UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RECHERCHES LOGIQUES ET PHILOSOPHIQUES SUR LE CONCEPT DE MÉTALANGAGE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR
NEIL KENNEDY

AOÛT 2006

REMERCIEMENTS

Si l'on en croit la page titre de cet ouvrage, son aboutissement est l'affaire d'une seule personne, en l'occurrence moi, Neil Kennedy. Mais je dois admettre, en toute humilité, que cette page biaise l'attribution du mérite en ma faveur, et il convient en cette seconde page de redistribuer les honneurs aux autres acteurs méritants ayant participé à la réalisation du projet.

D'abord, puisqu'il faut bien se nourrir, se loger et se vêtir lorsque nous étudions, je tiens à remercier le Fonds québécois de la recherche sur la société et la culture (FQRSC) pour son soutien financier tout au long de ma maîtrise.

Je tiens à remercier, par ailleurs, Alain Voizard et Serge Robert pour leur révision minutieuse des versions préliminaires de ce mémoire et pour les commentaires constructifs qui en ont résulté.

Mille mercis à mon directeur Mathieu Marion pour m'avoir orienté, conseillé et encadré dans toutes les sphères relatives à ma formation académique.

Reconnaissance infinie à mes parents, lesquels ont toujours su m'encourager dans la poursuite interminable de mes études.

Merci à Anouk, pour son soutien, particulièrement dans les moments difficiles.

Merci à vous tous et à tous ceux que j'oublie.

Comme disait le *King* : « Thank yuh, thank yuh very much ».

TABLES DES MATIÈRES

RÉSUMÉ

Ce mémoire a pour objectif principal l'analyse du concept de métalangage tel qu'il s'est développé en logique mathématique. L'introduction et la conclusion mises à part, chaque chapitre porte sur un auteur – logicien, mathématicien ou philosophe – ayant contribué de manière significative à l'évolution de ce concept. Ces auteurs sont, en ordre de présentation, Gottlob Frege, Bertrand Russell, Ludwig Wittgenstein, David Hilbert, Kurt Gödel et Alfed Tarski. Puisque la notion de métalangage s'est développée avec la formalisation progressive de la logique, une attention particulière est accordée à l'émergence des systèmes formels et à leur présentation. Trois périodes se dessinent dans la genèse de cette notion. Une première, que j'appelle « pré-météthéorique », où l'intervention d'une théorie externe au langage formel est rejetée catégoriquement, mais où certaines notions métathéoriques sont implicitement tracées. Une seconde, dite « hilbertienne », qui marque l'entrée en jeu de la métamathématique et qui consacre le métalangage dans l'étude des mathématiques, quoiqu'avec des moyens limités. Et une troisième, dite « tarskienne », où la notion moderne de métalangage est exposée. Par ailleurs, j'effectue une analyse détaillée de la preuve que Gödel donne de son second théorème d'incomplétude où je prétends qu'il commet une erreur conceptuelle entre langage et métalangage. Enfin, en conclusion, j'explore une conception fondationnelle de la logique compatible avec l'étude métathéorique.

Mots clés :
métalangage, logique, philosophie, métamathématique, Gödel, Tarski

INTRODUCTION

Historiquement, l'évolution du concept de métalangage en logique mathématique couvre une période qui s'étale sur plusieurs décennies, allant de la fin du dix-neuvième siècle jusqu'à la fin des années trente. Parmi les logiciens, mathématiciens et philosophes qui ont contribué de manière significative à la genèse de ce concept, nous retrouvons Frege, Russell, Wittgenstein, Hilbert, Gödel et Tarski. Chacun d'eux, avec la philosophie mathématique qu'ils incarnent, feront l'objet d'un chapitre détaillé dans le cadre de cet ouvrage.

Ici, comme ailleurs en philosophie de la logique, l'œuvre de Gottlob Frege s'est avéré incontournable. Il faut mentionner d'emblée que Frege, comme tous les logicistes, n'a pas de métalangage annoncé. Toute son œuvre consiste à construire un langage qui est conceptuellement autonome et en ce sens, faire appel à un métalangage marquerait l'échec de cette autonomie. D'un autre côté, l'introduction d'un nouveau langage se fait forcément à l'aide d'un autre langage, qui nous est familier, sans quoi nous n'arriverions pas à comprendre la signification des nouveaux termes introduits. Le langage de présentation, comme il se nomme dans l'œuvre de Frege, n'est pas censé être une structure permanente; il est conçu comme un outil pédagogique de transition, voué à disparaître en fin de parcours une fois que nous aurons maîtrisé le nouveau langage formalisé. Toutefois, dans le rapport conceptuel qu'il y a entre le langage de présentation et le langage formalisé, il se trace de nombreuses distinctions qui seront l'apanage de la relation entre métalangage et langage objet, et c'est pour cette raison que Frege est sans conteste un point de départ pour une analyse du concept de métalangage.

La première et la plus importante distinction métathéorique établie par Frege est sans doute celle entre l'usage et la mention d'un signe, qui est effectuée en pratique au moyen de guillemets. Ce stratagème donne à Frege un moyen d'expression puissant qui lui permet d'adopter, au besoin, une posture métathéorique à l'endroit du langage qu'il introduit. Ce besoin est particulièrement manifeste lorsqu'il est question d'associer une référence aux éléments du langage, une tâche qui exige absolument qu'on puisse considérer les expressions du langage sur le plan de la mention et non pas de l'usage. Lorsqu'on dit que 🚹 dénote la toilette des hommes, « 🚹 » est un symbole et non pas la toilette elle-même.[1] Bien qu'il ait formulé l'ensemble des règles syntaxiques nécessaires à la manipulation des propositions, Frege tient néanmoins à fixer rigoureusement la dénotation (en termes de valeurs de vérité) de chacune d'entre elles d'une manière qui est non sans rappeler la définition tarskienne de la vérité.[2] Par exemple, selon les stipulations sémantiques de la *Begriffsschrift*, la proposition $\mathcal{A} \supset \mathcal{B}$ dénote le faux seulement lorsque $\mathcal{A}$ dénote le vrai et $\mathcal{B}$ le faux, autrement elle dénote le vrai.

Si Frege n'est pas cité pour ses conceptions métathéoriques dans la littérature, c'est en raison de ses opinions sur son propre langage de présentation. Le langage de présentation est réduit à un appareil auxiliaire qui est jugé inessentiel à la réussite de l'entreprise dans son ensemble. Il y a deux explications, à mon sens, pour cette attitude. La première est qu'il veut éviter que son langage formel ait une dépendance conceptuelle envers le langage de présentation. Une telle dépendance serait vraisemblablement jugée comme un défaut : d'une part, parce qu'elle condamnerait le

---

[1] Que le lecteur en soit averti!

[2] La définition tarskienne de la vérité est évidemment postérieure aux travaux de Frege.

projet à la circularité ou à la régression à l'infini et, d'autre part, parce qu'elle serait une source potentielle de contradictions.[3] La deuxième explication tient à ce que le langage de présentation n'est pas le siège de la référence des symboles du langage conceptuel. La dénotation des termes du langage conceptuel se trouve dans la réalité « logique » telle qu'elle est postulée par les tenants du réalisme ou du platonisme mathématique. La postulation d'une réalité logique occulte donc le rôle joué par le métalangage.

Bertrand Russell partage sensiblement les mêmes convictions philosophiques que Frege en matière de fondements des mathématiques. Comme Frege, il vise l'élaboration d'un langage dépouillé des ambiguïtés du langage ordinaire, conçu sur mesure pour épouser la réalité logique et pour reproduire l'ensemble des mathématiques dans une formulation rigoureuse et exhaustive. Contrairement à Frege, Russell eut l'honneur de mettre au monde avec Whitehead le premier système cohérent pouvant réduire les mathématiques à la logique : les *Principia Mathematica*. L'exposition des *Principia* a nécessité, comme ce fut le cas pour la *Begriffsschrift*, d'un langage de présentation, mais Russell y accordera un statut temporaire et inessentiel au projet qu'il cherche à réaliser. Comme nous le verrons, la posture anti-métathéorique demeure la même.

D'importantes différences subsistent toutefois entre les conceptions métathéoriques – embryonnaires – de Frege et de Russell. On ne retrouvera pas dans le langage de présentation des *Principia* toutes les distinctions métathéoriques qu'on pouvait retrouver dans l'exposition de la *Begriffsschrift*. Russell est beaucoup moins scrupuleux sur cet aspect que Frege. La signification du symbolisme est présentée dans des discussions informelles, et parfois la frontière entre le symbolisme lui-même et le langage de présentation n'est pas claire. Sur des questions métathéoriques comme la vérité, nous verrons que Russell est à la fois incohérent et indécis, oscillant

[3] N'oublions pas qu'une des motivations pour développer le langage conceptuel formel est d'échapper aux ambiguïtés du langage ordinaire, celui-là même qui est jugé responsable d'introduire des antinomies dans la logique et les mathématiques du temps de Frege.

entre plusieurs définitions incompatibles de cette notion. En bout de ligne, le lecteur est amené à comprendre que la réalité logique est la référence et la motivation ultimes du symbolisme. Le langage de présentation est donc inessentiel et temporaire, il met en rapport le langage formel avec le monde idéel et, une fois accompli, on peut l'éliminer. C'est étonnant de voir Russell justifier les règles des *Principia* de la même manière que Frege le fit pour les lois des *Grundgesetze*, considérant qu'il a lui-même montré l'insuffisance de cette méthode en exhibant une contradiction dans le système frégéen. Il semblerait que même si la réalité logique est garante de la fiabilité de nos lois, notre accès partiel à cette réalité (un accès qui est forcément partiel car comment expliquerions-nous l'erreur sinon?) nous empêche de tirer des conclusions finales sur leur validité. Russell ne semble pas ignorer ce problème, comme en témoigne de nombreux passages où il laisse entendre que la validité d'un système est jugé sur sa capacité à répondre aux attentes logiques tout en évitant les contradictions (deux critères qui sont non sans rappeler la complétude d'un côté et la cohérence de l'autre). Bien qu'il pose cette question, toute métathéorique soit elle, il refuse qu'on y réponde autrement que par essai et erreur, c'est-à-dire de manière empirique. Tout autre procédé serait circulaire ou régressif.

Nous aboutissons donc à une certaine difficulté : d'un côté, l'échec des *Grundgesetze* nous montre que les motivations intuitives ou intellectuelles ne sont pas suffisantes pour cautionner un formalisme; d'un autre côté, des considérations philosophiques logicistes nous empêchent d'étudier ce formalisme avec toute la puissance des moyens logiques et arithmétiques pour déterminer s'il est complet ou non contradictoire. D'une certaine manière, il est vrai que le recours à une théorie, arithmétique ou autre, compromettrait la démarche fondationaliste envisagée par Russell. Fonder consiste à poser des assises sûres et certaines qui ne demandent aucune autre justification, aucune autre assise. Le problème est qu'il est bien difficile de théoriser sur la logique sans disposer d'une théorie, qu'il s'agisse seulement de la discussion philosophique ou de l'arithmétique élémentaire, et il semblerait que la plupart des tentatives fondationalistes ont connu les mêmes difficultés.

Malgré l'immense succès des *Principia*, le statut questionnable de certains de ses axiomes, dont principalement l'axiome de réductibilité, affaiblie considérablement le réalisme logique. Le logicisme frégéo-russellien est justifié par l'appel à une réalité essentielle composée d'objets et de relations (à caractère essentiel aussi).[4] La *Begriffsschrift* et les *Principia* sont des langages qui retirent leur universalité de cette réalité. Dans un effort de réforme magistrale, Wittgenstein tente dans le *Tractatus* de sauver l'universalisme en logique en concevant le rapport entre le langage et le monde sous un autre jour. Chez lui la logique n'est plus la description d'un monde essentiel, elle est la forme de représentation du langage.

Pour faire des approximations radicales, dans la philosophie du *Tractatus* il n'y a qu'un monde et ce monde est composé d'état de choses. Une représentation, qui est également un état de choses (donc un élément du monde), représente en vertu d'une corrélation de forme de représentation. La métaphore de la projection est souvent employée pour illustrer l'idée : c'est par une sorte de relation de projection du représenté vers la représentation que le représentation représente. La partition de la neuvième de Beethoven est une représentation de la neuvième parce qu'elle partage une forme de représentation en commun avec la neuvième, de même, les gravures sur un vinyle de la neuvième représente également la neuvième en vertu de cette forme qui leur est commune; ce sont deux exemples de projection de la neuvième. Une représentation peut représenter tout état de choses qui possède une forme commune avec elle. Certaines représentations peuvent représenter plusieurs états de choses, d'autres moins; aux extrémités de ces pôles on retrouve la forme purement logique, celle qui est commune à toutes les représentations, la tautologie, et celle qui est commune à aucune d'entre elles, la contradiction. La logique est donc la syntaxe du langage, la syntaxe de nos représentations.

Wittgenstein développe par ailleurs dans le *Tractatus* une philosophie de la limitation du pouvoir représentationnelle des représentations qui s'articule

[4] Certains logicistes, dont Carnap, ne font pas appel à la réalité essentielle pour justifier leur approche.

essentiellement autour de la distinction entre dire et montrer. Parmi les choses qui sont notoirement indicibles, on retrouve la syntaxe du langage, les limites du monde et les questions métaphysiques en général. Naturellement, l'indicibilité de la syntaxe est ce qui retiendra notre attention. Dans la philosophie wittgensteinienne, la logique ne peut pas être exprimée, elle peut seulement être montrée. En fonction de cette analyse, il dénonce les explicitations que Frege et Russell donnent de leurs formalismes respectifs comme un non-sens. Nous n'avons pas à *dire* par des lois que telle proposition logique découle de telle autre, la forme logique de ces propositions (une proposition est une représentation) devrait le *montrer*. Ainsi, Wittgenstein considère les stipulations métathéoriques comme des transgressions au pouvoir expressif des propositions, ce sont des confusions. Pour mettre en évidence la syntaxe, il faut non pas la dire par une métathéorie, il faut la montrer à l'aide d'une notation conceptuelle bien choisie.

Nous pouvons concevoir la distinction entre dire et montrer comme un renforcement de la distinction entre l'usage et la mention. Dans la philosophie du langage de Frege, il n'y a pas de contradiction dans le fait qu'un symbole puisse être à la fois utilisé et mentionné. Mon exemple de « 🚹 » donnée plus haut en témoigne. Pour Wittgenstein, dire et montrer sont incompatibles; une forme de représentation se montre mais ne se dit pas. La distinction entre dire et montrer est donc absolue au sens où il n'existe aucune manière de représenter la catégorie d'entités strictement montrables. Il y a un seul langage, ce langage a des limites expressives et la catégorie des entités strictement montrables est au-delà de ces limites.

Curieusement, c'est Russell qui répondra à Wittgenstein sur cette question en posant les bases d'une conception générale du métalangage. Pour échapper au sort de l'indicible, résultat inéluctable de la distinction entre dire et montrer, et pour parer les erreurs philosophiques qu'il avait pu faire en exposant les *Principia*, Russell fait l'hypothèse d'une hiérarchie de langages $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \ldots$ où chaque langage de la hiérarchie permet d'exprimer ce que son prédécesseur ne peut que montrer ($\mathcal{L}_{n+1}$

permet de dire ce que $\mathcal{L}_n$ ne peut que montrer). Russell prend donc au sérieux l'idée d'un métalangage.

Frege, Russell et Wittgenstein participent à ce que j'appelle la période pré-métathéorique. Ils partagent tous l'ambition de réaliser leur programme fondationaliste de manière anhypothétique, c'est-à-dire de concevoir, de présenter et d'évaluer un nouveau langage sans avoir à s'exprimer dans un autre ou, à tout le moins, à prendre l'activité de présentation au sérieux. Fonder est donc pris ici dans son sens le plus traditionnel en philosophie. Dans ce sens traditionnel, il est clair que la présupposition d'un métalangage mène à des problèmes philosophiques classiques : ou bien le langage $\mathcal{L}_0$ présuppose le langage $\mathcal{L}_1$ qui à son tour présuppose le langage $\mathcal{L}_2$ qui à ton tour en présuppose un autre, etc., et nous sommes condamnés à une régression à l'infini, ou bien le langage $\mathcal{L}$ présuppose le langage $\mathcal{L}'$ qui lui présuppose le langage $\mathcal{L}$ et nous tournons en rond. Mais au lieu de mettre le métalangage en cause, nous pourrions questionner le sens traditionnel (et platonicien dans une certaine mesure) de fonder. La métamathématique hilbertienne apporte justement une réforme au concept de fondation pour permettre une activité métathéorique plus abondante.

Hilbert ressuscite une forme de kantisme qui l'autorisera à employer des moyens métathéoriques que les logicistes s'étaient refusés. Ces moyens découlent d'une faculté arithmétique qui est donnée à la raison *a priori*. L'existence de cette faculté est déduite *à la Kant*, arguant qu'elle seule peut rendre l'arithmétique possible. La raison peut donc compter sur une forme d'arithmétique élémentaire dans ses recherches sur les fondements mathématiques. Jusqu'où vont les moyens de cette arithmétique élémentaire (contentuelle) à la disposition de notre raison est une question qui a des réponses variables dans l'œuvre de Hilbert; dans l'usage, il s'agirait de l'arithmétique (informelle) de Peano. Hilbert motive l'évidence intuitive de cette arithmétique en concevant un nombre comme une succession de symboles. Par exemple, si 1 est le symbole « ☝ », 7 serait le symbole « ☝☝☝☝☝☝☝ »; la somme du symbole « ☝☝☝ » (le nombre 3) avec le symbole « ☝☝☝☝☝ » (le nombre 5) est tout simplement la concaténation « ☝☝☝☝☝☝☝☝ » de ces symboles (le nombre 8); le nombre *n* est plus petit que le nombre *m* si *n* est un segment initial de *m*; le nombre *n* est égal au nombre *m* si en enlevant successivement un « ☝ » à chacun d'eux, nous arrivons au symbole vide en même temps; etc. Hilbert prétend même que le principe d'induction est intuitivement évident à partir de cette façon de concevoir les nombres.

Si Hilbert prend pour acquis l'arithmétique élémentaire, ce n'est pas pour achever la question des fondements, c'est plutôt pour développer une approche générale des fondements pour des théories qui pourraient difficilement être motivées par des considérations intuitives ou philosophiques, l'exemple classique étant l'infini, sous toutes ses formes. Comment justifier l'infini actuel ou l'infiniment petit? Hilbert abandonne toute forme de justification positive pour ce type d'entités, il propose une forme de justification instrumentaliste : si l'entité est non contradictoire, elle est admissible. Parler d'une entité de cette manière induit en erreur, car en réalité nous n'en avons pas une intuition directe. Parler d'une entité idéelle revient à poser des axiomes qui stipulent son comportement. Puisque les axiomes sont essentiellement des successions de symboles (ce sont des expressions d'un langage), il est donc possible d'appliquer la « raison » arithmétique à ces symboles pour déterminer, par des preuves rigoureuses, si des contradictions peuvent en résulter. Ainsi, démontrer l'irrationalité de $\sqrt{2}$ est essentiellement la même activité que démontrer un résultat métamathématique sur une théorie axiomatisée, l'unique différence entre les deux étant la forme et l'arrangement des symboles. Dans cette perspective, l'arithmétique élémentaire (contentuelle) constitue la seule et unique métathéorie, et les théories diffèrent seulement par leurs symboles, c'est-à-dire qu'elles sont entièrement caractérisées par leur syntaxe.

La recherche d'une preuve métamathématique de non contradiction pour l'arithmétique du premier et du second ordre (l'arithmétique et l'analyse réelle) reçut l'appellation de *programme de Hilbert*. Plusieurs éminents logiciens et mathématiciens y ont participé, le plus célèbre étant Kurt Gödel. Gödel amène ce programme à son point d'aboutissement avec le concept d'arithmétisation de la métamathématique. Par un stratagème ingénieux, il associe un nombre à chaque expression d'une théorie de sorte que les propriétés syntaxiques de l'expression soient lisibles à partir de la décomposition en facteurs premiers du nombre. Il est donc littéralement possible de transformer une analyse métathéorique en analyse arithmétique. L'arithmétisation est devenue l'outil indispensable du métamathématicien, mais ironiquement le premier usage que Gödel en fit a été de démontrer, d'une part, que l'arithmétique formelle souffrait d'une incomplétude « essentielle » et, d'autre part, que la non contradiction de l'arithmétique était indémontrable dans l'arithmétique. Ces deux théorèmes sont mieux connus comme étant le premier et le second théorème d'incomplétude de Gödel.

Le corps du chapitre V est consacré à exposer la démonstration originale de Gödel ainsi qu'une démonstration moderne de ses résultats. Les raisons de ce dédoublement sont multiples. D'une part, considérant que Gödel démontre ses résultats pour une extension des *Principia*, une théorie avec laquelle nous sommes peu familiers de nos jours, il m'a semblé intéressant de revoir la preuve pour l'arithmétique de Peano du premier ordre, d'autant plus que les arguments se sont affinés avec le temps. Mais surtout, j'expose dans ce chapitre une erreur dans l'argumentation de Gödel et je tenais à montrer que les théorèmes demeuraient néanmoins valides. L'erreur en question survient dans la preuve du second théorème d'incomplétude et a des conséquences graves, à un point tel qu'une preuve valide suivant les lignes directrices de l'argumentation de Gödel ne semble pas possible.

Quoiqu'il en soit de l'argumentation de Gödel, l'incomplétude a un impact important pour le programme de Hilbert. Le second théorème d'incomplétude démontre l'insuffisance des moyens arithmétiques élémentaires pour démontrer la non contradiction de l'arithmétique ou de l'analyse. La métamathématique dans sa formulation finitiste est donc sérieusement compromise. Naturellement, la solution la plus simple consiste à se donner de plus amples mesures dans notre métathéorie, mais malheureusement la philosophie mathématique néo-kantienne de Hilbert n'y survit pas. Les avantages d'une métamathématique élargie n'ont pas tardé : en éliminant la contrainte finitiste, Gentzen parvient à donner une preuve de non contradiction de l'arithmétique en utilisant l'induction transfinie jusqu'à l'ordinal ε_0 (l'ordinal limite des ordinaux dénombrables).

La morale du second théorème d'incomplétude est donc que les méthodes finitistes ne suffisent pas, ce qui nous amène vers un nouveau paradigme métathéorique dans l'ère post-hilbertienne : le métalangage tarskien. Tarski marque un point important dans la transition vers la notion métathéorique contemporaine. Il est le premier à formaliser la notion de métalangage et à considérer que le métalangage est relatif à un langage objet. Ainsi, le fait qu'une théorie donnée soit une théorie objet ou une métathéorie dépend du contexte et de l'utilisation que nous en faisons. La distinction entre métalangage et langage objet est pour Tarski non seulement fertile mais nécessaire; un langage qui ne l'observe pas risque d'être contradictoire. On ne s'étonnera donc pas qu'elle soit encore en service aujourd'hui dans des termes qui sont essentiellement les mêmes.

L'évolution du concept de métalangage semble donc converger vers une philosophie mathématique formaliste, un peu comme elle a été défendue par Carnap et von Neumann. Mais la position n'est pas sans difficultés, notamment parce que le formalisme n'est pas tant une thèse philosophique que l'absence d'une telle thèse. Cette perspective sera discutée en conclusion.

Quelques remarques sur le texte lui-même, ou plutôt les omissions dont il se rend coupable. Il est clair que de nombreux autres intervenants auraient pu être cités dans ce travail. Les intuitionnistes, entre autres, ne sont pas abordés dans cette analyse. J'ai essayé de me concentrer sur les auteurs qui ont contribué de manière significative à la question, mais il est possible que les intuitions et contributions de certains aient été négligées de ma part.

CHAPITRE I

Frege et l'émergence de la logique mathématique

0. Introduction

L'histoire du concept de métalangage est intrinsèquement liée à celle de la formalisation du langage. Avant l'apparition des langages formels, le seul modèle de langage était le langage ordinaire. Par la souplesse et la malléabilité de ses capacités expressives, ce type de langage a la remarquable capacité de parler de lui-même sans « sortir » de lui-même. En témoigne notoirement le fait que la grammaire d'un langage ordinaire est typiquement écrite dans ce même langage. Mais en matière de rigueur et de précision, le langage ordinaire laisse parfois à désirer; ses meilleures qualités font ses pires défauts. L'idée de concevoir un langage avec une syntaxe fixée et une sémantique rigoureuse fit donc son chemin. Avec un tel langage, nous quittons le confort d'une expressivité universelle et illimitée, et par conséquent l'idée d'une instance théorique extérieure à ce langage prend forme. La *Begriffsschrift* de Frege nous amène clairement dans l'ère du langage formalisé et perfectionné pour un usage spécifique. L'objectif de chapitre sera d'explorer le projet de formalisation de Frege et de voir son incidence sur le développement de certaines notions métathéoriques.

1. Le projet de l'idéographie

Gottlob Frege est considéré comme étant l'une des figures les plus importantes, voire la plus important, du développement de la logique moderne. Son œuvre aborde rigoureusement et analytiquement un ensemble de problèmes fondamentaux en philosophie et en logique, allant de la sémantique des énoncés à la formalisation de la déduction. Les innovations qu'il lègue à la postérité sont nombreuses et, malgré le peu de reconnaissance que Frege a eu de son vivant, elles marqueront le paysage logique (mathématique et philosophique) pour des générations à venir, autant celle de Russell et Wittgenstein que celle de Dummett. Au chapitre des innovations, on lui doit notamment : une conception vérifonctionnelle du calcul propositionnel, une théorie complète de la quantification logique, la réduction du rapport entre sujet et prédicat à celui entre objet et fonction – qui consacre d'ailleurs son divorce avec la logique aristotélicienne –, une définition de suite et de nombre qui ne s'appuie pas sur l'intuition, et (surtout) une notation conceptuelle où sont spécifiées toutes les étapes et les hypothèses sous jacentes à une déduction valide.

Ce travail s'étend sur plusieurs œuvres : Frege nous introduit à son formalisme dans la *Begriffsschrift* (1879), élabore la position logiciste et sa définition de nombre dans les *Grundlagen der Aritmetik* (1884) et parfait son système formel dans les *Grundgesetze der Arithmetik* (1893-1902), sans compter de nombreux articles qui clarifient certaines de ces positions dont les plus importants sont *Fonction et concept* (1891), *Sens et dénotation* (1892) et *Concept et objet* (1892).

La *Begriffsschrift*, ou l'idéographie,[1] est le point de départ de l'entreprise de Frege, dont l'objectif, nous dit-il, est de fournir une fondation immuable au raisonnement logique et, plus précisément, aux propositions de l'arithmétique. Afin de traiter les questions arithmétiques avec la rigueur qu'elles méritent, Frege veut remplacer les définitions psychologiques ou empiriques de la notion de suite – sur laquelle il compte fonder les nombres naturels – par une notion purement formelle, celle de conséquence logique. C'est en travaillant à cette tâche qu'il constate l'inaptitude et l'imprécision du langage ordinaire :

> [...] j'ai trouvé que l'inaptitude du langage constituait un obstacle; peu importe la longueur et la complexité des expressions utilisées, j'étais de moins en moins capable, au fur et à mesure que les relations devenaient de plus en plus complexes, d'atteindre la précision que la réalisation de mon objectif demandait. Cette déficience m'a amené à la présente idéographie. Son but principal, par conséquent, est de nous fournir une méthode fiable pour juger de la validité d'une chaîne d'inférences et pour expliciter toute présupposition qui tenterait de s'y glisser inaperçue, pour que son origine soit examinée. (van Heijenoort 1967 : p. 5-6, ma traduction)

Cette quête de la certitude cartésienne exige donc au préalable qu'on élabore un langage conceptuel construit sur mesure pour exposer, sans ambiguïtés et sans prémisses cachées, l'architecture d'une inférence valide.

Le langage ordinaire comporte en lui trop d'imprécisions qui rendent parfois l'expression de la pensée difficile. Frege conçoit le langage de la *Begriffsschrift* comme un outil qui multiplie notre acuité logique. De la même manière que le télescope de Galilée permit à celui-ci d'observer les lunes de Jupiter, des satellites jusque-là inobservables à l'oeil nu, l'idéographie permet une meilleure saisie des enchaînements d'une inférence logique et peut exposer ainsi certaines subtilités

[1] J'emploierai au cours de ce texte une terminologie qui, je l'espère, ne confondra pas le lecteur. Le terme « Begriffschrift » autant que le terme « idéographie » référeront tantôt à l'œuvre écrite en 1879 tantôt au système formel qu'il développe tout au long de son œuvre. Le contexte devrait rendre claire l'utilisation de ces mots.

logiques qui étaient ignorés auparavant. Même le mathématicien Cauchy, à qui l'on doit notamment une systématisation et une organisation accrues de l'analyse complexe, n'avait pas vu certaines subtilités importantes concernant l'ordre des quantificateurs, un ordre qui pouvait passer inaperçu dans le langage ordinaire. En effet, Cauchy n'avait pas vu la nuance, dans certains contextes, entre les locutions « pour tout x, il existe un y » et « il existe un x pour tout y ». Or, en mathématiques, il se cache derrière cette subtilité la différence – fondamentale pour la théorie – entre convergence simple et convergence uniforme. Tel que l'avait pressenti Leibniz, la science doit une partie importante de ses progrès à l'amélioration du langage dans lequel elle s'exprime.

Dans *De Arte Combinatoria*, Leibniz conçoit l'idée (bien avant Frege) de construire un langage universel sur mesure, une *characteristica universalis*, pour exprimer la pensée sans l'obscurité et les confusions du langage ordinaire. Dans ce langage, chaque phonème représenterait une pensée simple et l'on formerait des pensées complexes par la réunion de pensées simples. Frege rend hommage à Leibniz pour avoir vu l'importance du langage dans l'entreprise de la science, mais il considère que dans sa forme générale ce projet est irréalisable car trop imposant. Il croit néanmoins que le développement d'une caractéristique universel est ce dont a besoin les mathématiques pour ses fondements, mais cette caractéristique il veut la construire seulement pour les mathématiques, comme on a pu faire pour la physique avec le calcul différentiel et intégral.[2]

Si le besoin de cette caractéristique s'est fait sentir, selon Frege, c'est dû à un certain départ de la rigueur euclidienne. Avec l'arrivée de l'analyse et l'introduction de la notion de limite, de l'infiniment petit et des séries infinies, un laxisme s'est introduit dans les procédés de démonstrations qui est venu compromettre le statut des mathématiques comme paradigme de la certitude et de la rigueur scientifique. Si par cette rigueur vacillante les mathématiques ne peuvent plus faire figure de canon pour

[2] Frege mentionne à plusieurs endroits la notion de caractéristique telle que défendue par Leibniz, notamment dans l'introduction de la *Begriffsschrift* et dans *Sur le but de l'idéographie*.

la science, sur quoi pouvons-nous alors tabler pour atteindre une quelconque certitude en matière de connaissance? Il en va donc de la certitude de la connaissance dans son ensemble de développer un langage formel adéquat pour répondre au problème des fondements en mathématiques. Ce langage devra non seulement exprimer correctement les propositions mathématiques, il devra exprimer avec autant de précision les enchaînements logiques entre ces propositions. Dans les *Grundgesetze*, Frege décrit l'objectif qu'il cherche à atteindre comme suit :

> L'idéal d'une méthode strictement scientifique en mathématiques que j'ai tenté de réaliser ici et qui remonte à proprement parler à Euclide, j'aimerais la décrire comme suit. Il ne peut être demandé que tout soit démontré, car cela est impossible; mais nous pouvons exigé que toute proposition utilisée sans preuve soit explicitement déclarée comme telle, de sorte que nous puissions clairement voir ce sur quoi repose l'ensemble de l'édifice. Après cela, nous devons diminuer le nombre de ces lois primitives autant que possible, en démontrant tout ce qui peut être démontré. Par ailleurs, j'exige – et en ce sens je dépasse Euclide – que toutes les méthodes d'inférence employées soient spécifiées d'avance; autrement, nous n'aurions pas l'assurance d'avoir rempli le premier critère. (Frege 1964 : p. 2, ma traduction)

Dans les *Éléments* d'Euclide, on ne retrouve que des axiomes portant sur les objets géométriques, aucuns ne concernent le processus déductif. L'idéographie se distingue de ses prédécesseurs sur ce plan : la formalisation couvre l'ensemble de l'activité logique. Bien que l'art de la démonstration à l'œuvre dans les *Éléments* soit fort complexe et légiféré, ses règles ne sont jamais formulées telles quelles car les démonstrations sont effectuées dans le langage ordinaire. Dans la *Begriffsschrift* et les *Grundgesetze*, le langage ordinaire de la démonstration est remplacé par un système formel complet.

Le système logique de Frege est définitivement réformateur de la langue d'usage. Frege est l'initiateur d'une tradition de reconstruction logique qui sera pratiquée par Russell et qui culminera chez Carnap par la suite. Il est évident que la reconstruction souhaitée par Frege se veut indépendante de l'intuition et de la psychologie humaine, elle se veut aussi – et par conséquent – indépendante du langage d'usage, sans quoi l'objectif d'indépendance vis-à-vis les concepts incertains du parler commun ne serait pas atteint. Par exemple, la langue d'usage est linéaire parce que « l'espace sonore » a seulement une dimension. Or, Frege estime que cette écriture unidimensionnelle n'est pas apte à exprimer toute la complexité des rapports logiques entre propositions. Il n'hésite donc pas, entre autres choses, à sacrifier celle-ci pour adopter une écriture bidimensionnelle.[3] La rigueur et l'autonomie conceptuelle sont les deux grands pôles du développement de la *Begriffsschrift* : elle doit pouvoir exposer rigoureusement l'ensemble des éléments intervenant dans une preuve mathématique sans faire appel à quoique ce soit d'étranger au système. À propos des preuves de son système, il nous dit :

> Les preuves sont entièrement contenus dans les paragraphes intitulés « Construction », tandis que les paragraphes intitulés « Analyse » sont là pour faciliter la compréhension en apportant des indications préliminaires aux preuves qui suivent. Les preuves elles-mêmes ne contiennent pas de mots mais sont exécutées entièrement dans mon symbolisme; elles y figurent comme des suites de formules séparées par des lignes continues ou brisées ou par d'autres signes. Chacune de ces formules est une proposition complète incluant toutes les conditions nécessaires à sa validité. Cette complétude, qui ne permet pas l'occurrence d'hypothèses tacites de la pensée, me semble indispensable pour la rigueur et la conduite de la preuve. (Frege 1964 : pp. 1-2, ma traduction)

Les preuves sont donc des constructions symboliques formées de suites de formules, qui sont elles-mêmes formées de symboles, et la validité de la déduction est apparente

[3] L'écriture bidimensionnelle de Frege n'a manifestement pas fait le bonheur de tous – imprimeurs, philosophes ou mathématiciens – à en juger par sa postérité. Ceci s'explique sans doute par le fait que plusieurs stratagèmes peuvent être utilisés pour exprimer dans un langage linéaire les rapports logiques « bidimensionnels » de Frege : l'utilisation de points de « priorité » (Peano, Russell, etc.), l'écriture polonaise (Lukasiewicz, Tarski, etc.) ou l'utilisation de parenthèses.

à partir de cette construction sans avoir à être complétée par des remarques supplémentaires venant du langage ordinaire.

Bien que l'idéographie soit un langage épuré de l'influence « malsaine » de l'intuition, il ne faut pas croire, pour autant, qu'elle est une sorte d'algèbre dépouillée de signification :

> Je n'ai pas voulu donner en formules une logique abstraite, mais donner l'expression d'un contenu au moyen de signes écrits, et d'une manière plus précise et plus claire au regard que cela n'est possible au moyen des mots. En fait, je n'ai pas voulu créer seulement un *calculus ratiocinator* mais une *lingua charateristica* au sens de Leibniz, étant bien entendu que le calcul de la déduction est à mon sens une partie obligée d'une idéographie. (Frege 1971 : p.70-1)

Frege estime que les logiques de Boole et de Schröder[4] ne remplissent que la première condition d'une *lingua characteristica*, celle d'être un calcul logique formel (un *calculus ratiocinator*), mais qu'il leur manque à toutes les deux la capacité d'exprimer un contenu : « Si l'on prend une vue d'ensemble du langage formulaire de Boole, on voit qu'il consiste à habiller la logique abstraite du vêtement des signes algébriques; il n'est pas propre à l'expression d'un contenu et tel n'est pas son but » (Frege 1971 : p.73). D'abord, et dans une moindre mesure, parce qu'il lui manque la capacité d'exprimer la relation de subordination d'un individu à un concept.[5] Mais surtout, parce qu'il fait un usage relatif des signes de la logique. Autant les signes d'addition ou de multiplication que les signes 0 ou 1 ont des significations logiques et des significations arithmétiques qui dépendent du contexte. $A+B=0$ signifie tantôt que la disjonction des propositions A et B est fausse, et tantôt que la somme des nombres A et B est nulle. D'autre part, dans leurs sens logiques, 0 désigne l'extension du concept sous lequel ne tombe rien (de l'univers de discours) et 1 désigne

[4] *The Laws of Thought* de Boole et *Algebra der Logik* de Schröder.

