th(\overline{\ulcorner A \Rightarrow B \urcorner}) \Rightarrow \Big(Th(\overline{\ulcorner A \urcorner}) \Rightarrow Th(\overline{\ulcorner B \urcorner}) \Big)$;

(HB3) $\mathscr{T} \vdash Th(\overline{\ulcorner A \urcorner}) \Rightarrow Th\Big(\overline{\ulcorner Th(\overline{\ulcorner A \urcorner}) \urcorner}\Big)$.

Pour manipuler plus facilement ces trois conditions, on a l'habitude d'introduire la notation suggestive

$$\Box A \;\equiv\; Th_{\mathscr{T}}(\overline{\ulcorner A \urcorner}) \qquad (\text{« } A \text{ est prouvable »})$$

pour chaque formule close A. Avec cette notation, les conditions de dérivabilité de Hilbert-Bernays s'écrivent alors plus simplement [10] :

(HB1) Si $\mathscr{T} \vdash A$, alors $\mathscr{T} \vdash \Box A$;

(HB2) $\mathscr{T} \vdash \Box(A \Rightarrow B) \Rightarrow (\Box A \Rightarrow \Box B)$;

(HB3) $\mathscr{T} \vdash \Box A \Rightarrow \Box\Box A$;

(où A et B sont des formules closes quelconques).

10. On retrouve là trois principes bien connus en logique modale, à savoir la règle de nécessitation (HB1), la loi de distributivité (HB2) et la loi de transitivité (HB3), ici restreints aux formules closes.

La condition (HB1) exprime que si A est prouvable dans $\mathscr{T}$, alors la formule $\Box A$ qui exprime que « A est prouvable » est elle aussi prouvable dans $\mathscr{T}$.

On notera que si la formule $Th(x)$ est définie par $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z,x)$, où $Dem_{\mathscr{T}}(z,x)$ est une représentation (au sens de la Déf. 6) de l'ensemble $\mathbf{Dem}_{\mathscr{T}}$ défini à la section 4.7, alors la condition (HB1) est automatiquement satisfaite, car si une formule (close) A est démontrable dans $\mathscr{T}$ et si D est une démonstration de cette formule dans $\mathscr{T}$, alors on a $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{\ulcorner D \urcorner}, \overline{\ulcorner A \urcorner})$ (d'après la propriété de représentation) et par conséquent $\mathscr{T} \vdash Th(\overline{\ulcorner A \urcorner})$ (règle $\exists$-intro).

On peut cependant remarquer que ce raisonnement n'utilise que la « partie positive » de la propriété de représentation, c'est-à-dire le critère (i) de la Déf. 6 :

Si $(n,m) \in \mathbf{Dem}_{\mathscr{T}}$, *alors* $\mathscr{T} \vdash Dem_{\mathscr{T}}(\overline{n}, \overline{m})$.

On pourra donc choisir la formule $Dem_{\mathscr{T}}(z,x)$ de façon à faciliter la vérification du premier critère sans se soucier du second [11].

La condition (HB2) exprime dans la théorie $\mathscr{T}$ que la prouvabilité d'une implication et la prouvabilité de son membre gauche entraînent la prouvabilité de son membre droit. Cette condition n'est donc rien d'autre que le principe du *modus ponens*

Si $\mathscr{T} \vdash A \Rightarrow B$ *et si* $\mathscr{T} \vdash A$, *alors* $\mathscr{T} \vdash B$

qu'on a internalisé dans la théorie $\mathscr{T}$ à l'aide de la modalité $\Box A$:

$$\mathscr{T} \vdash \Box(A \Rightarrow B) \Rightarrow (\Box A \Rightarrow \Box B)\,.$$

La condition (HB3) procède du même esprit que les deux autres conditions, et n'est en réalité rien d'autre que la condition (HB1)

Si $\mathscr{T} \vdash A$, *alors* $\mathscr{T} \vdash \Box A$

qu'on a internalisée à son tour dans la théorie $\mathscr{T}$:

$$\mathscr{T} \vdash \Box A \Rightarrow \Box\Box A\,.$$

Nous verrons à la section 6.3 que les conditions de Hilbert-Bernays capturent les propriétés nécessaires pour démontrer le second théorème d'incomplétude. Pour pouvoir utiliser ces conditions dans l'arithmétique de Peano et dans ses extensions récursives, il nous faut donc d'abord démontrer le résultat suivant :

Proposition 29 (Critères de Hilbert-Bernays pour PA) — *Toute théorie $\mathscr{T}$ qui est une extension récursive de l'arithmétique de Peano (i.e.* PA $\subseteq \mathscr{T}$*) satisfait les critères de Hilbert-Bernays pour la formule $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z,x)$, où $Dem_{\mathscr{T}}(z,x)$ est une représentation de l'ensemble récursif $\mathbf{Dem}_{\mathscr{T}}$ dans* PA$^-$ *(au sens de la Déf. 6).*

