UNE NOTE SUR LA TENTATIVE D'INTERNALISATION DU PRÉDICAT DE VÉRITÉ DANS LES *PRINCIPLES OF MATHEMATICS REVISITED* DE HINTIKKA[1]

François Rivenc

1. La question de savoir si la notion de vérité est ou non définissable est une question traditionnelle de la philosophie. On peut considérer que le travail de Tarski a montré définitivement que la réponse était positive, du moins pour les langages formalisés, dont les ressources expressives et la syntaxe sont exactement spécifiées. Mais avec une réserve, qu'on ne formulera ici que de façon très générale. Une définition adéquate de la notion de vérité pour un langage ne peut être construite dans ce même langage, mais exige un « métalangage » plus riche en moyens d'expressions. La « révolution » que propose Hintikka dans *The Principles of Mathematics Revisited* (1996) consiste globalement à montrer qu'on peut surmonter ce résultat d'impossibilité. Les lignes qui suivent n'ont d'autre but que de spécifier les différentes « briques » que dispose Hintikka dans cet ouvrage, pour construire pas à pas une définition de la vérité pour certains langages *dans* ces mêmes langages. J'éviterai toute technicité non indispensable à la compréhension de cette construction. De même, je laisserai de côté les multiples implications philosophiques que Hintikka développe ou suggère au passage. L'unique sujet de cette étude est donc la définition de la vérité d'un énoncé dans un modèle[2].

L'importance d'une telle définition ne saurait être surestimée. Sans la notion de vérité, remarque Hintikka, « il y a peu d'espoir de capturer des concepts logiques de base tels que la validité (vérité dans tous les modèles) et la conséquence logique » (p. 15)[3]. De plus, les conditions de vérité d'un énoncé *S* qu'articule une telle définition donnent d'une certaine façon le sens de cet énoncé : affirmer *S* revient à dire que ses conditions de vérité sont réalisées, et c'est là connaître ce que signifie *S*.

Malheureusement, la logique du premier ordre, ordinairement tenue pour le cœur de la logique, notre logique fondamentale et naturelle, ne permet pas de définir les

---

[1] Je remercie Philippe de Rouilhan, qui m'a aidé à corriger les erreurs d'une première version.

[2] Je reprends la terminologie de Hintikka. Il serait sans doute préférable de parler de structure d'interprétation pour un langage, et de réserver le terme de modèle d'un énoncé ou d'un ensemble d'énoncés pour une structure d'interprétation où cet énoncé (ou cet ensemble d'énoncés) est vrai. C'est en ce sens qu'on parlait d'un *modèle* d'une géométrie non-euclidienne dans la géométrie euclidienne.

[3] Les numéros de page renvoient à l'édition révisée du texte *(The Principles of Mathematics Revisited*, Revised ed., Cambridge, Cambridge University Press, 1999).

concepts métalogiques fondamentaux de vérité, de validité, ni de conséquence logique, et cela suffit à montrer l'insuffisance de la logique du premier ordre. La théorie des modèles pour cette logique est reléguée en second ordre, ou plus ordinairement dans la théorie des ensembles. Telle est, dans le style volontiers railleur de Hintikka, la malédiction de Tarski (« *Tarski's curse* » ; ce n'est pas la lettre des écrits de Tarski sur la question, mais on peut en effet estimer que c'en est bien l'esprit). Selon Hintikka, de tels résultats pavent la voie à la thèse de *l'ineffabilité* de la sémantique, c'est-à-dire à un certain défaitisme concernant la possibilité de la théorie des modèles (défaitisme qui serait au fond la pensée profonde de la plupart des logiciens du XX^{e} siècle). Le but des *Principles of Mathematics Revisited* est de combattre ce défaitisme, en montrant qu'on peut s'affranchir des résultats d'impossibilité de Tarski (en particulier).

2. Commence alors, avec le Chapitre II des *Principles of Mathematics Revisited*, un détour, qui n'est que le rappel de la sémantique en termes de théorie des jeux (« *game-theoretical semantics* », en abrégé GTS). Soit *S* un énoncé d'un langage L du premier ordre interprété dans un modèle M de domaine do(M). Le jeu sémantique met aux prises deux joueurs opposés (qui peuvent être des équipes), le vérificateur et le falsificateur. Le jeu G(*S*) commence avec l'énoncé *S*, et se poursuit selon des règles qui, pas à pas, traitent chaque constante logique (connecteurs et quantificateurs). On remarquera que ces règles de vérification et de falsification, dont voici la liste, ont un caractère obvie, à condition qu'on soit au clair sur l'ambiguïté de ces notions de vérification et de falsification[4].