[5] Pour exprimer ce type de relation, il faut la notion de quantificateur, ce que Boole n'a pas. Il peut comparer et subordonner les extensions de concepts, mais il ne peut pas parler d'individus.

l'extension de l'univers de discours. Si nous substituons un univers de discours pour un autre, les dénotations de 0 et 1 vont varier. Pour Frege, la logique ne peut pas être seulement un outil auxiliaire pour faciliter la déduction – à l'aide d'équations algébriques par exemple –, elle n'est pas seulement la science de la manipulation des entités symboliques, comme si les nombres étaient les pièces d'un jeu d'échecs. Les signes logiques dénotent quelque chose : « $3^2 + 4^2 = 5^2$ » exprime une pensée, tandis qu'une configuration de pièces d'échecs n'exprime rien (Frege 1967 : p. 10-11). Les symboles de la *Begriffsschrift* ont donc ceci de particulier qu'ils parlent/dénotent la « réalité » logique, et cette réalité logique n'est pas une stipulation arbitraire ou conventionnelle, elle est permanente.

Pour comprendre cette « réalité », il faut retourner aux arguments fondateurs du réalisme logicomathématique. De manière générale, le réalisme en logique (et en mathématiques) est la thèse à l'effet que la logique (et les mathématiques) ne porte ni sur le monde empirique ni sur l'esprit humain mais sur des entités essentielles et « réelles », c'est-à-dire des entités indépendantes du sujet connaissant et du monde empirique. Chez Frege et Russell, ce réalisme justifiera la réduction des mathématiques à la logique, thèse que l'histoire a convenu d'appeler le logicisme. C'est principalement dans les *Grundlagen der Arithmetik* qu'on retrouve une défense de ce réalisme logique. Elle commence par le constat suivant :

> Si, malgré plusieurs tentatives de deux parties, la collaboration de ces deux sciences n'est pas aussi fructueuse qu'il serait possible et souhaitable, cela tient, semble-t-il, à ce que la méthode psychologique prévaut en philosophie et tend à s'introduire en logique. Or les mathématiques sont étrangères à ce courant de pensée et l'on comprend aisément la répugnance que suscite[nt] chez bon nombre de mathématiciens les considérations philosophiques. (Frege 1969 : p. 118)

L'argument pour le réalisme exposé dans les *Grundlagen* est essentiellement un raisonnement par l'absurde de la forme : si la logique était psychologique (ou empirique) alors la logique ne posséderait pas ses attributs caractéristiques, c'est-à-dire son immutabilité, sa certitude, son indépendance, etc; or, puisqu'il est impossible que la logique soit dépourvue de ces attributs, elle ne peut-être psychologique ou empirique. Les arguments antipsychologistes de Frege sont notoires et je ne les repasserai pas en revue un à un. On peut cependant les regrouper de façon générale de la manière suivante :[6]

(1) ARGUMENT DE LA VARIABILITÉ. Si le sens (ou la dénotation) d'une expression logique ou mathématique est une représentation mentale, alors le sens (ou la dénotation) est quelque chose de résolument subjectif et, en conséquence, il varie considérablement d'un sujet à l'autre rendant la communication entre locuteurs impossible.

(2) ARGUMENT DE LA DISSOLUTION. La logique ne saurait être une discipline immuable si elle portait sur des représentations mentales, des entités subjectives, car les représentations sont sujettes aux variations des individus. Les répercussions seraient graves : la logique serait dissoute dans le magma psychologique des individus.

(3) ARGUMENT DU CONTEXTE DE DÉCOUVERTE. Il ne faut pas confondre d'un côté les dispositions psychologiques nécessaires à la compréhension d'une proposition logique ou mathématique, et de l'autre, les conditions objectives de vérité d'une proposition. C'est précisément ce que font les philosophes psychologistes selon Frege.

(4) ARGUMENT DE L'ERREUR. Si les lois logiques sont des lois naturelles, dans le sens où elles décrivent une réalité mentale universelle, comment ferions-nous pour expliquer l'erreur de raisonnement? Ce ne sont évidemment pas toutes les opérations psychologiques qui mènent à des jugements, des déductions ou des inférences correctes. Le fait qu'un, deux ou la majorité des êtres humains

[6] Je m'inspire ici en grande partie du chapitre 1 de Engel, *Philosophie et psychologie*, et du chapitre trois de Kusch, *Psychologism : A case study on the sociology of philosophical knowledge*.

fassent des inférences correctes n'a aucune incidence sur la vérité de la logique.

Frege accuse la plupart des logiciens d'avoir commis, à un moment ou un autre, une erreur de cette sorte : Stuart Mill, pour avoir suggérer que les lois logiques étaient des généralisations empiriques; Husserl : parce qu'il confond contexte psychologique et contexte de justification dans sa *Philosophie der Arithmetik*; Cantor, parce qu'il fait appel à une faculté d'abstraction pour définir ses ensembles; Schröder, pour la même raison que Husserl; etc.

Ainsi, la critique frégéenne du psychologisme (et du naturalisme) établit que la logique ne peut porter sur des représentations et qu'elle n'est pas la science des facultés psychologiques nécessaires à l'aboutissement d'une déduction correcte. En conséquence, la logique est la science des entités essentielles et intemporelles. La forme la plus concise et achevée de cet argument se retrouve dans *La Pensée* :

> Il en résulte, semble-t-il, que les pensées ne sont ni des choses du monde extérieur ni des représentations.
>
> Il faut admettre un troisième domaine. Ce qu'il renferme s'accorde avec les représentations en ce qu'il ne peut pas être perçu par les sens, mais aussi avec les choses en ce qu'il n'a pas besoin d'un porteur dont il serait le contenu de conscience. Telle est par exemple la pensée que nous exprimons dans le théorème de Pythagore, vraie intemporellement, vraie indépendamment du fait que quelqu'un la tienne pour vraie ou non. (Frege 1971 : p. 184)

Par « pensée » [Gedanke], Frege entend « non pas l'acte subjectif de penser mais son contenu objectif, lequel peut être la propriété commune de plusieurs sujets » (Frege 1971 : p. 108, note en bas de page). La pensée retient, si j'ose dire, le meilleur des deux mondes : elle est immatérielle comme les représentations et elle est indépendante de la conscience comme l'est un objet matériel. Le troisième monde est donc un ingrédient inévitable pour celui qui voudra préserver le caractère *a priori* de la logique tout en évitant le formalisme linguistique (la logique comme jeu d'échecs).

Nous aboutissons ainsi à la thèse centrale que défend Frege : la logique est la discipline qui analyse les propriétés essentielles des pensées conçues comme entités du troisième domaine, elle ne porte ni sur les représentations ni sur les objets matériels mais sur des entités immuables d'une réalité essentielle. Pour le réaliste, la logique ne résulte pas des structures réceptives de la représentation, comme l'intuition spatio-temporelle kantienne, elle ne peut venir que des propriétés de la pensée elle-même. De ce fait, l'*a priori* et l'analytique sont synonymes et les mathématiques forment un ensemble de propositions analytiques, c'est-à-dire que les mathématiques sont réductibles à la logique. Le réaliste est donc logiciste.

2. Le formalisme

Quiconque jette un coup d'œil sur l'idéographie frégéenne pour la première fois éprouve certaines difficultés à s'y retrouver. Aucun élément du langage ordinaire ou des langages formels de l'époque (celui de Boole par exemple) ne semble y figurer, il n'y a que des schémas complexes et intriqués de lignes horizontales et verticales à l'intérieur duquel nous retrouvons des lettres latines et grecques. Le caractère novateur de l'entreprise explique sans doute l'étrangeté de l'écriture frégéenne, après tout il n'y avait aucune notation disponible sur le marché pour les quantificateurs. Toutefois, l'écriture bidimensionnelle n'étant pas très commode, on ne s'étonnera pas d'apprendre qu'elle n'a pas connu une très grande popularité. Quoiqu'il en soit, les éléments primitifs de ce symbolisme méritent d'être exposés afin de cerner les pierres d'assises du système.

L'ontologie frégéenne postule deux sortes d'entités dans le troisième domaine : des objets et des fonctions, indéfinissables les uns autant que les autres par leur nature primitive. Sur le plan sémantique, à partir de *Sens et dénotation*,[7] les

[7] *Sens et dénotation*, dans Frege 1971, pp. 102-26. Avant cet article, les termes de la *Begriffsschrift* désignaient des pensées.

termes de la *Begriffsschrift* possèdent une dénotation [*Bedeutung*] et un sens [*Sinn*] : le terme *dénote* un objet, sa dénotation, et il *exprime* un sens, le mode de donation de l'objet (qui n'est pas plus une entité psychologique que l'objet).[8] Il est donc très important de remarquer que la dénotation est toujours un objet, jamais une fonction. La *Begriffsschrift* est un langage dont les termes portent sur des objets. Ainsi, il n'y a pas de symbole de fonction dans l'idéographie à proprement parler, c'est-à-dire un symbole pour décrire une fonction *en tant que fonction* et non pas en tant que valeur indéterminée de la fonction. Les symboles qui décrivent des fonctions appartiennent au langage de présentation. Lorsque Frege veut parler de ϕ en tant que fonction, il utilise une variable grecque : $\phi(\xi)$; et lorsqu'il veut parler d'une valeur indéterminée de ϕ (donc un objet indéterminé), il utilise une variable latine : $\phi(x)$. Seule la deuxième expression est une expression de la notation conceptuelle. Même si les fonctions n'ont pas de place officiel dans le formalisme, on dira néanmoins que les symboles de fonction *indiquent* une fonction, de la même qu'un nom dénote un objet.

Pour décrire la logique propositionnelle, Frege retient trois fonctions de vérité primitives : le tirait de contenu, la négation et l'implication. La plus primitive des fonctions est le tirait de contenu. D'une certaine manière, cette fonction est la fonction caractéristique du vrai : pour tout A, la fonction

$$—\!\!—\, A$$

dénote le Vrai si A dénote le Vrai, et le Faux si « A » dénote n'importe quoi d'autre que le Vrai. Le fait de juger ou d'avancer le contenu A est représenté par

$$\vdash\!\!—\, A.$$

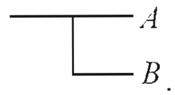
La négation d'un objet A est représentée par le symbole

$$—\!\!\top\!\!—\, A\,,$$

lequel dénote le Faux si A dénote le Vrai et le Vrai si « A » dénote autre chose que le Vrai. Pour affirmer ou juger la négation du contenu de A, il suffit, comme

[8] Il est possible qu'une expression ait un sens sans qu'il dénote quoique ce soit, comme c'est le cas pour « L'actuel roi de France est chauve ».

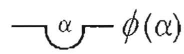
précédemment, de placer une barre verticale devant ce symbole. Enfin, l'implication (formelle) de A par B est donnée par le symbole



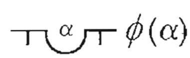
À noter qu'il s'agit de l'implication de A par B : l'antécédent est B et le conséquent est A (en écriture moderne : $B \supset A$). Selon la définition de Frege : ce symbole dénote Faux uniquement lorsque B dénote le Vrai et A autre chose que le Vrai.

Les question relatives au tirait de contenu et à la barre de jugement mises à part, on remarquera que la logique propositionnelle de Frege est générée par les fonctions de négation et d'implication seulement. Il est facile retrouver les autres connecteurs – qu'il s'agisse de la conjonction ou de la disjonction (inclusive ou exclusive) – par une combinaison de ceux-ci.

Afin de représenter les relations entre les individus et les concepts, il faut ajouter d'autres fonctions au système, dont les quantificateurs et l'égalité. Si $\phi(\xi)$ est une fonction à une variable, alors le symbole



dénote le Vrai si $\phi(x)$ dénote le Vrai quelque soit l'objet x. Autrement dit, il dénote la quantification universelle de $\phi(x)$. Pour représenter le quantificateur existentiel, il suffit de deux négations bien placées :



dénote le Vrai dans le seul cas où ce ne sont pas tous les $\phi(x)$ qui ne dénotent pas le Vrai, c'est-à-dire s'il existe un x pour lequel $\phi(x)$ est le Vrai. Il y a également des quantificateurs de second ordre pour des fonctions de fonctions (à une variable et à deux variables). L'égalité joue un double rôle dans l'idéographie, il est à la fois la relation d'identité entre d'objets et l'équivalence entre propositions. Puisque les valeurs de vérité sont des dénotations, ce double rôle s'énonce dans une seule définition : $A = B$ dénote le Vrai si et seulement si « A » et « B » ont la même dénotation.

Il manque deux éléments au système pour qu'il soit complet : une fonction pour parler du graphe[9] d'une fonction et une autre qui fait figure d'article défini.[10] Pour une fonction $\phi(\xi)$ à une variable x, le symbole

$$\dot{\varepsilon}(\phi(\varepsilon))$$

dénote son graphe. De manière analogue,

$$\backslash(\phi(\varepsilon))$$

dénote le x qui satisfait $\phi(\xi)$ lorsqu'un tel x existe, autrement il dénote le graphe de $\phi(\xi)$.

Toutes les autres fonctions du système seront définies à partir de ces fonctions primitives à l'aide du double trait de définition. Dans les *Grundgesetze*, les axiomes qui dictent le comportement de ces fonctions sont au nombre de huit, dont le célèbre axiome V qui est à l'origine de la contradiction dans les *Grungesetze*. D'autre part, Frege recense plusieurs règles inférentielles à l'œuvre dans les *Grundgesetze*, mais on peut les réduire au nombre de deux : soient la règle du *modus ponens* et une règle de généralisation (qui en réalité est composée de trois règles différentes selon le type de la variable).[11]

3. L'universalité de la *Begriffsschrift*

L'idéographie frégéenne a certaines prétentions à l'universalité qu'il nous faut examiner à présent, d'autant plus que cette question est directement reliée aux

[9] On parle souvent de « parcours de valeurs » dans les traductions françaises. Dans la terminologie mathématique moderne, il serait plus juste de traduire par le terme « graphe », le graphe d'une fonction $f(x)$ étant l'ensemble formé des couples $(x, f(x))$.
[10] Ces deux dernières fonctions sont introduites seulement dans les *Grundgesetze*.
[11] Il faudrait peut-être mentionner aussi la règle de l'amalgamation des horizontaux et, surtout, les règles de substitution. Ces dernières ne sont jamais formulées en tant que telles par Frege.

conceptions métalinguistiques de Frege. En effet, la métathéorie se manifeste généralement dans un contexte où plusieurs alternatives théoriques sont disponibles au niveau objet : si plusieurs univers de discours sont possibles, le métalangage devient alors un langage hôte qui permet de les comparer et de les évaluer entre eux sous un dénominateur commun. Rien n'empêche évidemment que métalangage et universalité puissent cohabiter. Mais, en pratique, les logiciens universalistes ont tendance à limiter l'importance de toute métathéorie. En fait, comme nous le verrons plus tard, pour eux, la métathéorie – si métathéorie il y a – n'est qu'un outil de transition, éliminable en fin de parcours.

L'universalité chez Frege se manifeste d'abord et avant tout par les quantificateurs et leur portée illimitée. Grâce aux quantificateurs, on le sait, la *Begriffsschrift* peut exprimer un ensemble de relations qui sont inaccessibles aux calculs logiques algébriques. L'algèbre logique de Boole et de Schröder ne peut pas être un langage logique complet, car il agit seulement à titre d'outil de calcul pour un raisonnement qui autrement nécessite l'intervention de la langue d'usage pour être complet. L'autonomie conceptuelle de la *Begriffsschrift* est certainement l'un des éléments qui joue en faveur de son universalité. Un autre élément, plus important encore que celui-ci, est le caractère illimité de la référence des variables.

Si Frege n'a jamais circonscrit un ensemble d'individus ou un univers de discours sur lequel l'idéographie était censée porter, ce n'était pas par négligence mais par conviction « idéographique » :

> Veiller à ce qu'aucune expression ne puisse être dépourvue de dénotation, à ce qu'on ne puisse jamais calculer sans y prendre garde sur des signes vides tout en croyant opérer sur des objets, c'est là ce qu'exige la rigueur scientifique. On a fait récemment d'effroyables expériences avec des suites infinies divergentes. Il est donc nécessaire d'établir des règles d'où l'on puisse tirer ce que dénote par exemple :
>
> « $\odot + 1$ »
>
> où $\odot$ dénoterait le soleil. (Frege 1971 : p. 93)

L'expérience « effroyable » des séries divergentes à laquelle Frege fait référence est le fait que des résultats contradictoires puissent être démontrés lorsque les symboles employés n'ont pas de dénotation. Par exemple, l'absence de dénotation de l'expression $1-1+1-1+1-1+1\cdots$ est, dirait Frege, la raison pour laquelle nous pouvons démontrer qu'elle est simultanément égale à 0 et à 1. À chaque fois qu'un terme est dépourvu de dénotation, il prive l'expression dans laquelle il figure de dénotation et il peut en résulter des antinomies. Ainsi, il est impératif de définir la fonction $\xi+1$, entre autres, pour toutes les entités possibles, qu'elles soient empiriques, psychologiques ou idéelles, si on veut que l'addition par 1 ait toujours une dénotation.[12] On remarquera que c'est ce qu'il fit avec les fonctions primitives :

$$—\!\!—\,A$$

est définie pour tout argument A, qu'il soit une tautologie, une contradiction, une pomme, une patate, le nombre 7, la galaxie d'Andromède ou l'arrière grand-mère de Bertrand Russell. Cette conception de la super variable qui parcourt l'ensemble du répertoire ontologique est l'expression la plus manifeste de l'universalité à l'œuvre dans l'idéographie... elle est également son talon d'Achille.

Il est bien connu que le système décrit dans les *Grundgesetze* est contradictoire et ce, parce qu'il n'y aucune limitation sur la référence d'une variable. C'est en exploitant cette faiblesse que le logicien anglais Bertrand Russell, jeune et inconnu à l'époque, démontra une contradiction qui sonna le glas de l'entreprise frégéenne. Il fit connaître cette nouvelle malheureuse à Frege en 1902 alors que le deuxième tome des *Grundgesetze* achevait d'être imprimé. L'origine du problème se trouve principalement[13] au niveau de l'axiome V,[14] un axiome à l'endroit duquel

[12] Il est vrai que rien n'empêche tous les termes du système aient d'avoir une dénotation si le domaine est limité. Seulement, cette option ne paraît pas valable pour Frege parce que le signe d'addition, par exemple, changerait de signification chaque fois que l'on changerait de domaine. Il y aurait un signe pour les naturels, un pour les entiers, et ainsi de suite pour les rationnels, les réels et les complexes.

[13] Je dis *principalement* parce qu'on doit faire appel à d'autres axiomes pour démontrer la contradiction, l'axiome V à lui seul ne suffit pas.

[14] Puisqu'il est question de l'axiome V, je devrais peut-être l'énoncer. Soient $f(x)$ et $g(x)$ des fonctions (propositionnelles). L'axiome V stipule alors que si ces fonctions sont équivalentes pour tout x, alors

Frege a toujours émis des réserves, et de l'imprédicativité des expressions de la *Begriffsschrift*. [15] La contradiction procède essentiellement comme suit :

(1) On définit le concept « x n'appartient pas à l'extension (au graphe ou au parcours de valeurs) de x ».

(2) On considère X son extension.

(3) On démontre que si X est dans son extension, il n'est pas dans son extension; et que si X n'est pas dans son extension, il est dans son extension.

(4) On en déduit la contradiction suivante : X appartient et n'appartient pas à son extension.

Toutes les solutions qui ont été apportées à ce paradoxe subséquemment ont, d'une manière ou d'une autre, limité le champ des variables.

Outre les quantificateurs, il y a une autre raison encore pour laquelle on peut qualifier le système de Frege d'universel. La sémantique[16] de l'idéographie est telle qu'un terme, pour être bien formé, doit posséder une dénotation et cette dénotation, nous l'avons vu, est unique. Le principe directeur pour la formation des noms est : « Pour les définitions, j'énonce maintenant le principe directeur suivant : les noms bien formés doivent toujours posséder une dénotation » (Frege 1967 : p. 83, ma traduction). Dans l'approche modèle-théorétique,[17] on s'emploie à séparer l'aspect syntaxique, qui est strictement réservé à la combinatoire des symboles, de l'aspect sémantique, qui est réservé à la signification. Cette distance est motivée par une sorte de relativité du domaine de discours. Un même langage peut être « interprété » dans

l'extension composée des éléments qui satisfont $f(x)$ est la même que celle composée des éléments qui satisfont la fonction $g(x)$. L'axiome présuppose implicitement que toute fonction définit une extension.

[15] Ce sur quoi Russell a insisté.

[16] J'entends ici par sémantique la signification donnée aux termes et non pas le sens étroit tarskien dont il est question normalement en logique mathématique. Nous reviendrons sur ces distinctions plus loin lorsque nous analyserons les lectures « sémantiques » et « anti-sémantiques » de Frege.

[17] C'est Tarski qui jettera les bases de l'analyse modèle-théorétique. Toutefois, cette orientation théorique est présente, sous forme embryonnaire, chez Löwenheim.

deux domaines différents. Par exemple, le terme « droite » ne signifie pas la même chose dans la géométrie euclidienne que dans les géométries non euclidiennes. Par contre, chez Frege, les termes jouissent d'une sémantique intégrée et invariable; un nom propre, c'est-à-dire un symbole dénotant (signifiant), n'a qu'une et une seule dénotation. Le vocabulaire de l'idéographie est ainsi fixé définitivement et il réfère déjà par définition (ou par construction) à des objets d'un univers réel et immuable.

La définition de la vérité est un autre lieu où on peut voir la fixité de la sémantique frégéenne. Le vocabulaire de la *Begriffsschrift* est composé de noms propres (ou saturés) et de noms insaturés qui sont des noms pour des objets et des fonctions respectivement, les deux seules catégories d'entités dans l'ontologie de Frege. Par exemple, l'expression « le pays à l'est d'Haïti » est un nom pour la République Dominicaine, tandis que « le pays à l'est de ξ » est un nom insaturé (un nom de fonction) qui devient un nom propre lorsqu'on substitut un nom de pays à ξ : le Botswana, si ξ = Namibie; Slovaquie, si ξ = Tchéquie; etc. Cette structure est transposée intégralement à la proposition et sa valeur de vérité :

> Nous sommes donc conduits à identifier la *valeur de vérité* d'une proposition avec sa dénotation. Par valeur de vérité d'une proposition, j'entends le fait qu'elle est vraie ou fausse. Il n'y a pas d'autre valeur de vérité. J'appellerai plus brièvement l'une le vrai et l'autre le faux. Toute proposition affirmative, quand on considère la dénotation des mots qui la constituent, doit donc être prise comme un nom propre; sa dénotation, si elle existe, est le vrai ou le faux. (Frege 1971 : p. 110)

Une proposition est un nom propre d'une valeur de vérité et les prédicats (et relations) sont des fonctions qui, lorsque saturées, donnent des valeurs de vérité. Le vrai et le faux sont donc des objets au même titre que les nombres. Ce ne sont pas des prédicats qui s'appliquent au langage, ce sont des dénotations de propositions.

Plus tard, dans les *Recherches logiques*,[18] il abandonne toute possibilité de faire de la vérité un prédicat, et ainsi, toute possibilité de relativiser la vérité d'une proposition. Son raisonnement procède en considérant la possibilité que la vérité puisse être « l'accord d'un tableau avec son objet ».[19] Or, selon Frege, on ne saurait concevoir cet accord comme une identité stricte car un tableau est toujours différent de l'objet qu'il représente. Il faut donc parler d'une identité selon un certain point vue. Cela dit, la difficulté subsiste, comment définir l'accord selon un certain point de vue sans retomber dans le même dilemme? Frege ne se fait guère plus optimiste pour définir « l'être vrai » d'une autre manière et prétend que « toute autre tentative pour définir l'être vrai échoue également ».[20] Son argument se résume essentiellement au fait que le prédicat « être vrai » sera toujours nécessaire afin de définir « être vrai » et il en conclut que :

> On tourne en cercle. Il est donc vraisemblable que le contenu du mot « vrai » est unique en son genre et indéfinissable. (Frege 1971 : pp. 171-2)

La vérité est reléguée au second plan. Il semble que le seul véritable prédicat de proposition soit l'affirmation :

> La reconnaissance de la vérité est enfin exprimée dans la forme de la proposition affirmative. Il n'est nul besoin pour cela du mot « vrai ». Quand bien même l'emploierait-on, la force proprement affirmative ne réside pas en lui mais dans la forme de la proposition affirmative; si la proposition perd sa forme affirmative, le mot « vrai » ne peut pas la lui rendre. (Frege 1971 : p. 176)

Le concept de Frege est pour ainsi dire déflationniste, le vrai comme prédicat d'un énoncé n'ajoute rien à l'assertion d'une proposition.

---

[18] Frege 1971 : pp. 170-234.
[19] Frege 1971 : p. 172.
[20] Frege 1971 : p. 172.

Tous ces éléments participent à l'universalité de l'idéographie : les quantificateurs, par l'autonomie qu'ils confèrent au langage; les super variables, par leur capacité à dénoter sans restriction toutes les entités de l'univers; la sémantique, par son unicité et son intégration à la syntaxe; et la vérité, parce qu'elle est conçue de manière objective et non relative.

4. Le métalangage et la *Begriffsschrift*

En scrutant l'œuvre de Frege, nous ne trouvons qu'une seule occurrence claire et explicite du concept de métalangage, et elle apparaît dans un écrit incomplet qui devait être la quatrième recherche logique. Parlant de l'usage des guillemets, il dit :

> Les propositions de mon langage auxiliaire [*Hilfssprache*] sont des objets dont il est parlé dans le langage d'exposition [*Darlegungssprache*]. Je dois pouvoir les désigner dans mon langage d'exposition tout comme, dans un traité d'astronomie, les planètes sont désignées par leur nom propre « Vénus » ou « Mars ». J'obtiens les noms propres des propositions de la langue auxiliaire en mettant celles-ci entre guillemets. (Frege 1971 : p. 19)[21]

Sinon, bien qu'il ait maîtrisé la plupart des subtilités relatives à la manipulation d'un langage objet (comme l'usage des guillemets par exemple), il ne fait aucune mention de la notion de métalangage tout au long de son œuvre. En conséquence, pour dresser un portrait des conceptions métalinguistiques de Frege, il faut lire entre les lignes.

Cette tâche est rendue d'autant plus difficile pour deux raisons : le sort du métalangage est intimement lié à celui de la sémantique et de nombreuses divergences existent sur la sémantique de Frege, particulièrement lorsque la sémantique est prise dans un sens modèle-théorétique. On peut regrouper les

[21] Ce passage a été rédigé entre 1923 et 1925 et se trouve à l'origine dans *Nachgelassene Schriften*, pp. 208-281. Claude Imbert le traduit dans l'introduction de Frege 1971.

divergences en deux clans : ceux qui font une lecture « sémantique » de Frege d'une part, parmi lesquels on retrouve Furth et Dummett,[22] et ceux en font une lecture « anti-sémantique », dont van Heijenoort, Dreben, Weiner, Goldfarb et Ricketts.[23] En gros, les premiers soutiennent que Frege a développé une théorie sémantique qui contient, en germe, les concepts sémantiques développés plus tard par Tarski et Gödel, et les autres prétendent que le système de Frege est étranger à ces concepts.[24]

La sémantique modèle-théorétique est une idée qui prend ses origines chez Tarski. Dans sa formulation définitive, le langage (formel) est une combinatoire récursive de symboles, le domaine (ou le modèle) est conçu comme un ensemble de base muni de relations et le langage exprime le domaine par une application qu'on nomme « interprétation ». Avec cette dernière, il est possible de définir les notions de satisfaction et de vérité ainsi que la notion de complétude. Il est clair que ces concepts ne sont jamais formulés par Frege dans toute leur généralité. Toutefois, il est légitime de se demander : 1) si Frege les comprendrait; 2) si Frege serait d'accord pour leur donner l'importance dont ils jouissent aujourd'hui; et 3) s'ils se trouvent dans son système sous une forme embryonnaire.

À ces trois questions, Jean van Heijenoort répond par la négative. Son point de vue est essentiellement que Frege, comme Russell, en cherchant à développer un langage logique universel et autosuffisant, a une conception de la logique si radicalement différente de celle de Tarski et Gödel qu'il ne parviendrait même pas à comprendre le pourquoi (ni même le comment) d'une preuve de la complétude de l'arithmétique formelle. Cette différence tient à ce que la logique de Frege (et de Russell) est un langage (une *lingua characteristica*) tandis que celle de Tarski est un

---

[22] Montgomery Furth, dans l'introduction à sa traduction (partielle) des Grundgesetze (Frege 1964 : pp. *v-lx*); Michael Dummett, dans *Frege : Philosophy of Language* (Dummett 1973).
[23] Jean van Heijenoort, dans *Logic as Calculus and Logic as Language* (van Heijenoort 1985 : pp. 11-16); Warren Goldfarb, dans *Logic in the Twenties : The Nature of the Quantifier* (Goldfarb 1977); Thomas Ricketts, dans *Frege, the* Tractatus, *and the Logocentric Predicament* (Ricketts 1985).
[24] Il y a par ailleurs trois auteurs qui me guident dans la lecture de cette polémique, il s'agit de Richard Heck Jnr, Stanford Shieh et Juliet Floyd. Juliet Floyd présente un excellent résumé de ce débat dans (Marion & Voizard 1998 : pp. 137-169).

calcul (un *calculus ratiocinator*) avec un univers de discours relatif.[25] Or, la logique comme langage aspire à l'universalité, ce qui signifie, entre autres choses, que rien ne peut être dit à l'extérieur du langage. Par conséquent, les questions de la sémantique modèle-théorétique, qui nécessitent le cadre externe du métalangage, ne peuvent pas être formulées. Frege ne pourrait donc pas comprendre ou cautionner une preuve de complétude parce qu'elle constitue une sorte d'hérésie au dogme logiciste frégéo-russellien, une fausse conception qui contrevient à sa philosophie universaliste. Van Heijenoort s'étonne, d'ailleurs, des propos de Gödel à l'effet que la complétude est une question qui se pose *naturellement* dès qu'on formule un système d'axiomes. Cette question n'est pas naturelle pour Frege, nous dit van Heijenoort, car elle est le résultat d'une posture théorique incompatible avec sa philosophie.

À l'autre extrême du spectre interprétatif, Floyd nous fait remarquer les propos de Montgomery Furth à l'endroit des conceptions sémantiques de Frege :

> [L]'explicitation du fondement primitif de son système de logique [...] est entreprise au regard d'une interprétation sémantique profondément réfléchie qui est, elle-même, le reflet d'une véritable philosophie du langage. L'influence de cette dernière sur la structure sémantique ainsi que la syntaxe du langage ainsi développé se ressent de façon constante à travers tout le développement de l'explicitation. (Frege 1964 : p. *vi-vii*; traduction prise dans (Marion & Voizard 1998 : p. 137))

Pour Furth, l'œuvre de Frege introduit de manière embryonnaire des concepts syntaxiques et sémantiques qui seront clairement formulés ultérieurement par Tarksi. Selon la lecture de Furth, il n'y a pas de rupture conceptuelle entre la logique de Frege et la logique modèle-théorétique, la dernière suit naturellement de la première. Dans son livre *Frege : Philosophy of Language*, Michael Dummett défend aussi l'idée selon laquelle Frege démontre une connaissance implicite des notions

[25] Tarski n'est pas le premier à avoir pensé la logique comme un calcul, ses précurseurs sont évidemment Boole, De Morgan, Schröder ainsi que Löwenheim.

modernes de syntaxe et de sémantique. Il va même jusqu'à prétendre que Frege avait en possession les moyens conceptuels pour formuler la question de la complétude mais qu'il ne le fit tout simplement pas. S'ensuit alors une controverse où certains (Goldfarb, Ricketts, etc.) accusèrent Furth et Dummett de projeter à tort des conceptions modernes de la logique dans l'œuvre de Frege.

Mais de quelles notions méta-systémiques parle-t-on? On peut classer les remarques « extra-idéographiques » en quatre catégories, selon qu'elles sont relatives : 1) à l'explication d'une preuve; 2) à la manipulation des fonctions; 3) à la fixation d'une dénotation; ou 4) à la spécification de règles formelles. De façon générale, Frege cherchera à minimiser le rôle du langage ordinaire dans toutes ces sphères. Ainsi, dit-il d'emblée au §0 des *Grundgesetze* :

> Il ne sera pas toujours possible de donner une définition régulière à tout, précisément parce que notre tâche est de remonter vers ce qui est logiquement simple, lequel n'est pas définissable à proprement parler. Je dois donc me satisfaire d'indiquer ce que je vise par des indices. (Frege 1964 : p. 32, ma traduction)

Plus loin, il émet des réserves sur l'importance des remarques du premier type :

> Les preuves sont entièrement contenus dans les paragraphes intitulés « Construction », tandis que les paragraphes intitulés « Analyse » sont là pour faciliter la compréhension en apportant des indications préliminaires aux preuves qui suivent. (Frege 1964 : p. 1)

Les analyses ne sont là que pour faciliter la compréhension, elles ne font pas parti du système et ne sont en aucun cas nécessaires à la complétude de la déduction. Elles forment une sorte d'heuristique pour aborder le systèmc, mais ne sont ni essentielles ni fondamentales pour son maintien. Frege affirme à nouveau dans *Concept et Objet* une conception similaire des indications métathéoriques. Se défendant contre Kerry d'avoir fondé ses principes logiques sur des distinctions linguistiques, Frege affirme qu'on ne peut pas procéder autrement qu'en analysant ainsi le langage, car :

> Sans le langage, nous ne pourrions pas nous comprendre et nous en serions réduits à un acte de foi, à croire qu'autrui comprend les mots, les formes, et les constructions comme nous les comprenons nous-mêmes. Comme je l'ai déjà dit, je n'ai pas voulu donner une définition mais quelques indications qui font appel au sens de la langue que partage tout Allemand. (Frege 1971 : p. 130)[26]

La langue usuelle (l'Allemand dans ce cas) ne sert que de guide pour indiquer le sens précis qu'on veut donner à un symbole du langage formel. L'article défini « $\backslash\xi$ » est introduit conceptuellement par l'article défini du langage ordinaire, mais il n'est pas la formalisation de l'article « le/la » (« the » ou « die ») dans l'idéographie, car ce dernier possède plusieurs autres usages que « $\backslash\xi$ » ne possède pas.[27]

La plupart des lectures anti-sémantiques se basent sur les propos de Frege à l'égard de ses élucidations pour infirmer l'idée qu'il accorde la moindre importance au métalangage. Les propos de Frege à ce sujet me semblent sans équivoque : il est clair qu'il a toujours cherché à minimiser le rôle du langage d'exposition et ce, pour des raisons qui sont parfaitement cohérentes avec les ambitions théoriques qu'il entretient envers son projet. Mais ce que je veux faire valoir ici est qu'il existe un certain décalage entre l'idéographie et la conception (philosophique) que Frege en a. Même si Frege ne théorise pas sur les relations entre sa caractéristique et la langue d'exposition, il n'en demeure pas moins qu'il démontre la maîtrise d'un ensemble important de concepts métalinguistiques. Commençons par le plus simple d'entre eux : la distinction entre l'usage et la mention d'un terme. Ainsi, nous dit-il :

> On s'étonnera peut-être de l'emploi fréquent des guillemets. Ils servent à distinguer le cas où je parle du signe lui-même de celui où je parle de sa dénotation. (Frege 1964 : p. 32)[28]

---

[26] On remarquera que ce passage est cité de *Concept et objet* (1892) qui est un écrit antérieur aux *Grungesetze* (1893 & 1902).

[27] Par exemple, « le » dans « le cheval est un quadrupède » ne joue pas le rôle de l'article défini de la *Begriffsschrift*.

[28] Traduction de Claude Imbert dans (Frege 1971 : p. 19).

On ne verra jamais Frege confondre le nom d'une chose et la chose elle-même. Sans la distinction pratique entre l'usage et la mention, Frege n'aurait pas su délimiter aussi clairement les concepts de son système, il n'aurait même pas pu le faire de manière cohérente. N'oublions pas, par ailleurs, que la *Begriffsschrift* comme les *Grundgesetze* sont écrits dans la langue d'usage (l'Allemand), ils ne sont pas exprimés *dans* le langage de l'idéographie, l'idéographie est seulement leur objet.