11. Le critère (ii) de la Déf. 6 (« *Si* $(n,m) \notin \mathbf{Dem}_{\mathscr{T}}$, *alors* $\mathscr{T} \vdash \neg Dem_{\mathscr{T}}(\overline{n}, \overline{m})$ ») sert en fait à nous assurer que la relation $Dem_{\mathscr{T}}(z,x)$ n'accepte pas « trop de démonstrations » (c'est-à-dire des entiers ne correspondant à aucune démonstration), et est essentielle pour nous convaincre que la formule $Cons\mathscr{T} \equiv \neg\exists z\, Dem_{\mathscr{T}}(z, \ulcorner \bot \urcorner)$ traduit la propriété de cohérence. D'un point de vue technique cependant, le critère (ii) est inutile dans la démonstration du second théorème d'incomplétude, et le lecteur pourra vérifier que si on choisit une formule $Dem_{\mathscr{T}}(z,x)$ qui est toujours vraie dans $\mathscr{T}$, alors les conditions de dérivabilité de Hilbert-Bernays sont trivialement satisfaites pour la formule $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z,x)$ (toujours vraie dans $\mathscr{T}$). Avec ce choix erroné de la formule $Th(x)$, l'énoncé « Si $\mathscr{T}$ est cohérente, alors la formule $Cons\mathscr{T}$ n'est pas démontrable dans $\mathscr{T}$ » reste non seulement vrai, mais devient même une lapalissade dans la mesure où la formule $Cons\mathscr{T}$ est *par construction* équivalente à la formule $\bot$.

Démonstration. Nous avons déjà vu que la condition (HB1) est automatiquement satisfaite d'après le choix de la formule $Th(x)$. Le critère (HB2) se déduit du fait que :

$$\text{PA} \vdash \forall x \, \forall y \, \forall z \, (Imp(x,y,z) \wedge Th(z) \wedge Th(x) \Rightarrow Th(y)) \, ,$$

où $Imp(x,y,z)$ représente dans PA la fonction récursive $\mathsf{imp}(a,b) : \mathbb{N}^2 \to \mathbb{N}$ qui code l'implication, et qui est définie par $\mathsf{imp}(a,b) = \langle 2, a, b \rangle$. On ne construira pas ici la démonstration formelle (dans PA) de la formule ci-dessus, qui dépend du choix de la formule $Dem_{\mathscr{T}}(z,x)$. Enfin, le critère (HB3) se déduit du fait que :

$$\text{PA} \vdash \forall x \, \forall y \, (SubstN(\overline{\ulcorner Th(x_0) \urcorner}, \overline{\widetilde{x_0}}, x, y) \wedge Th(x) \Rightarrow Th(y)) \, ,$$

où $SubstN(x_1, x_2, x_3, y)$ est la formule introduite à la section 5.2. Là encore, on ne construira pas ici la démonstration formelle (dans PA) de la formule ci-dessus. □

On notera que la démonstration de ce résultat permet de capturer les deux théorèmes dont on a besoin de construire une démonstration formelle dans PA, à savoir :

1. $\text{PA} \vdash \forall x \, \forall y \, \forall z \, (Imp(x,y,z) \wedge Th(z) \wedge Th(x) \Rightarrow Th(y))$
2. $\text{PA} \vdash \forall x \, \forall y \, (SubstN(\overline{\ulcorner Th(x_0) \urcorner}, \overline{\widetilde{x_0}}, x, y) \wedge Th(x) \Rightarrow Th(y))$

La démonstration de ces théorèmes dépendra évidemment du choix de la formule $Dem_{\mathscr{T}}(z,x)$, et certains choix pourront se révéler plus judicieux que d'autres...

6.3 Le théorème de Löb

On commence par démontrer une version abstraite du second théorème d'incomplétude de Gödel, due à Löb [4] :

Théorème 4 (Löb) — *Soit $\mathscr{T}$ une théorie récursive dans laquelle toutes les fonctions récursives sont représentables, et satisfaisant les conditions de Hilbert-Bernays pour une certaine formule $Th(x)$. Si $\mathscr{T} \vdash Th(\overline{\ulcorner C \urcorner}) \Rightarrow C$ (où C est une formule close quelconque), alors $\mathscr{T} \vdash C$.*