G ($S_1 \vee S_2$) : le vérificateur choisit S_i (i = 1 ou 2). Le jeu continue avec S_i.

G ($S_1 \wedge S_2$) : le falsificateur choisit S_i (i = 1 ou 2). Le jeu continue avec S_i.

G ($(\exists x)$ S[x]) : le vérificateur choisit un élément de do(M) ; si le nom de cet élément est b, le jeu se poursuit avec S[b].

G ($(\forall x)$ S[x]) : c'est le falsificateur qui choisit un élément de do(M) ; si le nom de cet élément est b, le jeu se poursuit avec S[b].

G ($\sim S$) : le jeu est identique à G(*S*), sauf que les deux joueurs ont échangé leur rôle.

[4] Il est évident qu'au sens usuel, dans « la vie réelle », on ne vérifie pas un énoncé en appliquant une règle sémantique, mais par des démarches scientifiques de preuve, de déductions, ou d'inférences inductives. Mais avant de savoir si tel énoncé est vrai, il faut bien avoir quelque idée de ce qu'est la vérité. C'est pourquoi Hintikka déclare que les procédures scientifiques de vérification et de falsification (par exemple d'hypothèses) sont « *parasitic* » des règles sémantiques (p. 33-36).

Règle pour les énoncés atomiques : si *A* est un énoncé atomique vrai, le vérificateur initial gagne et le falsificateur initial perd, et l'inverse si *A* est faux.

Au bout d'un nombre fini de pas, les constantes logiques sont éliminées, et on atteint une situation où la règle pour les énoncés atomiques s'applique. Un scrupule de circularité pourrait venir à l'esprit du fait que la vérité est mentionnée dans cette règle. Cependant, comme le fait remarquer Hintikka, la vérité ou la fausseté des formules atomiques est décidée dès lors que L est interprété sur un modèle M. Les autres règles ne font qu'étendre la notion de vérité aux autres énoncés du langage en question. Plus importante est la remarque selon laquelle l'analyse des quantificateurs selon GTS procède en sens inverse de la définition tarskienne de la vérité. Cette dernière part de l'intérieur des énoncés, où figurent des formules ouvertes, ce qui oblige à passer par la notion de satisfaction, pour procéder pas à pas vers l'extérieur par une définition récursive de la vérité (p. 29). Cette démarche présuppose le principe de compositionnalité (qui dit que les attributs sémantiques d'une expression complexe sont fonction des attributs sémantiques des expressions composantes). Selon Hintikka, c'est la confiance aveugle de Tarski en la compositionnalité qui l'a conduit à une définition récursive de la vérité, et à la procédure de « l'intérieur vers l'extérieur », qui oblige de plus la définition de la vérité à être préfacée par une définition de la satisfaction. Du coup, Hintikka est en devoir de montrer qu'une logique peut se passer du principe de compositionnalité (p. 106-110).

Revenons au point de vue de GTS. L'idée intuitive est que l'énoncé *S* est vrai si et seulement si le vérificateur initial a une stratégie gagnante, *i.e.* une stratégie qui lui assure le succès quels que soient les coups joués par son adversaire. Cette idée heuristique mène à une première définition de la vérité en termes de stratégie gagnante :

Def 1 : *S* est vrai dans le modèle M si et seulement s'il existe une stratégie gagnante pour le vérificateur dans le jeu G(*S*) joué sur M[5].

Et de même, *S* est faux dans M si et seulement s'il existe une stratégie gagnante pour le falsificateur dans le jeu G(*S*) joué sur M.

---

[5] Qu'il existe une stratégie gagnante n'implique pas nécessairement qu'elle soit connue du vérificateur. Mais il est possible d'ajouter des conditions restrictives en plus de l'existence en soi, de manière à obtenir une sémantique « game-theoretical » constructiviste. En quel sens et jusqu'à quel point, ces questions font l'objet du Chapitre 10 des *Principles* de Hintikka. Je ne poursuivrai pas ici ces questions.