La distinction entre langage objet et métalangue n'est pas la seule subtilité métalinguistique à l'œuvre chez Frege, son langage d'exposition en possède d'autres qui méritent d'être soulignées ici, dont la manipulation des signes de fonction. On rappellera que l'ontologie frégéenne est composée de fonctions et d'objets, mais que l'idéographie n'a de noms que pour les objets. Les noms de fonctions – en tant que fonctions – ne font pas partie de l'idéographie, ils font partie du métalangage : il y a une différence entre « $\phi(x)$ » qui est considéré comme un nom du langage objet, et $\phi(\xi)$ qui est le nom d'une fonction et qui appartient au métalangage. Les noms de fonctions sont utilisés abondamment pour parler du système et exposer des règles. On citera par exemple les règles de formation des noms A1 et B1 au §30 :

(A1) Un nom propre peut être formé d'un nom de fonction d'une variable et d'un nom propre;

(B1) Un nom de fonction d'une variable peut être formé d'un nom propre et d'un nom de fonction de deux variables.

Les noms de fonctions apparaissent de manière essentielle dans la formulation de A1 et B1.

Malgré l'importance de ces fonctions sur le plan métathéorique, Frege leur réserve le même sort que toutes les autres entités du langage d'exposition :

> « μ_β » et « $\mu_{\beta\gamma}$ » ne sont, toutefois, pas plus des signes de la *Begriffsschrift* que les lettres précédentes [les lettres ϕ, ψ, ξ et ζ sont des variables de fonction et d'objet respectivement], ils ont une utilité provisoire seulement. (Frege 1964 : p. 79, ma traduction)

Mais en quoi l'utilisation des lettres grecques est-elle provisoire ici? Frege croit-il aussi que les règles de formation des noms sont provisoires, en ce sens qu'un locuteur versé en « idéographais » n'en aurait plus besoin pour construire des noms? Ce locuteur aura toujours besoin des règles de formation des noms parce que le stock d'expressions de l'idéographie est infini, il est donc impossible de le générer en une seule ou plusieurs applications de ces règles. Cette remarque s'applique autant aux autres méta-règles du système de Frege, dont nous avons un échantillon au §48 des *Grundgesetze*. Toutes ces lois nécessitent qu'on parle de l'idéographie comme objet pour être énoncées. Prenons au hasard les lois :

(4) *L'amalgamation des prémisses identiques* (règle de contraction)

Si une prémisse apparaît deux fois dans une implication, on peut éliminer l'une de ses occurrences, c'est-à-dire de « $A \supset (A \supset B)$ » on peut inférer « $A \supset B$ », où « A » et « B » sont des propositions quelconques.

(5) *Mode d'inférence (a)* (règle du *modus ponens*)

Si la prémisse d'une implication est affirmée, alors on peut affirmer la conséquence de cette implication, c'est-à-dire de « $A \supset B$ » et de « A » on peut inférer « B ».

Pour énoncer ces lois, on remarquera qu'il a fallu employer des méta-variables (« A » et « B ») et des noms de termes (l'usage des guillemets), un vocabulaire qui est étranger à l'idéographie. Pour être charitable avec Frege, on pourrait lui concéder que ce qui est provisoire n'est pas tant l'usage des règles mais plutôt leur énonciation.

Frege a beau jeu sur ce point car les règles de la *Begriffsschrift* sont fixes et immuables, elles n'ont pas à être reformulées selon le contexte de discours car celui-ci est unique et universel. Le menuisier expérimenté n'a pas besoin de connaître l'énoncé des règles de l'art une fois que celles-ci sont assimilées, elles se manifestent seulement dans sa pratique. De la même manière, seuls les apprentis novices de la *Begriffsschrift* ont besoin de *lire* les méta-règles du système, les maîtres en font seulement *usage*. Il reste cependant qu'il y a une métathéorie à l'œuvre dans l'exposition de la *Begriffsschrift*, qu'elle soit provisoire ou pas.

Les stipulations métathéoriques ne se limitent pas seulement à la syntaxe de l'idéographie, elles précisent également sa sémantique, c'est-à-dire la dénotation des termes. Il est à noter que le fonctionnement syntaxique de l'idéographie ne dépend pas de la signification qu'on donne aux termes. Sans les stipulations sémantiques, Frege aurait pu dériver les mêmes théorèmes, mais ces théorèmes ne représenteraient pas des propositions vraies, ils ne seraient que les symboles d'un *calculus ratiocinator* vide de sens. Les stipulations sémantiques ont une autre fonction : elle font du *calculus ratiocinator* une *lingua characteristica*. Lorsqu'il introduit les fonctions primitives, Frege ne se contente pas de donner quelques indications floues sur leurs dénotations, il formule des règles de dénotation précises :

- « $\neg A$ » dénote le Vrai si et seulement si « A » dénote le Faux (§6);
- « $A \supset B$ » dénote le Faux si et seulement si « A » dénote le Vrai et « B » dénote le Faux (§12);
- « $A = B$ » dénote le Vrai si et seulement si « A » et « B » ont la même dénotation (§7);
- « $\forall x \Phi(x)$ » dénote le Vrai si et seulement si, pour tout argument, la fonction « $\Phi(\xi)$ » dénote le Vrai (§8); etc.[29]

[29] Les paragraphes indiqués sont ceux des *Grundgesetze*.

Si l'on remplace l'expression « dénote le Vrai/Faux » par « est Vrai/Faux », nous retrouvons donc quelque chose de très similaire au concept de vérité tarskien, à la différence près que Frege n'a pas de concept de satisfaction.

Il est également curieux de voir Frege justifier ses règles de déduction à l'aide de son langage d'exposition. En introduisant les règles et la plupart des axiomes, Frege effectue une démonstration de leur validité. En guise de justification de la règle du *modus ponens*, Frege nous dit : si $A \supset B$ et A étaient le Vrai et que B ne l'était pas le Vrai, alors, par la définition de « $\supset$ » et le fait que A est le Vrai, on aurait que $A \supset B$ serait le Faux, une contradiction.[30] Une justification similaire nous est donnée pour l'axiome « $A \supset (B \supset A)$ » : si $A \supset (B \supset A)$ était le Faux, on aurait que A serait le Vrai et que $B \supset A$ serait le Faux, auquel cas A serait le Faux, une contradiction.[31] Il est curieux que Frege fasse de telles démonstrations, considérant qu'il n'a pas une très grande estime (conceptuelle) pour le langage d'exposition et qu'il n'a de cesse de marteler le rôle indicateur, métaphorique et provisoire de celui-ci. Quoi qu'il en soit, nous, lecteurs contemporains, ne pouvons que remarquer la similarité entre ces démonstrations et certaines démonstrations modèle-théorétiques à l'effet, d'une part, que le *modus ponens* est une règle qui préserve la dénotation du Vrai et, d'autre part, que la formule « $A \supset (B \supset A)$ » est une tautologie (dans le sens où elle dénote le Vrai quel que soit la dénotation des formules « A » et « B »). Si chaque règle et chaque axiome reçoit une preuve de validité de cette sorte (et si ces preuves sont correctes), nous obtenons une démonstration à l'effet que tout ce qui est démontré dans le système dénote Vrai.

L'activité métathéorique dans les *Grundgesetze* est à son comble aux §§28-32. Ces paragraphes sont consacrés à la logique derrière la formation des noms. Ici, Frege veut prouver que tous les noms de son système possèdent une dénotation pour montrer, selon le critère du §28,[32] qu'ils sont bien formés. Il énonce aux §28-30 les

[30] (Frege 1964 : p. 57)
[31] (Frege 1964 : p. 69)
[32] C'est-à-dire : les noms correctement formés doivent avoir une dénotation.

règles de formation des noms pour ensuite démontrer au §31 (par une preuve plus ou moins inductive) que tout nom formé par ces règles possède une dénotation. Ces deux tâches se font en exploitant le caractère compositionnel des noms qui s'articule comme suit :

> [...] tout nom formé de noms ayant une dénotation dénote quelque chose. Cette formation procède de la manière suivante : un nom remplit les arguments d'un autre nom pour lesquels il convient. [...] Les noms formés ainsi peuvent être utilisés de la même manière pour la formation d'autres noms, et tous les noms qui résultent de ce procédé dénoteront si les noms simples primitifs dénotent quelque chose. (Frege 1964 : p. 85, ma traduction)[33]

Le caractère inductif de la preuve est apparent dans le passage suivant :

> Appliquons les propos précédents afin de montrer que les noms propres et les noms de fonctions de premier niveau, que nous pouvons former de cette manière avec les noms simples introduits précédemment, dénotent toujours quelque chose. Par ce qui a été dit, il est suffisant de montrer de nos noms primitifs qu'ils dénotent quelque chose. (Frege 1964 : p. 87, ma traduction)

La preuve de Frege procéderait donc par induction sur la complexité des expressions, c'est-à-dire sur le nombre de connecteurs qui figure dans une expression. L'intention est juste, mais la preuve est fausse. En effet, selon Heck,[34] l'induction achoppe sur la nature imprédicative de sa logique du second ordre combinée à l'instanciation des variables du second ordre : en général, l'instanciation d'une fonction ne diminue pas la complexité d'une expression (on ne passe pas d'une expression contenant n connecteurs à une autre qui contient moins de n connecteurs). Peu importe qu'elle soit fausse, ce qui est intéressant est le fait que Frege confie au métalangage le soin de

[33] C'est ici d'ailleurs que sont énoncées les règles A1 et B1 citées plus haut.
[34] Heck 1996.

démontrer par induction une propriété métathéorique importante du système de la *Begriffsschrift*.

Pour que l'idéographie soit une caractéristique, il est clair que les termes doivent tous avoir une dénotation. Mais cette démonstration, si elle avait été correcte, aurait-elle d'autres conséquences? Le passage du §31 est souvent pris comme une preuve de non contradiction de la *Begriffsschrift*, sinon, à tout le moins, comme une preuve que l'ensemble des théorèmes prouvables à l'intérieur de l'idéographie dénotent le vrai.[35] La relation entre la preuve du §31 et la non contradiction de l'idéographie aurait été connue de Frege, selon Heck, d'après son interprétation de la contradiction de Russell :

> Votre découverte de la contradiction m'a étonné au plus haut point [...] Il semblerait en conséquence que la transformation de la généralisation d'une identité en identité d'extensions [ou de graphes] [...] ne soit pas toujours permise, que ma loi V [...] est fausse, et que mes explications à la section 31 ne suffisent pas à garantir une référence pour mes signes dans tous les cas. (Frege 1980 : p. 32, ma traduction)

Frege souligne l'incompatibilité entre la validité de l'argument présenté au §31 et la contradiction dans son système. Voyait-il dans cet argument une preuve de consistance? Probable. À tout le moins, il semblerait que Frege n'ait pas été si étranger à des notions métathéoriques de haut niveau, contrairement à ce que prétend van Heijenoort et à ce qu'affirment les lectures anti-sémantiques.

Mon objectif ici n'était pas tant de trancher le débat opposant les lectures sémantiques et anti-sémantiques de Frege que de dépister des notions métathéoriques présentes dans son œuvre. Il est évident que les intentions théoriques et philosophiques de Frege sont très différentes de celles que partageront les logiciens

[35] Car : 1) tous les termes auraient une dénotation et, en particulier, tous les noms de propositions auraient une valeur de vérité définie; 2) tous axiomes dénoteraient le Vrai; et 3) toutes les règles d'inférence préserveraient la dénotation du Vrai. Le fait que le système soit non contradictoire découlerait ensuite du fait qu'il existe au moins une proposition qui est fausse.

modèle-théorétiques, mais il est moins clair, pour moi, qu'il n'ait pas développé un ensemble des notions métathéoriques (parfois de manière partielle et embryonnaire) qui seront utilisées plus tard pour d'autres finalités et avec plus de raffinement. Ainsi, j'estime qu'une part du mérite dans l'élaboration du concept de métalangage lui revient malgré qu'il n'ait jamais souhaité lui accorder tant d'importance.

CHAPITRE II

Bertrand Russell et les *Principia Mathematica*

0. Introduction

Comme Frege, Bertrand Russell marque un point tournant dans l'histoire de la notion de langage formel. Les *Principia Mathematica*, rédigés en collaboration avec Whitehead, ont longtemps incarné l'idéal d'un langage logiquement parfait à l'intérieur duquel les mathématiques pouvaient être logiquement reconstruites. L'objectif, encore une fois, étant de se dissocier du langage ordinaire pour échapper à son influence malsaine. L'entreprise n'est pas seulement logique, elle est philosophique également : l'ensemble du symbolisme procède d'une réflexion philosophique approfondie qui prétend arriver aux sources de la logique elle-même. Il va donc sans dire que Russell vise une sorte d'autonomie conceptuelle où le métalangage n'a pas sa place. En fait, nous verrons comment la philosophie russellienne (à ses premières heures) est plutôt réfractaire à ce concept.

1. L'émergence de la philosophie analytique

G. E. Moore et Bertrand Russell sont considérés comme étant les fondateurs de la philosophie analytique, une philosophie née d'une opposition à l'endroit de l'idéalisme de Bradley.[1] Les grands thèmes de cette philosophie comprennent, notamment : l'objectivité et l'indépendance de la proposition, l'objectivité et l'externalité des relations, le caractère absolu de la vérité et de la fausseté, et la séparation nette et absolue entre la connaissance et ce qui est connu.[2] On peut dire qu'ils procèdent, d'une part, de la postulation d'une réalité indépendante et atemporelle qui est plus vaste que l'ensemble des entités existantes et, d'autre part, de la distinction entre l'*acte* de jugement et l'*objet* du jugement. En gros, la connaissance n'est pas un acte créatif, il est passif; connaître consiste à nous familiariser avec une réalité préexistante[3] qui ne dépend pas de notre façon de l'appréhender.[4] De même, la vérité ne procède pas par degrés, allant de la contradiction jusqu'à l'absolu (qui nous est ultimement inaccessible), elle est franche et nette : une proposition est soit vraie soit fausse, rien ne se situe entre le deux.[5] Russell et Moore font, par ailleurs, valoir que la réalité est décomposable en éléments « atomiques » et que le véritable travail de la connaissance consiste à isoler, c'est-à-dire analyser, les éléments simples qui composent une relation, à l'inverse des idéalistes pour qui la totalité est primitive et les éléments isolés d'un complexe sont conçus comme le résultat d'une abstraction. Pour l'ensemble de ces raisons, la thèse philosophique défendue par Moore et Russell est souvent qualifiée comme étant un *atomisme platonicien* (ou logique) et leur méthode est dite *analytique*.

---

[1] Bien que Bradley ait été le porte-parole principal de l'idéalisme en Angleterre, McTaggart (à Cambridge) est celui qui a vraisemblablement introduit l'idéalisme à Russell et Moore.
[2] Hylton, 1990 : p. 130.
[3] Cette réalité n'est pas préexistante dans le sens d'une existence empirique perçue, elle est préexistante dans le sens où elle *est* indépendamment de nous.
[4] Le terme « familiarisation » est une traduction libre du terme russellien « acquaintance », qui décrit selon lui la relation épistémique qui nous avons à l'endroit de la réalité.
[5] Moore et Russell prétendent que la vérité par degrés doit ultimement reposer sur un concept de vérité « binaire ».

Il revient à Moore d'avoir développer les grandes lignes de l'atomisme platonicien et à Russell d'avoir appliquer cette réforme philosophique à la logique. Dans la préface des *Principles of Mathematics*, Russell reconnaît la dette qu'il a à l'endroit de son collègue : « Sur les questions fondamentales de la philosophie, toutes les caractéristiques principales de ma position proviennent de celle de M. G. E. Moore » (Russell, 1989 : p. 5). De sa philosophie, Russell nous dit :

> J'en ai retenu la nature non existentielle des propositions [...], et leur indépendance par rapport à l'esprit connaissant; et aussi le pluralisme qui considère le monde, tant celui des existants que celui des entités, comme composé d'un nombre infini d'entités mutuellement indépendantes et de relations qui sont fondamentales et irréductibles à des adjectifs de leurs termes ou du tout que ceux-ci constituent. (Russell, 1989 : p. 6)

Russell prend comme confirmation de ces thèses le fait qu'elles rendent possible une philosophie des mathématiques, ce que l'idéalisme, avec ses tendances psychologisantes ou panpsychiques, ne pouvait faire. Le psychologisme en philosophie, autant chez les empiristes que chez les idéalistes, est responsable, selon lui, des errements philosophiques en matière de logique et de mathématiques, car il induit à penser que les mathématiques sont fondées dans la pensée ou l'intuition du sujet connaissant, qu'elles sont un objet de sa psyché. Comme nous l'avons vu chez Frege, ceci mène à certaines difficultés, dont l'impossibilité de rendre compte des vérités analytiques. L'analytique chez Russell (comme chez Frege) est sauvé par le réalisme logique : une proposition n'est pas vraie ou fausse en vertu de l'existence d'un fait ou d'une pensée (dans le sens non frégéen), elle l'est en vertu de la réalité (de l'être). Le logicisme est, en quelque sorte, le corollaire de cette réalité. Il représente l'aboutissement d'une tendance vers la réduction progressive des mathématiques à la logique : après la réduction de l'analyse à des opérations logiques sur des ensembles de nombres réels par Cauchy et Weierstrass, la réduction des réels à des ensembles de rationnels par Dedekind et la réduction des rationnels à des ensembles d'entiers naturels, il semblait naturel de croire qu'on pouvait traduire la notion d'ensemble et de nombre naturel pour faire des mathématiques un chapitre de la logique. C'était là l'objectif de Russell.

Les *Principles of Mathematics* (1902) marque le début de l'élaboration de sa philosophie mathématique. Les *Principles* sont suivis, notamment, de *On Denoting*, où Russell expose sa théorie des description définies qui sera essentielle à la réduction de la notion de classe à la logique, et de *Mathematical Logic as Based on the Theory of Types*, où l'on trouve une solution au paradoxe du menteur. Tout ce travail culmine par les *Principia Mathematica*, rédigés avec A. N. Whitehead,[6] qui constituent – si j'ose dire – la première réduction complète des mathématiques à la logique.[7] Sur le plan logique, Russell est influencé surtout par Cantor et Peano, Frege n'apparaît dans son paysage que plus tard, seulement après qu'il ait terminé la plus grande partie des *Principles*. L'influence de Peano est manifeste au premier coup d'oeil dans les *Principia*, le symbolisme est le même : l'emploi des points pour la ponctuation, l'utilisation du fer à cheval pour l'implication, d'un « v » pour la disjonction, etc. L'ensemble de l'œuvre est traversé par la même conviction logiciste qui, je trouve, est bien résumé au paragraphe **1** des *Principles* :

> **1.** La mathématique pure est la classe de toutes les propositions de la forme « *p* implique *q* », où *p* et *q* sont des propositions contenant une ou plusieurs variables, les mêmes dans les deux propositions, et où ni *p* ni *q* ne contiennent d'autre constantes que des constantes logiques. Et les constantes logiques sont toutes ces notions qui peuvent être définies au moyen de l'implication, de la relation d'un terme à une classe dont il est membre, de la notion *tel que*, de la notion de relation, et de toutes les autres notions que peut impliquer celle, générale, de proposition de cette forme. En outre la mathématique *fait usage* d'une autre notion qui n'est pas un constituant des propositions qu'elle considère, à savoir celle de vérité. (Russell, 1989 : p. 22)

[6] Dans la suite, je ne mentionnerai que Russell lorsqu'il sera question des *Principia*, puisqu'il est l'auteur de ses conceptions philosophiques.

[7] Cette affirmation ne fait pas l'unanimité du fait qu'on retrouve, d'après certains, des axiomes non logiques dans les *Principia* tels que l'axiome de l'infini et l'axiome de réductibilité.

Il y a essentiellement deux choses qui distinguent les *Principia* des *Principles* : d'une part, un plus grand aboutissement technique, et d'autre part, une hiérarchie d'entités imposée par le paradoxe du barbier. La logique mathématique des *Principia* vise les mêmes objectifs que la *Begriffsschrift* de Frege, en plus de vouloir corriger les paradoxes :

> En premier lieu, elle vise à fournir l'analyse la plus poussée possible des idées dont elle traite et des procédés de démonstrations qu'elle utilise, ainsi que diminuer autant que faire se peut le nombre d'idées non définies et de propositions non démontrées [...] En second lieu, elle est conçue de manière à fournir, au moyen de ses symboles, une expression parfaitement précise de propositions mathématiques [...] En troisième lieu, le système est spécialement conçu pour résoudre les paradoxes [...] (Russell 1989 : p. 221)

Prouver qu'un système formel est dénué de contradictions (le troisième point) prendra une importance considérable par la suite, Hilbert en avait déjà fait le deuxième problème à résoudre par la communauté mathématique au Congrès de Paris en 1900. Pour Russell, la solution passe par la théorie des types, laquelle aura pour conséquence de complexifier considérablement son symbolisme. Nous verrons comment cette exigence de résolution des paradoxes sera une porte d'entrée à ses conceptions métalinguistiques.

2. Le cercle vicieux et la théorie des types

Puisque la théorie des types a un impact majeur sur la philosophie de Russell, commençons notre analyse des *Principia* par le problème du cercle vicieux et par sa solution. D'abord, contre quoi Russell part-il en croisade? Le problème peut être formulé de la manière suivante. Considérons les concepts θ et ψ définis comme

$\theta(\varphi).\equiv.\varphi \in \hat{x}(\varphi(x))$ et $\psi(\varphi).\equiv.\sim \varphi \in \hat{x}(\varphi(x))$,[8] c'est-à-dire $\theta(\varphi)$ si et seulement si le concept φ appartient à sa propre extension, et $\psi(\varphi)$ si et seulement si φ n'y appartient pas. Par le tiers exclu (et le caractère illimité de φ), nous avons $\psi(\psi)$ ou $\sim \psi(\psi)$. Dans le premier cas, par l'identité $\sim \varphi \in \hat{x}(\varphi(x)).\equiv.\varphi \in \hat{x}(\sim \varphi(x))$, il s'ensuit que $\sim \psi(\psi)$. Dans le deuxième, nous avons que $\sim \psi(\psi).\equiv.\psi \in \hat{x}(\psi(x))$ qui entraîne à son tour, par la définition de l'opérateur $\hat{x}(-)$,[9] que $\psi(\psi)$. Ensembles, ces deux conclusions mènent à une contradiction. Bien que Russell ait exposé ce paradoxe dans les *Principles*, il ne fait qu'esquisser les grande lignes de ce qui devrait constituer une solution :

> La solution suggérée consiste à reconnaître qu'il est nécessaire de distinguer divers types d'objets, à savoir les termes, les classes de termes, les classes de classes, les classes de couples de termes, etc.; et aussi qu'une fonction propositionnelle φx exige en général, pour être dotée d'un sens, que x appartienne à un type déterminé. (Russell 1989 : p. 156)

Plus tard dans les *Principia*, Russell arrivera à une thèse plus affinée sur les paradoxes :

> L'analyse des paradoxes à éviter montre qu'ils résultent tous d'une certaine espèce de cercle vicieux. Ces cercles vicieux surgissent de la supposition qu'une collection d'objets peut contenir des membres qui ne peuvent être définis qu'au moyen de la collection prise comme tout. (Russell 1989 : p. 270)

En conséquence, le principe – dit du cercle vicieux – auquel il faudrait se soumettre est le suivant :

---

[8] J'emploie ici la notation russellienne. Pour plus de précisions sur la lecture des points, consulter le premier chapitre de l'introduction des *Principia*.

[9] J'utilise une notation russellienne mais la définition de l'extension à laquelle je fais référence est celle de l'axiome V des *Grundgesetze* de Frege.

> « Tout ce qui met en jeu *tout* d'une collection ne doit pas être un des éléments de la collection »; ou, inversement : « si, une certaine collection ayant un total, certains de ses membres ne sont définissables qu'au moyen de ce total, alors cette collection n'a pas de total ». (Russell 1989 : p. 271)

Énoncé tel quel, ce principe est encore vague. Sa formulation précise et définitive sera incarnée par la hiérarchie des types ainsi que les contraintes de prédicativité imposées aux fonctions propositionnelles du système.

Nous avons vu que le talon d'Achille de Frege était le caractère omnipotent et universel de ses variables. Ce pouvoir référentiel absolu est présent chez Russell aussi. La notion de variable qui ressort des *Principles* est essentiellement l'idée que la variable est ce qui dénote n'importe quel objet. Malgré les paradoxes, cette idée ne quitte pas Russell dans les *Principia* :

> L'idée de variable, telle qu'elle figure dans cet ouvrage, est plus générale que celle dont fait explicitement usage la mathématique ordinaire. Dans la mathématique ordinaire, une *variable* représente généralement un nombre indéterminé ou une quantité. En logique mathématique, n'importe quel symbole dont le sens n'est pas déterminé est appelé une variable, et les diverses déterminations dont ce sens est susceptible sont appelées les *valeurs* de la variable. Les valeurs peuvent être, selon les circonstances, un ensemble quelconque d'entités, de propositions, de fonctions, de classes ou de relations. [...] Nous pouvons appeler une variable une variable *restreinte*, quand ses valeurs sont limitées à certaines seulement de celles qu'elle peut avoir; autrement nous disons qu'elle est *sans restriction*. [...] Du point de vue de la logique, la variable sans restriction est plus commode que la variable restreinte, et nous n'emploierons qu'elle. [...] [L]es limitations auxquelles est sujette la variable sans restriction n'ont pas besoin d'être explicitement indiquées, puisque ce sont les limites de la signification de l'énoncé dans laquelle il figure, et que par conséquent elles sont intrinsèquement déterminées par cet énoncé. (Russell 1989 : p. 225-6)

Russell minimise dans ce passage (à la fin) l'importance qu'a la théorie des types sur la limitation des variables, ce qui est pour le moins étonnant, considérant les répercussions qu'elle a sur la complexité de son symbolisme. Russell transfert à la proposition (ou la fonction propositionnelle) le sort de décider de la limitation des variables.

En réalité, la limitation se fait davantage par l'ontologie sous-jacente aux *Principia*. L'idée est que l'univers est stratifié par des ordres, allant de zéro à n'importe quel nombre naturel, que les entités d'ordre supérieur sont ontologiquement dépendantes des entités d'ordre inférieur, et que, par conséquent, il ne faut pas définir les dernières en fonction des premières sans quoi nous risquons de tourner en rond. En effet, si une fonction propositionnelle contient une variable apparente du même rang que le sien, elle est une valeur possible de cette variable, donc elle est définie en fonction d'elle-même : selon l'exemple préféré de Russell, « avoir toutes les qualités d'un excellent général » est elle-même une qualité d'un excellent général, pour la définir sans compromettre le principe du cercle vicieux il aurait fallu que cette qualité soit déjà définie. Pour remédier à ce problème, il faut stratifier les qualités que peut posséder un général : on aurait, à la base, des qualités de premier ordre, ensuite des qualités de second ordre, etc.

On peut comprendre la hiérarchie des fonctions en procédant par étapes. D'abord, convenons d'appeler matrice une fonction propositionnelle n'ayant aucune variable apparente. Les matrices seront importantes pour définir les fonctions appartenant à un ordre. Au plus bas de l'échelle, nous retrouvons les matrices de premier ordre, c'est-à-dire les matrices $\theta(x, y,...)$ ne prenant que des individus pour arguments. Toute fonction obtenue d'une matrice de premier ordre $\theta(x, y,...)$ en quantifiant sur ses variables, telles que $\exists x.(y).\theta(x, y,...)$ ou $(x).\theta(x, y,...)$ par exemple, sera également dite de premier ordre. Le symbole $\varphi!\hat{x}$ sera utilisé comme variable pour une fonction de premier ordre. On remarquera que de telles fonctions présupposent au plus des totalités d'individus. Ensuite viennent les matrices de second ordre qui sont des matrices ayant une ou plusieurs variables qui ont pour arguments des fonctions de premier ordre (et possiblement des variables d'individus). Les fonctions obtenues d'une matrice du second ordre en quantifiant soit sur des variables de fonctions (de premier ordre) soit sur des variables d'individus seront également de second ordre. On remarquera ici que ces fonctions présupposent au plus des totalités de fonctions de premier ordre. On remarquera également qu'elles peuvent avoir des variables d'individus ou des variables de fonctions de premier ordre. Cette situation fait en sorte qu'il existe plusieurs sortes de variables de second ordre : par exemple, $f!(\varphi!\hat{x})$ sera le symbole de variable pour une fonction de second ordre d'une seule variable d'une fonction de premier ordre; $f!(\varphi!\hat{x},x)$ pour une fonction de second ordre de deux variables, l'une qui prend des fonctions de premier ordre et l'autre des individus; $f!(\varphi!\hat{x},\psi!\hat{x},x)$ pour une fonction (du second ordre) de trois variables, deux qui prennent des fonctions de premier ordre et l'une qui prend des individus; etc. La notion de type intervient ici pour différencier ces cas : les variables $f!(\varphi!\hat{x})$, $f!(\varphi!\hat{x},x)$ et $f!(\varphi!\hat{x},\psi!\hat{x},x)$ sont toutes de même ordre, mais elles définissent par contre chacune un type différent. Pour la définition des fonctions de troisième ordre et plus, on procède de manière analogue.

La notion d'ordre débouche sur un autre concept clé des *Principia* qui sert à défendre la logique mathématique contre l'apparition des contradictions : la notion de prédicativité. Une fonction propositionnelle d'une seule variable sera dite prédicative si elle est d'un ordre immédiatement supérieure à celui de sa variable. Pour une fonction de plusieurs variables, si n est l'ordre le plus élevé parmi les variables, elle sera dite prédicative si elle est d'ordre $n+1$. Par exemple, la matrice de second ordre $F(\varphi!\hat{x},x)$ est prédicative car $\varphi!\hat{x}$, sa variable ayant l'ordre le plus élevé, est de premier ordre. Par contre, la fonction du second ordre $(\varphi).F(\varphi!\hat{x},x)$ n'est pas prédicative car sa variable (libre) d'ordre le plus élevé, x en l'occurrence, est d'ordre nul. D'ailleurs, $(\varphi).F(\varphi!\hat{x},x)$ est précisément le type de fonction qu'on souhaite éviter parce qu'elle présuppose (par le quantificateur (φ)) une totalité de fonctions du premier ordre à laquelle elle appartient. Il va sans dire que les seules fonctions admissibles dans les *Principia* sont les fonctions prédicatives.

Quelques mots sur le paradoxe du barbier. En hiérarchisant son ontologie par des ordres, Russell se trouve à empêcher l'écriture même du paradoxe. La (pseudo) fonction propositionnelle $\psi(\varphi)$ définie par $\sim \varphi \in \hat{x}(\varphi(x))$ (qui est à l'origine du paradoxe) n'en est pas une dans le système des *Principia* parce $\varphi \in \hat{x}(\varphi(x))$ est équivalent à dire $\varphi(\varphi)$, ce qui ne respecte le fait que la variable d'une fonction est d'ordre strictement inférieur à cette fonction. Le paradoxe est donc impossible à formuler. Il est important d'observer que la contrainte de prédicativité n'est pas nécessaire pour prévenir ce paradoxe. La prédicativité vient s'ajouter comme défense supplémentaire contre l'apparition de futurs paradoxes pouvant naître du cercle vicieux. Exiger que les fonctions propositionnelles soient prédicatives n'est pas sans conséquences : certaines définitions mathématiques fondamentales sont impossibles à formuler de manière prédicative, comme les notions de borne supérieure ou inférieure d'un ensemble de nombres réels qui sont essentielles au développement de l'analyse. La capacité des *Principia* à inclure l'ensemble des mathématiques est sérieusement compromise par cette situation, et Russell fera appel à un étrange compromis pour sauver les apparences : l'axiome de réductibilité. Ainsi, Russell nous dit :

> L'axiome de réductibilité est introduit pour légitimer une grande masse de raisonnements dans lesquels nous avons à première vue affaire à des notions telles que « toutes les propriétés de *a* » ou « toutes les fonctions-*a* », et dans lesquelles, néanmoins, il semble difficilement possible de déceler une quelconque erreur d'importance. (Russell 1989 : p. 295).

L'axiome de réductibilité s'énonce comme suit : $(\exists \psi) : (x). \varphi x \equiv \psi ! x$, c'est-à-dire, pour toute fonction $\varphi \hat{x}$, il existe une fonction $\psi \hat{x}$ dont l'ordre est immédiatement supérieure à l'ordre de sa variable (ce qui revient à dire qu'elle est prédicative) et telle que $\varphi x \equiv \psi x$ pour toute valeur de x. (Intuitivement, il affirme que toute combinaison de prédicats est équivalente à un prédicat unique.) Il y a dans cet axiome une certaine ambiguïté volontaire sur les types, l'important étant les types relatifs entre x et $\psi \hat{x}$ et non pas leurs types absolus. Il faut donc voir que l'axiome vient en tout ordre et en tout type, l'ambiguïté étant une façon de rassembler sous une même expression les différentes possibilités.

L'axiome de réductibilité est sans l'ombre d'un doute l'élément le plus controversé des *Principia*. Il vient discréditer la portée des arguments philosophiques contre l'imprédicativité et mine sérieusement la posture « anti-classe » de Russell. Pour sa défense, Russell avance que l'axiome est une hypothèse moins forte que l'existence des classes, qu'il ne soutient aucunement d'ailleurs. En effet, si l'ensemble des classes et des relations existaient, on pourrait facilement associer à un prédicat imprédicatif $F(x)$ un prédicat prédicatif simple $G(x)$ en considérant la classe A des objets qui satisfont $F(x)$ et en définissant $G(x)$ comme $x \in A$. Ainsi, sur le plan du rasoir d'Occam, il faut donner raison à Russell : l'axiome est une aubaine ontologique. Mais ce qui est plus déconcertant est la justification philosophique de l'axiome qui, à mon sens, compromet la conception réaliste dans son ensemble :

> La raison pour laquelle nous acceptons un axiome, comme n'importe quelle proposition, est toujours largement inductive, c'est-à-dire que de nombreuses propositions, qui sont presque indubitables, peuvent en être déduites, et que, si l'axiome est faux, on ne connaît rien de plus plausible qui pourrait rendre ces propositions vraies, et enfin que rien de probablement faux ne peut en être déduit. [...] On ne peut jamais atteindre l'infaillibilité, et par conséquent chaque axiome est toujours entaché de quelque élément de doute, et toutes ses conséquences aussi. (Russell 1989 : p. 300)

Cette conception pragmatique et inductive de la justification des axiomes tend à remettre en question le concept de vérité réaliste et ouvrir le chemin vers une approche plus instrumentaliste ou formaliste de la logique.

3. Le roi de France et sa calvitie

Nous avons vu jusqu'à présent certaines particularités de l'ontologie russellienne. Elle est composée d'entités hiérarchisées par un ordre : à l'ordre zéro, au plus bas de l'échelle, il y a les individus, et viennent ensuite les fonctions propositionnelles d'ordre un, deux, trois, etc. dans lesquelles figurent des variables d'ordre de plus en plus élevé. Toute fonction appartient à un et un seul ordre et il est possible de discriminer les fonctions à l'intérieur d'un ordre selon leur type, c'est-à-dire selon leurs arguments. D'autre part, les fonctions d'ordre plus élevé entretiennent une dépendance ontologique envers les individus ou les fonctions d'ordre moins élevé. L'ontologie frégéenne est plus simple : elle est constituée d'objets qui sont tous du même rang et ordre, et ils sont les seules entités susceptibles d'être dénotées par des termes du symbolisme. Ce que sont dans l'ontologie frégéenne les concepts (l'équivalent frégéen des fonctions propositionnelles) n'est pas très clair, et Frege s'épargne la difficulté de préciser ce qu'elles sont en développant un langage portant seulement sur des extensions.[10] Pour Russell, il est hors de question d'accorder une réalité aux classes et aux relations, parce qu'il estime qu'elles sont en partie responsables des paradoxes mathématiques. D'autant plus que l'élimination des classes est une condition *sine qua non* de la réduction des mathématiques à la logique.

La croisade contre les classes commence par une analyse des descriptions définies. L'idée est d'éliminer le concept de description définie pour éviter, d'une

[10] Les concepts sont néanmoins chez Frege des entités indépendantes de l'expérience et de l'esprit par le simple fait qu'ils expriment un sens.

part, d'avoir des entités constantes dans l'ontologie et faire en sorte, d'autre part, qu'une proposition dans laquelle apparaît un terme sans dénotation est fausse au lieu d'être sans signification. Frege soutenait que la dénotation d'un nom est fonction de la dénotation des noms qui le composent et qu'ainsi, il suffisait de démontrer que les noms primitifs du système avait une dénotation pour s'assurer que tous les noms du système en avait une. Par exemple, selon la sémantique frégéenne, la proposition *L'actuel roi de France est chauve* aura une dénotation seulement si le concept « x est actuellement un roi de France » possède un et un seul objet dans son extension. Sinon, la proposition en question est sans dénotation, ce qui veut dire qu'elle est ni vraie ni fausse (puisque le vrai et le faux sont des objets de dénotation).[11] Russell développe une conception qui évite cette tierce position, il reconstruit les propositions dans lesquelles figurent des descriptions définies de sorte qu'elles aient toujours une valeur de vérité et que cette valeur soit le faux dans le cas où la description définie n'ait pas de dénotation. Le stratagème de Russell consiste à reconstruire la proposition « L'actuel roi de France est chauve » en remplaçant la description définie par une combinaison astucieuse de quantificateurs :

« Il existe un x tel que x est actuellement un roi de France et, pour tout y, si y est actuellement un roi de France, $x = y$, et x est chauve ».