Démonstration. Soit F une formule close telle que $\mathscr{T} \vdash F \Leftrightarrow (Th(\overline{\ulcorner F \urcorner}) \Rightarrow C)$, dont l'existence nous est donnée par le lemme du point fixe (Lemme 11) appliqué à la formule $A(y) \equiv (Th(y) \Rightarrow C)$. Avec la notation $\Box A \equiv Th(\overline{\ulcorner A \urcorner})$, on vérifie que :

1. $\mathscr{T} \vdash \Box C \Rightarrow C$ — par hypothèse
2. $\mathscr{T} \vdash F \Leftrightarrow (\Box F \Rightarrow C)$ — par construction de F
3. $\mathscr{T} \vdash F \Rightarrow (\Box F \Rightarrow C)$ — d'après 2.
4. $\mathscr{T} \vdash \Box(F \Rightarrow (\Box F \Rightarrow C))$ — d'après 3. et (HB1)
5. $\mathscr{T} \vdash \Box F \Rightarrow \Box(\Box F \Rightarrow C)$ — d'après 4. et (HB2)
6. $\mathscr{T} \vdash \Box F \Rightarrow (\Box\Box F \Rightarrow \Box C)$ — d'après 5. et (HB2)
7. $\mathscr{T} \vdash \Box F \Rightarrow \Box\Box F$ — d'après (HB3)
8. $\mathscr{T} \vdash \Box F \Rightarrow \Box C$ — d'après 6. et 7.
9. $\mathscr{T} \vdash \Box F \Rightarrow C$ — d'après 1. et 8.
10. $\mathscr{T} \vdash F$ — d'après 2. et 9.
11. $\mathscr{T} \vdash \Box F$ — d'après 10. et (HB1)
12. $\mathscr{T} \vdash C$ — d'après 9. et 11. □

On notera que ce théorème n'impose pas de choix particulier sur la formule $Th(x)$. On pourra donc prendre $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z,x)$, $Th(x) \equiv \exists z\, Dem'_{\mathscr{T}}(z,x)$ (en reprenant la formule $Dem'_{\mathscr{T}}(z,x)$ de la section 5.5) ou même $Th(x) \equiv \top$ (où $\top$ désigne une formule trivialement vraie).

De ce résultat général combiné avec la Prop. 29 on déduit le

Théorème 5 (Second théorème d'incomplétude) — *Si $\mathscr{T}$ est une théorie cohérente et récursive contenant l'arithmétique de Peano (*PA*), alors la formule*

$$\mathrm{Cons}\mathscr{T} \;\equiv\; \neg Th(\overline{\ulcorner \bot \urcorner}) \;\equiv\; \neg\exists z\, Dem_{\mathscr{T}}(z, \overline{\ulcorner \bot \urcorner}) \qquad (\text{« } \mathscr{T} \textit{ est cohérente } \text{»})$$

(où la formule $Dem_{\mathscr{T}}(z,x)$ est une représentation de l'ensemble récursif $\mathbf{Dem}_{\mathscr{T}}$ dans PA$^-$*) n'est pas démontrable dans la théorie $\mathscr{T}$.*

Démonstration. On raisonne par contraposition en supposant que $\mathscr{T} \vdash \mathrm{Cons}\mathscr{T}$, c'est-à-dire : $\mathscr{T} \vdash Th(\overline{\ulcorner \bot \urcorner}) \Rightarrow \bot$. D'après la Prop. 29, la théorie $\mathscr{T}$ munie de la formule $Th(x) \equiv \exists z\, Dem_{\mathscr{T}}(z,x)$ satisfait les conditions de Hilbert-Bernays, ce qui permet d'appliquer le théorème de Löb avec $C \equiv \bot$ et d'en déduire que $\mathscr{T} \vdash \bot$. □

(On notera que dans le cas où $C \equiv \bot$, la formule F qui est utilisée dans la démonstration du théorème de Löb est précisément la formule G de Gödel.)

Références

[1] R. Cori et D. Lascar. *Logique mathématique 1 - Calcul propositionnel ; algèbre de Boole ; calcul des prédicats* (2e édition). Dunod. ISBN : 2100054538. 2003.

[2] R. Cori et D. Lascar. *Logique mathématique 2 – Fonctions récursives ; théorème de Gödel ; Théorie des ensembles* (2e édition). Dunod. ISBN : 2100054538. 2003.

[3] K. Gödel. *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I.* Monatshefte für Math. und Physik 38 :173–98. 1931

[4] E. Mendelson. *Introduction to Mathematical Logic, 4th Edition.* Chapman & Hall, 1997.

[5] E. Nagel, J. R. Newman, K. Gödel et J.-Y. Girard. *Le théorème de Gödel*, Éditions du Seuil (coll. *Points*), ISBN : 2-02-032778-3. 1989