Quels que soient les mérites de GTS, elle ne donne pas tout ce qu'on est en droit de demander. Elle ne satisfait pas en particulier l'exigence d'une définition de la vérité pour un langage interne à ce langage : le problème soulevé en 1 reste donc ouvert.

3. Un pas de plus est alors fait (Chapitre II, alinéa (iv)), qui consiste à analyser logiquement le concept de stratégie (on admettra désormais que les énoncés sont en forme normale pour la négation, i.e. que la négation est confinée à ne porter que sur des énoncés atomiques ou des identités ; mais ce point n'est nullement une restriction puisqu'il y a une procédure effective pour la mise en forme normale). Une stratégie pour le vérificateur est un ensemble fini de fonctions, dites fonctions de choix (ou fonctions de Skolem), dont les arguments sont les éléments du domaine du modèle M, et les valeurs les éléments de l'image de ce domaine par ces fonctions. L'existence ou la non-existence de telles fonctions peut être exprimée par un énoncé S^* du second ordre, qui peut être considéré comme la traduction de l'énoncé S exprimant sa condition de vérité. Pour prendre un exemple simple, soit S le schéma d'énoncé, où S (disons) ne contient pas de quantificateurs :

$$(\forall x)(\exists y)\ \mathrm{S}\ [x, y]$$

Sa traduction en second ordre est :

$$(\exists f)\ (\forall x)\ \mathrm{S}\ [x, f(x)].$$

De manière générale, S^* est obtenu à partir de S de la manière suivante :

(a) Soit $(\exists x)$ un quantificateur existentiel ayant une occurrence dans S dans la portée des quantificateurs universels $(\forall y_1)$, $(\forall y_2)$, …. $(\forall y_k)$. Remplacer chaque occurrence de x liée par $(\exists x)$ par $f(y_1, y_2, …., y_k)$, où f est une nouvelle variable de fonction, différente pour chaque quantificateur existentiel différent, et omettre $(\exists x)$.

(b) Soit $(S_1 \vee S_2)$ ayant une occurrence dans la portée de $(\forall y_1)$, $(\forall y_2)$, …. $(\forall y_k)$. Remplacer la disjonction par $((S_1 \wedge g(y_1, y_2, …. y_k) = 0) \vee (S_2 \wedge g(y_1, y_2, …. y_k) \neq 0))$, où g est pour chaque disjonction une nouvelle variable, différente des fonctions f mentionnées en (a)[6].

[6] Hintikka reconnaît que dans (b), qui paraît peu naturel, il y a un élargissement de la notion de fonction de Skolem, mais il ne s'attarde pas sur ce point.

(c) Préfixer les formules ainsi obtenues par $(\exists f_1)$ $(\exists f_2)$ … $(\exists g_1)$ $(\exists g_2)$ …, où f_1, f_2… sont les fonctions introduites en (a), et g_1, g_2… les fonctions introduites en (b).

Hintikka commente ainsi le résultat obtenu (p. 31) :

> « Le résultat S^* sera appelé la *traduction en second ordre* de S. Il exprime la condition de vérité de S en termes de théorie des jeux. Il établit comment la vérité de S est connectée avec les jeux sémantiques décrits plus haut ».

4. Après le mythe de la compositionnalité, Hintikka (Chapitre 3) débusque un autre préjugé qui nous vient de Frege aussi bien que de Russell : le mythe de la dépendance d'un quantificateur dans la portée d'autres quantificateurs à l'égard de ces derniers, ou plus exactement la limitation de la logique à l'étude de ce mode de dépendance. Se libérer de cette limitation, admettre l'indépendance relative des quantificateurs, ouvre la voie à une nouvelle logique du premier ordre, la logique IF (IF pour « independence-friendly », la logique qui admet l'indépendance (p. 50)). Le moyen privilégié pour extraire un quantificateur de la portée d'un autre quantificateur est la « *slash-notation* », /. Par exemple, dans :

$$(\forall x)(\exists y/\forall x)\ S\ [x, y],$$

$(\exists y)$ est indépendant de $(\forall x)$. Autrement dit, le choix par le vérificateur d'une valeur de y est fait dans l'ignorance de la valeur de x. Cet énoncé est donc équivalent à :