La proposition ci-dessus se passe non seulement de l'opération de description définie, mais elle possède une valeur de vérité dans toutes les situations : elle est vraie quand il existe effectivement un et un seul individu tel que ... et elle est fausse lorsqu'il n'existe aucun individu ou lorsqu'il en existe plusieurs tels que ... Formellement, si $\varphi(x)$ et $\psi(x)$ sont des prédicats et $\iota(x)\varphi(x)$ représente l'unique individu qui satisfait $\varphi(x)$, la proposition « L'unique individu qui satisfait φ est un ψ » sera représentée

[11] Une proposition peut avoir un sens même si elle n'a pas de dénotation.

par $\psi(\iota(x)\varphi(x))$. La reconstruction précédente appliquée à $\psi(\iota(x)\varphi(x))$ prendra donc la forme $\exists x : \varphi(x).\psi(x).(y).\varphi(y) \supset x = y$.[12]

S'il est vrai qu'un des objectifs de ce stratagème est de faire des économies ontologiques, il n'est pas le seul ni le plus important. Ce que Russell recherche dans cet exemple est une justification de la thèse sémantique à l'effet que les fonctions non propositionnelles (de manière général) n'ont pas de signification isolée (et qu'elles n'en ont pas besoin). La reconstruction du paragraphe précédent nous permet de faire usage de l'expression $\iota(x)\varphi(x)$ dans $\psi(\iota(x)\varphi(x))$ sans la définir à l'extérieur du contexte propositionnel. Cette thèse, qui manifeste certaines ressemblances avec le principe du contexte chez Frege,[13] sera appliquée intégralement aux classes :

> Les symboles des classes, comme ceux des descriptions, sont dans notre système des symboles incomplets : leurs *usages* sont définis, mais eux-mêmes sont supposées ne rien vouloir dire du tout. C'est-à-dire que les usages de ces symboles sont définis de telle sorte que, quand le *definiens* est substitué au *definiendum*, il ne reste aucun symbole qui puisse être supposé représenter une classe. Aussi les classes, dans la mesure où elles sont introduites, ne le sont que comme des commodités purement symboliques ou linguistiques, et non comme des objets authentiques tels que le sont leurs membres quand ce sont des individus. (Russell 1989 : p. 317)

4. Le système

On aura donc compris que les fonctions propositionnelles occupent une place d'importance dans les *Principia*, mais il est temps maintenant de voir comment le symbolisme s'articule autour d'elles. Trois calculs sont nécessaires à l'accomplissement de l'entreprise russellienne : le calcul des propositions, le calcul

[12] En notation moderne : $\exists x(\varphi(x) \wedge \psi(x) \wedge \forall y(\varphi(y) \supset x = y))$.

[13] « On doit chercher ce que les mots veulent dire non pas isolément mais pris dans leur contexte » (Frege 1969 : p. 122).

des fonctions propositionnelles et le calcul des classes et des relations. Le premier calcul introduit les constantes logiques : la négation « $\sim$ », la somme logique « $\vee$ », le produit logique « $\cdot$ » et l'implication « $\supset$ », les deux premières étant les constantes primitives du système.[14] Les fonctions de propositions jouissent d'un statut particulier dans le système : elles sont les seules fonctions constantes, les seules fonctions qui prennent des propositions comme arguments et elles ne sont pas typées.[15] La règle d'inférence fondamentale peut être formulée dans ce calcul : de « $\vdash p$ » et « $\vdash p \supset q$ » (ou « $\vdash \sim p \vee q$ »), on peut déduire $\vdash q$.

Des fonctions propositionnelles, Russell nous dit qu'elles « sont l'espèce fondamentale de fonction à partir desquelles les espèces les plus courantes, telles que 'sin x', ou 'log x', ou 'le père de x', sont dérivées ».[16] En guise de définition, Russell nous donne :

> Par « fonction propositionnelle » nous entendons quelque chose qui contient une variable, et exprime une *proposition* aussitôt qu'une valeur est assignée à x. C'est-à-dire qu'elle diffère d'une proposition par cela seul qu'elle est ambiguë : elle contient une variable dont la valeur n'est pas assignée. (Russell 1989 : p. 272)

Le calcul des fonctions propositionnelles, en plus d'incorporer le calcul des propositions, comprend les quantificateurs existentiel « $(\exists x)$ » et universel « (x) ». Avec le quantificateur universel vient une règle d'inférence supplémentaire, la règle de généralisation, laquelle stipule que de « $\vdash . \varphi x$ » on peut déduire « $\vdash : (x). \varphi x$ ». La présence d'une variable réelle dans l'assertion « $\vdash . \varphi x$ » oblige Russell à introduire une autre ambiguïté (volontaire) dans son système : l'assertion ambiguë. Le contenu

---

[14] $p \supset q$ se traduit évidemment par $\sim p \vee q$ et $p \cdot q$ par $\sim(\sim p \vee \sim q)$. Les constantes supplémentaires sont préservées dans les *Principia* pour faciliter l'écriture et la lecture des propositions.

[15] Le fait que les constantes logiques ne soient pas typées participe, aux dires de Russell, à l'ambiguïté typique des *Principia*. En fait, l'ambiguïté ici est plus fondamentale que dans le cas de l'axiome de réductibilité. Il n'aurait pas été possible de développer un calcul des propositions viable avec des propositions typées.

[16] Russell 1989 : p. 240.

affirmé dans cette assertion n'est pas une proposition définitive, il est la valeur indéterminée de la fonction propositionnelle.[17]

Le calcul des fonctions propositionnelles introduit aussi la notion d'égalité « = ». En apparence simple, celle-ci pose certaines difficultés dans les *Principia* par le fait que les fonctions propositionnelles sont typées. D'une part, la hiérarchie des types nous contraint de définir une relation d'égalité pour chaque type, puisqu'une fonction propositionnelle donnée a des variables d'un type donné. Donc, en matière d'égalité, une ambiguïté typique flottera aussi. D'autre part, afin de rendre à l'égalité toute la généralité du principe des indiscernables de Leibniz, il faut appliquer l'axiome de réductibilité. Le principe des indiscernables défini l'égalité de la manière suivante :

$$x = y . =_{Df} : (\varphi) . \varphi x \supset \varphi y ,$$

où φ est une fonction quelconque d'une variable (pas nécessairement prédicative). Une variable de fonction aussi générale que φ n'existe pas dans les *Principia*, le mieux qu'on puisse faire comme définition de l'égalité est :

$$x = y . =_{Df} : (\varphi) . \varphi ! x \supset \varphi ! y .$$

C'est-à-dire x et y sont égaux si et seulement si, lorsqu'un prédicat prédicatif est vrai de x, il est vrai de y aussi. Si l'axiome de réductibilité était faux, il pourrait exister un prédicat imprédicatif vrai de x mais faux de y. Sinon, en utilisant l'axiome, on peut démontrer que les deux définitions coïncident : si ψ est un prédicat quelconque, il existe un prédicat prédicatif équivalent à ψ qui est dans le champ (prédicatif) de la variable φ.

---

[17] Russell se débarrasse de ce subterfuge philosophique dans l'introduction à la seconde édition des *Principia*.

Nous arrivons enfin à la théorie des classes et des relations qui est, quantitativement parlant, la plus importante de l'ensemble des *Principia* et ce, malgré le fait que les classes ne disposent d'aucune réalité ontologique. C'est à l'intérieur de ce calcul qu'on retrouve les opérations de description définie « $\iota(x)\varphi(x)$ » et d'extension « $\hat{x}(\varphi(x))$ » ainsi que la notion d'appartenance « $\in$ » à partir desquelles seront définies notamment les notions d'inclusion « $\subset$ », d'union « $\cup$ », d'intersection « $\cap$ » et de complémentation « $-$ ». Ces dernières seront fondamentales par la suite pour le développement de l'arithmétique cardinale et ordinale, puisque Russell a une arithmétique basée sur les classes. En effet, la définition du nombre 1 au paragraphe *52.01, pour ne donner qu'un aperçu, est essentiellement ensembliste :

***52.01** $1 = \hat{\alpha}\{(\exists x).\alpha = \iota' x\}$ (où $\iota' x = \hat{\beta}(\beta = x)$).

(Intuitivement, 1 est défini comme étant l'ensemble de toutes les classes ayant un seul élément.)

N'oublions point toutefois que les classes sont éliminables, c'est-à-dire qu'il est possible de reconstruire l'ensemble des propositions dans lesquelles elles figurent en faisant seulement appel à des fonctions propositionnelles. Par exemple, l'opérateur d'extension « ^ » et la relation d'appartenance « $\in$ » sont définis contextuellement par l'expression $x \in \hat{z}(\varphi(z)). \equiv .\varphi(x)$, ce qui les rend éliminables en tout temps au profit de $\varphi(x)$. Les pseudo-classes de Russell ont seulement une valeur heuristique, elles facilitent l'expression de l'arithmétique.

5. L'assertion et la vérité

Dans le premier paragraphe des *Principles*, Russell affirme que « la mathématique *fait usage* d'une autre notion qui n'est pas un constituant des propositions qu'elle considère, à savoir celle de la vérité ».[18] Contrairement à toutes les autres notions fondamentales au développement des *Principles* – et plus tard des *Principia* – qui, elles, sont des composantes de la proposition ou de la fonction propositionnelle, la notion de vérité est externe. On ne trouvera malheureusement pas dans les *Principles* une analyse détaillée du concept de vérité, malgré que de nombreuses allusions au concept y soient faites. De la relation entre vérité et proposition, on peut lire dans l'introduction à la première édition que[19]

> les propositions sont habituellement considérées comme (1) vraies ou fausses, (2) mentales. Considérant que ce qui est vrai ou faux n'est généralement pas mental, j'ai besoin d'un nom pour le vrai et le faux en tant que tels, et ce nom peut difficilement être autre chose que *proposition*. (Russell 1989 : p. 7)

Ainsi, la proposition serait, selon ce passage, un nom pour le vrai ou le faux ou, inversement, le vrai ou le faux est ce qui est nommé par une proposition. Cette conception épouse de près celle de Frege, pour qui la valeur de vérité est un objet dénoté par un type particulier de nom propre, la proposition.

Dans l'appendice A des *Principles*, Russell rejette la conception frégéenne de la vérité et de l'assertion, et en particulier l'idée que les valeurs de vérité sont indiquées par des propositions. Deux arguments sont donnés contre cette conception. D'abord, il prétend que la distinction entre vérité et assertion, sur laquelle cette dernière repose, est psychologique. Psychologiquement, on peut choisir d'affirmer ou non une proposition vraie (l'on peut même affirmer une proposition fausse); mais

[18] Russell 1989 : p. 21.
[19] Des *Principles*.

dans le sens logique de l'assertion, toutes les propositions vraies et seulement les propositions vraies sont affirmées. Si Frege a cru bon de séparer le jugement de la dénotation de la vérité, c'est qu'il a été leurré par une conception psychologisante du jugement, mais en réalité

> [i]l est quasiment impossible, du moins à mes yeux, de séparer l'assertion de la vérité ainsi que le fait Frege. Une proposition affirmée, me semble-t-il, doit être identique à une proposition vraie. [...] [S]éparer l'assertion de la vérité ne semble possible qu'à la condition de prendre l'assertion dans son sens psychologique. (Russell 1989 : p. 164)

Ainsi, il semblerait selon cette critique que *vérité* et *assertion*, dans leurs sens logiques, soient des concepts interchangeables.

L'autre argument que Russell avance contre la conception frégéenne de la vérité se trouve dans le passage suivant :

> La théorie de Frege selon laquelle les suppositions [les propositions] sont des noms propres du vrai et du faux, selon les cas, me semble également intenable. Un examen direct montre apparemment que la relation entre une proposition et le vrai et le faux est tout à fait différente de celle qu'il y a entre (par exemple) « l'actuel roi d'Angleterre » et Edouard VII. De plus, si le point de vue de Frege est correct, il nous faut dire que, dans une proposition affirmée, c'est le sens et non pas l'indication qui est affirmé, car sinon toutes les propositions affirmées affirmeraient la même chose à savoir le vrai (car les propositions fausses ne sont pas affirmées). Aussi les propositions ne différeraient-elles en aucune façon l'une de l'autre et seraient toutes rigoureusement et parfaitement identiques. (Russell 1989 : p. 164)

Si la théorie dénotationnelle de la vérité est défendable pour Frege, c'est que la proposition est un nom propre et la diversité propositionnelle est une affaire de langage seulement.[20] Pour Russell, la situation est radicalement différente, car il existe une diversité propositionnelle à même son ontologie. Chaque proposition étant une entité individuelle, dire qu'il y a seulement le vrai et le faux comme objets propositionnels reviendrait à dire qu'on ne peut exprimer que deux propositions, ce qu'il juge « intenable ». Russell conclut son argumentation en affirmant qu'il « semble inutile d'introduire ici la valeur de vérité, et suffisant de dire qu'une proposition affirmée est une proposition dont le sens est vrai, et dire que le sens est vrai, c'est dire la même chose que le sens est affirmé ».[21] On retrouve donc, encore une fois, le rapprochement entre vérité et assertion.

Curieusement, le Russell des *Principia* revient à une position plus conservatrice. En se référant à Frege, il reprend la notion de valeur vérité qu'il utilisera notamment pour introduire les fonctions de propositions : « La 'valeur de vérité' d'une proposition est le *vrai* si elle est vraie, et le *faux* si elle est fausse ».[22] Sur la question de l'assertion, il adopte une vue similaire à celle des *Principles* mais plus discrète quant à son rapport avec la vérité :

> [Le signe d'assertion] est nécessaire pour distinguer une proposition complète que nous affirmons de toutes les propositions subordonnées qu'elle peut contenir mais qui ne sont pas affirmées. Dans l'écriture en langage ordinaire, une phrase encadrée par des points dénote une proposition affirmée, et si elle est fausse, le livre où elle figure est dans l'erreur. Le signe « $\vdash$ » préfixé à une proposition joue le même rôle dans notre symbolisme. (Russell 1989 : p. 231)

Le signe d'assertion « $\vdash$ » sert seulement à distinguer la proposition qui est affirmée des autres propositions constituantes. L'équivalence entre l'assertion et la vérité est donnée sous une forme métaphorique : si ce signe est placé devant une proposition

[20] Pour être honnête à l'endroit de Frege, il faut dire que la notion de sens (*Sinn*) est un moyen de rendre compte de la diversité propositionnelle. Toutefois, comme nous le savons, Frege n'élabore guère sur cette notion après la *Begriffsschrift*.
[21] Russell 1989 : p. 164.
[22] Russell 1989 : p. 230.

fausse, les auteurs responsables sont dans l'erreur. On pourrait croire que Russell tient pour implicite ce qu'il avait établi dans l'appendice des *Principles* et qu'il ne juge pas nécessaire de le répéter, mais ce n'est pas le cas. À la section A des *Principia*, Russell affirme que « [l]e signe '⊢' est le appelé le signe d'assertion*; on peut le lire comme 'il est vrai que' (même si philosophiquement parlant ce n'est pas sa signification exacte) ».[23] Qui plus est, l'astérisque renvoie à la note suivante : « Nous [Russell et Whitehead] avons adopté à la fois l'idée et le symbole d'assertion de Frege ».[24] Non seulement on affirme ici une différence philosophique entre l'assertion et la vérité, mais on dit de ces deux notions qu'elles proviennent directement de Frege.

Malgré les remarques philosophiques, dans l'usage, le concept de vérité est interchangeable avec le concept d'assertion. En témoignent les explications données des différentes définitions et lois élémentaires suivantes :

> (5) *Négation.* Si p est une proposition, la proposition « non-p », ou « p est fausse », sera représentée par « $\sim p$ ». [...]
>
> (6) *Disjonction.* Si p et q sont des propositions quelconques, la proposition « p ou q », c'est-à-dire « ou bien p est vraie ou q est vraie », où les deux alternatives ne sont pas exclusives, sera représenté par « $p \vee q$ ». (Russell 1962 : p. 93, ma traduction)[25]
>
> ***1.1.** Tout ce qui est impliqué par une proposition élémentaire vraie est vrai. Pp. [...]
> ***1.11.** Lorsque ϕx peut être affirmé [...] et $\phi x \supset \psi x$ peut être affirmé [...], alors ψx peut être affirmé [...]. Pp. (Russell 1962 : p. 94-5, ma traduction)

Dans la définition de la disjonction, affirmer « p ou q » est équivalent à dire que « ou bien p est vraie ou q est vraie ». De même, dans la présentation du *modus ponens*, on

[23] Russell 1962 : p. 92, ma traduction de l'anglais.
[24] Russell 1962 : p. 92, ma traduction de l'anglais.
[25] Normalement la disjonction inclusive est traduite dans le langage ordinaire par « p est vraie ou q est vraie ». Ici, j'ai traduit la disjonction inclusive par « ou bien p est vraie ou q est vraie » car l'expression que Russell emploie est « either p is true or q is true ».

emploie la notion de vérité pour formuler 1.1, mais dans 1.11, qui est une adaptation du premier principe pour les fonctions propositionnelles, on utilise la notion d'assertion. La même chose peut être observée dans la présentation des axiomes :

> ***1.2.** $\vdash : p \vee p . \supset . p$ Pp.
> Cette proposition énonce que : « Si p est vraie ou p est vraie, alors p est vraie ». [...]
> ***1.3.** $\vdash : q . \supset . p \vee q$ Pp.
> Cette proposition énonce que : « Si q est vraie, alors 'p ou q' est vraie ». [...]
> ***1.4.** $\vdash : p \vee q . \supset . q \vee p$ Pp.
> Cette proposition énonce que « p ou q » implique « q ou p ». [...]
> (Russell 1962 : ma traduction)

La notion de vérité est employée pour expliciter la signification des propositions primitives 1.2 et 1.3, mais elle n'apparaît pas dans la proposition 1.4. L'interchangeabilité de la vérité et de l'assertion institue chez Russell une sorte de conception déflationniste de la vérité : affirmer p revient à affirmer que p est vraie.

L'ambiguïté en matière de vérité ne s'arrête pas là. Des propositions primitives, Russell nous dit qu'elles « doivent être supposées vraies sans que l'on dispose pour elles de preuves ».[26] Mais la signification du terme « vrai » est compromise par une remarque à l'effet que le choix des propositions primitives « est jusqu'à un certain point arbitraire ».[27] Son concept de vérité est altéré encore une fois dans le passage suivant :

> La preuve de la vérité d'un système logique réside dans son adéquation et sa cohérence. C'est-à-dire : (1) que tout système doit embrasser parmi ses déductions toutes les propositions que nous croyons vraies et capables d'être déduites à partir de seules prémisses logiques, quoiqu'il faille peut-être y apporter une légère restriction en

[26] Russell 1989 : p. 236.
[27] Russell 1989 : p. 236-7.

> leur donnant une expression plus rigoureuse; et (2) que le système ne doit conduire à aucune contradiction[.] (Russell 1989 : p. 237)

La vérité prend ici la forme d'un prédicat qui s'applique à un système, et non plus seulement à une proposition individuelle. Pour être vrai, un système logique doit répondre à deux critères : être non contradictoire et « embrasser » les propositions que nous croyons vraies. Après le déflationnisme, viennent donc le cohérentisme et une conception épistémique de la vérité, une conception qui tient compte des attentes formulées à l'endroit du système. Ajoutez à cela les remarques faites sur l'axiome de réductibilité à l'effet que les raisons pour accepter un axiome sont toujours largement inductives et le portrait est plutôt bigarré.

Pourtant, en matière de propositions individuelles, le correspondantisme est, semble-t-il, la thèse que Russell adopte officiellement dans les *Principia*. Sa définition exacte de la vérité apparaît dans une discussion sur la nature imprédicative de la vérité. Le problème est introduit en considérant la fonction propositionnelle $\hat{p}\ \textit{est fausse}$ et la proposition $(p).\ p\ \textit{est fausse}$, obtenue de cette dernière par généralisation. Puisqu'il existe des propositions vraies, il est légitime d'affirmer que $\big((p).\ p\ \textit{est fausse}\big)\textit{est fausse}$. Mais ce faisant, nous nous trouvons en plein délit d'imprédicativité : la variable de la fonction $\hat{p}\ \textit{est fausse}$ prend des valeurs d'ordres différents. Pour empêcher les cercles vicieux de s'introduire par l'entremise du concept de vérité, Russell propose de le hiérarchiser : au lieu d'avoir un seul et unique prédicat, il y aurait en réalité une famille de tel prédicats, un pour chaque ordre propositionnel. Par exemple, la proposition $\big((p).\ p\ \textit{est fausse}\big)\textit{est fausse}$ devrait s'écrire plutôt comme

$$\big((p).\ p\ \textit{est fausse}_0\big)\textit{est fausse}_1,$$

où $fausse_0$ serait un prédicat s'appliquant aux propositions élémentaires, tandis que $fausse_1$ s'appliquerait aux propositions d'ordre un. Le correspondantisme se manifeste au moment où il définit la vérité élémentaire, c'est-à-dire la vérité d'ordre nul. L'explication commence par : «[l]'univers comprend des objets possédant diverses qualités et entretenant diverses relations »,[28] certains objets dans l'univers sont complexes et un objet complexe est essentiellement un objet « qui est fait de parties reliées entre elles ».[29] Si *a* et *b* sont des objets simples reliés par une relation *R*, l'entité unifiée qu'ils forment est un objet complexe (un « objectif » selon la terminologie meinongienne) qu'on pourrait nommé « *a*-est-dans-la-relation-*R*-avec-*b* ». La définition procède alors comme suit :

> En fait, on peut définir la vérité, pour des jugements de cette espèce [les jugements élémentaires], comme le fait qu'il y a un complexe *correspondant* à la pensée discursive qu'est le jugement. C'est-à-dire que quand nous jugeons que « *a* est dans la relation *R* avec *b* », notre jugement est dit *vrai* quand il y a un complexe « *a*-est-dans-la-relation-*R*-avec-*b* », et il est dit *faux* quand ce n'est pas le cas. (Russell 1989 : p. 278-9)

C'est essentiellement la vérité correspondance appliquée aux propositions « élémentaires » : un jugement *élémentaire* est donc vrai quand il y a un complexe correspondant, et faux quand il n'y pas de complexe correspondant. Naturellement, ce concept de vérité n'est valide que pour les jugements élémentaires car les jugements généraux sont d'une forme différente. La vérité du jugement « Tous les hommes sont mortels », ne tient pas à l'actualité d'un complexe, elle tient à la vérité « élémentaire » d'une collection (infinie) de jugements élémentaires. Le concept de vérité qui s'applique à ce type de jugement est donc de premier ordre, et de manière générale, nous pouvons définir des concepts de vérité pour chaque ordre subséquent.

[28] Russell 1989 : p. 278.
[29] *Idem*

La position de Russell n'est donc pas facile à suivre, oscillant entre un correspondantisme amendé par la théorie des types, le cohérentisme, le déflationnisme et un pragmatisme épistémologique. Plus tard, dans la préface de la seconde édition des *Principles*, il reviendra sur sa notion de vérité. Considérant que la logique doit permettre des systèmes mutuellement inconsistants – les exemples les plus notoires étant les géométries non euclidiennes – Russell soutient que « nous devons [...] nous contenter d'affirmer que les axiomes impliquent les propositions, et non que les axiomes sont vrais et que par conséquent les propositions sont vrais aussi ».[30] Cette remarque préfigure le virage linguistique qu'il prendra par suite dans *The Philsophy of Logcial Atomism*, mais jusqu'aux *Principia*, il semble avoir voulu préserver la notion de vérité, au prix, nous l'avons vu, de certaines imprécisions.

6. Les *Principia* et le métalangage

Les quelques citations plus hauts concernant la définition des constantes logiques et l'introduction des propositions primitives montrent bien que les *Principia*, comme la *Begriffsschrift*, font usage d'un langage d'exposition pour présenter ses notions : on donne à l'expression « $p \vee q$ » le sens de « ou bien p est vraie ou q est vraie » (Russell 1989 : p. 93, ma traduction), on dit de la proposition primitive « $\vdash : p \vee p . \supset . p$ » qu'elle énonce que si p ou p est vraie alors p est vraie, et les termes qui sont utilisés pour expliciter ces significations appartiennent au langage de présentation. Sur la question de l'importance du langage d'exposition, Russell partage essentiellement la même position que Frege :

> Puisque toutes les définitions de termes sont effectuées à l'aide d'autres termes, tout système de définitions qui n'est pas circulaire doit commencer à partir d'un certain éventail de termes indéfinis. [...]

[30] Russell 1989 : p.11.

> Suivant Peano, nous appellerons les idées indéfinies et les propositions sans démonstration les idées *primitives* et les propositions *primitives* respectivement. Les idées primitives sont *expliquées* à l'aide de descriptions qui ont pour but d'indiquer au lecteur ce qui est signifié; mais les explications ne constituent pas des définitions, parce qu'elles utilisent les idées qu'elles expliquent. (Russell 1962 : p. 91, ma traduction)

La démarche fondationnelle de Russell exige que les *Principia* emploient des idées indéfinies sans quoi l'entreprise serait vouée à la circularité. En ce sens, les explications données dans le langage d'exposition sont des indications tout au plus, elles sont éliminables une fois que le sens précis du formalisme est compris par le lecteur.

Cette attitude contre l'intervention d'une théorie externe aux *Principia* est présente sous une autre forme dans le même paragraphe dans un commentaire sur le nombre minimal d'idées primitives nécessaires au système :

> Le choix des idées indéfinies en mathématiques est d'une certaine manière optionnelle; les critères qui guideront ce choix seront (1) faire en sorte que le nombre d'idées indéfinies est aussi petit que possible, (2) entre deux systèmes avec un même nombre d'idées indéfinies, choisir celui qui semble être le plus simple et le plus facile. *Nous ne connaissons aucune manière pour démontrer que tel ou tel système d'idées indéfinies en contienne aussi peu pour donner tel ou tel résultat.**
> * *Les méthodes reconnues pour démontrer l'indépendance ne sont pas applicables, sans réserves, aux idées fondamentales.* Cf. *Principles of Mathematics*, § 17. Ce qui y est dit concernant les propositions primitives s'applique avec une plus grand force aux idées primitives. (Russell 1962 : p. 91 et note en bas de page, mes italiques, ma traduction)

Au § 17 des *Principles*, on trouve notamment la remarque :

> [O]bservons que la méthode qui consiste à supposer faux un axiome et à déduire les conséquences de cette supposition, et qui, dans des cas

> comme celui de l'axiome des parallèles, s'est révélée admirable, n'est pas ici universellement utilisable. Car tous nos axiomes sont des principes de déduction; et s'ils sont vrais, les conséquences qui semblent devoir découler de l'emploi d'un principe opposé en réalité n'en découleront pas, de telle sorte que les arguments qui recourent à la supposition de la fausseté d'un axiome sont ici sujets à des erreurs particulières. (Russell 1989 : p. 37)

Russell suggère donc qu'une preuve métathéorique n'est pas valide dans la sphère des « idées fondamentales » car elle repose ultimement sur des idées et des principes qui sont les objets de son étude. Cette conclusion s'applique encore plus aux preuves d'indépendance, selon le §17, qui raisonnent sur des principes contraires à ceux qu'elles emploient, une situation qui, en plus d'être circulaire, ne peut conduire qu'à l'erreur. Sans barrer pour autant la route aux preuves d'indépendances, pour lesquelles il affiche quand même son admiration, il limite considérablement leur portée dans le débat philosophique des fondements de la logique et des mathématiques.

Les remarques de Russell sur l'acceptation de l'axiome de réductibilité et sur la vérité d'un système logique donnent un portrait similaire pour les preuves de non contradiction. La justification d'un axiome, ou d'un système d'axiomes, étant une affaire principalement inductive, il est seulement possible de démontrer empiriquement qu'un système est contradictoire, comme Russell a lui-même fait pour les *Grundgesetze* en exposant explicitement une contradiction. Démontrer qu'un système est non contradictoire est hors de portée pour cette méthode, à moins que l'ensemble des preuves soit fini, ce qui n'est pas le cas dans les *Principia*. Les moyens métathéoriques dans les *Principia* vont toutefois plus loin que la simple méthode empirique. N'oublions pas que Russell nous démontre (méta-théoriquement) pourquoi un certain type de paradoxe ne peut pas apparaître dans son formalisme typé, ce qui en soi constitue une preuve partielle de non contradiction.[31] Cela dit, les

[31] Naturellement, puisqu'il s'agit seulement d'un type particulier de paradoxe, il existe toujours la possibilité que son système subisse le même sort que celui de Frege.

outils métathéoriques de Russell ne dépassent guère l'argumentation philosophique. La situation est la même pour la complétude : la vérité d'un système, nous dit Russell, dépend de sa capacité à épouser l'ensemble des propositions logiques que nous croyons être vraies. Là encore, les méthodes sont principalement empiriques.

Dans le débat sur la logique comme calcul par opposition à la logique comme langage, j'avais présenté certaines réserves sur l'interprétation anti-sémantique de Frege qui, soutenais-je, faisait fi des nombreuses conceptions métalinguistiques présentes dans son œuvre. Certes, l'objectif philosophique de Frege était de faire une langue logique universelle, mais nous retrouvons sur le plan technique (en germe parfois) plusieurs éléments qui seront caractéristiques des méthodes modèle-théorétiques, qu'il s'agisse des définitions sémantiques ou des preuves de complétude. Par contre, en ce qui concerne Russell, je trouve les propos de van Heijenoort sont tout à fait justes. Russell élabore une langue logique complète dans laquelle les interventions métathéoriques se situent exclusivement au niveau de la présentation ou de l'explication, et pour laquelle il serait difficile d'appliquer, sans modifications importantes, les conceptions tarskiennes de la logique.

Ce qui rend Frege plus adaptable que Russell aux conceptions modernes est à la base une question d'ontologie. D'abord, l'univers de Frege est extensionnel : il est composé d'objets et de parcours de valeurs de la même manière qu'un modèle est constitué par un ensemble d'éléments et de relations. L'univers de Russell est, quant à lui, peuplé principalement de fonctions propositionnelles, c'est-à-dire d'entités non extensionnelles. D'autre part, dans les *Principia*, une affirmation est l'affirmation qu'un certain état de choses *est*, c'est-à-dire que la proposition ou la fonction propositionnelle décrite par le symbole *est*. Par opposition, la proposition de Frege n'est pas une entité de l'ontologie, elle est un élément de langage, un nom pour le vrai ou le faux. Par la force des choses, cette conception linguistique de la proposition amène Frege à manipuler le langage comme un objet, et à développer, par conséquent, quelques rudiments de la relation entre langage objet et métalangage, ce dont Russell, avec son correspondantisme littéraliste, n'a pas à se préoccuper.

Le concept de vérité est un autre point de divergence majeure entre Frege et Russell qui contribue à distinguer davantage les *Principia* des approches modernes en logique. S'il était pour nous aisé de voir la parenté entre la vérité frégéenne et la vérité tarskienne, ce n'est pas le cas de la vérité russellienne (s'il est possible de parler d'une seule et unique conception russellienne de la vérité). Le concept de vérité pour lequel nous pourrions peut-être constater une parenté serait la vérité correspondance, mais encore faudrait-il que les éléments en correspondance avec le langage soient des éléments extensionnels, ce qui n'est pas forcément le cas. En ce qui concerne le concept « assertif » de vérité, qui est le concept de vérité le plus fréquemment associé aux *Principia*, il n'y a aucune filiation possible : il est à l'antithèse de l'approche modèle-théorétique qui a pour vache sacrée la séparation nette entre vérité et affirmation (c'est-à-dire prouvabilité).

Un autre élément qui tend à effacer (ou plutôt masquer) la frontière entre langage objet et métalangage dans les *Principia* est l'utilisation « élargie » que Russell fait des connecteurs logiques dans des contextes extérieurs au formalisme lui-même. La plupart de ces usages pénètrent dans le système par le biais d'abréviations. Par exemple, au lieu d'écrire « $\vdash.q$ découle de $\vdash.p$ et de $\vdash:p.\supset.q$ », on abrége le tout en écrivant seulement « $\vdash.p.\supset.\vdash.q$ ». Le symbole d'affirmation « $\vdash$ » se trouve maintenant à apparaître à l'intérieur d'une implication. Au départ, il s'agit seulement d'une abréviation, mais plus loin Russell accordera plus de signification à cette expression qu'une simple économie d'écriture : « Ainsi '$\vdash:p.\supset.q$' signifie 'il est vrai que p implique q', tandis que '$\vdash.p.\supset.\vdash.q$' signifie '$p$ est vraie; par conséquent q est vraie' » (Russell 1962 : p. 92, ma traduction). L'expression « $\vdash.p.\supset.\vdash.q$ » est donc pour lui une manière de distinguer l'implication matérielle de l'implication formelle à l'intérieur du formalisme des *Principia*.[32]

[32] La question de la distinction entre l'implication formelle et matérielle a beaucoup préoccupé Russell dans les *Principles*.

Il y a d'autres occurrences de symboles logiques dans des descriptions de preuves, dont notamment le symbole de conjonction « . ». Pour signifier que le théorème « $\vdash . p \vee \sim p$ » est démontrable à partir du théorème *$\mathbf{2.08}$ et de la définition *$\mathbf{1.01}$, on indique $[*2.08.(*1.01)]$ en marge de la démonstration, c'est remarquer que le théorème est le résultat de la conjonction de *$\mathbf{2.08}$ et de *$\mathbf{1.01}$. Ces nombreuses règles supplémentaires régissant l'écriture des preuves et l'abréviation des formules constituent une part importante du formalisme « officieux » des *Principia*; en fait, un simple coup d'œil aux pages des *Principia* nous fait voir l'importance de ce formalisme « extra symbolique ». Je cite par exemple la proposition *$\mathbf{2.16}$:

$\vdash :\sim p \supset \sim q . \supset . q \supset p$

Dém.

$\left[*2.03 \dfrac{\sim q, p}{p, q}\right]$	$\vdash :\sim q \supset \sim p . \supset . p \supset \sim (\sim q)$	(1)
$[*2.14]$	$\vdash :\sim (\sim q) \supset q :\supset$	
$[*2.05]$	$\vdash : p \supset \sim (\sim q) . \supset . p \supset q$	(2)
$[Syll]$	$\vdash :. (1).(2). \supset \vdash . Prop$	

Se trouvent à l'œuvre dans cette démonstration des notions qui dépassent le formalisme de base des *Principia*. D'abord, les remarques entre crochets précisent généralement les propositions utilisées et au besoin les substitutions effectuées, la dernière remarque signale l'utilisation d'une forme de déduction, en l'occurrence le syllogisme. La numérotation en marge gauche est un outil de référence à l'intérieur de la preuve. Enfin, on remarque la présence du signe d'assertion à l'intérieur des formules. Tous ces symboles expriment non pas des éléments de l'ontologie russellienne, mais plutôt le langage des *Principia*.

Il existe donc une tension entre les convictions philosophiques universalistes de Russell et ce que l'élaboration d'un langage formel exige. D'un côté, l'idée même d'un langage universel demande une forme d'indépendance conceptuelle. Mais de l'autre elle implique la limitation des pouvoirs expressifs. Russell tente de concilier ces deux contraintes comme nous l'avons vu plus haut, mais ne réussit pas à contourner le recours inévitable à un métalangage. Le métalangage devient inévitable, notamment, en raison de la faillibilité de notre intuition ou de notre perception « logique ». Pour juger de la validité d'un langage ou pour en ériger un autre, il faut se placer dans un médium extérieur à ce langage. Ainsi, malgré les bonnes intentions philosophiques, Russell ne peut pas exclure le recours éventuel à une instance métathéorique.

7. Épilogue

Il reviendra à Bertrand Russell de formuler de manière plus précise la notion de métalangage, mais seulement à la suite de sa lecture du *Tractatus* de Wittgenstein. En gros, le *Tractatus* gravite autour d'une distinction fondamentale à notre débat, celle entre dire et montrer.[33] Cette distinction est parallèle à celle entre l'usage d'un terme et sa mention. Wittgenstein développera une thèse sur les limites expressives du langage avec cette distinction, et s'emploiera à montrer comment la philosophie porte parfois sur des sujets sur lesquels elle ne peut rien dire. En particulier, il est impossible de dire la forme logique, selon Wittgenstein, il est seulement possible de la montrer. Ainsi, Frege et Russell se trouvent dans la mire de ses critiques, eux qui ont tenté de donner des définitions aux éléments de leurs systèmes logiques (par la porte de derrière) alors qu'une telle chose transgresse les limites du dicible. Russell reconnaît la pertinence de la distinction dire/montrer, mais au lieu de sombrer dans une forme d'agnosticisme, il propose explicitement de faire appel à une hiérarchie de langages :

[33] Le *Tractatus* de Wittgenstein fera l'objet du prochain chapitre.

> Ces difficultés me font penser à quelque possibilité comme celle-ci : que chaque langage a, ainsi que le dit Wittgenstein, une structure au sujet de laquelle, dans le langage, on ne peut rien dire, mais qu'il doit y avoir un autre langage traitant de la structure du premier langage, et possédant lui-même une nouvelle structure, et qu'à cette hiérarchie de langages il ne doit pas avoir de limite. (*Tractatus* 1961 : p. 32)

Certes, le langage ne peut pas s'exprimer dans lui-même, mais cette situation peut être rectifiée si, dans la logique symbolique, il y a une suite de langages formels emboîtés, chacun servant à exprimer les conceptions métalinguistiques du précédent.

CHAPITRE III

Wittgenstein sur l'indicibilité dans le *Tractatus*

0. Introduction

Avec Wittgenstein, la notion de langage formalisé a suffisamment vécue pour que certaines difficultés philosophiques liées à son exposition apparaissent. L'intention chez Frege et Russell était de fonder la logique et les mathématiques dans un formalisme sûr pour échapper aux problèmes posés par le langage ordinaire. Mais nous avons vu que l'exposition des formalismes respectifs de Frege et de Russell exigeait, tantôt pour la syntaxe tantôt pour la sémantique, l'intervention d'un métalangage. Wittgenstein partage avec Frege et Russell une aversion pour les notions métathéoriques; seulement chez lui, cette aversion ne résulte pas tant de considérations sur l'autonomie conceptuelle que d'une thèse philosophique sur l'indicibilité de la forme logique du langage. Il reprochera même à Frege et à Russell, malgré leur opposition au métalangage, d'avoir transgressé les limites du dicible en parlant « informellement » de la structure du langage et de sa relation au monde. Dans un cadre ontologique et sémantique nouveau, le *Tractatus* cherchera à relever le défi universaliste de développer un langage qui épouse la forme logique du monde tout en évitant les problèmes que pose une métathéorie.

1. Le *Tractatus Logico-Philosophicus*

La philosophie du *Tractatus* est une critique radicale de la philosophie frégéo-russellienne de laquelle elle est issue. Elle est en rupture avec de nombreuses conceptions traditionnelles, allant de l'ontologie jusqu'à la épistémologie. L'entreprise philosophique du *Tractatus* est donc considérable. Dans cette section, il sera question d'exposer les principales thèses du *Tractatus* relatives à l'ontologie, la sémantique, la logique et les mathématiques en prenant soin d'identifier clairement en cours de route la thèse de l'indicibilité des relations sémantiques. Pour ce faire, je suivrai autant que possible l'ordre de présentation des idées dans le *Tractatus*, c'est-à-dire en partant de la première section et en poursuivant jusqu'à la septième.[1]

1.1. L'ontologie

Selon Wittgenstein, il n'y a pas lieu de parler de dualisme ou d'un troisième domaine (ni même de monisme) lorsqu on parle du monde. Il n'y a que *le* monde et c'est dans ce monde que réside tout ce qui arrive, ou tout ce qui est le cas (1). Ce monde n'est pas une sommation d'objets, comme le stipule la plupart des ontologies traditionnelles, il est plutôt l'ensemble des faits (1.1). Le fait est l'actualité ou l'existence d'un état de choses (2) : la chose, ou l'objet, est l'élément simple et irréductible de cette ontologie (2.02); l'état de choses, ou la configuration, est ce qui est instable et changeant (2.027).

---

[1] Mon interprétation du *Tractatus* est guidée notamment par (Marion 1998), (Marion 2004) et (Black 1966).

Les propriétés des objets ne sont pas des « choses » au sens où les relations sont des entités dans l'ontologie russellienne. Sur ce plan, Wittgenstein opère le même renversement qu'Aristote : il incarne les formes essentielles. En effet, l'idée incarnée prend la forme de la possibilité de l'occurrence d'un objet dans un état de choses. Ainsi, « si la chose *peut* arriver dans un état de choses, il faut que la possibilité de l'état de choses soit préalablement inscrite dans la chose » (2.012). Connaître l'objet c'est connaître la possibilité de son occurrence dans un état de choses (2.0123). La forme logique d'un objet est précisément la possibilité de son occurrence dans des états des choses. L'expression « espace logique » est parfois utilisée pour décrire la forme logique : dire que l'objet est dans un certain état de choses revient à dire qu'il occupe une position (ou une région) dans son espace logique, son espace de possibilités. Par exemple, si nous avons affaire à un objet temporel, nous savons déjà que ses possibilités seront contraintes par la temporalité, il se situera forcément dans l'espace temporel; s'il est coloré, dans l'espace coloré; etc. L'objet est donc lié de manière intrinsèque à la possibilité de ses occurrences.

1.2. La représentation

La deuxième section du *Tractatus* est là où Wittgenstein pose les premiers éléments de sa théorie de la représentation. Le « tableau »[2] fait figure de représentation archétypique. L'activité de représentation consiste à faire des tableaux de faits (2.1). Naturellement, dans l'esprit « moniste » du *Tractatus*, le tableau est lui-même un fait (2.141). Le tableau est essentiellement une transposition de la réalité (2.12) et il est lié à l'état de choses qu'il représente par une sorte d'isomorphisme de faits. Plus précisément, les objets de l'état de choses correspondent aux éléments du tableau de cet état de choses (2.13 et 2.131) et

[2] Le mot « *Bild* » qui est traduit ici par « tableau » peut également être traduit par « image ».

> Le fait que les éléments du tableau ont des rapports déterminés les uns avec les autres tient à ce que les choses [les objets de l'état de choses que le tableau représente] se comportent de la même manière les unes vis-à-vis des autres. Cette connexion des éléments du tableau, nous la nommerons sa structure, et la possibilité de sa structure la forme de représentation. (2.15)

La forme de représentation (ou la forme logique de représentation) d'un tableau est au cœur de la sémantique wittgensteinienne, elle est ce qui permet au tableau de représenter. On pourrait qualifier cette sémantique de « correspondantisme structurel », car c'est en vertu d'une corrélation de forme qu'un tableau d'état de choses représente cet état de choses : ce « que le tableau doit avoir en commun avec la réalité, afin de pouvoir la représenter à sa manière – avec justesse ou fausseté – c'est la forme de la représentation » (2.17). Par ailleurs, seule la forme de représentation est pertinente dans la relation de représentation car « [l]e tableau peut représenter chaque objet dont il a la forme » (2.171). Ainsi, le « tableau étendu dans l'espace peut représenter tout ce qui est spatial, le tableau coloré tout ce qui est coloré, etc. » (2.171). De même, il y a un dénominateur commun à tous les tableaux : tout tableau est à la fois un tableau logique (2.182).

La forme de représentation confère donc au tableau une capacité représentationnelle, elle fait partie intégrante du tableau (2.1513, 2.16) et détermine ses liens sémantiques avec le monde. Ces liens sémantiques font en sorte que le tableau « atteint » le monde « comme un étalon de mesure qui 'colle' avec la réalité » (2.1511, 2.1512). Bien que la forme de représentation appartienne au tableau, elle n'est jamais représentée par le tableau, elle peut seulement être montrée :

> 2.172 – Cependant le tableau ne saurait représenter sa propre forme de représentation : il ne fait que la montrer.
>
> 2.173 – Le tableau représente son objet du dehors (son point de vue constitue sa forme de représentation); c'est pourquoi le tableau représente son objet justement ou faussement.

> 2.174 – Le tableau cependant ne saurait se représenter en dehors de sa forme de représentation.

L'ineffabilité de la forme de représentation implique que les relations sémantiques ne peuvent jamais être représentées. Il est clair que cette thèse est motivée dans un premier temps par la distinction entre l'usage et la mention : une représentation ne peut jamais représenter simultanément l'objet de sa représentation et la manière dont elle représente cet objet. Dans un deuxième temps, il y a l'idée que la relation de représentation implique toujours, à une extrémité de la relation, un élément non linguistique. Nous verrons dans la suite comment cette distinction est centrale au *Tractatus*.

La deuxième section termine sur un concept embryonnaire de vérité (qui sera affiné plus loin). Puisque le tableau représente des faits, c'est-à-dire l'existence ou l'inexistence d'états de choses, il s'ensuit que la vérité d'un tableau se décide avec la réalité (2.223 et 2.224). Cette thèse a une conséquence importante : il n'y a pas de vérités *a priori* (2.225).

1.3. La pensée et la proposition

La troisième section s'ouvre sur la relation entre la pensée, la proposition et le monde. Le principe directeur de cette relation est que la pensée est elle-même un tableau et qu'elle trouve une expression perceptible dans la proposition (3, 3.1). En un mot, la proposition est l'outil représentationnel de la pensée. Ce qui a été dit du tableau et de sa forme de représentation est appliqué aux propositions avec une nouvelle terminologie. De la même manière qu'une représentation est un état de choses muni d'une relation de représentation, une proposition est la donnée d'un signe propositionnel et d'une projection (3.11-3.13). La projection prend donc le relais de la forme de représentation. Elle est la manière dont le signe propositionnel « colle » à sa représentation.

Bien que le signe propositionnel n'inclue pas sa relation de projection, il ne faut pas croire qu'il est comme un assemblage de mots. Le signe propositionnel est un fait et un fait exprime un sens (3.14). Il exprime son sens comme le fait un état de choses : le signe complexe « *aRb* » ne dit pas que « *a* » se trouve dans la relation *R* à « *b* », « *aRb* » est dit par le fait que « *a* » se trouve dans une certaine relation à « *b* » (3.1432).[3]

Comme le tableau, le signe propositionnel est décomposable en éléments simples, en signes simples. Les signes simples sont des noms et ont pour signification des objets (3.2). La décomposition d'un signe propositionnel en signe simples (ou en noms) est unique,[4] et le nom « ne saurait être décomposable par aucune autre définition : il est un signe originel » (3.25, 3.26). Les relations sémantiques entre le signe propositionnel et ce qu'il représente (lesquelles relations sont dictées par la projection) sont essentiellement les mêmes que celle entre un tableau et ce qu'il représente : « [à] la configuration du signe simple dans le signe propositionnel correspond la configuration des objets dans l'état de choses » (3.21). Ce qui ne peut pas s'exprimer par les signes se manifeste par leur application ou par l'usage (3.262). La signification des signes originels peut toutefois être expliquée par des élucidations, en autant que la signification de ces signes soit connue au préalable (3.263).[5]

Wittgenstein soutient, comme le fait Russell, que l'unité minimale de signification est la proposition. Ceci l'amène d'ailleurs à souscrire au principe du contexte de Frege : « [l]a proposition seule a un sens; et ce n'est que dans le contexte d'une proposition qu'un nom a une signification. » (3.3). Rappelons qu'une expression ou un nom est à la proposition ce que l'objet est à l'état de choses. Par conséquent, l'expression contient sa possibilité d'occurrence dans les propositions :

---

[3] On remarquera la subtilité entre les deux : dans le premier cas, le signe dit, tandis que dans le second, le fait dit/montre.

[4] On retrouve ici une des prémisses de l'atomisme logique.

[5] Il est donc impossible de définir tous les signes sans commettre une pétition de principe.

« [l]'expression présuppose les formes de toutes les propositions dans lesquelles elle peut intervenir », elle « est une marque caractéristique commune d'une classe de propositions » (3.311). Ceci mène à une conception singulière de la variable : « [l]'expression est donc représentée par une variable dont les valeurs sont les propositions qui contiennent l'expression » (3.313).[6] Enfin, l'architecture compositionnelle de la proposition est reprise dans le *Tractatus* : « [l]a proposition est une fonction des expressions contenues en elle » (3.318).

On voit également apparaître dans la section 4 le concept de syntaxe logique. La syntaxe logique est ce qui s'approche le plus chez Wittgenstein de la logique comme discipline théorique, un rapprochement que Carnap n'a pas manqué de faire. Celle-ci se manifeste dans une discussion sur les ambiguïtés du langage quotidien :

> (3.323) – Dans le langage quotidien il arrive très fréquemment que le même mot désigne d'une manière différente – donc appartienne à différents symboles – ou que deux mots, qui désignent de manière différente, soient utilisés extérieurement de la même manière dans la proposition. [...]
> (3.324) – C'est ainsi que se produisent facilement les confusions fondamentales (dont toute la philosophie est remplie).
> (3.325) – Pour échapper à ces erreurs nous devons utiliser un langage de signes qui les exclut [...] Par conséquent un langage de signes qui obéit à la grammaire *logique*, donc à la syntaxe logique.

Les systèmes de Frege et de Russell sont de tels langages, même si de toute évidence ils ne parviennent pas à exclure toutes les erreurs (3.33). Par ailleurs, les syntaxes de Frege et de Russell contreviennent à l'indicibilité de certaines relations. Par exemple, la justification de la théorie des types est erronée selon Wittgenstein car elle en appelle à la signification des signes (3.331). Au lieu du principe du cercle vicieux, il aurait fallu une justification de la forme suivante : « [a]ucune proposition ne peut énoncer quelque chose sur elle-même, parce que le signe propositionnel ne peut être

[6] Le concept de variable propositionnelle est complexe dans le *Tractatus*, il se veut le pendant du concept formel dans les systèmes de Frege et Russell (Voir 4.53).

contenu en lui-même » (3.332).[7] Cette justification doit cependant aller de soi car les « règles de la syntaxe logique doivent se comprendre d'elles-mêmes, pourvu que l'on sache la manière dont chaque signe désigne » (3.334).

1.4. Le langage

L'objet de la quatrième section est spécifiquement le langage. De manière générale, celui-ci est simplement défini comme étant l'ensemble de toutes les propositions. L'être humain a la capacité de construire des langages sans avoir pour autant une connaissance de la manière dont un langage signifie (4.002). Le langage est une institution naturelle qui va de soi : « [l]e langage quotidien est une partie de l'organisme humain, et pas moins compliqué que ce dernier » (4.002). Toutefois, notre connaissance de la logique du langage ne va pas de soi. Le langage travestit souvent la pensée, il est comme un vêtement qui voile les formes réelles du corps (4.002). Cette conception l'amène à formuler une thèse notoire de la philosophie analytique à l'effet que « toute philosophie est 'critique du langage' » (4.0031) et que la « plupart des propositions et des questions des philosophes viennent de ce que nous ne comprenons pas la logique de notre langage » (4.003). Wittgenstein accorde d'ailleurs à Russell le mérite d'avoir montré que la forme réelle du langage n'est pas sa forme apparente (4.0031).[8]

Comprendre la logique du langage, c'est donc « dévoiler » les formes réelles des propositions. Pour ce faire, Wittgenstein développe sur la relation de représentation ou de projection :

> 4.014 – Le disque de phonographe, la pensée musicale, les notes, les ondes sonores, tous se trouvent les uns par rapport aux autres dans

[7] Une justification qui selon Wittgenstein ne fait pas appel à la signification des signes mais seulement à leur forme logique.

[8] Il fait sans doute référence à l'analyse des descriptions définies de Russell.

> cette relation interne de représentation qui existe entre le langage et le monde.
>
> Leur structure logique leur est commune à tous.
>
> 4.0141 – Qu'il existe une règle générale qui permette au musicien de déchiffrer la symphonie dans la partition, qu'il en soit une qui permette de reconstituer à partir du sillon de disque la symphonie et d'après la première règle derechef la partition, – voilà en quoi consiste la similitude intérieure de ces formulations en apparence si dissemblables les unes des autres. Et cette règle est la loi de la projection, qui projette la symphonie dans le langage des notes. Elle est la règle de la traduction du langage des notes dans le langage du disque phonographique.[9]

Selon Wittgenstein, l'écriture hiéroglyphique, par exemple, est un langage qui expose bien sa relation interne de représentation, une relation qui est moins apparente dans l'écriture alphabétique (4.016).[10] La critique du langage sert à mettre en évidence la forme de représentation pour montrer le sens. Nous devons pouvoir comprendre « le sens du signe propositionnel, sans qu'il nous ait été expliqué » (4.02). Le sens d'une proposition est ce qui se montre seulement, il n'est pas représentable. Il en va de même pour la forme logique : « les 'constantes logiques' ne représentent pas » et « la logique des faits ne se laisse pas représenter » (4.0312).

Si la logique n'est pas une discipline et la philosophie est critique du langage, c'est-à-dire s'il n'y a pas de propositions logiques ou philosophiques, quelle discipline a le monopole des propositions (vraies)? La réponse à cette question est une conséquence immédiate de l'impossibilité des vérités *a priori*, c'est-à-dire du fait que la proposition « ne peut être vraie ou fausse qu'en étant une image de la réalité » (4.06). Les propositions sont donc des propositions sur la réalité, par conséquent l'ensemble des propositions vraies constitue les sciences de la nature (4.11). La philosophie n'est pas une science, elle est au-dessus ou en dessous de la science :

---

[9] Ce passage facilite la compréhension de 3.1432.

[10] Malgré qu'il a bien fallu la pierre de Rosette pour rendre la structure de ce langage « apparente »!

> 4.112 – Le but de la philosophie est la clarification logique de la pensée.
> La philosophie n'est pas une doctrine mais une activité.
> Une œuvre philosophique consiste essentiellement en élucidations.
> Le résultat de la philosophie n'est pas un nombre de « propositions philosophiques », mais le fait que des propositions s'éclaircissent.
> La philosophie a pour but de rendre claires et de délimiter rigoureusement les pensées qui autrement, pour ainsi dire, sont troubles et floues.[11]

En clarifiant la pensée, la philosophie se trouve à délimiter « le domaine discutable des sciences de la nature » (4.113), elle limite « le concevable, et, de la sorte, l'inconcevable » de l'intérieur (4.114). Le convenable ou le dicible a par ailleurs une frontière très nette : tout « ce qui peut être en somme pensé, peut être clairement pensé », « tout ce qui se laisse exprimer, se laisse clairement exprimer » (4.116).[12] La distinction entre dire et montrer prend ici sa forme la plus achevée :

> 4.121 – La proposition ne peut représenter la forme logique, celle-ci se reflète dans la proposition.
> Ce qui se reflète dans le langage, le langage ne peut le représenter.
> Ce qui s'exprime *soi-même* dans le langage, *nous-mêmes* ne pouvons l'exprimer par le langage.
> La proposition *montre* la forme logique de la réalité. Elle l'exhibe.
> [...]
> 4.1212 – Ce qui *peut* être montré *ne peut pas* être dit.

Wittgenstein s'attarde maintenant à développer un symbolisme qui lui permettra de « montrer » les possibilités de vérité d'une proposition complexe en fonction des possibilités de vérité des propositions élémentaires qui la constituent. Ce

[11] Comparer avec 4.0031.
[12] On peut lire ici une version de la vérité des logicistes : ou bien c'est vrai, ou bien c'est faux, pas d'entre deux.

symbolisme est essentiellement celui des tables de vérité.[13] Pour représenter les possibilités d'accord ou de désaccord, c'est-à-dire les possibilités de vérité, d'une proposition complexe en fonction de ses propositions élémentaires, on utilise un signe propositionnel comme le suivant :

p	*q*
V	V
F	V
V	F
F	F

lequel montre les possibilités de vérité de deux propositions élémentaires. Le signe propositionnel

p	*q*	
V	V	V
F	V	V
V	F	F
F	F	V

montre la forme logique de l'implication de la proposition q à partir de la proposition p. Bien qu'il considère ces tables comme des signes propositionnels, Wittgenstein nous interdit d'associer une signification à certaines de leurs composantes : « [i]l est clair que nul objet (ou complexe d'objets) ne correspond au complexe de signes « F » et « V »; pas plus qu'aux lignes horizontales et verticales ou aux parenthèses », car il « n'y a point d' 'objets logiques' ».[14]

La capacité de parler des possibilités de vérité nous permet de définir la tautologie et la contradiction :

[13] On retrouve la notion chez Frege et indépendamment chez Peirce.

[14] Le fait que rien ne corresponde à V et F résulte également de l'indicibilité de la relation entre une proposition et la réalité (ce qui fait sa vérité ou sa fausseté).

> Dans le premier cas la proposition est vraie pour la totalité des possibilités de vérité des propositions élémentaires. Nous disons que les conditions de vérité sont *tautologiques*.
> Dans le second cas la proposition est fausse pour la totalité des possibilités de vérité : les conditions de vérités sont contradictoires.
> Dans le premier cas nous nommons la proposition une tautologie, dans le second cas une contradiction. (4.46)

La tautologie et la contradiction sont donc indifférentes aux possibilités de vérité : l'une est invariablement vraie de tout état de choses et l'autre fausse. « La proposition montre ce qu'elle dit, la tautologie et la contradiction montrent qu'elles ne disent rien » (4.461). Ce n'est pas dire pour autant que la tautologie et la contradiction sont des non sens, c'est seulement dire qu'elles sont vides de sens (4.4611) :

> La tautologie laisse à la réalité tout l'espace de logique infini; la contradiction remplit tout l'espace logique et ne laisse aucun point à la réalité. Aucune des deux ne peut déterminer la réalité d'aucune manière. (4.463)
> La vérité de la tautologie est certaine, celle de la proposition possible, celle de la contradiction impossible. (4.464)

1.5. Les relations entre propositions

La cinquième section est consacrée aux relations entre propositions. Le principe directeur cette fois-ci est le fait qu'une « proposition est une fonction de vérité des propositions élémentaires » (5). Seront particulièrement importantes les possibilités de vérité qui rendent une proposition vraie. Wittgenstein rendra compte des relations entre propositions en termes des possibilités de vérité de ces propositions. L'idée est assez simple : si les possibilités de vérité qui rendent un certain nombre de propositions vraies sont toutes des possibilités de vérité qui rendent une autre proposition vraie, « nous dirons que la vérité de cette proposition résulte de la vérité de ces propositions » (5.11). « En particulier la vérité d'une proposition '*p*' résulte de la vérité d'une autre 'q' », si les possibilités de vérité rendant la première vraie rendent la seconde vraie (5.12). L'intérêt de traduire les liens de conséquence entre propositions de cette manière vient de ce que les possibilités de vérité sont « montrées » par la structure des propositions (5.13). En effet :

> Lorsque nous inférons q de $p \vee q$ et $\sim p$, il se trouve que l'écriture cache la relation des formes de proposition de « $p \vee q$ » et de « $\sim p$ ». Mais si par exemple au lieu de « $p \vee q$ » nous écrivons « $p|q.|.p|q$ » et au lieu de « $\sim p$ » nous écrivons « $p|p$ » ($p|q$ = ni p, ni q) la connexion interne devient évidente. (5.1311)[15]

Une écriture appropriée (comme celle de Scheffer) nous montre immédiatement la structure logique des propositions et par conséquent les liens inférentiels entre propositions. Dans une telle situation, les règles d'inférence frégéo-russelliennes sont complètement caduques, voire dénuées de sens (5.132).

Les relations entre propositions étant seulement limitées aux relations entre leurs possibilités de vérité respectives, il s'ensuit que toute « inférence se fait *a priori* » (5.133). Si les propositions ne sont pas liées par les possibilités de vérité, elles ne sont pas liées du tout. Les conséquences de cette position sont assez radicales, notamment :

> 5.1361 – Nous ne *pouvons* inférer les événements de l'avenir des événements présents.
>
> La croyance au rapport entre cause et effet est une *superstition.*
>
> 5.1362 – Le libre arbitre consiste en ce que des actes futurs ne peuvent être sus maintenant. Nous ne pourrions les savoir que si la causalité constituait une nécessité intérieure telle que celle de la conclusion logique [ce qu'elle n'est pas]. – La connexion du savoir et de ce qui est su est celle de la nécessité logique.

Il en résultera plus loin que la seule nécessité dans le monde est la nécessité logique.

[15] J'ai modifié la traduction ici pour des fins de clarté.

Wittgenstein pousse ensuite la structure propositionnelle un peu plus loin. Il veut en arriver à la forme générale de la proposition. L'idée est que « les structures des propositions se trouvent mutuellement en relations internes » (5.2) et que nous « pouvons souligner dans notre mode d'expression ces relations internes par le fait que nous représentons une proposition comme résultant d'une opération qui la produit à partir d'autres propositions » (5.21). « L'opération est l'expression d'une relation entre les structures de son résultat » et de son argument (5.22), elle est « ce qui doit se produire dans une proposition pour en faire une autre à partir d'elle » (5.23). La relation interne ordonnant une suite est donc précisément « l'opération par laquelle un terme procède de l'autre » (5.232). Par exemple, $[a, x, O'x]$ est le terme général de la suite

$$a, O'a, O'O'a, O'O'O'a, \ldots,$$

c'est-à-dire qu'il précise l'argument de base, a, et le procédé qui transforme un terme de la suite en son successeur, $x \mapsto O'x$. La thèse est donc que « [t]outes les propositions sont des résultats d'opérations de vérité sur des propositions élémentaires »,[16] que l'opération « de vérité est la manière dont la fonction de vérité procède des propositions élémentaires » (5.3).

L'opération qui définie la relation interne de la proposition apparaît à même le paragraphe 5.5 :

> Chaque fonction de vérité est un résultat de l'application successive de l'opération $(-----V)(\xi,\ldots)$.
>
> Cette opération nie l'ensemble des propositions dans la parenthèse de droite, et je la nomme la négation de ces propositions.

[16] Le nombre d'opérations ou le nombre d'applications de ces opérations est toujours fini, voir 5.32.

Si une variable ξ admet un certain nombre de variables p, q, r, ... alors $N(\overline{\xi})$ sera « la négation de l'ensemble des valeurs des variables propositionnelles », c'est-à-dire la proposition $(-----V)(\xi,...)$ (5.502).[17] Si ξ n'a qu'une seule valeur p, on aura $N(\overline{\xi}) = \sim p$; si ξ a deux valeurs p et q, alors $N(\overline{\xi}) = p \mid q = \sim p . \sim q$ (5.51). De même, si ξ prend toutes les valeurs d'une fonction fx, alors $N(\overline{\xi}) = \sim (\exists x).fx$ (5.52). Le terme générale de l'ensemble des propositions est donné par $[\,\overline{p}, \overline{\xi}, N(\overline{\xi})]$, où $\overline{p}$ est l'ensemble des propositions élémentaires et $\overline{\xi}$ une sélection de ces propositions prises ensembles (6). Ainsi, l'ensemble des propositions est obtenu des propositions élémentaires en appliquant le procédé $\overline{\xi} \mapsto N(\overline{\xi})$.

1.6. Suite et fin

Nous avons vu que la forme générale de la proposition était $[\,\overline{p}, \overline{\xi}, N(\overline{\xi})]$. L'opération qui définit les nombres sera construite de manière analogue. On commence par se donner l'opération (quelconque) suivante : $x = \Omega^{0}{}'x$ Déf. et $\Omega'\Omega^{\nu}{}'x = \Omega^{\nu+1}{}'x$ Déf (6.02). D'après ces règles, la suite $x, \Omega'x, \Omega'\Omega'x, \Omega'\Omega'\Omega'x,...$ serait donc

$$\Omega^{0}{}'x, \Omega^{0+1}{}'x, \Omega^{0+1+1}{}'\Omega^{0+1+1+1}{}'x,....$$

Le nombre est conçu comme l'exposant d'une telle opération (6.021) :

$$0+1=1 \text{ Déf.},\ 0+1+1=2 \text{ Déf.},\ 0+1+1+1=3 \text{ Déf., etc.}$$

[17] C'est la fonction : $\sim p . \sim q . \sim r . \cdots$.

Nous avons donc que la « forme générale du nombre entier est : $[0, \xi, \xi + 1]$ » (6.03). Des mathématiques, Wittgenstein nous dit qu'elles « sont une méthode logique » (6.2). De même, la « logique du monde, que les propositions de logique montrent dans les tautologies, les mathématiques la montrent dans les équations » (6.22). Sur la méthode des mathématiques on peut lire :

> 6.24 – La méthode des mathématiques pour en arriver à leurs équations, est la méthode de substitution.
>
> Car les équations expriment la substituabilité de deux expressions et nous progressons d'un nombre d'équations à de nouvelles équations, en remplaçant des expressions par d'autres, conformément aux équations.[18]

Sur les propositions de la logique, on apprend qu'elles sont des tautologies (6.1) et qu'elles ne disent rien, qu'elles sont analytiques (6.11). Wittgenstein applique sa thèse de transparence à la logique toute entière :

> 6.113 – C'est la caractéristique particulière aux propositions logiques que l'on puisse reconnaître au symbole seul qu'elles sont vraies, et ce fait renferme toute la philosophie de la logique. Et ainsi c'est également l'un des faits les plus importants que la vérité ou la fausseté des propositions non-logiques *ne* se puisse reconnaître à la seule proposition.

Le fait qu'une proposition est une tautologie est montré par les propriétés formelles ou logiques du langage. Les propositions logiques à proprement parler sont éliminables parce qu'une notation adéquate nous permet de « reconnaître les propriétés formelles des propositions à la simple vue de ces propositions » (6.122).

---

[18] Il faut mentionner au passage que Wittgenstein considère les propositions mathématiques comme des pseudo-propositions parce qu'elles font intervenir le signe d'égalité. Dans le *Tractatus*, il propose une mathématique sans égalité qui rend compte de l'identité d'une chose par l'identité de son symbole et la multiplicité des choses par la multiplicité des symboles. Voir à cet effet les paragraphes 4.24, 4.241, 4.242, 4.243, 5.5302, 5.5302, 5.531, 5.532, 5.5321, 5.533 et 5.534.

La preuve d'une proposition logique est intrinsèquement liée à ses relations internes, elle n'est rien d'autre que « le fait que nous les faisons procéder d'autres propositions logiques par l'application successive de certaines opérations qui engendrent sans cesse des tautologies à partir des premières propositions » (6.126). Ainsi, les notions de preuve et de vérité se fondent l'une dans l'autre : « processus et résultat sont équivalents » (6.1261). Wittgenstein favorisera en dernière instance la procédure ou l'opération : « [l]a proposition ayant un sens énonce quelque chose et sa preuve montre qu'il en est ainsi; en logique chaque proposition est la forme d'une preuve » (6.1246). Enfin, la « logique n'est pas une théorie, mais une image réfléchie du monde », la « logique est transcendantale » (6.13).

2. L'universalité du *Tractatus*

De tous les universalistes, Wittgenstein est celui qui est allé le plus loin et jusqu'aux dernières conséquences. Il en résulte une conception qui fait de la logique *la* syntaxe – ou la grammaire – des propositions. La logique est l'expression de la mécanique interne du monde, ce à quoi doit se plier toute représentation en tant qu'elle participe à ce monde. Un des tours de force majeurs du *Tractatus* a été d'éviter le troisième domaine et l'objet « logique » que Frege et Russell avaient tout naturellement postulés dans le but explicite d'échapper au psychologisme et à l'idéalisme en philosophie de la logique. Par une manœuvre ontologique et sémantique très habile, il réussit à développer une conception « réaliste » de la logique en ne quittant pas le seul et unique monde. La logique est la forme des possibilités des états de choses, une forme qui délimite de quelle manière un état de choses peut se manifester dans le monde.[19]

[19] Le stratagème rappelle la philosophie transcendantale de Kant; seulement chez Kant c'est la raison qui contient la forme des possibilités et non le monde.

La notion de variable était l'expression la plus manifeste de l'universalité dans les systèmes de Frege et de Russell. Toutefois, la situation n'est pas vraiment la même dans le *Tractatus*, étant donnée la conception radicalement différente de la variable. La transition d'objet logique à forme logique amène Wittgenstein à voir la variable comme la forme d'un objet prototypique, le concept d'un objet. Par exemple, la variable d'un entier naturel est son terme général, $[0, \xi, \xi + 1]$, et pareillement la variable d'une proposition est $[\bar{p}, \bar{\xi}, N(\bar{\xi})]$. Une telle variable ne dénote pas toutes les entités dans le même sens que les variables de Frege et Russell le font, elle exhibe seulement la structure commune à *toutes* les entités d'une certaine sorte : « Chaque variable est le signe d'un concept formel » (4.1271).

Outre la variable, nous avons vu que l'unicité du domaine ou du monde était une caractéristique de l'universalité chez Frege et Russell. Cette caractéristique est omniprésente dans le *Tractatus* : il n'existe qu'une seule et unique syntaxe logique invariable pour un seul et unique monde. La notation ou le symbolisme logique ne font que mettre en évidence les structures logiques existantes, elles n'en établissent aucune et, par conséquent, il ne peut exister de logiques différentes. Wittgenstein n'admet même pas de lois logiques que nous pourrions qualifier de vraies ou fausses comme le fait Russell, qui souscrit pourtant lui aussi à l'idée d'une logique universelle. Énoncer des propositions comme l'axiome de l'infini ou l'axiome de réductibilité comme hypothèse « logique » constitue un non sens; si elles sont vraies c'est qu'une notation logique bien conçue le montrera. L'axiome de l'infini, s'il s'agit bien d'un axiome logique, n'a pas à être une hypothèse qui tantôt peut être admise, tantôt être rejetée; la même chose s'applique à l'axiome de réductibilité. Le cadre de Wittgenstein est donc encore plus rigide que celui de Russell sur ce plan, la logique ne peut dépendre de nos choix, elle est indépendante, immuable et va sans dire.

Le concept de vérité est naturellement un autre lieu où se manifeste l'universalité d'un système. Celui du *Tractatus* s'articule essentiellement de la manière suivante : 1) d'abord, un tableau (une pensée ou une proposition), qui est un fait, représente un état de choses lorsque le tableau a la même forme de représentation que l'état de choses représenté; 2) une forme de représentation commune assure que les éléments du tableau sont en correspondance avec les objets de l'état de choses et qu'ils partagent essentiellement la même structure; 3) si l'état de choses représenté existe, alors le tableau est vrai, s'il n'existe pas le tableau est faux. Comme il a été dit plus haut, la représentation est une sorte d'isomorphisme (structurel) entre deux états de choses, avec d'un côté le tableau et de l'autre ce qui est représenté, et la représentation est vraie lorsque ce qui est représenté est un fait, c'est-à-dire un état de choses existant. Il est important de remarquer que le tableau, ou la proposition, est un fait, qu'il n'appartient pas à une catégorie d'entités séparées du monde – des entités linguistiques par exemple –, il appartient au monde. Cette conception simplifie énormément la sémantique de Wittgenstein par la symétrie qu'elle induit entre le tableau et l'état de choses : le sens d'une proposition est l'état de choses qu'elle représente, et la relation entre ces deux états de choses est l'isomorphisme induit par la projection. La vérité est une similarité de faits ou un identité de structure, et cette identité de structure se décide par la constitution du monde, et non par des stipulations linguistiques quelconques. Il n'y a qu'une seule vérité.

La fixité du troisième domaine de Frege menait également à la fixité de la vérité, mais nous avons vu comment il était simple d'adapter ses définitions à des stipulations modèle-théorétiques en relativisant le troisième domaine. Pour arriver au même résultat avec le *Tractatus*, il faudrait relativiser la structure logique du monde. Mais même en changeant la structure du monde, on ne retrouvera pas les définitions modèle-théorétiques aussi simplement. La raison en est que la vérité n'est pas transformable en prédicat dans le *Tractatus*.[20]

[20] Car la vérité ne figure pas dans un état de choses. Elle ne peut donc pas former un concept formel (un prédicat).

3. Le métalangage

Un métalangage est utilisé principalement pour préciser la sémantique et la syntaxe d'une théorie. Sur le plan sémantique, l'absence du prédicat de vérité dans le *Tractatus* bloque une porte d'entrée importante pour des notions métathéoriques. La vérité n'étant pas prédiquée aux propositions, il n'a pas lieu d'employer un métalangage dans lequel les propositions seraient des objets. Sur le plan syntaxique, l'harmonisation entre les notions de vérité et d'inférence dans le *Tractatus*, contribue à élimination des lois syntaxiques et, par conséquent, à l'élimination du rôle que joue le métalangage dans ce contexte. Rappelons que, selon Wittgenstein, ce n'est pas une loi déductive ou inférentielle qui assure la déduction, mais bien la forme logique de la proposition. Les possibilités de vérité d'une proposition déterminent entièrement les propositions qu'elle implique ou, inversement, les propositions qui l'impliquent. La présentation syntaxique d'une notation est non nécessaire et éliminable si elle met correctement en évidence – montre, pour ainsi dire – les possibilités de vérité de ses propositions. Il semblerait donc qu'un langage correctement construit montre ce que le métalangage dit.