$$(\exists y)(\forall x)\ S\ [x, y].$$

Dans ce cas simple, la *slash notation* n'apporte rien de nouveau, puisque le dernier énoncé appartient à un langage ordinaire du premier ordre. Mais ce n'est pas toujours le cas. Considérons l'énoncé du second ordre :

$$(\exists f)(\exists g)(\forall x)(\forall z)\ S\ [x, f(x), z, g(z)]$$

Ici $f(x)$ ne dépend que de la valeur donnée par le falsificateur à x, et $g(z)$ de la valeur donnée à z. Grace à la *slash-notation*, cet énoncé peut être exprimé en premier ordre dans un langage IF :

$$(\forall x)(\forall z)\ (\exists y/\forall z)(\exists u/\forall x)\ S\ [x, y, z, u]^7$$

qui n'a pas d'équivalent dans la logique usuelle du premier ordre.

[7] Il y a d'autres moyens de noter l'indépendance des quantificateurs, par exemple avec la notation des « *branching quantifiers* » de Henkin. On peut exprimer le dernier énoncé par :

$$\left.\begin{matrix}(\forall x)\ (\exists y)\\ (\forall z)\ (\exists u)\end{matrix}\right\}\ S\ [x, y, z, u]$$

Il est à noter que la logique IF est plus forte que la logique usuelle du premier ordre. L'indépendance des quantificateurs signifie en termes d'information qu'une stratégie gagnante est fondée sur une moindre information quant aux arguments des fonctions qui composent une stratégie. Donc l'affirmation qu'il existe une telle stratégie est plus forte que l'existence d'une stratégie fondée sur plus d'information. En termes de capacités représentationnelles, la plus grande force de la logique IF tient à ce qu'elle peut isoler des classes de structures que la logique usuelle du premier ordre ne peut capturer (p 58-59).

Comme pour la logique usuelle du premier ordre, les traductions des énoncés d'un langage IF appartiennent au fragment Σ_1^1 du second ordre, dont tous les énoncés sont constitués d'une séquence éventuelle de quantificateurs existentiels du second ordre, suivie d'une formule du premier ordre (au sens où toutes les occurrences de variables du second ordre y sont libres). Mais pour la logique IF, la réciproque est également vraie : tout énoncé du fragment Σ_1^1 a une traduction (logiquement équivalente) dans un langage IF (p. 61). Mieux que la démonstration générale, un exemple aidera à saisir le point. Considérons le schéma de formule :

$$(\exists f)(\forall x_1)(\forall x_2)\ S\ [f(x_1), f(x_2)]$$

Ici la même fonction f est appliquée à deux (suites d') arguments possiblement distinct(e)s. Pour parvenir à la traduction dans un langage IF, il faut tout d'abord introduire une seconde fonction, chaque fonction ayant ses propres arguments. Ce qui donne :

$$(\exists f)\ (\exists g)\ (\forall x_1)\ (\forall x_2)\ (((x_1 = x_2) \supset (f(x_1) = g(x_2))) \wedge S\ [f(x_1), g(x_2)])$$

dont la traduction dans un langage IF est :

$$(\forall x_1)\ (\forall x_2)\ (\exists y_1/\forall x_2)\ (\exists y_2/\forall x_1)\ (((x_1 = x_2) \supset (y_1 = y_2)) \wedge S\ [y_1, y_2]),$$

puisque la fonction f ne s'applique qu'à l'argument x_1 indépendamment du quantificateur ($\forall x_2$), et que *ceteris paribus*, il en est de même de la fonction g, qui ne s'applique qu'à x_2[8].

La possibilité d'une traduction d'un énoncé du fragment Σ_1^1 dans un langage IF va se révéler cruciale quand il s'agira de montrer que le prédicat de vérité pour un langage IF peut être formulé dans ce même langage.

---

[8] Je résume, sans prendre en compte toutes les situations qui peuvent se présenter : fonctions qui s'appliquent à la valeur d'une autre fonction (« *nesting of functions* »), convention (ou non) selon laquelle les quantificateurs existentiels sont indépendants les uns des autres (p. 62-64). À mon sens, ces difficultés à surmonter montrent que la traduction du fragment Σ_1^1 dans IF est plus complexe et moins naturelle que la traduction dans l'autre sens.