La morale anti-métathéorique du *Tractatus* est claire : on ne peut pas dire la logique du monde, on peut seulement la montrer. La logique ne peut pas être un objet de représentation, donc aucun métalangage ne saurait en parler. Mais comment montrer la logique? Avec une notation logique qui respecte la structure logique du monde? Difficile de pénétrer à l'intérieur de ce cercle. Comment reconnaître et différencier la pseudo-notation logique qui mène à des antinomies – comme celle de Frege –, de la notation authentique qui révèle la structure du monde telle qu'elle est réellement? Comment évalue-t-on une notation, comment l'analyse-t-on sans la mentionner? Sur ce point, aucune innovation de la part de Wittgenstein, il justifie ses considérations de la même manière que ses prédécesseurs : par des motivations philosophiques d'ordre métathéorique. Pour ne prendre qu'un exemple, sa thèse à l'effet qu'une équation est une pseudo-proposition n'est pas justifiée en pointant une région précise de la réalité avec l'index de sa main droite, elle l'est grâce à des considérations faites sur la relation d'égalité qui sont *dites* et non pas *montrées*. En fait, l'ensemble du *Tractatus* s'emploie à dire ce qui, selon ses propres thèses, ne peut être que montré, de la syntaxe logique à la vérité d'une proposition en passant par les limites du monde et le solipsisme. Rien n'est hors de portée de cette philosophie qui pourtant s'érige en porte parole de l'indicibilité. L'excuse que Wittgenstein nous offre pour ses errements se trouve dans l'avant dernier paragraphe :

> 6.54 – Mes propositions servent d'élucidations de la manière suivante : celui qui me comprend reconnaît éventuellement qu'elles sont insensées lorsqu'il les a utilisé – comme des marches – pour les surpasser. (Il doit, d'une certaine façon, jeter l'échelle après l'avoir grimpée).
>
> Il doit transcender ces propositions, et il verra ensuite le monde correctement.

Ainsi, le *Tractatus* serait un non sens, mais pas n'importe lequel : un non sens capable de véhiculer rien de moins que la structure de la réalité dans son ensemble.

Il faut donner le mérite à Wittgenstein d'avoir voulu relever le défi universaliste de fonder la logique universelle du monde sans introduire les concepts primitifs par derrière, en marge ou en note de bas de page (5.452). Il est difficile toutefois de croire qu'il aurait réussi, eut-il choisi d'élaborer son système jusqu'au bout. Pour décrire les fonctions de vérité du langage, Wittgenstein, comme les autres, est obligé de *nommer* leurs conditions de vérité. Il est même amené à considérer le signe

p	q	
V	V	V
F	V	V
V	F	F
F	F	V

comme un signe propositionnel, c'est-à-dire comme une partie perceptible d'un tableau (du monde). Il y a, me semble-t-il, une tension manifeste entre le fait que la forme logique soit indicible et le fait qu'on puisse représenter les possibilités de vérité d'une proposition. Ces outils représentationnels dépassent largement le cadre du dicible fixé par la philosophie du *Tractatus*, il en va de même pour la philosophie logique qui en découle.

4. Épilogue carnapien

Le *Tractatus* eut un impact retentissant sur les penseurs de son époque, en particulier sur le cercle de Vienne, dont Rudolf Carnap faisait partie. Dans *The Logical Syntax of Langage*, Carnap emprunte une des idées directrices du *Tractatus* : la syntaxe ou la grammaire logique. Mais la syntaxe logique de Carnap n'a de commun avec celle du *Tractatus* que le nom. Carnap développe une position résolument formaliste et conventionnaliste soutenant que la logique reçoit sa signification par des postulats arbitrairement choisis de la syntaxe. Ainsi le rôle de la syntaxe logique n'est pas de dépister les véritables principes logiques, selon le principe de tolérance carnapien, il est d'établir des conventions sur la logique.[21] Qui plus est, en utilisant l'arithmétisation de la syntaxe, qui est reprise de Gödel,[22] Carnap prétend que la syntaxe peut être décrite dans le langage décrit par la syntaxe et ce, sans mener à des contradictions.[23] En conséquence, en projetant le métalangage dans le langage objet, Carnap se trouve à contredire non seulement Wittgenstein (pour qui la syntaxe est indicible), mais aussi le Russell de l'introduction du *Tractatus* (qui,

[21] Carnap 1937 : p. 51.
[22] Nous verrons les détails de l'arithmétisation de la syntaxe au chapitre V.
[23] Carnap 1937 : p. 67.

rappelons-le, a suggéré l'utilisation d'une suite de métalangages pour décrire la syntaxe) en se débarrassant de la hiérarchie de méta-syntaxes.

CHAPITRE IV

La métamathématique hilbertienne

0. Introduction

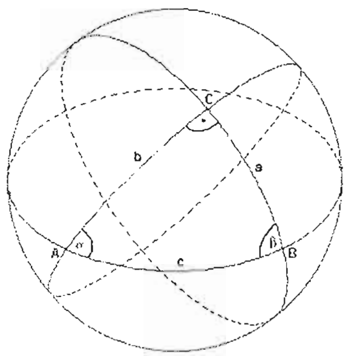
Le mérite d'avoir formalisé l'analyse métathéorique revient très certainement à Hilbert. On lui doit notamment la métamathématique, laquelle est l'avant précurseur du métalangage tel qu'on le retrouvera plus tard en logique mathématique. En un mot, la métamathématique est une discipline qui utilise des méthodes fiables pour évaluer la fiabilité des théories mathématiques. Les méthodes fiables en question sont celles de l'arithmétique intuitive, une théorie qui nous est donné *a priori*, et évaluer une théorie consiste à démontrer qu'elle est non contradictoire. La philosophie mathématique hilbertienne ne mise pas sur des essences platoniciennes pour garantir la certitude d'une théorie, elle se contente plutôt d'avoir une preuve de cohérence. Ainsi, pour Hilbert, du moment qu'un système ne mène à aucune contradiction, il est jugé admissible. Fonder dans le sens hilbertien consiste donc à donner des preuves de non contradiction, c'est là le mandat principal de la métamathématique.

1. Le deuxième problème de Hilbert

La géométrie euclidienne a toujours été un modèle de rigueur pour l'ensemble des mathématiques par la clarté de ses conclusions et la sévérité des ses méthodes. Ainsi, d'aucuns seront étonnés d'apprendre que la quête hilbertienne des fondements mathématiques commence, dans les années 1898-99, par une revue systématique de cette géométrie.[1] Cette revue comportait plusieurs objectifs, les plus importants étant de clarifier certaines preuves des *Éléments* pour exposer plus distinctement les hypothèses sur lesquelles elles reposaient, d'étudier les relations de dépendance entre les axiomes, et d'analyser les conséquences de l'abandon ou de la modification d'un ou de plusieurs de ces axiomes pour étudier les géométries qui pouvaient en résulter. Sur le plan des fondements mathématiques, c'est le dernier point qui retient notre attention : comment s'assurer que l'ensemble d'axiomes qui résulte de ces modifications est non contradictoire et définit, par conséquent, une géométrie?[2] Une façon de faire consiste à trouver une structure mathématique (préexistante) qui, à l'aide d'une interprétation adéquate du point et de la droite, rend les axiomes vrais. C'est ainsi que nous avons démontré l'« existence » des géométries non euclidiennes, en interprétant les notions de point et de droite de manière non conventionnelle dans des structures mathématiques autres que le plan euclidien.

[1] Ces recherches ont été exposées dans le cadre d'un séminaire sur la géométrie euclidienne à l'Université de Göttingen et furent publiées plus tard sous le titre de *Les fondements de la géométrie*.
[2] Le fait qu'un système d'axiomes définisse une géométrie découle de la conception hilbertienne de l'existence : exister c'est être non contradictoire.

Les géométries non euclidiennes les plus célèbres sont issues d'une modification du cinquième postulat d'Euclide, lequel stipule que par tout point P extérieur à une droite Δ passe une et une seule droite parallèle à Δ. Dans la géométrie sphérique, le cinquième postulat subit la modification suivante : au lieu qu'il y ait une et une seule droite passant par P et parallèle à Δ, il n'y en a aucune. Il existe un modèle très simple de la géométrie sphérique : on pose S la surface d'une certaine sphère, et on interprète les *points* de la géométrie comme étant les points de la sphère et les *droites* comme étant les géodésiques[3] de la sphère.[4] Sous cette interprétation, il est possible de démontrer que les axiomes de la géométrie sphérique sont vrais. Ainsi, ce modèle établit que : 1) l'ensemble des axiomes de la géométrie sphérique est non contradictoire, et 2) le postulat des parallèles est indépendant des autres axiomes géométriques.



La surface S sur laquelle sont représentés les points A, B et C, sommets du triangle formé par les droites a, b et c (des grands cercles).

Il faut remarquer que les résultats métathéoriques 1) et 2) dépendent de l'existence préalable de la sphère, ce qui n'est pas forcément donné comme hypothèse. Il est vrai toutefois que la sphère se laisse définir en fonction des nombres réels, le nombre réel comme un ensemble de rationnels, et le rationnel par une relation d'équivalence sur les entiers, de telle sorte que le support ultime de notre modèle soit l'arithmétique (avec la quantification de premier et de second ordre). Pour assurer l'existence de la géométrie sphérique (et bien d'autres modèles d'ailleurs), il « suffirait » donc de démontrer que l'arithmétique (du premier et du second ordre) est dénuée de contradictions. Pour ce faire, une preuve de cohérence

[3] Les géodésiques d'une sphère sont des grands cercles, c'est-à-dire des cercles sur la sphère de même rayon que celle-ci.
[4] Il faudrait également définir une notion de distance et d'angle pour que le tableau soit complet.

relative comme celle que nous avons donnée ci-dessus pour la géométrie sphérique ne nous est d'aucun secours, car une telle preuve dépendrait ultimement de l'arithmétique pour construire un modèle de l'arithmétique. Ce qu'il faut pour assurer la non contradiction de l'arithmétique est une preuve absolue de non contradiction. La nécessité d'une telle preuve pour consolider les fondements des mathématiques a été constatée par Hilbert. En 1900, lors du second congrès des mathématiques à Paris où il énonce les vingt-trois problèmes qui devaient marquer les recherches mathématiques du siècle à venir, le problème de prouver la cohérence des axiomes de l'arithmétique[5] arrive en deuxième position.[6] Ce problème de la cohérence préoccupera Hilbert tout au long de sa longue carrière, atteignant un sommet d'activité dans les années vingt et trente où il sera accompagné à Göttingen par Ackermann, Bernays, Gödel, Herbrand et Weyl.

2. Le finitisme

La philosophie des mathématiques de Hilbert est une conciliation entre l'intuitionnisme et l'instrumentalisme (ou le formalisme). Pour assurer leurs fondements, Hilbert crut bon de diviser les mathématiques en deux groupes : les mathématiques réelles, qui procèdent de notre intuition, et les mathématiques idéelles, qui la dépassent. La validité des mathématiques du premier groupe n'est pas problématique, elle est immédiate par la nature même de notre faculté de représentation. Par contre, les mathématiques du second groupe portent sur des objets qui dépassent les limites de notre intuition et en ce sens, leur validité est loin d'être assurée. L'intuitionnisme de Brouwer interdit toute forme de méthode mathématique qui ne porte pas sur un objet de l'intuition. L'intuitionnisme que Hilbert propose

[5] En réalité, il faut préciser que Hilbert entend parfois « arithmétique » dans un sens général qui comprend l'analyse réelle.

[6] Curieusement, la numérotation de Hilbert suggère que la démonstration de l'hypothèse du continu était pour lui un problème encore plus important (?!).

constitue un compromis : les mathématiques idéelles sont admissibles comme théorie mathématique dans la mesure où elles n'introduisent aucune contradiction. Cette thèse philosophique est connue sous le nom de *finitisme*, étant donné la nature *finie* des objets et des méthodes de la mathématique réelle.

La formulation la plus achevée que Hilbert a donné du finitisme se trouve dans *Uber das Unendliche* (*Sur l'infini*). Le problème des objets réels et idéaux en mathématiques est introduit par l'entremise de l'infini. On peut dire que les mathématiques font appel à deux sortes d'infinis : l'un actuel et l'autre potentiel. L'infini potentiel est l'infini d'Archimède, le fait que l'ensemble des nombres ne connaît aucune limite car quelque soit le nombre, il est toujours possible de lui ajouter un. L'infini actuel est l'idée d'un infini complété, d'un infini pris comme totalité achevée. Fonder l'infini actuel sur l'infini potentiel serait une manière de résumer brièvement l'entreprise finitiste. Hilbert entreprend cette tâche en élaborant une théorie de l'esprit qui rappelle l'architecture kantienne de la raison. Hilbert nous dit que

> Kant a déjà enseigné [...] que les mathématiques ont à leur disposition un contenu assuré indépendamment de toute logique et, par conséquent, qu'elles ne peuvent jamais être fondées par l'entremise de la logique seule [...]. Plutôt, comme condition pour l'utilisation des inférences logiques et l'exécution d'opérations logiques, quelque chose doit déjà être donnée à notre faculté de représentation, des objets extralogiques concrets qui sont présents intuitivement comme une expérience immédiate précédant toute forme de pensée. (van Heijenoort 1967 : p. 376, ma traduction)

En conséquence,

> Si une inférence logique est pour être fiable, il doit être possible d'analyser exhaustivement ces objets dans toutes leurs parties, et le fait qu'ils se produisent, qu'ils diffèrent l'un de l'autre, et qu'ils se suivent l'un l'autre, ou qu'ils soient concaténés, est immédiatement donné intuitivement, de la même manière que les objets, comme quelque

> chose qui ne peut être réduit à autre chose et qui ne nécessite aucune réduction. [...] [E]n mathématiques, en particulier, ce que nous considérons sont les signes concrets eux-mêmes, dont la forme, selon la conception que nous avons adoptée, est immédiatement claire et reconnaissable. (van Heijenoort 1967 : p. 376, ma traduction)

Hilbert construit sa philosophie des mathématiques sur l'idée d'un contenu intuitif (*inhaltlich*) qui donne à l'entendement des objets et des lois *a priori*. Le contenu intuitif ne suffit pas à explorer l'ensemble des mathématiques, mais seulement une portion limitée, soit l'arithmétique finitiste ou « contentuelle ». Les propriétés finitistes des nombres sont construites à partir d'observations ou de saisies immédiates d'un nombre. Hilbert définit le nombre comme étant (informellement) une suite quelconque de « 1 » (un code dans sa terminologie), un terme dans la suite

$$1, 11, 111, 1111, 11111...$$

Ces termes sont abrégés par des noms plus commodes et plus couramment employés : « 2 » pour « 11 », « 3 » pour « 111 », « 4 » pour « 1111 », etc. Il est donc possible à partir de cette perception intuitive (*a priori*) de définir les relations et les opérations de base sur lesquelles l'arithmétique est fondée. Par exemple, la relation d'égalité est définie comme l'identité entre deux suites, la relation $a < b$ comme le fait que le nombre b (en tant que suite de « 1 ») est un segment initial propre du nombre a. Ainsi, il est intuitivement évident que la loi de la trichotomie est valide, c'est-à-dire qu'une et une seule des trois possibilités $a < b$, $a = b$ ou $a > b$ est toujours le cas. Ensuite, l'addition de deux nombres a et b est la concaténation de la suite de « 1 » de a et de celle de b. De cette manière, il est intuitivement évident que $2 + 3 = 3 + 2$, car le fait d'ajouter trois fois un « 1 » au nombre 2 donne le même nombre que le nombre 3 augmenté de deux « 1 ». De manière générale, en utilisant l'induction intuitive (le pendant matériel du principe formel d'induction) il est possible de démontrer que l'addition est commutative, $n + m = m + n$, et associative, $(l + m) + n = l + (m + n)$.

L'induction intuitive, matérielle ou « contentuelle » est, soit dit au passage, un principe qui découle selon Hilbert de la construction des nombres et de notre manière de les représenter. En ce qui concerne la multiplication d'un nombre n par un nombre m : $a \cdot b$ s'obtient en remplaçant chaque « 1 » dans a par b. On peut montrer que cette opération est elle aussi commutative et associative et qu'elle est distributive sur « + » en employant l'induction matérielle. Il ne sera donc d'aucun étonnement d'apprendre qu'il existe un algorithme euclidien sur ces nombres et que l'ensemble des résultats de la théorie élémentaire des nombres s'ensuit, dont le théorème fondamental de l'arithmétique.

Cette arithmétique est celle qui procède directement de notre intuition et de la construction des nombres. Elle est irréductible et elle ne nécessite de toute manière aucune réduction. Cette arithmétique est toutefois limitée. Il lui manque d'importants contingents de la mathématique, notamment, tout ce qui est relatif à l'infini actuel. Parmi les concepts cardinaux qui touchent à l'infini, on retrouve les notions de quantification existentielle et universelle ainsi que la loi du tiers exclu; parmi les objets infinis en mathématiques, on retrouve les ensembles de tous ordres et en particulier tous les systèmes de nombres (actualisés). Sur le plan finitiste, les quantificateurs se définissent au mieux comme suit :

> Un *jugement universel* sur les codes [les nombres] ne peut être interprété de façon finitiste que dans un sens hypothétique, c'est-à-dire, comme un énoncé sur n'importe quel code donné. Un tel jugement exprime une loi qui doit se vérifier dans chaque cas isolé qui se présente.
>
> Un *théorème d'existence* sur des codes, donc un théorème de la forme « il existe un code n ayant la propriété $A(n)$ », se conçoit de manière finitiste comme un jugement partiel [...] qui se compose, soit de la donnée directe d'un code ayant la propriété $A(n)$, soit de la donnée d'un procédé pour la production d'un tel code[.] (Hilbert & Bernays 2001 : tome I, p.87)[7]

[7] On pourrait contester que Hilbert a une définition plus « généreuse » des quantificateurs dans l'arithmétique contentuel tel qu'il la pratique dans (Hilbert & Bernays 2001).

La notion de vérité est ainsi identifiée à un calcul ou à une vérification : la proposition p est vraie si et seulement si elle est vérifiée. Mais, puisque l'intellect finitiste ne peut accomplir qu'un nombre fini d'opérations dans une vérification, il est incapable de vérifier un énoncé comme « $A(n)$ est vrai pour tout n » et en conséquence, il est incapable d'affirmer sa vérité. Le sort du tiers exclu est intimement relié à celui des propositions universelles et existentielles. Si « être vrai » c'est « être vérifié », il n'est plus très plausible de considérer inéluctable une loi qui s'énoncerait alors comme

Pour tout nombre n, $A(n)$ ou $\neg A(n)$ est vérifié

car cela présupposerait que l'intellect mathématique a scruté un ensemble infini de propositions, à savoir $A(0), A(1), A(2),\ldots$ et $\neg A(0), \neg A(1), \neg A(2),\ldots$ Hilbert se trouve donc à épouser les conclusions de Brouwer à propos de la validité du principe du tiers exclu : « Du point de vue finitiste, cette non-validité existe en fait, dans la mesure où on ne réussit pas à y trouver, pour le jugement existentiel aussi bien que pour l'universel, une négation de contenu [*Inhalt*] finitiste satisfaisant au principe du tiers exclu ».[8]

3. La métamathématique

Contrairement à Brouwer ou à Kronecker, Hilbert ne prend pas la condition intuitionniste comme finale. Il considère que le mathématicien n'a pas à s'aliéner les outils fondamentaux que sont, entre autres, le principe du tiers exclu, l'infini actuel et la théorie des ensembles transfinis, outils qui font selon lui la fertilité des

[8] Hilbert & Bernays 2001 : tome I, p. 89.

mathématiques. Malgré le fait que l'infini n'existe ni dans la réalité ni dans la psychologie du mathématicien, il se pourrait très bien qu'il joue un rôle dans notre manière de penser et qu'il soit indispensable à la théorie mathématique. Comment introduire l'infini sans se commettre à l'infini? Comment parler de l'infini si l'infini n'est pas un objet d'intuition valable? La solution consiste à traiter l'infini comme une entité idéelle, une entité qui n'a pas de contenu intuitif réel et qui se définit seulement par le rôle qu'elle possède à l'intérieur d'une théorie. On se souviendra que Russell et Whitehead avaient introduit la notion de classe par l'entremise des symboles incomplets. La sémantique que Hilbert développe pour les entités idéelles comporte certaines ressemblances; elle renonce à leur donner un contenu déterminé, se contentant seulement de préciser les axiomes qui les définissent.

Ainsi, le nombre imaginaire i n'est qu'un objet de pensée fictif (c'est-à-dire un symbole) et pour s'en servir, il suffit de le considérer comme un objet arithmétique avec la propriété supplémentaire $i^2 = -1$. Le fait qu'aucun contenu mental ou mondain ne soit associé à i n'interfère pas dans nos calculs. En effet, en supposant que la commutativité, l'associativité et la distributivité tiennent avec i, nous avons en main le nécessaire pour faire la multiplication suivante :

$$(2+3i)(1-i) = 2-2i+3i-3i^2 = 5+i.$$

L'infini a été introduit dans la théorie des ensembles de Zermelo-Fraenkel essentiellement de la même manière. Un ensemble infini n'est rien d'autre qu'un objet X dont l'usage est régi par les deux axiomes :

(1) L'ensemble vide appartient à X, $\varnothing \in X$;

(2) Si $x \in X$ alors le successeur de x appartient à X, c'est-à-dire $x \cup \{x\} \in X$.

C'est grâce à l'ensemble infini (aux axiomes 1 et 2) qu'il est possible de définir l'ensemble des nombres naturels $\mathbb{N}$ comme totalité achevée dans cette théorie :

$$\mathbb{N} = \bigcap \{ E \mid \varnothing \in E \wedge \forall x (x \in E \supset x \cup \{x\} \in E) \},$$

$\mathbb{N}$ est donc le plus petit ensemble ayant la propriété « inductive » : $\varnothing \in E$ et $\forall x (x \in E \supset x \cup \{x\} \in E)$.

La méthode axiomatique permet ainsi d'introduire quantité d'objets fictifs avec une très grande facilité. Pour prendre un autre exemple, ajoutons aux nombres naturels une borne supérieure ∞. Encore une fois, il suffit seulement de dicter le comportement de ∞ à l'endroit de ses comparses : pour $n \in \mathbb{N}$, posons

(1) $n + \infty = \infty + n = \infty$ et $\infty + \infty = \infty$;
(2) $0 \cdot \infty = \infty \cdot 0 = 0$, $\infty \cdot \infty = \infty$ et si $n \geq 1$ alors $n \cdot \infty = \infty \cdot n = \infty$;
(3) $n < \infty$.

Puisque ces règles précisent comment manipuler ∞ avec les opérations « + » et « · » et les relations « = » et « < », la vérité de toute proposition arithmétique portant sur les objets de $\mathbb{N} \cup \{\infty\}$ pourra être déterminée en dépit de l'inexistence de ∞.

La facilité avec laquelle nous pouvons introduire un objet fictif comme ∞ est minée par l'exigence de démontrer qu'il ne mène à aucune contradiction, c'est-à-dire de démontrer qu'aucune contradiction ne peut être déduite des axiomes ajoutés pour définir le nouvel objet. Si la condition « $0 \cdot \infty = \infty \cdot 0 = 0$ » dans le point (2) du paragraphe précédent était remplacée par « $0 \cdot \infty = \infty \cdot 0 = 1$ » (et qu'on supposait valide l'associativité de la multiplication à nouveau), on obtiendrait que

$$1 = (n \cdot 0) \cdot \infty = n \cdot (0 \cdot \infty) = n$$

quelque soit *n*. Ce qui est évidemment une contradiction lorsque $n \neq 1$. Le finitisme doit donc être doublé d'une méthode métathéorique pour démontrer la non-contradiction des systèmes d'axiomes, une méthode qui par ailleurs ne peut pas dépasser les limites de l'arithmétique contentuelle. La méthode de Hilbert consiste à considérer la théorie mathématique idéelle comme une entité symbolique (finie) que l'on scruterait à l'aide de l'arithmétique intuitive :

> Si l'arithmétique usuelle est formalisée, c'est-à-dire si ses hypothèses et ses modes d'inférence sont traduits en formules initiales et en règles de déduction, une preuve arithmétique se présente comme une succession, intuitivement perceptible d'un coup d'œil, de processus dont chacun appartient à une liste, donnée à l'avance, des traitements qui entrent en ligne de compte. (Hilbert & Bernays 2001 : t. I, p. 98-9)

Les propositions de la mathématique formalisée ne sont rien d'autre que des successions de symboles, ce à quoi peut s'appliquer l'arithmétique contentuelle (à quelques modifications près). Une preuve de non-contradiction d'une théorie formalisée n'aurait donc rien d'essentiellement différent d'une preuve de l'irrationalité de $\sqrt{2}$. L'analyse des théories mathématiques formalisées par l'arithmétique intuitive est ce que Hilbert nomme la *théorie de la preuve* ou la *métamathématique*.

4. Une première preuve de cohérence

Hilbert a illustré très tôt (en 1904, à la conférence d'Heidelberg) sa démarche métamathématique avec une preuve de cohérence pour un fragment de théorie mathématique très simple. C'est n'est pas tant le résultat que la méthode qui est intéressante dans cet exemple, celle-ci contenant en germe plusieurs idées directrices de la théorie de la preuve. Le fragment de théorie arithmétique dont il est question porte sur les objets composés des symboles primitifs suivants : 1, = , (,), f , f' et u , un terme de la théorie étant une combinaison quelconque de ces sept symboles. Par exemple, les expressions « 11= », « ((11)(1))(1) », « 1(= 1)f' » et « (1)(===)uf » sont des termes. Les termes seront éventuellement divisés en deux catégories disjointes – les entités et les non entités – par les axiomes de la théorie, et ces catégories serviront à caractériser le vrai et le faux.

Les cinq axiomes de la théorie sont :

1. $x = x$,
2. $\big(x = y \wedge F(x)\big) \supset F(y)$,
3. $f(ux) = u(f'x)$,
4. $f(ux) = f(uy) \supset ux = uy$,
5. $\overline{f(ux) = u1}$,[9]

auxquels on ajoutera quelques rudiments du calcul des prédicats, de quoi faire des déductions. Intuitivement, u signifie un ensemble infini, ux un élément de l'ensemble u , f est la fonction successeur et f' , sa fonction accompagnatrice.[10] Le premier et le second axiomes définissent une propriété familière de l'égalité; le troisième stipule que tout élément ux possède un successeur $f(ux)$ et ce successeur est égal à un élément de l'ensemble u , à savoir $f'(ux)$; le quatrième axiome stipule que si deux éléments de u possèdent le même successeur alors ces éléments sont égaux; et le dernier affirme que 1 n'est le successeur d'aucun élément (qu'il est le premier élément).[11]

---

[9] La barre horizontale signifie la négation.

[10] Cette fonction est nécessaire afin de dire que le successeur de x appartient à u tout en préservant l'homogénéité des équations.

[11] La négation est représentée par la barre horizontale sur l'expression.

Les expressions qui résulteront de l'application des axiomes 1 à 4 aux termes de la théorie (en substituant pour x des termes quelconques et en déduisant les conséquences) formeront la catégorie des entités, et les expressions qui résulteront de l'application de l'axiome 5 à des termes formeront les non-entités. Nous considérerons les expressions dans la catégorie des entités comme des propositions vraies et les expressions dans la catégorie des non-entités comme des propositions fausses. Nous voudrions prouver qu'il est impossible de démontrer une contradiction, c'est-à-dire prouver qu'il n'y a pas d'expression qui soit à la fois une entité et une non-entité.

La preuve de non contradiction utilisera une propriété invariante des entités, soit l'homogénéité : une expression sera dite homogène si, en faisant abstraction des parenthèses, il y a un nombre égal de symboles à gauche et à droite de l'égalité. Par exemple, toutes les expressions issues des axiomes 1 et 2 seront de la forme $\alpha = \alpha$ où α est un terme quelconque. Ces axiomes engendrent donc des expressions homogènes. On observera que l'application des axiomes 3 et 4 engendre également des expressions homogènes car les expressions qui y figurent sont elles aussi homogènes.[12] Il en résulte donc que toutes les entités sont homogènes.

Toutes les non-entités sont de la forme $f(ut) = u1$, où t est un terme quelconque. Si une contradiction existe dans cette théorie, c'est qu'une expression de la forme $f(ut) = u1$ est une entité. Or, $f(ut) = u1$ n'est pas homogène, donc elle ne peut pas être une entité. La cohérence de la théorie s'ensuit par conséquent.

Cette preuve est emblématique de la métamathématique hilbertienne : la théorie est décrite par une famille d'axiomes portant sur des objets symboliques et l'arithmétique élémentaire est ensuite utilisée pour démontrer qu'aucune contradiction ne peut survenir dans la théorie. L'arithmétique élémentaire est nécessaire, en dépit des apparences, afin de donner une forme rigoureuse à

---

[12] Pour simplifier l'exemple, on ne considérera que le conséquent dans les axiomes 2 et 4.

l'argumentation, une constatation que Poincaré n'a pas manqué de faire.[13] Les affirmations à l'effet que 1) les entités sont toutes homogènes, 2) que les non entités ne le sont pas et 3) qu'aucune expression de la forme $f(ut) = u1$ ne peut résulter des axiomes 1 à 4 doivent toutes êtres démontrées par le principe d'induction. Poincaré trouve téméraire de penser qu'une preuve de cohérence de l'arithmétique est possible par cette voie, alors qu'elle dépendrait ultimement de ce dont elle tente de démontrer la non-contradiction : les principes de l'arithmétique. Selon lui, une telle entreprise serait circulaire, sans mentionner le fait que les théories pour lesquelles ces démonstrations sont faites ne signifient rien, qu'elles ne sont que des combinatoires de symboles.

5. Le calcul epsilon

Dans le cadre de l'arithmétique contentuelle, nous avons vu que les quantificateurs universel et existentiel ont un sens très particulier : démontrer une proposition générale consiste à donner une méthode permettant de démontrer toutes ses instances (comme c'est le cas pour $n + m = m + n$), et démontrer une proposition existentielle consiste à donner une façon de construire une valeur satisfaisant le prédicat quantifié existentiellement. Malheureusement, les quantificateurs contentuels sont loin d'épouser toute la généralité que possèdent les quantificateurs de la théorie mathématique ordinaire; notamment, il est impossible de déduire à partir de leurs définitions respectives que $(x)\big(A(x) \vee \neg A(x)\big)$ devrait être valide ou que $(Ex)\neg A(x)$ découlerait de $\neg(x)A(x)$. En gros, l'arithmétique contentuelle n'endosse pas la version transfinie du principe du tiers exclu, un principe si cher au développement des mathématiques ordinaires. Les quantificateurs existentiel et universel pris dans leurs

[13] Poincaré 1906.

significations respectives constituent donc des éléments idéels, et la métamathématique devra démontrer qu'ils ne mènent à aucune contradiction.

Pour ce faire, Hilbert se donne une extension du calcul des prédicats, le calcul epsilon. L'idée est d'ajouter une expression au calcul des prédicats (à l'aide d'un axiome) pour unifier le traitement des axiomes portant sur la quantification universelle et existentielle (les axiomes transfinis). L'axiome en question est

$$A(x) \to A(\varepsilon_x A(x))$$

où $\varepsilon_x A(x)$ doit être interprété comme étant un élément (s'il en existe un) qui satisfait le prédicat $A(x)$, le choix de cet élément n'étant pas déterminé. L'opération ε est une généralisation de l'opérateur de description définie de Russell et Whitehead ι défini par Hilbert comme suit :

$$(Ex)\big(A(x) \wedge (y)\big(A(y) \to x = y\big)\big) \to A(\iota_x A(x)),$$ [14]

c'est-à-dire que $\iota_x A(x)$ est l'unique élément qui satisfait $A(x)$ (s'il existe). On se souviendra que Russell avait réduit l'opérateur ι aux quantificateurs existentiel et universel. Hilbert fait le contraire avec l'opérateur ε : il réduit les quantificateurs à des expressions contenant le symbole ε. En effet, on peut démontrer que

$$(Ex)A(x) \equiv A(\varepsilon_x A(x)) \text{ et } (x)A(x) \equiv A(\varepsilon_x \neg A(x)).$$

Dans le calcul des prédicats augmenté de l'opérateur ε, il est donc possible de remplacer toutes les formules idéelles du calcul des prédicats par des formules qui font seulement intervenir le symbole ε de Hilbert.[15]

[14] Comme d'habitude, je reste fidèle aux notations des auteurs.

Les éléments idéels ont pour but de faciliter la démonstration des théorèmes de la mathématique finitiste réelle. Selon la philosophie mathématique hilbertienne (qui prêche la pureté des méthodes), ils devraient être éliminables de la démonstration d'un théorème qui ne fait pas intervenir le symbole ε dans ses hypothèses ou sa conclusion. Cette conviction est confirmée par le premier théorème epsilon :[16]

(ε1) Si F et $A_1,...A_n$ sont des formules ne contenant pas le symbole ε et si F est déductible de $A_1,...A_n$ dans le calcul epsilon, alors il existe une preuve de F à partir de $A_1,...A_n$ dans le calcul des prédicats sans quantificateurs (sans symbole ε).[17]

Avec ce théorème, il est possible d'illustrer une preuve de non-contradiction pour une théorie arithmétique dont les axiomes ne contiennent pas de variables liées. Il s'agit de l'arithmétique définie par les axiomes :

$(=_1)$ $x = x$

$(=_2)$ $x = y \rightarrow (A(x) \rightarrow A(y))$

$(<_1)$ $\neg x < x$

$(<_2)$ $(x < y \wedge y < z) \rightarrow x < z$

$(<_3)$ $x < x'$

(P_1) $x' \neq 0$

(P_2) $x' = y' \rightarrow x = y$.[18]

---

[15] Il faut mentionner aussi que le calcul epsilon possède une seule règle inférentielle relative aux prédicats, la règle de substitution : de $A(x)$ on peut inférer $A(t)$ où t est un terme quelconque.

[16] Hilbert & Bernays 2001b : p. 68.

[17] Ce théorème établit donc que les mathématiques réelles sont démontrables dans les mathématiques réelles.

[18] C'est la règle de substitution (de l'avant avant dernière note) qui rend possible l'utilisation d'un système d'axiomes à variables libres.

Appelons ce système d'axiomes (S). Le métathéorème s'énonce plus précisément comme suit : *une formule numérique*[19] *déductible des axiomes de* (S) *dans le calcul epsilon est forcément vraie.*[20]

La preuve de ce métathéorème découle de $(\varepsilon 1)$ et de deux autres observations métathéoriques :

(1) Une preuve d'une formule numérique n'utilisant que le calcul des prédicats sans quantificateurs peut être transformée (par le transfert à rebours des substitutions) en une preuve ne contenant aucune variable libre, et où chaque formule initiale est soit une substitution des axiomes de (S), soit une instance numérique d'une tautologie du calcul propositionnel.
(2) Toute formule numérique obtenue par substitution d'une tautologie est vraie; toute formule numérique résultant d'une substitution des axiomes de (S) est également vraie; et si A et $A \to B$ sont vraies alors B est vraie aussi.[21]

La preuve de non-contradiction procède ainsi : si F est une formule numérique déduite dans le calcul epsilon à partir de nos axiomes arithmétiques, le premier théorème epsilon nous assure qu'il existe une preuve de F dans le calcul des prédicats sans quantificateurs. Par la remarque (1), cette preuve peut être transformée davantage en une preuve qui ne contient que des formules numériques et qui possède comme formules initiales des instances numériques d'axiomes ou de tautologies propositionnelles. Par (2), toutes les formules initiales de cette preuve sont vraies et toutes les conséquences des ces formules obtenues par *modus ponens* le sont également. Or, puisque les formules de la preuve sont toutes numériques, la seule

[19] Une formule numérique est une formule ne contenant ni variable libre ni variable liée.
[20] Cet énoncé n'empêche pas toutefois qu'une formule numérique fausse puisse être déduite à partir d'hypothèses transfinies.
[21] Hilbert & Bernays 2001 : tome II, p. 85.

règle applicable dans celle-ci est le *modus ponens*. Donc, toutes les formules de la preuve sont vraies, et en particulier la formule F est vraie.

Il va sans dire que l'argument décrit ci-dessus n'est qu'une esquisse de preuve et que son exposition rigoureuse demanderait qu'on examine l'ensemble des affirmations contenues dans $(\varepsilon 1)$, (1) et (2), dont celles qui se rapportent au transfert des substitutions et à la transformation des preuves. En dépit de ces lacunes, la preuve donne cependant une mesure des techniques et des méthodes employées par Hilbert et son école pour résoudre le problème des fondements mathématiques (ce qui était mon objectif ici). On remarquera que la théorie arithmétique décrite par les axiomes $=_1$, $=_2$, $<_1$, $<_2$, $<_3$, P_1 et P_2 n'a pas de principe d'induction, elle est donc un fragment très faible de l'arithmétique ordinaire. Hilbert et Ackermann ont généralisé ces méthodes afin de tenter une preuve de non-contradiction pour l'arithmétique dans sa totalité, ce que Ackermann parvint à faire en 1940 avec certaines difficultés.[22]

6. Le programme de Hilbert et le métalangage

En instituant la métamathématique comme méthode pour résoudre le problème des fondements, Hilbert fait du métalangage une notion cardinale pour l'étude de toute théorie mathématique. Ce qui rend cette approche possible est notamment une conception symbolique de l'objet et de la théorie mathématiques. La preuve de non contradiction illustrée à la section 4 ne porte pas sur une théorie mais sur des symboles. Elle ne dépend que des propriétés symboliques du langage, et ne fait jamais référence à la structure définie par les axiomes. De manière générale, toute théorie mathématique qui dépasse le cadre finitiste est également dépourvue de

[22] Ackermann crut atteindre le but en 1924, mais suite à la publication des théorèmes d'incomplétudes de Gödel, on en a déduit que sa preuve était forcément erronée. En 1940, Ackermann complète une preuve de non contradiction avec le calcul epsilon en s'inspirant de la preuve par induction transfinie de Gentzen (1936). Il va sans dire que cette preuve dépasse le cadre de l'arithmétique contentuelle.

référence, les seules véritables propriétés qu'elle possède sur le plan contentuel sont les propriétés symboliques de sa formalisation. Le contenu procède du symbole à la théorie mathématique et non l'inverse, car le symbole est l'objet par excellence de la « raison arithmétique contentuelle ». Dans la sémantique hilbertienne, le symbole n'a pas de signification, il *est* la signification.

Cette curieuse sémantique semble bousculer le rapport traditionnel entre le symbole et sa signification. Chez Frege et Russell, l'objet logique est ce qui est désigné par l'expression symbolique et non pas l'expression elle-même, l'objet est donc indépendant de notre façon de le dénoter. Mais pour Hilbert, la logique est une question de manipulations symboliques, objet et expression sont fondus en un. Cette curiosité sémantique n'est pas sans incidence sur sa conception métalinguistique. La métamathématique n'est pas vraiment une métathéorie, elle est une seule et même théorie, la seule véritable théorie : l'arithmétique contentuelle. De la même manière, les objets de la métamathématique sont toujours les mêmes, ce sont des symboles, en dépit du fait qu'on les appelle « axiomes de l'analyse réelle », « axiomes de la géométrie » ou encore « nombres naturels ». Hilbert ne formule pas la relation entre une théorie objet et son métalangage de manière générale, il n'a qu'une seule théorie – l'arithmétique réelle – qui porte sur un seul type d'objet – le symbole – qui prend parfois plusieurs formes.

Enfin, le programme de Hilbert, dans sa forme originelle, s'avérera impossible à réaliser comme le montreront les théorèmes d'incomplétudes de Gödel (particulièrement le deuxième) que nous verrons au prochain chapitre. Le deuxième théorème d'incomplétude montre que l'arithmétique ne peut démontrer sa propre non-contradiction; ainsi, l'arithmétique contentuelle se trouve à être insuffisante pour réaliser le programme. La notion de métalangage telle que pratiquée dans la métamathématique devra donc être révisée.

CHAPITRE V

L'analyse métamathématique de Gödel

0. Introduction

Les contributions de Kurt Gödel à la logique sont formidables. Parmi les plus remarquables, on retrouve des résultats sur la complétude du calcul des prédicats, sur l'incomplétude de l'arithmétique, sur l'impossibilité du programme de Hilbert et sur la compatibilité de l'axiome du choix avec l'hypothèse du continu. En ce qui nous concerne, Gödel est un artisan important pour la notion de métalangage. D'abord, pour la méthode de l'arithmétisation de la métamathématique, qui est une manière de projeter le métalangage dans le langage, et d'autre part, pour avoir démontré l'insuffisance de la métamathématique comme métathéorie pour l'arithmétique.

Les questions de la cohérence et de la complétude seront les thèmes dominants de la recherche de Gödel. Déjà en 1930, la complétude est au centre de ses préoccupations :

> Évidemment, lorsqu'une telle procédure [la déduction dans un système formel] est suivie la question se pose à savoir si le système d'axiomes et de principes inférentiels postulé à l'origine est complet ou pas, c'est-à-dire s'il est effectivement suffisant pour la dérivation de toute proposition logicomathématique vraie, ou s'il est possible qu'il y ait des propositions vraies (qui pourraient être démontrées avec d'autres principes) qui ne peuvent pas être dérivées dans le système en question. (van Heijenoort 1967 : p. 583, ma traduction)[1]

En bon hilbertien, Gödel préconisera l'approche métamathématique pour répondre à la question de la complétude, et la réponse prendra la forme des célèbres théorèmes d'incomplétude.

Ce chapitre sera consacré en grande partie à revoir les preuves originales de Gödel de ces théorèmes. Outre l'intérêt pour les résultats en tant que tels, nous serons particulièrement attentifs aux méthodes métathéoriques à l'œuvre dans ces démonstrations,[2] des méthodes qui caractérisent ce que j'appellerais la métamathématique gödelienne.

1. Un avant goût

Si P est la théorie composée des *Principia* auxquels on a ajouté les axiomes de l'arithmétique de Peano, les théorèmes d'incomplétude s'énoncent comme suit :

THÉORÈME GR (Gödel-Rosser).[3] Si P est cohérent (non contradictoire), il existe un énoncé de P qui est indécidable.[4]

---

[1] Ce passage fait sourciller van Heijenoort ailleurs (van Heijenoort 1985 : p. 14).

[2] Nous nous attarderons également à l'erreur dans la démonstration du second théorème d'incomplétude que j'ai signalé dans l'introduction.

[3] La formulation originale de Gödel (que nous verrons plus loin) était moins forte que celle-ci : il fallait supposer que P était ω-cohérent. L'amélioration est due à Rosser.

[4] Une formule close F est indécidable dans un système si F et $\sim F$ ne sont pas démontrables dans ce système.

THÉORÈME G2. Si P est cohérent, une preuve de la cohérence de P ne peut être formalisée dans P.

Qui plus est, le premier théorème n'affecte pas seulement la théorie P, mais toute théorie récursivement axiomatisable qui contient suffisamment d'arithmétique. Ainsi, la théorie des ensembles de Zermelo-Fraenkel ou celle de von Neumann-Bernays-Gödel souffrent également de cette pathologie. Le deuxième théorème est évidemment responsable de la chute du programme de Hilbert.

Pour arriver à ces théorèmes, Gödel a dû pousser l'analogie finitiste entre les nombres et les expressions symboliques plus loin que la simple allégorie philosophique. Plus précisément, si $\mathfrak{E}$ est l'ensemble des expressions de P, Gödel définit une fonction Γ qui associe à chaque $\mathcal{E}$ dans $\mathfrak{E}$ un nombre naturel $\Gamma(\mathcal{E})$, et ce nombre naturel est tel que les propriétés syntaxiques de $\mathcal{E}$ sont lisibles à partir de la décomposition en nombres premiers de $\Gamma(\mathcal{E})$. En transformant les expressions en nombres, Gödel permet la réalisation de l'idéal finitiste : transformer l'analyse métathéorique des systèmes formels en arithmétique élémentaire, c'est-à-dire arithmétiser la métamathématique. Le fait de transformer les énoncés métathéoriques en énoncés arithmétiques par la fonction de Gödel Γ est le nerf de l'arithmétisation de la métamathématique.

Avec cette description sommaire de l'arithmétisation, il est possible de donner une description sommaire de la démonstration du premier théorème d'incomplétude. Nous construisons d'abord la relation $Dém(x,y)$ ainsi que les deux fonctions $N(x)$ et $Sub(x,y,z)$ suivantes :

- $Dém(x,y) \equiv x$ est le nombre de Gödel d'une preuve de la formule dont le nombre de Gödel est y;
- $N(x)$ = le nombre de Gödel associé au nombre x;

- $Sub(x,y,z)$ = le nombre de Gödel de la formule obtenue de la formule dont le nombre de Gödel est x en remplaçant la variable dont le nombre de Gödel est y par le nombre z.[5]

Nous posons ensuite $Q(y) \equiv \forall x \sim Dém(x, Sub(y, 19, N(y)))$ où 19 est le nombre associé à la variable y. $Q(y)$ est une formule de P, elle est donc associée à un nombre, disons le nombre q. Considérons $\mathcal{G} \equiv Q(q) \equiv \forall x \sim Dém(x, Sub(q, 19, N(q)))$. Nous allons montrer informellement que $\mathcal{G}$ est indécidable. Supposons que $\mathcal{G}$ est démontrable, alors $\mathcal{G}$ est vraie et donc $Q(q)$ est vraie. Si $Q(q)$ est vraie c'est qu'il n'existe aucun nombre x tel que x est le nombre associé à une démonstration de la formule associée au nombre $Sub(q, 19, N(q))$, c'est-à-dire que la formule associée au nombre $Sub(q, 19, N(q))$ est indémontrable. Or la formule indémontrable en question est la formule obtenue de la formule associée à q en remplaçant la variable y par le nombre q, c'est-à-dire que $Sub(q, 19, N(q))$ est le nombre de la formule $Q(q)$ (la formule $\mathcal{G}$). Ce qui de toute évidence est contradictoire. $\mathcal{G}$ n'est donc pas démontrable. Par ailleurs, $\mathcal{G}$ ne peut pas être réfutable non plus car, si tel était le cas, $\mathcal{G}$ serait fausse et il existerait un n tel que $Dém(n, Sub(q, 19, N(q)))$ serait vraie. Ainsi la formule associée au nombre $Sub(q, 19, N(q))$, en l'occurrence $\mathcal{G}$, serait démontrable ce qui est contradictoire encore une fois.

Fait remarquable s'il en est un, la démonstration formelle de ce résultat ne fera pas appel à la vérité d'une formule ou à sa signification, elle sera formulée uniquement dans les termes de la théorie formelle. Ainsi, l'incomplétude est un défaut purement syntaxique de la théorie.

[5] Inutile de spécifier quelles définitions métathéoriques cette relation et ces fonctions arithmétisent.

2. La métamathématique gödelienne

2.1. Le langage *P*

Les célèbres résultats de Gödel sont démontrés, à la base, pour une extension *P* du système des *Principia* auquel on a ajouté les axiomes arithmétiques de Peano. Que l'on choisisse d'inclure ou non l'axiome de réductibilité et l'axiome de l'infini ne change rien, puisque nous avons à notre disposition les axiomes de Peano pour définir l'arithmétique. Les symboles primitifs de *P* sont composés (1) de constantes primitives « $\sim$ » (non), « $\vee$ » (ou), « $\forall$ » (pour tout), « 0 » (zéro), « ' » (fonction successeur), « (» et «) » (parenthèses), et (2) de variables pour tous les types, $x_1, x_2, x_3,...$ pour le premier type, $y_1, y_2, y_3,...$ pour le second, $z_1, z_2, z_3,...$ pour le troisième, etc. Les *termes* de premier type seront tous les termes $a, a', a'', a''',...$ où a est soit 0 soit une variable de type 1. Les termes de type $n > 1$ sont les variables de type *n*. Une expression de la forme $a(b)$, où b est un terme de type n et a un terme de type $n+1$, est une *formule élémentaire*. L'ensemble des *formules* est défini comme le plus petit ensemble X satisfaisant les contraintes (1) X contient tous les formules élémentaires et (2) si $a, b \in X$ et x est une variable quelconque alors $\sim a$, $(a \vee b)$ et $\forall x\, a$ sont dans X. Si n est le nombre de *variables libres*[6] de a, on dira que a est une *formule close* si $n = 0$ et que a est une *relation* si $n > 0$ (parfois on dira *prédicat* pour $n = 1$).

Les *axiomes* de cette théorie sont relativement simples et en petit nombre (les symboles « $\wedge$ », « $\supset$ », « $\equiv$ », « $\exists$ » et « = » seront utilisés pour abréger leur écriture) :

[6] La définition de la variable libre ressemble à celle qui de la formule.

I. **1.** $\sim x_1{}'=0$ [7]

2. $x_1{}'=x_2{}'\supset x_1=x_2$

3. $\left(y_1(0)\wedge\forall x_1\, y_1(x_1)\supset y_1(x_1{}')\right)\supset\forall x_1\, y_1(x_1)$

II. Toutes les formules qui résultent des schèmes un à quatre suivants, en substituant p, q ou r par des formules, sont des axiomes :

1. $(p\vee p)\supset p$

2. $p\supset(p\vee q)$

3. $(p\vee q)\supset(q\vee p)$

4. $(p\supset q)\supset\left((r\vee p)\supset(r\vee q)\right)$

III. Toutes les formules qui résultent des schèmes suivants sont des axiomes ($Subst(a,x,c)$ est la formule obtenue de la formule a en substituant la variable x pour c, c'est une relation métathéorique) :

1. $\forall x\, a\supset Subst(a,x,c)$

2. $\forall x(b\vee a)\supset b\vee\forall x\, a$

où x est une variable quelconque et b est une formule ne contenant pas x comme variable libre.

IV. Les formules qui résultent du schème suivant sont des axiomes :

1. $\exists y\left(\forall x\left(y(x)\equiv a\right)\right)$

où a est une formule qui ne contient pas x comme variable libre (et sujet à la condition que les types relatifs des variables x et y soit corrects). C'est l'axiome de compréhension.

V. Les formules qui résultent du schème suivant en variant le type relatif sont des axiomes :

1. $\forall x_1\left(y_1(x_1)\equiv y_2(x_1)\right)\supset y_1=y_2$

Il s'agit de l'axiome d'extension.

---

[7] La relation d'égalité n'est pas primitive dans P, elle est définie en fonction des opérateurs primitifs du calcul des classes des *Principia*.

Les règles inférentielles du langage sont le *modus ponens* ainsi que la généralisation. Une formule c est dite une *conséquence immédiate* des formules a et b si a est la formule $(\sim b \vee c)$ ou si a est $\forall x a$ où x est une variable quelconque. La classe des formules prouvables est le plus petit ensemble X tel que (1) les axiomes sont dans X et (2) si $a, b \in X$ les conséquences immédiates de a et b sont dans X.

2.2. Les nombres de Gödel

Le procédé d'arithmétisation de Gödel procède en associant un nombre à chaque symbole primitif du langage :

0	'	$\sim$	$\vee$	$\forall$	(	)
1	3	5	7	9	11	13

Il en va de même pour chaque symbole de variable : si p_n est le n-ième premier

x_1	x_2	$\cdots$	x_n	y_1	y_2	$\cdots$	y_n	z_1	z_2	$\cdots$	z_n	$\cdots$
17	19	$\cdots$	p_{n+6}	17^2	19^2	$\cdots$	p_{n+6}^2	17^3	19^3	$\cdots$	p_{n+6}^3	$\cdots$

Le nombre associé à une expression $S_1 S_2 \ldots S_k$ composée des symboles primitifs S_i sera

$$2^{n_1} \times 3^{n_2} \times 5^{n_3} \times \cdots \times p_k^{n_k},$$

où $n_1, n_2, \ldots, n_k$ sont les nombres associés aux symboles $S_1, S_2, \ldots, S_k$. Ainsi, par exemple, le nombre de $y_1(x_1) \vee \sim y_1(x_1)$ sera $2^{17^2} 3^{11} 5^{17} 7^{13} 11^7 13^5 17^{17^2} 19^{11} 23^{17} 29^{13}$.[8] Si les expressions de la suite $E_1, E_2, \ldots, E_k$ possèdent respectivement les nombres $N_1, N_2, \ldots, N_k$, le nombre associé à la suite sera $2^{N_1} \times 3^{N_2} \times 5^{N_3} \times \cdots \times p_k^{N_k}$. Le nombre associé à la suite de (deux) formules $y_1(x_1) \vee \sim y_1(x_1)$ et $y_1(0') \vee \sim y_1(0')$ sera

$$2^{2^{17^2} 3^{11} 5^{17} 7^{13} 11^7 13^5 17^{17^2} 19^{11} 23^{17} 29^{13}} 3^{2^{17^2} 3^{11} 5^1 7^3 11^{13} 13^7 17^5 19^{17^2} 23^{11} 29^1 31^3 37^{13}}.$$ [9]

Le nombre associé à une expression E selon cette procédure sera appelé le nombre de Gödel (le *ndg*) de *E*, ou pour abréger $\Gamma(E)$. Par le théorème fondamental de l'arithmétique, un nombre possède une et une seule décomposition en nombres premiers, ce qui nous assure que Γ est une fonction injective. Par ailleurs, ceci nous dit que le *ndg* d'une expression est suffisant pour déterminer la suite de symboles primitifs qui composent *E*.

2.3. Les fonctions récursives

Nous avons vu que le finitisme pose certaines limitations à ce que l'entendement peut effectuer comme raisonnements mathématiques. Les seules opérations que l'arithmétique contentuelle admet – et par conséquent la métamathématique aussi – sont des opérations constructives. La notion de récursivité est la formalisation que Gödel donne à l'exigence d'effectivité dans l'arithmétique contentuelle, et par conséquent les constructions métathéoriques doivent être récursives. Nous verrons plus loin que les relations métathéoriques sur *P* (être un

[8] Un nombre de 539 chiffres.
[9] Un nombre ayant environ 10 exposant 540 chiffres.

terme, une formule, une conséquence immédiate, etc.) sont traduisibles en des relations récursives par l'entremise de l'arithmétisation gödelienne, d'où l'importance de la récursivité dans ce qui suit.

On dira qu'une fonction arithmétique $\phi(x_1,...,x_n)$ (une fonction dont les variables $x_1,...,x_n$ sont des nombres naturels) est définie *récursivement* en fonction des fonctions arithmétiques $\psi(x_1,...,x_{n-1})$ et $\mu(x_1,...,x_n,x_{n+1})$ si :

$$\phi(0,x_2,...,x_n) = \psi(x_2,...,x_n)$$
$$\phi(k+1,x_2,...,x_n) = \mu(k,\phi(k,x_2,...,x_n),x_2,...,x_n).$$

Une fonction arithmétique est *récursive* s'il existe une suite finie de fonctions arithmétiques $\phi_1,...,\phi_n$ qui termine avec ϕ telle que chaque fonction ϕ_i de la suite est (1) définie récursivement en fonction de deux de ses prédécesseurs, ou (2) obtenue d'une fonction précédente par substitution, ou (3) la fonction successeur ou (4) une constante.[10] Une relation $R(x_1,x_2,...,x_n)$ sera dite récursive s'il existe une fonction récursive $\phi(x_1,...,x_n)$ telle que

$$R(x_1,x_2,...,x_n) \equiv \big(\phi(x_1,x_2,...,x_n) = 0\big).$$

$\phi(x_1,...,x_n)$ est une sorte de fonction caractéristique de la relation.[11]

Afin de faciliter la reconnaissance et la construction des fonctions et relations récursives, certains résultats nous seront utiles :[12]

---

[10] Il existe de nombreuses formalisations de la notion de récursivité. Je donne ici celle qui est présentée dans (Gödel 1931) et qui correspond à la récursivité primitive dans la terminologie moderne.

[11] En réalité, la fonction caractéristique devrait valoir 0 (resp. 1) si la relation tient et 1 (resp. 0) si elle ne tient pas. Ici, la relation tient si et seulement si la fonction vaut 0.

[12] Les preuves de ces théorèmes sont données dans (Gödel 1931). On peut trouver une exposition plus pédagogique dans (Kleene 1952 : pp. 217 et suivantes) ou n'importe quel manuel standard en logique mathématique.

I. Toute fonction (ou relation) obtenue de fonctions (ou relations) récursives par substitution des variables par des fonctions récursives est récursive. De même, toute fonction obtenue de fonctions récursives par une définition récursive est récursive.

II. Si R et S sont des relations récursives, alors $\sim R$ et $(R \vee S)$ sont des relations récursives.

III. Si les fonctions $\phi(X)$ et $\psi(Y)$ sont récursives alors la relation $\phi(X) = \psi(Y)$ est récursive, où X et Y sont des variables ou des suites de variables quelconques.[13]

IV. Si la fonction $\phi(X)$ et la relation $R(Y)$ sont récursives, alors les relations S, T et la fonction $\psi(X,Y)$ définies comme suit le sont également :

a) $S(X,Y) \equiv \exists x(x \leq \phi(X) \wedge R(x,Y))$,

b) $T(X,Y) \equiv \exists x(x \leq \phi(X) \supset R(x,Y))$,

c) $\psi(X,Y) \equiv \varepsilon x(x \leq \phi(X) \wedge R(x,Y))$,

où $\varepsilon x F(x)$ est le plus petit nombre x telle que $F(x)$ et 0 le cas échéant.[14]

Ces théorèmes nous permettent de prouver que les fonctions $x + y$, $x \cdot y$ et x^y de même que les relations $x < y$ et $x = y$ sont récursives, un fait que nous utiliserons abondamment.

Pour la suite de cette analyse, nous aurons besoin d'une notion qui lie le concept de récursivité à P : la représentabilité dans P. On dira qu'une relation arithmétique $R(x_1, x_2,..., x_n)$ est *représentable* dans P s'il existe une relation $\mathcal{R}(u_1, u_2,.., u_n)$ de P telle que pour tout n-tuplet de nombres naturels $k_1, k_2,..., k_n$

[13] Les variables X et Y sont des variables ou des suites (finies) de variables quelconques.

[14] L'opération ε est seulement (primitivement) récursive lorsque la variable x est bornée.

(i) Si $R(k_1,k_2,...,k_n)$ alors $\vdash_P \mathcal{R}(\overline{k_1},\overline{k_2},...,\overline{k_n})$

(ii) Si $\sim R(k_1,k_2,...,k_n)$ alors $\vdash_P \sim \mathcal{R}(\overline{k_1},\overline{k_2},...,\overline{k_n})$

où $\overline{k_1},\overline{k_2},...,\overline{k_n}$ sont les symboles de P dénotant les nombres $k_1,k_2,...,k_n$ respectivement. Un mot sur la notation. Pour faciliter le saut entre une relation récursive et sa représentation, convenons de la chose suivante : 1) si n est un nombre alors $\overline{n}$ sera le symbole qui le représente; 2) pour distinguer les variables des relations récursives des variables de P, convenons simplement de changer leur nom (par exemple, changer $x_1,x_2,...,x_n$ pour $u_1,u_2,..,u_n$); et 3) si $R(x_1,x_2,...,x_n)$ est une relation récursive, $\overparen{R(x_1,x_2,...,x_n)}(u_1,u_2,...,u_n)$ sera la relation de P qui la représente (si elle existe).

Le théorème suivant que je cite sans preuve nous assure qu'une relation récursive est toujours représentable dans P.

THÉORÈME A. Toute relation récursive est représentable dans P.

La preuve de ce théorème utilise la célèbre fonction $\beta(x,y,i)$, une fonction arithmétique de P qui a la particularité suivante : pour toute suite finie $a_1,a_2,...,a_n$ de nombres naturels (n est variable aussi), il existe des entiers c et d tels que $\beta(c,d,i)=a_i$ pour i variant de 1 à n. [15]

[15] Pour une preuve complète du théorème, voir (Cori & Lascar 1992 : tome II, pp. 76 et suivantes), (Kleene 1952 : pp. 217 et suivantes) ou encore (Mendelson 1987 : pp. 116-145).

2.4. L'arithmétisation de la métathéorie

Nous allons maintenant utiliser Γ pour construire récursivement quarante-cinq fonctions et relations qui terminent par l'arithmétisation de la relation métathéorique « *a* est (le *ndg* d') une démonstration de (la formule dont le *ndg* est) *b* » (45) avec laquelle nous pourrons définir la relation (non récursive) « *a* est (le *ndg* d'une formule) démontrable » (46). L'exercice est fastidieux et procède ainsi :

1. x divise y, $x \mid y$: $x \mid y \equiv \exists z(z \leq y \wedge y = z \cdot x)$

2. x est un nombre premier, $\Pr(x)$: $\Pr(x) \equiv \forall y((y \leq x \wedge y \mid x) \supset (y = 1 \vee y = x))$

3. $\Pr(n, x)$ est le n-ième premier dans la suite des premiers qui divise x :

$$\Pr(0, x) = 0 \text{ et } \Pr(n+1, x) = \varepsilon y(y \leq x \wedge \Pr(y) \wedge y \mid x \wedge y > \Pr(n, x))$$

4. La fonction factorielle : $0! = 1$ et $(n+1)! = (n+1) \cdot n!$

5. $P(n)$ est le n-ième premier dans la suite des nombres premiers :

$$P(n) = 0 \text{ et } P(n+1) = \varepsilon y(y \leq x \wedge P(n)! + 1 \wedge \Pr(y) \wedge y > P(n))$$

6. $Terme(n, x)$ est le n-ième terme dans la suite numérique x (pourvu que n soit plus grand que 0 et soit au plus égal à la longueur de x) :

$$Terme(n, x) = \varepsilon y(y \leq x \wedge \Pr(n, x)^{y} \mid x \wedge \sim \Pr(n, x)^{y+1} \mid x)$$

7. $l(x)$ est la longueur de la suite x :

$$l(x) = \varepsilon y(y \leq x \wedge \Pr(y, x) > 0 \mid x \wedge \Pr(y+1, x) = 0)$$

8. $x * y$ est la concaténation des deux suites (de symboles) x et y :

$$x * y = \varepsilon z\big(z \leq (P(l(x) + l(y)))^{x+y} \wedge \forall n(n \leq l(x) \supset Terme(n, x) = Terme(n, z)) \wedge \forall n(0 < n \leq l(y) \supset Terme(n, y) = Terme(n + l(y), z))\big)$$

9. $R(x)$ est une suite du nombre x seulement : $R(x) = 2^{x}$

10. $E(x)$ est l'opération qui place x entre parenthèses : $E(x) = R(11) * x * R(13)$

11. x est une variable de type n, $Var(n, x)$:

$$Var(n,x) \equiv \exists z\left(13 < z \leq x \wedge \Pr(z) \wedge x = z^n\right)$$

12. x est une variable, $Var(x)$:

$$Var(x) \equiv \exists n(n \leq x \wedge Var(n,x))$$

13. $Nég(x)$ est la négation de x : $Nég(x) = R(5) * x$

14. $Dis(x,y)$ est la disjonction de x et y : $Dis(x,y) = E(x * R(7) * y)$

15. $Gén(x,y)$ est la généralisation de y par x : $Gén(x,y) = R(9) * R(x) * y$

16. $N(n,x)$ est l'opération qui place n symboles « ' » devant x :

$$N(0,x) = x \text{ et } N(n+1,x) = R(3) * N(n,x)$$

17. $N(x)$ est le nombre qui désigne x : $N(x) = N(x,1)$

18. x est un signe de type un, $type1(x)$:

$$type1(x) \equiv \exists m \exists n(m,n \leq x \wedge (m = 1 \vee Var(1,m)) \wedge x = N(n, R(m)))$$

19. x est un signe de type n, $type_n(x)$

$$type_n(x) \equiv ((n = 1 \wedge type1(x)) \vee (n > 1 \wedge \exists y(y \leq x \wedge Var(n,x) \wedge x = R(y))))$$

20. x est une formule élémentaire, $FÉ(x)$:

$$FÉ(x) \equiv \exists y \exists z \exists n(y,z,n \leq x \wedge type_n(y) \wedge type_{n+1}(z) \wedge x = z * E(y))$$

21. x est obtenue des formules y et z par disjonction ou de la formule y par négation ou généralisation, $Obt(x,y,z)$:

$$Obt(x,y,z) \equiv (x = Dis(y,z) \vee x = Nég(y) \vee \exists v(v \leq x \wedge Var(v) \wedge x = Gén(v,y)))$$

22. x est une suite de formules, chacune obtenue d'une ou de deux formules précédentes dans la suite, $SForm(x)$:

$$\begin{aligned} SForm(x) \equiv l(x) > 0 \wedge \forall n(0 < n \leq l(x) \supset FÉ(Terme(n,x)) \vee \\ \exists p \exists q(0 < p,q < n \wedge Obt(Terme(n,x), Terme(p,y), Terme(q,z))) \end{aligned}$$

23. x est une formule, $Form(x)$:

$$Form(x) \equiv \exists n\left(n \leq \left(P(l(x)^2)\right)^{x \cdot l(x)^2} \wedge SForm(n) \wedge x = Terme(l(n),n)\right)$$

24. La variable v est liée à la n-ième place dans x, $Liée(v,n,x)$:

$$Liée(v,n,x) \equiv \big(Var(v) \wedge Form(x) \wedge \exists a \exists b \exists c \big(a,b,c \le x \wedge x = a * Gén(v,b) * c \wedge Form(b) \wedge l(a)+1 \le n \le l(a) + l(Gén(v,b))\big)\big)$$

25. La variable v est libre à la n-ième place dans x, $Libre(v,n,x)$:

$$Libre(v,n,x) \equiv \big(Var(v) \wedge Form(x) \wedge v = Terme(n,x) \wedge n \le l(x) \wedge \sim Liée(v,n,x)\big)$$

26. La variable v est libre dans x, $Libre(v,x)$:

$$Libre(v,x) \equiv \exists n \big(n \le l(x) \wedge Libre(v,n,x)\big)$$

27. $Su(x,n,y)$ est le résultat de la substitution du n-ième terme de x par y :

$$Su(x,n,y) = \varepsilon z\big(z \le \big(P(l(x)+l(y))\big)^{x+y} \wedge \exists u \exists v \big(u,v \le x \wedge x = u * R(Terme(n,x)) * v \wedge z = u * y * v \wedge n = l(u)+1\big)\big)$$

28. $Lib(k,v,x)$ est la $k+1$-ième place où v est libre dans x :

$$Lib(0,v,x) = \varepsilon n\big(n \le l(x) \wedge Libre(v,n,x) \wedge \forall p\big(n < p \le l(x) \supset \sim Libre(v,p,x)\big)\big)$$

$$Lib(k+1,v,x) = \varepsilon n\big(n \le Lib(k,v,x) \wedge Libre(v,n,x) \wedge \forall p\big(n < p < Lib(k,v,x) \supset \sim Libre(v,p,x)\big)\big)$$

29. $NLib(v,x)$ est le nombre de places où v est libre dans x :

$$NLib(v,x) = \varepsilon n\big(n \le l(x) \wedge Lib(n,v,x) = 0\big)$$

30. $Sub_k(x,v,y)$ est le résultat de la substitution de v par y à la $1^{\text{ère}}$, $2^{\text{ième}}$,..., k-ième places dans x :

$$Sub_0(x,v,y) = x$$

$$Sub_{k+1}(x,v,y) = Su\big(Sub_k(x,v,y), Lib(k,v,x), y\big)$$

31. $Sub(x,v,y)$ est le résultat de la substitution de v par y dans x :

$$Sub(x,v,y) = Sub_{NLib(v,x)}(x,v,y)$$

32. $imp(x,y)$ est l'implication de y par x : $imp(x,y) = Dis(Nég(x),y)$

$con(x, y)$ est la conjonction de x et y : $con(x, y) = Nég(Dis(Nég(x), Nég(y)))$

$éq(x, y)$ est l'équivalence entre x et y : $con(imp(x, y), imp(y, x))$

$exist(v, x)$ est la quantification existentielle de x par v : $Nég(Gén(v, Nég(x)))$

33. $ÉType(n, x)$ est l'élévation de x de n types :

$$ÉType(n,x) = \varepsilon y\left(y \leq x^{(x^n)} \wedge \forall k\left(k \leq l(x) \supset (Terme(k,x) \leq 13 \wedge Terme(k,x) = Terme(k,y)) \vee (Terme(k,x) > 13 \wedge Terme(k,y) = Terme\left(k, x \cdot \Pr(1, Terme(k,x))^n\right)\right)\right)$$

34. x est un des axiomes du groupe I, $AxI(x)$: $AxI(x) \equiv (x = n_1 \vee x = n_2 \vee x = n_3)$, où n_1, n_2 et n_3 sont les *ndg* des axiomes I.1, I.2 et I.3 respectivement.

35. x est une formule qui résulte du schème d'axiome II.1 par substitution, $AxII1(x)$:

$$AxII1(x) \equiv \exists y(y \leq x \wedge Form(y) \wedge x = imp(Dis(y, y), y))$$

Des prédicats similaires peuvent être donnés pour les axiomes II.2, II.3 et II.4 qui porteront les noms $AxII2(x)$, $AxII3(x)$ et $AxII4(x)$ respectivement.

36. x résultat de l'application d'un axiome du deuxième groupe, $AxII(x)$:

$$AxII(x) \equiv (AxII1(x) \vee AxII2(x) \vee AxII3(x) \vee AxII4(x))$$

37. z ne contient pas de variable liée dans y à l'endroit où v est libre, $Q(z, y, v)$:

$$Q(z,y,v) \equiv \sim \exists n \exists m \exists w(n \leq l(y) \wedge m \leq l(z) \wedge w \leq z \wedge w = Terme(m,z) \wedge Liée(w,n,y) \wedge Libre(v,n,y))$$

38. x est une formule qui résulte de l'axiome III.1 par substitution, $AxIII1(x)$:

$$AxIII1(x) \equiv \sim \exists v \exists y \exists z \exists n(v, y, z, n \leq x \wedge Var(n,v) \wedge type_n(z) \wedge Form(y) \wedge Q(z,y,v) \wedge x = imp(Gén(v,y), Sub(y,v,z)))$$

39. x est une formule qui résulte de l'axiome III.2 par substitution, $AxIII2(x)$:

$$AxIII2(x) \equiv \exists v \exists q \exists p(v, q, p \leq x \wedge Var(v) \wedge Form(p) \wedge \sim Libre(v,p) \wedge$$

$$Form(q) \wedge x = imp\big(Gén(v, Dis(p,q)), Dis(p, Gén(v,q))\big)\big)$$

40. x est une formule qui résulte de l'axiome IV.1 par substitution, $AxIV1(x)$:

$$AxIV1(x) \equiv \exists u \exists v \exists y \exists n\big(u, v, y, n \leq x \wedge Var(n,v) \wedge Var(n+1,u) \wedge \sim Libre(u,y)$$

$$\wedge Form(y) \wedge x = existe\big(u, Gén\big(v, éq(R(u) * E(R(v)), y)\big)\big)\big)$$

41. x est résulte de l'axiome V.1, $AxV1(x)$: $AxV1(x) \equiv \exists n\big(n \leq x \wedge x = ÉType(n, n_4)\big)$, où n_4 est le *ndg* l'axiome V.1.

42. x est un axiome, $Ax(x)$:

$$Ax(x) \equiv \big(AxI(x) \vee AxII(x) \vee AxIII1(x) \vee AxIII2(x) \vee AxIV1(x) \vee AxV1(x)\big)$$

43. x est une conséquence immédiate de y et z, $Cons(x,y,z)$:

$$Cons(x,y,z) \equiv \big(y = imp(z,x) \vee \exists v\big(v \leq x \wedge Var(v) \wedge x = Gén(v,y)\big)\big)$$

44. x est une démonstration, $Démon(x)$:

$$Démon(x) \equiv \big(l(x) > 0 \wedge \forall n\big(0 < n \leq l(x) \supset \big(Ax(Terme(n,x)) \vee$$

$$\exists p \exists q\big(0 < p, q < n \wedge Cons(Terme(n,x), Terme(p,x), Terme(q,x))\big)\big)\big)\big)$$

45. x est une démonstration de y, $Dém(x,y)$:

$$Dém(x,y) \equiv \big(Démon(x) \wedge y = Terme(l(x),x)\big)$$

46. x est démontrable, $D(x)$: $\exists y Dém(y,x)$

Quelques remarques s'imposent sur ces définitions : (a) ces fonctions et relations n'appartiennent pas à proprement parler à P (parce qu'elles sont définies récursivement); (b) les définitions 1, 2, 11, 12, 18-26 et 34-46 sont des relations, et les autres (3-10, 13-17 et 27-33) sont des fonctions; (c) les fonctions ou relations 1-45 sont récursives, seulement 46 ne l'est pas; (d) la justification des bornes particulières qui sont données aux variables quantifiées (5, 8, 23, 27 et 33) sont omises ici mais sont faciles à démontrer; (e) à partir de la définition 6, les fonctions ou les relations sont toutes des traductions d'opérations ou de relations métathéoriques; (f) les traductions d'opérations ou de relations métathéoriques sont des fonctions ou des relations entre nombres de Gödel, c'est-à-dire que la relation $Dém(x,y)$ signifie en réalité : « x est le nombre de Gödel d'une démonstration qui a pour dernière formule la formule dont le nombre de Gödel est y », de même pour $Cons(x,y,z)$: « x est le *ndg* d'une formule qui est la conséquence immédiate de deux formules de *ndg* y et z respectivement ».