En attendant d'en venir à ce point, qui est le but vers lequel tendent toutes ces considérations, notons un certain nombre de traits qui distinguent la logique IF de la logique usuelle du premier ordre (p. 65) :

(a) La logique ordinaire du premier ordre est un cas spécial de la logique IF, au sens où elle n'admet pas l'indépendance des quantificateurs les uns par rapport aux autres.

(b) La logique IF est une extension conservative de la logique du premier ordre, au sens où il n'y a pas d'énoncé du langage du premier ordre usuel qui ne soit démontrable (à supposer que cette notion soit définie ; Hintikka n'en dit rien) que *via* la logique IF.

(c) La logique IF n'admet pas une axiomatisation complète (l'ensemble des formules valides de la logique IF n'est pas récursivement énumérable)[9].

(d) La logique IF n'admet pas une définition de la vérité à la Tarski (en raison de l'échec de la compositionnalité, puisque les conditions de vérité d'une sous-formule dépendent en général de son contexte).

(e) La loi du Tiers Exclu n'est pas valide dans la logique IF (à vrai dire, ce point dépend de la définition de la vérité et de la fausseté (voir Def.1) selon GTS, où il est possible qu'aucun des joueurs n'ait une stratégie gagnante).

Pour chaque énoncé du premier ordre (IF ou non), sa traduction en second ordre exprime ses conditions de vérité en termes de stratégies (ensembles de fonctions de choix). La tâche qui reste à accomplir, on l'a dit, est d'intégrer ces conditions de vérité dans une unique définition générale de la vérité. Or, « il va se révéler, nonobstant Tarski, qu'une telle définition peut être donnée pour un langage IF représentatif dans ce langage lui-même » (p. 112) : c'est au Chapitre 6 que la preuve va en être donnée.

---

[9] Dans le langage de Hintikka, cela revient à dire qu'on a affaire à une logique du premier ordre sémantiquement incomplète (l'ensemble des vérités logiques n'est pas récursivement énumérable). Cette « situation conceptuelle » nouvelle due à l'existence de la logique IF contrevient à la leçon qu'on a cru pouvoir tirer des résultats négatifs de Gödel : l'incomplétude déductive de l'arithmétique élémentaire (il y a un énoncé *S* tel que ni *S*, ni sa négation ne sont prouvables) serait un fait mathématique plutôt que logique, et entraînerait l'incomplétude descriptive de l'arithmétique (l'impossibilité d'isoler les modèles attendus, et *a fortiori* la catégoricité, i.e. l'existence d'un seul modèle à un isomorphisme près). En revanche, si la logique sous-jacente est incomplète (sémantiquement), une théorie T peut être descriptivement complète sans l'être déductivement (c'est, si l'on peut dire, la « faute à la logique », qui ne permet pas de décider entre *S* et sa négation). Et selon Hintikka, ce qui compte du point de vue mathématique, c'est surtout la recherche de la complétude descriptive d'une théorie. De sorte que l'incomplétude (sémantique) de la logique IF serait la bienvenue (p. 88-96).

5. 1ère étape : on montre tout d'abord comment un prédicat de vérité pour un langage arithmétique usuel du premier ordre peut être formulé dans le langage arithmétique IF correspondant[10]. La syntaxe du langage objet peut être codée dans l'arithmétique qu'il contient, et les propriétés des formules peuvent être représentées par des prédicats « number-theoretical » dans le langage objet usuel comme dans le langage IF, puisqu'il est assumé que les deux contiennent l'arithmétique élémentaire. Le nombre de Gödel d'un énoncé S est désigné par $\ulcorner S \urcorner$, et le numéral qui désigne le nombre n sera écrit **n**. On commence par formuler le prédicat de vérité pour le langage usuel du premier ordre, non directement dans le langage IF, mais dans le fragment Σ^1_1 du second ordre. Après quoi, il n'y aura plus qu'à traduire ce prédicat dans le langage IF correspondant (on a vu que cette traduction en retour est toujours possible).

Le prédicat de vérité en second ordre a une signification évidente : il dit qu'appliqué au nombre de Gödel $y = \ulcorner S \urcorner$, il existe un prédicat monadique X qui se comporte comme un prédicat de vérité en termes de stratégie gagnante doit le faire, et que ce prédicat s'applique à y. En d'autres termes, le prédicat de vérité en second ordre est :

$$(\exists X)(TR[X] \wedge X(y)),$$

où TR[X] résume les « bonnes » propriétés de X :

$$(\forall x)(\forall y)(\forall z)\ ((x = \ulcorner(S_1 \wedge S_2)\urcorner \wedge y = \ulcorner S_1 \urcorner \wedge z = \ulcorner S_2 \urcorner) \supset (X(x) \supset (X(y) \wedge X(z))))$$

$$(\forall x)(\forall y)(\forall z)\ ((x = \ulcorner(S_1 \vee S_2)\urcorner \wedge y = \ulcorner S_1 \urcorner \wedge z = \ulcorner S_2 \urcorner) \supset (X(x) \supset (X(y) \vee X(z))))$$[11].

$$(\forall y)(\forall z)(\forall u)\ ((y = \ulcorner(\forall x)S[x]\urcorner \wedge u = \ulcorner S[\mathbf{z}]\urcorner \wedge X(y)) \supset X(u))$$

$$(\forall y)(\forall z)(\forall u)\ ((y = \ulcorner(\exists x)S[x]\urcorner \wedge u = \ulcorner S[\mathbf{z}]\urcorner \wedge X(y)) \supset X(u))$$

Si R est une relation primitive à deux places, $(\forall x)(\forall y)\ (X(R(\mathbf{x}, \mathbf{y})) \leftrightarrow R(x, y))$; et de même pour les autres prédicats primitifs.

$$(\forall x)(\forall y)\ ((x = \ulcorner(S_1)\urcorner \wedge y = \ulcorner S_2 \urcorner \wedge N(x, y)) \supset (X(x) \leftrightarrow (X(y))),$$

où $N(x, y)$ est la relation du nombre de Gödel d'un énoncé au nombre de Gödel de la forme normale pour la négation de cet énoncé.

La définition donnée ci-dessus « *does its job* », affirme Hintikka. La propriété d'être vrai (attribuée aux nombres de Gödel des énoncés du langage usuel du premier ordre) satisfait TR[X], et donc si $y = \ulcorner S \urcorner$ est vrai, y satisfait la définition. Inversement, si y satisfait

[10] « Correspondant » veut dire que les deux langages ont les mêmes primitifs non logiques (p. 114).
[11] Après la formule pour la conjonction, Hintikka se contente d'écrire « *likewise for disjunction* ». Je précise.

la définition, X se comporte comme le prédicat intuitif de vérité doit le faire (il y a une stratégie gagnante pour le vérificateur). Mais son affirmation selon laquelle le prédicat de vérité ainsi défini « n'est guère plus qu'une façon d'épeler [« *a way of spelling* »] de manière plus complète la définition de la vérité en termes de théorie des jeux », me semble devoir être prise *cum grano salis*. Il ne transparaît rien, dans la définition de TR[X], du fait que pour la conjonction comme pour le quantificateur universel, c'est l'adversaire (le falsificateur) qui fait le choix. Pourtant l'existence d'une stratégie gagnante pour le vérificateur suppose qu'il peut répondre positivement (*i.e.* continuer le jeu de manière à gagner) à ces choix arbitraires. Quoi qu'il en soit, Hintikka commente ainsi ce résultat :

> « La chose la plus remarquable concernant ce prédicat de vérité est qu'il est de la forme Σ_1^1. (Le seul quantificateur d'ordre supérieur y est le quantificateur existentiel initial du second ordre (∃X)). Il s'ensuit qu'il [le prédicat de vérité] peut être traduit dans le langage IF du premier ordre correspondant. Ainsi un prédicat de vérité pour un langage ordinaire du premier ordre donné peut être défini dans le langage IF du premier ordre correspondant ; en d'autres termes sans faire appel à des quantificateurs d'ordre supérieur. » (p. 116)

2ème et dernière étape[12] : il s'agit à présent de montrer que le prédicat de vérité peut être défini pour un langage IF dans ce même langage. L'essentiel consiste à enrichir TR[X] d'une clause capable de traiter les phénomènes d'indépendance des quantificateurs (sous la forme des quantificateurs dits « de Henkin » (selon la *slash-notation*). Elle peut être formulée ainsi :

$$(\exists x)(x = \ulcorner(\forall y)(\forall u)(\exists z/\forall u)(\exists t/\forall y)S[y, z, u, t]\urcorner \wedge X(x)) \supset$$
$$(\exists f)(\exists g)(\forall y)(\forall u)\ (\exists w)(w = \ulcorner S[y, f(y), u, g(u)]\urcorner \wedge X(w))^{13}$$

Hintikka commente ainsi ce qu'il advient de la définition du prédicat de vérité après ajout de cette clause (*i.e.* pour un langage objet IF) :

> « La principale différence est que la définition [il s'agit toujours de (∃X)(TR[X] ∧ X(*y*))] ne contient plus un unique quantificateur initial du second ordre (∃X). Des quantificateurs du second ordre peuvent aussi être introduits par (g)[14], comme le

[12] Pour ne pas obscurcir la ligne générale de la démonstration, je néglige la généralisation (p. 116) de la définition du prédicat de vérité pour des langages qui, bien que contenant l'arithmétique, ne sont pas purement arithmétiques. Le problème provient du fait qu'on n'a pas forcément un nom pour chaque individu du domaine du modèle en question.

[13] Le texte écrit « X(*x*) » dans le conséquent, ce qui est visiblement une erreur typographique. Le conséquent affirme, sous la condition exprimée par l'antécédent, l'existence d'un nombre de Gödel *w* d'un certain énoncé, tel que *w* a la propriété X (tel que *w* est vrai).

[14] Je suppose que « (g) », dont c'est la première occurrence, renvoie à la clause qui vient d'être introduite.

> montrent les quantificateurs (∃*f*) et (∃*g*) ici. Mais quand ces quantificateurs sont déplacés en position initiale [*i.e.* l'énoncé est mis en forme prénexe], ils restent existentiels. Donc la définition dans son ensemble est de la forme Σ_1^1. Donc elle peut être retraduite dans le langage IF du premier ordre originel. En un mot, il s'ensuit que le prédicat de vérité pour le modèle donné d'un langage IF du premier ordre est exprimable dans le langage IF du premier ordre lui-même. » (p. 118)

On a enfin atteint le but recherché, et levé l'interdit (ou la « malédiction ») de Tarski. Un prédicat de vérité pour un langage peut être défini dans ce même langage. Voilà qui devrait suffire à justifier la logique IF : elle met fin au mouvement de pensée qui conduit à la crainte de l'ineffabilité de la sémantique, et partant à l'impossibilité de la théorie des modèles.

6. Il y a quand même ici matière à doute, et d'ailleurs Hintikka s'adresse à lui-même l'objection, – avant naturellement de la lever. Présentée en second ordre, la définition de la vérité est « aussi claire qu'on peut le souhaiter ». Pourtant le résultat de sa traduction dans la logique IF du premier ordre est « un énoncé complexe et difficile, tel que je n'ai même pas eu le courage de tenter de l'écrire explicitement. […] Pourquoi cette préférence intuitive pour la formulation en second ordre ? » (p. 122). Réponse : parce que la définition de la vérité en termes de stratégie gagnante nous conduit naturellement à parler en termes (d'ensembles finis) de fonctions, et donc en second ordre (dans Σ_1^1). Mais cela ne change rien à la situation théorique, affirme Hintikka, car un langage Σ_1^1 peut en principe être pensé comme une variante notationnelle, une abréviation, du langage IF du premier ordre correspondant (p. 123).

Variante pour variante, on pourrait aussi bien renverser la situation, et tenir le langage IF pour une variante notationnelle d'un langage Σ_1^1. De ce point de vue, la définition du prédicat de vérité pourrait être tenue pour essentiellement du second ordre. En désespoir de cause, Hintikka en appelle à la distinction aristotélicienne entre ce qui est premier *en soi* (le langage IF), et ce qui est premier *pour nous* (le fragment Σ_1^1). Mais comment harmoniser cette distinction avec les innombrables passages où Hintikka affirme que la logique IF est notre vraie logique élémentaire (p. 50), notre logique de base (p. 74)), « *the basic area of our logic* » (p. 88), notre logique naturelle ? Il y a là sinon une contradiction en forme, du moins une discordance.