2.5. Les théorèmes de Gödel

Pour la démonstration des théorèmes de Gödel, il nous faudra un dernier théorème qui découle de ce qui précède et du théorème A plus haut :

THÉORÈME B.[16] Si $R(x_1, x_2, ..., x_n)$ est une relation récursive sur les nombres naturels alors il existe une relation dans P dont le nombre de Gödel est r (et qui possède les variables libres $u_1, u_2, ..., u_n$) telle que, pour tout n-tuplet $k_1, k_2, ..., k_n$ de nombres naturels,

(i) $R(k_1, k_2, ..., k_n) \supset D(Sub(r, u_i, N(k_i)))$

(ii) $\sim R(k_1, k_2, ..., k_n) \supset D(Nég(Sub(r, u_i, N(k_i))))$

où $Sub(r, u_i, N(k_i))$ est le résultat de la substitution des variables $u_1, u_2, ..., u_n$ par les nombres $N(k_1), N(k_2), ..., N(k_n)$ respectivement.

Le théorème B est une conséquence directe de l'application du théorème A et de la traduction par Γ, comme l'illustre le schéma suivant :

[16] C'est une partie du théorème V de Gödel (van Heijenoort 1967 : p. 606-7).

$$\begin{tikzcd} & \textbf{(2)}\ \vdash_P \overline{R(x_1,\dots,x_n)}(\overline{k}_1,\dots,\overline{k}_n) \arrow[dr, "\Gamma"] & \\ \textbf{(1)}\ R(k_1,\dots,k_n) \arrow[ur, "\text{Th. A}"] & & \textbf{(3)}\ D\big(Sub(r,u_i,N(k_i))\big) \end{tikzcd}$$

(1) $R(x_1,\dots,x_n)$ est une relation récursive (vraie de $k_1,k_2,\dots,k_n$);

(2) $\overline{R(x_1,\dots,x_n)}(u_1,\dots,u_n)$ est sa représentation dans P (selon le théorème A);

(3) $D\big(Sub(r,u_i,N(k_i))\big)$ est l'arithmétisation par Γ de l'énoncé métathéorique $\vdash_P \overline{R(x_1,\dots,x_n)}(\overline{k}_1,\dots,\overline{k}_n)$ et où $r=\Gamma\left(\overline{R(x_1,\dots,x_n)}(u_1,\dots,u_n)\right)$.

Derniers détails avant la pièce maîtresse. Les théorèmes de Gödel sont formulés non seulement pour P mais pour un ensemble de systèmes obtenus de P en ajoutant aux axiomes de celui-ci un ensemble κ de formules (des hypothèses ou des postulats spécifiques à une théorie). Pour peu que l'ensemble soit définissable récursivement, les définitions 45 à 46 peuvent être adaptées pour prendre en considération le fait qu'une démonstration utilise une formule de κ comme axiome. Dans ce cas, un indice sera ajouté aux relations de 45 et 46 pour éviter la confusion.

Par ailleurs, les théorèmes de Gödel s'appuieront sur une autre hypothèse, celle à l'effet que $P\cup\kappa$ est ω-cohérente : $P\cup\kappa$ est ω-cohérente si et seulement si aucune formule $\mathcal{F}(x)$ de $P\cup\kappa$ n'est telle que

$$\vdash_P \mathcal{F}(0)\,,\ \vdash_P \mathcal{F}(0')\,,\ \vdash_P \mathcal{F}(0'')\,,\ \vdash_P \mathcal{F}(0''')\ \dots \text{ et } \vdash_P \sim\forall x\mathcal{F}(x)\,.$$

À l'aide des définitions 1 à 46, cette condition peut s'énoncer comme : il n'y a pas de a telle que

$$\forall n D_{\kappa}\big(Sub(a,x,N(n))\big) \wedge D_{\kappa}\big(Nég(Gén(x,a))\big).$$

Il est facile de démontrer qu'un système ω-cohérent est automatiquement cohérent.

Le premier théorème d'incomplétude :

THÉORÈME G1. Sous l'hypothèse que $P \cup \kappa$ est ω-cohérent, il existe une relation récursive $\mathcal{R}(x_1)$ dans $P \cup \kappa$ d'une seule variable libre x_1 et de nombre de Gödel r telle que $\sim D_{\kappa}(Gén(17,r))$ et $\sim D_{\kappa}\big(Nég(Gén(17,r))\big)$, c'est-à-dire $Gén(17,r)$ est indécidable.

PREUVE. La preuve du théorème procède de la relation $Q(x,y)$ définie par $\sim Dém_{\kappa}\big(x, Sub(y,19,N(y))\big)$.[17] Par le théorème B, il existe une relation de P dont le *ndg* est q avec les variables libres x_1 et x_2 (de *ndg* 17 et 19 resp.) telle que

(*) $\quad \sim Dém_{\kappa}\big(x, Sub(y,19,N(y))\big) \supset D_{\kappa}\big(Sub(q;17,N(x);19,N(y))\big),$

(**) $\quad Dém_{\kappa}\big(x, Sub(y,19,N(y))\big) \supset D_{\kappa}\big(Nég\big(Sub(q;17,N(x);19,N(y))\big)\big).$

($Sub(q;17,N(x);19,N(y))$ est le résultat de la substitution des variables 17 et 19 par les termes $N(x)$ et $N(y)$.) Posons

$$p = Gén(17,q) \text{ et } r = Sub(q,19,N(p)).$$

(Intuitivement, p est le *ndg* de la généralisation de (la formule dont le *ndg* est) q par x_1, c'est-à-dire le *ndg* de la formule signifiant qu'il n'y a aucune démonstration de la

[17] On remarquera que x_1 est une variable de P tandis que x est une variable du métalangage de P.

formule dont le *ndg* est x_2. Le nombre r est le *ndg* de la formule signifiant que x_1 n'est pas (le *ndg* d') une démonstration de (la formule dont le *ndg* est) p.) On remarque ensuite que

$$(\#) \qquad Sub(p,19,N(p))) = Sub(Gén(17,q),19,N(p)) \\ = Gén(17,Sub(q,19,N(p))) = Gén(17,r).$$

(Intuitivement, $Sub(p,19,N(p))$ ou $Gén(17,r)$ signifient que (la formule dont le *ndg* est) p est indémontrable.) Par ailleurs, d'après la définition de r

$$(\#\#) \qquad Sub(q;17,N(x);19,N(p)) = Sub(r,17,N(x)).$$

En substituant y par p dans (*) et (**) et en utilisant (#) et (##), nous obtenons

$$(|) \qquad \sim Dém_\kappa(x,Gén(17,r)) \supset D_\kappa(Sub(r,17,N(x)))$$
$$(||) \qquad Dém_\kappa(x,Gén(17,r)) \supset D_\kappa(Nég(Sub(r,17,N(x)))).$$

Pour démontrer rigoureusement que $Gén(17,r)$ est (le *ndg* d'une formule) indécidable, il faut montrer (i) $\sim D_\kappa(Gén(17,r))$ et (ii) $\sim D_\kappa(Nég(Gén(17,r)))$:

(i) Supposons le contraire, $D_\kappa(Gén(17,r))$. Il existe donc un nombre naturel n tel que $Dém_\kappa(n,Gén(17,r))$. Par (||), nous avons $D_\kappa(Nég(Sub(r,17,N(n))))$. Mais de la κ-démontrabilité de $Gén(17,r)$ découle celle de $Sub(r,17,N(n))$ (le dernier est une instance du premier), c'est-à-dire que nous avons $D_\kappa(Sub(r,17,N(n)))$, contredisant la (ω-) cohérence de $P \cup \kappa$.[18]

[18] Seule la cohérence de $P \cup \kappa$ est utilisée pour cette partie.

(ii) De la non démontrabilité de $Gén(17,r)$ en (i) découle $\sim \forall n Dém_{\kappa}(n, Gén(17,r))$. Par (|), nous en déduisons donc que $\forall n D_{\kappa}(Sub(r,17,N(n)))$. Si $Nég(Gén(17,r))$ était démontrable, nous aurions que $D_{\kappa}(Nég(Gén(17,r)))$, en contradiction avec l'ω-cohérence de $P \cup \kappa$.

Dans G1, l'indécidabilité de $Gén(17,r)$ dépend de l'ω-cohérence de $P \cup \kappa$. Or, Rosser a réussi à réduire cette hypothèse à la cohérence de $P \cup \kappa$ seulement :

THÉORÈME G-R. Sous l'hypothèse que $P \cup \kappa$ est cohérent, il existe une relation récursive $\mathcal{R}(x_1)$ de $P \cup \kappa$ d'une variable libre x_1 et de nombre de Gödel r telle que $\sim D_{\kappa}(Gén(17,r))$ et $\sim D_{\kappa}(Nég(Gén(17,r)))$, c'est-à-dire $Gén(17,r)$ est indécidable.

La preuve utilise essentiellement les mêmes arguments en remplaçant la formule $Q(x,y)$ pour la formule $R(x,y)$ définie comme suit :

$$R(x,y) \equiv Dém_{\kappa}(x, Sub(y,19,N(y))) \supset \exists z(z \leq x \wedge Dém_{\kappa}(z, Nég(Sub(y,19,N(y))))).$$

C'est-à-dire que $R(x,y)$ est équivalente à : si x est le *ndg* d'une démonstration de la formule dont le *ndg* est $Sub(y,19,N(y))$, alors il existe z tel que z est le *ndg* d'une démonstration de la formule dont le *ndg* est $Nég(Sub(y,19,N(y)))$ et z est un nombre plus petit que x.

Le deuxième théorème d'incomplétude de Gödel est celui qui a scellé le sort du programme de Hilbert :

THÉORÈME G2. Sous l'hypothèse que $P \cup \kappa$ est cohérent, il n'existe pas de preuve de la cohérence de $P \cup \kappa$ dans $P \cup \kappa$ (ou formalisable dans $P \cup \kappa$).

PREUVE. La preuve de ce théorème repose sur les observations suivantes : 1) dans la preuve de G1, l'argument qui démontre l'indémontrabilité de $Gén(17,r)$ ne dépend que de la cohérence de $P \cup \kappa$; et 2) cet argument est (ou plutôt serait[19]) entièrement formalisable dans $P \cup \kappa$. L'énoncé métathéorique à l'effet que $P \cup \kappa$ est cohérent est formalisable dans $P \cup \kappa$ par

$$Coh(\kappa) \equiv \forall y \sim \left(\exists x Dém_{\kappa}(x,y) \wedge \exists x Dém_{\kappa}(x, Nég(y))\right).$$

Ainsi, selon (2), $Coh(\kappa) \supset \forall x {\sim}\, Dém_{\kappa}(x, Gén(17,r))$. Par (#), nous avons que $Gén(17,r) = Sub(p,19,N(p))$ donc

$$Coh(\kappa) \supset \forall x {\sim}\, Dém_{\kappa}(x, Sub(p,19,N(p))),$$

c'est-à-dire que $Coh(\kappa) \supset \forall x Q(x,p)$ d'après la définition de la formule $Q(x,y)$. Or, puisque les méthodes de cette démonstration sont elles aussi formalisables, en posant c comme étant le *ndg* de $Coh(\kappa)$ et en remarquant que $Gén(17,r)$ est le *ndg* de $\forall x Q(x,p)$,[20] nous avons que $imp(c, Gén(17,r))$ est κ-démontrable, c'est-à-dire que $D_{\kappa}(imp(c, Gén(17,r)))$. Si $Coh(\kappa)$ était démontrable, nous aurions $D_{\kappa}(c)$ et par *modus ponens* $D_{\kappa}(Gén(17,r))$, ce qui est en contradiction avec le théorème G1. Nous avons donc $\sim D_{\kappa}(c)$. Mais $\sim D_{\kappa}(c)$ affirme que la formule dont le nombre de Gödel est c est non démontrable, à savoir $Coh(\kappa)$, que la cohérence de $P \cup \{\kappa\}$ est indémontrable dans $P \cup \{\kappa\}$.

[19] Nous verrons pourquoi cette affirmation n'est pas juste.
[20] Nous verrons que cette affirmation n'est pas tout à fait juste.

3. L'incomplétude de l'incomplétude

J'aimerais explorer ici une critique de la démonstration de ces théorèmes qui met en lumière une confusion systématique chez entre l'usage et la mention et qui, par ailleurs, compromet sérieusement la validité du second théorème d'incomplétude. Que le lecteur soit rassuré d'emblée, la démonstration du deuxième théorème par le théorème de Löb (que nous verrons plus loin) n'est pas affectée par cette critique, donc il reste valide. Toutefois, telle qu'elle est présentée dans l'article original de 1931, il est difficile de voir comment la démonstration pourrait être modifiée pour demeurer valide. On pourrait imaginer que Gödel aurait réalisé ces difficultés eut-il développé tous les détails de sa preuve, un travail qui devait faire l'objet d'un second article[21] (d'où le « *I* » dans *Sur les propositions formellement indécidables des Principia Mathematica et systèmes apparentés – I*), mais curieusement cet article n'a jamais vu le jour.

La principale erreur ou confusion que Gödel a commise est de ne pas distinguer entre l'usage et la mention d'un symbole du langage. La théorie $P \cup \{\kappa\}$ est la théorie que nous étudions, c'est-à-dire le langage objet, et c'est pour cette théorie que sont démontrés les théorèmes d'incomplétude. Pour étudier les fonctions et les relations arithmétiques de $P \cup \{\kappa\}$, nous faisons appel à un métalangage qui contient notamment la notion de récursivité. La récursivité n'est donc pas un concept de $P \cup \{\kappa\}$ – on ne trouvera aucune description formelle de la récursivité dans la description de $P \cup \{\kappa\}$ –, c'est un type de fonction ou de relation qui appartient au métalangage. Relisons maintenant les hypothèses du premier théorème d'incomplétude : il existe une relation récursive de nombre de Gödel r telle que $\sim D_\kappa(Gén(17,r))$ et $\sim D_\kappa\left(Nég(Gén(17,r))\right)$. Puisqu'elle possède un nombre de Gödel, la relation qui a pour nombre de Gödel r est forcément une relation de

[21] van Heijenoort, 1967 : p. 596.

$P \cup \{\kappa\}$; il est donc étonnant de voir l'expression « relation récursive » – un qualificatif pour les fonctions et relations du métalangage – lui être attribuée. C'est précisément ce type de confusion qui est présente dans l'article de Gödel. Dans ce cas, il est clair qu'il suffit de supprimer l'expression « relation récursive » et la confusion, au niveau de l'énonciation, disparaît. Toutefois, nous verrons comment une confusion de cet ordre apparaît dans la démonstration du second théorème, et cette fois-ci elle n'est pas éliminable. Pour vendre la mèche, la confusion survient au moment où Gödel prétend que la partie (i) de la preuve du premier théorème est formalisable dans $P \cup \{\kappa\}$, c'est-à-dire que $Coh(\kappa) \supset \forall x \sim Dém_{\kappa}(x, Gén(17, r))$.

La mise en évidence des difficultés signalées dans le paragraphe précédent sera grandement facilitée par l'introduction d'un métalangage plus formel que ne l'est celui de Gödel. L'objet de la prochaine section sera précisément de mettre sur pied un tel métalangage pour $P \cup \{\kappa\}$.

3.1. Le métalangage de Gödel

Notre tâche est maintenant de construire un métalangage Π pour P (ou $P \cup \kappa$).[22] Pour définir ce métalangage, il faut le formaliser dans un méta-métalangage informel (celui de la présente discussion). Suivant Tarski,[23] le métalangage doit contenir au moins le langage P lui-même, sinon un langage isomorphe à celui-ci. Ce langagc a été formalisé plus haut à la section 2.1, inutile de répéter l'exercice ici. Le langage P ne suffit pas à la tâche à lui seul, car il n'inclut pas les notions de récursivité.[24] Il faut donc ajouter les différentes notions de récursivité telles qu'elles apparaissent à la section 2.3. Typiquement, la formalisation

[22] Nous supposerons, sans perte de généralité, que les P peut contenir κ ou non.
[23] Nous verrons en détail la présentation tarskienne au prochain chapitre.
[24] Il est important de remarquer, si ce n'est pas déjà fait, que les fonctions et relations récursives définies plus hauts n'appartiennent pas à P mais au métalangage de P.

de ces notions se fait en ajoutant deux symboles primitifs « λ » et « $\Re$ » dont le comportement est dicté par les axiomes suivants :

(R1) Si $t(x_1,...,x_n)$ est un terme de type 1 (un terme qui ne fait intervenir que des nombres et des variables de type 1 ainsi que des fonctions arithmétiques), alors

a) $\lambda[t(x_1,...,x_n)]$ est une fonction arithmétique des variables $x_1,...,x_n$;

b) $\lambda[t(x_1,...,x_n)](t_1,...,t_n) = t(t_1,...,t_n)$, où $t_1,...,t_n$ sont de type 1.

(R2) Si f et g sont des fonctions arithmétiques de n et de $n+2$ variables respectivement, alors

a) $\Re[f,g]$ est une fonction arithmétique de $n+1$ variables;

b) $\Re[f,g](x_1,...,x_n,0) = f(x_1,...,x_n)$ et

$\Re[f,g](x_1,...,x_n,y+1) = g(x_1,...,x_n,y,\Re[f,g](x_1,...,x_n,y))$.[25]

Les autres définitions de la section 2.3 peuvent maintenant se traduire aisément dans les termes de « λ » et « $\Re$ ».

Suivant Tarski toujours, il nous faut une théorie dans Π qui fait figure de description de P. Cette théorie sera essentiellement une formalisation de la description donnée à P à la section 2.1 (on remarquera bien la nuance entre formaliser P et formaliser une description de P), elle devra pouvoir reproduire les définitions métathéoriques de la section 2.1 (à savoir être un terme, un nombre, une formule élémentaire, une formule, un axiome, etc.). Il faut d'abord ajouter des symboles qui feront figure de noms pour les symboles primitifs de P : $\bar{\sim}$ pour « ~ », $\bar{\vee}$ pour « ∨ », $\bar{0}$ pour « 0 », $\bar{'}$ pour « ' », $\bar{\forall}$ pour « ∀ », $\bar{(}$ pour « (» et $\bar{)}$ pour «) » de même que $\bar{x}$ pour toute variable « x » de P. Ensuite, nous ajoutons aux symboles primitifs du métalangage un symbole de fonction « ^ » qui représentera essentiellement la

[25] Le lecteur aura compris qu'on a ajouté les symboles « [» et «] » également.

concaténation. La seule chose que nous devons spécifier est la manière dont « ^ » et les nouvelles constantes se comportent avec l'égalité :[26]

(D1) Les constantes sont toutes distinctes;

(D2) $x\,\hat{}\,y = s\,\hat{}\,t \equiv \big((x = s \wedge y = t) \vee \exists z(x = s\,\hat{}\,z \wedge t = z\,\hat{}\,y) \vee \exists w(s = x\,\hat{}\,w \wedge y = w\,\hat{}\,t)\big)$.[27]

De cette manière, nous pouvons construire des descriptions structurales de toutes les notions de *P*, comme le fit Tarski d'ailleurs.

Il reste maintenant la fonction Γ. Cette fonction est définissable aisément dans Π. Notons, au passage, que Γ est une fonction un peu particulière dans Π parce qu'elle met en relation des éléments des termes descriptifs avec des termes numériques.

Reprenons par étapes les constructions qui mènent à Π :

- L'ensemble des symboles primitifs de Π est constitué des symboles primitifs de *P*, des noms pour les symboles primitifs de *P*, les symboles « λ » et « $\Re$ » pour la récursion et le symbole « ^ » pour la concaténation.
- Les constantes « primitives » de type 1 de Π sont 0, $\overline{\sim}$, $\overline{\vee}$, $\overline{0}$, $\overline{'}$, $\overline{\forall}$, $\overline{(}$ et $\overline{)}$ de même que $\overline{x}$ pour toute variable « x » de *P*. (Nous pourrions considérer que $\lambda[t]$ et $\Re[f,g]$ sont des constantes de type 2 ou nous pourrions tout simplement les ajouter de manière auxiliaire.)[28]
- Les constantes numériques (qui sont de type 1) sont les constantes qui ne font intervenir que « 0 » et « ' »; les constantes descriptives (qui sont également de type 1) sont celles qui ne font intervenir que les constantes primitives $\overline{\sim}$,

[26] N'oublions que l'égalité dans *P* n'est pas un symbole primitif.

[27] C'est inspiré directement de Tarski 1956 : p. 173.

[28] Cette préoccupation avec les types n'est pas essentielle à la construction d'un métalangage. Si elle l'est dans le présent contexte, c'est que notre langage objet *P* est typé et nous avons choisi d'exprimer Π comme une certaine extension de *P*, donc comme un langage typé.

$\overline{\vee}$, $\overline{0}$, $\overline{{}^{\top}}$, $\overline{\forall}$, $\overline{(}$, $\overline{)}$ et $\overline{x}$ avec « ^ » (par le procédé suivant bien entendu : si X et Y sont des constantes descriptives, alors $X \wedge Y$ en est une également).

- Nous pouvons définir ensuite un terme de type 1, et faire la différence entre un terme numérique de type 1 et un terme descriptif de type 1.
- La définition d'un terme de type n ($n > 1$) est très similaire à celle de P (en prenant soin de décider si nous voulons que $\lambda[t]$ et $\Re[f, g]$ soient des constantes de type 2).
- Les définitions d'une formule, d'une variable libre, d'une conséquence immédiate, etc. sont essentiellement les mêmes.
- Pour les axiomes, nous reprenons les axiomes de P et nous ajoutons les axiomes nécessaires à la récursivité (dont (R1) et (R2)) ainsi que les axiomes nécessaires à au comportement de « ^ » avec « = » ((D1) et (D2)).
- Nous pouvons ensuite reconstruire aisément les définitions de P dans le métalangage en utilisant les termes descriptifs de type 1. Par exemple, nous commençons par définir des classes élémentaires comme : la classe des noms de nombres : $\bigcap\left\{y \mid \overline{0} \in y \wedge \forall x(x \in y \supset x \wedge \overline{{}^{\top}} \in y)\right\}$; la classe des noms de variables; la classe des noms de termes numériques; etc. Éventuellement, nous arrivons à la classe des formules et celle des démonstrations.
- Finalement, la fonction Γ est définissable à l'aide du calcul des classes en établissant une relation entre les termes descriptifs de type 1 et les termes numériques de type 1.

Quelques remarques brèves s'imposent sur la présente construction : (a) Π n'est pas P, c'est une extension assez complexe de P; (b) le fait que P soit une extension du système des *Principia* n'est pas essentiel, nous aurions pu faire la même construction avec l'arithmétique élémentaire, mais dans ce cas il aurait fallu ajouter des axiomes logiques d'ordre général, dont quelques éléments du calcul des classes (ou de la théorie des ensembles), pour construire Γ et démontrer certains théorèmes essentiels à la formalisation de G1 et G2 (je pense notamment aux théorèmes A et B).

3.2. La première preuve reconsidérée

Dans l'ensemble, la preuve du premier théorème de Gödel est valide. Il y a toutefois certains passages où la distinction entre langage et métalangage n'est pas claire. D'abord, entendons-nous sur le fait que les déductions de Gödel ont lieu dans Π et non dans P. Cette observation, si évidente fût-elle, peut être obscurcie par l'utilisation métathéorique du symbolisme de P. Par exemple, dans les énoncés (|) et (||), qui sont des énoncés métathéoriques (n'oublions pas que $Dém_{\kappa}(x,y)$ est une relation de la métathéorie malgré le fait qu'elle soit arithmétisée), le symbole « $\supset$ » apparaît et pourtant c'est un symbole de P.

Reprenons la partie (i) de la preuve du premier théorème :

1.	$D_{\kappa}(Gén(17,r))$	Hyp.
2.	$\exists n Dém_{\kappa}(n, Gén(17,r))$	Déf. 46
3.	$Dém_{\kappa}(x, Gén(17,r)) \supset D_{\kappa}(Nég(Sub(r,17,N(x))))$	Par (||)
…		
4.	$\exists n Dém_{\kappa}(x, Gén(17,r)) \supset \exists n D_{\kappa}(Nég(Sub(r,17,N(n))))$	Déd. dans Π
5.	$\exists n D_{\kappa}(Nég(Sub(r,17,N(n))))$	*MP* sur 3 et 4
6.	$D_{\kappa}(Gén(17,r)) \supset D_{\kappa}(Sub(r,17,N(x)))$	Théorème
7.	$D_{\kappa}(Sub(r,17,N(x)))$	*MP* sur 1 et 6
…		
8.	$\exists n(D_{\kappa}(Nég(Sub(r,17,N(n)))) \wedge D_{\kappa}(Sub(r,17,N(n))))$	Déd. dans Π
9.	$\sim \forall n \sim (D_{\kappa}(Nég(Sub(r,17,N(n)))) \wedge D_{\kappa}(Sub(r,17,N(n))))$	Déf. de $\exists$
10.	$\sim (D_{\kappa}(Nég(Sub(r,17,N(n)))) \wedge D_{\kappa}(Sub(r,17,N(n))))$	Cohérence de Π

11.	$\forall n \sim (D_\kappa(Nég(Sub(r,17,N(n)))) \wedge D_\kappa(Sub(r,17,N(n))))$	Gén. 10 par n
...		
12.	$\sim Coh(\kappa)$	Déd. de 11
13.	$D_\kappa(Gén(17,r)) \supset \sim Coh(\kappa)$	Élimination de 1.
14.	$Coh(\kappa) \supset \sim D_\kappa(Gén(17,r))$	Par contraposition

D'abord, cette déduction est une déduction dans Π et non pas dans P. D'autre part, elle ne peut pas être traduite dans P parce que la fonction Γ ne peut pas être arithmétisée et elle est nécessaire à la déduction des énoncés (|) et (||) (qui sont des énoncés fondamentaux à la preuve). Quoiqu'il en soit, ces omissions n'affectent pas tellement le théorème, et il demeure valide du moment qu'on ne prétend pas qu'il soit démontrable dans P.

Arrêtons-nous un moment sur Γ avant de terminer ici. Il faut comprendre qu'en disant que Γ rend l'arithmétisation possible, on veut dire qu'en l'absence de Γ, « $Dém_\kappa(x,y)$ » ainsi que toutes les autres relations métathéoriques arithmétisées ne sont que des relations arithmétiques arbitrairement choisies sans aucun pouvoir expressif. Γ assure les équivalences suivantes (dans le métalangage Π) :

F est une formule si et seulement si $Form(f)$

D est une preuve si et seulement si $Démon_\kappa(d)$

D est une preuve de F si et seulement si $Dém_\kappa(d,f)$ [29]

où D et F sont des expressions de P, $d = \Gamma(D)$ et $f = \Gamma(F)$. Γ induit des équivalences similaires pour les autres traductions arithmétiques de notions

[29] Les équivalences sont écrites ici de manière informelles.

métathéoriques,[30] et c'est en vertu de ces équivalences que les résultats de Gödel ont le moindre sens. Sans elles, la κ-prouvabilité de (la formule dont le *ndg* est $Gén(17,r)$) n'aurait aucune implication ni pour le programme de Hilbert ni pour la métalogique. À titre d'expérience, on n'aura qu'à comparer la signification de la définition 9 avec et sans les équivalences induites par Γ : « $R(x)$ est la suite composé du seul nombre x » *vs* « $R(x) = 2^x$ ».

3.3. La deuxième preuve reconsidérée

Passons maintenant au deuxième théorème d'incomplétude. La démonstration de ce théorème commence ainsi (en continuation avec la démonstration de la section précédente) :

1.	$Coh(\kappa) \supset \sim D_\kappa(Gén(17,r))$	À partir de G1
2.	$Coh(\kappa) \supset \forall x \sim Dém_\kappa(x, Gén(17,r))$	Déf. 46
3.	$Gén(17,r) = Sub(p,19,N(p))$	Par déf. de p et q
4.	$Coh(\kappa) \supset \forall x \sim Dém_\kappa(x, Sub(p,19,N(p)))$	Propriétés de =
5.	$Q(x,y) \equiv \sim Dém_\kappa(x, Sub(y,19,N(y)))$	Déf.
6.	$\forall x Q(x,p) \equiv \forall x \sim Dém_\kappa(x, Sub(p,19,N(p)))$	Gén. et Sub. sur 5
7.	$Coh(\kappa) \supset \forall x Q(x,p)$	Par 6

Les mêmes remarques faites à propos de la preuve du premier théorème s'appliquent au deuxième : cette déduction est dans Π et non dans P, et elle n'est pas réductible à une déduction dans P due à l'intervention de Γ: La prochaine étape de la preuve du

[30] C'est-à-dire les définitions 6 à 46 de la section 2.4.

second théorème est là où se manifestent deux erreurs importantes. Gödel fait ensuite les affirmations suivantes :[31]

(I) $Coh(\kappa)$ peut être « exprimé » par une relation dont le nombre de Gödel est c;

(II) $Q(x,y)$ est « exprimé » par q;

(III) $Q(x,p)$ est « exprimé » par r;

(IV) $\forall x Q(x,p)$ est « exprimé » par $Gén(17,r)$ car $r = Sub(q,19,N(p)))$;

(V) $Coh(\kappa) \supset \forall x Q(x,p)$ est « exprimé » par $impl(c,Gén(17,r))$;

(VI) $impl(c,Gén(17,r))$ est le *ndg* d'une formule démontrable dans P.

Par le terme « exprimable », Gödel semble signifier que les énoncés en question peuvent être représentés dans P. Il semble vouloir appliquer le théorème B à ces formules en négligeant le fait que certaines d'entre elles ne constituent pas des relations récursives (ce que les hypothèses du théorème B exigent). Par ailleurs, le fait que $Q(x,y)$ soit représentable par $Q(x_1,x_2)$ (dont le *ndg* soit q) n'entraîne pas que 1) $Q(x,y)$ ait un nombre de Gödel, 2) que $\forall x Q(x,p)$ soit représentable par $\overline{\forall}\wedge x_1 \wedge Q(x_1,\overline{p})$, 3) que la démontrabilité de $Coh(\kappa) \supset \forall x Q(x,p)$ dans Π entraîne que $impl(c,Gén(17,r))$ est le *ndg* d'un théorème de P, etc. Cette erreur prend clairement ses sources dans une confusion entre le langage et le métalangage, entre P et Π. C'est comme s'il avait affirmé que :

(I) $\Gamma\big(Coh(\kappa)\big) = c$;

(II) $\Gamma\big(Q(x,y)\big) = q$;

(III) $\Gamma\big(Q(x,p)\big) = r$;

(IV) $\Gamma\big(\forall x Q(x,p)\big) = Gén(17,r)$;

---

[31] van Heijenoort 1967 : p. 615.

(V) $\Gamma(Coh(\kappa) \supset \forall x Q(x,p)) = impl(c, Gén(17,r))$;

(VI) $D_{\kappa}(impl(c, Gén(17,r)))$.

(I) à (VI) sont de toute évidence des infractions commises à l'endroit des niveaux de langage : la fonction Γ n'est pas définie pour les formules de Π, elle ne s'applique qu'aux termes descriptifs de Π (les noms pour les expressions de *P*).

Passons temporairement par-dessus les difficultés (I) à (VI) pour voir comment la preuve devrait se poursuivre :

1.	$D_{\kappa}(impl(c, Gén(17, Sub(q,19,N(p)))))$	
2.	$Gén(17, Sub(q,19,N(p))) = Gén(17,r)$	Par déf.
3.	$D_{\kappa}(impl(c, Gén(17,r)))$	Déc. de 1 et 2
…		
4.	$(Dém_{\kappa}(x, impl(s,t) \wedge Dém_{\kappa}(y,s)) \supset Dém_{\kappa}(x * y * R(t), t)$	Découle de *Dém*
…		
5.	$(D_{\kappa}(impl(s,t)) \wedge D_{\kappa}(s)) \supset D_{\kappa}(t)$	Découle de 4
6.	$(D_{\kappa}(impl(c, Gén(17,r))) \wedge D_{\kappa}(c)) \supset D_{\kappa}(Gén(17,r))$	Subst. dans 5
7.	$D_{\kappa}(c)$	Hyp.
…		
8.	$(D_{\kappa}(impl(c, Gén(17,r)) \wedge D_{\kappa}(c))$	3 et 7
9.	$D_{\kappa}(Gén(17,r))$	*MP* sur 6 et 8
10.	$D_{\kappa}(Gén(17,r)) \supset \sim Coh(\kappa)$	De G1
11.	$\sim Coh(\kappa)$	*MP* sur 9 et 10
12.	$D_{\kappa}(c) \supset \sim Coh(\kappa)$	Élimination 7

Ce tronçon de preuve (l'enchaînement 1 à 12) est entièrement représentable dans Π sous l'hypothèse 1. Mais 1 n'est pas démontrable car il découle de (I) à (VI) qui sont des affirmations fausses. L'argument que Gödel suggère dans l'article de 1931 pour la démonstration du second théorème d'incomplétude n'est donc pas valide.

Afin de prendre le nombre de Gödel d'une formule de Π, il faudrait construire un métalangage de Π, disons $\mathfrak{P}$. Il suffirait de faire le même travail pour $\mathfrak{P}$ que nous avons fait pour Π. Ce langage serait donc une extension de Π qui comprendrait des notions descriptives pour Π. Π est essentiellement la réunion de a) *P*, b) l'arithmétique récursive et c) une théorie descriptive pour *P*. En conséquence, $\mathfrak{P}$ serait la réunion de a) *P*, b) l'arithmétique récusive, c) d'une description de l'arithmétique récursive, d) d'une description de *P* et e) une description de la description de *P*. Comme symboles primitifs, il nous faudrait :

- « $\sim$ », « $\vee$ », « 0 », « ' », « $\forall$ », « (», «) », « λ », « $\mathfrak{R}$ » et des variables « x » pour la partie *P*;
- « $\bar{\sim}$ », « $\overline{\vee}$ », « $\bar{0}$ », « $\overline{\prime}$ », « $\overline{\forall}$ », « $\bar{(}$ », « $\bar{)}$ », « $\bar{\lambda}$ », « $\overline{\mathfrak{R}}$ », « ^ » et des noms de variables « $\bar{x}$ » pour la description de *P* et de l'arithmétique récursive;
- « $\breve{\sim}$ », « $\breve{\vee}$ », « $\breve{0}$ », « $\breve{\prime}$ », « $\breve{\forall}$ », « $\breve{(}$ », « $\breve{)}$ », « $\breve{\lambda}$ », $\breve{\mathfrak{R}}$, « $\breve{\wedge}$ » et des noms de noms de variables « $\breve{x}$ » pour la description de description de *P*.

La construction de la section 4.1 peut être reprise intégralement ici. Voici quelques exemples de définitions dans $\mathfrak{P}$:

- $\breve{0}$, $\breve{0}{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime}$, $\breve{0}{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime}{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime}$, $\breve{0}{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime}{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime}{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime}$, ... sont des symboles de symboles de nombres. La classe de tels symboles est définie comme étant le plus petit ensemble X tel que 1) $\breve{0} \in X$ et 2) si $x \in X$ alors $x{}^\wedge\breve{\wedge}{}^\wedge\breve{\prime} \in X$.
- De manière analogue, $\bar{0}$, $\bar{0}{}^\wedge\overline{\prime}$, $\bar{0}{}^\wedge\overline{\prime}{}^\wedge\overline{\prime}$, $\bar{0}{}^\wedge\overline{\prime}{}^\wedge\overline{\prime}{}^\wedge\overline{\prime}$, ... sont des symboles de nombres. La classc dcs symboles de nombre est définie comme étant le plus petit ensemble X tel que 1) $\bar{0} \in X$ et 2) si $x \in X$ alors $x{}^\wedge\overline{\prime} \in X$.

- Enfin, 0, 0', 0'', 0''', ... sont des nombres. La classe des nombres est définie comme étant le plus petit ensemble X tel que 1) $0 \in X$ et 2) si $x \in X$ alors $x' \in X$.
- Si la relation $SSVAR(n,x)$ signifie que x est un symbole de symbole de variable de type n, nous pouvons définir les symboles de symboles de formules élémentaires, le prédicat $SSFORMÉ(x)$, comme suit :

$$\exists n \exists y \exists z \left(SSVAR(n,y) \wedge SSVAR(n+1,z) \wedge x = z{}^{\wedge}\breve{\bar{{}^{\wedge}}}{}^{\wedge}\breve{(}{}^{\wedge}\breve{\bar{{}^{\wedge}}}{}^{\wedge}y{}^{\wedge}\breve{\bar{{}^{\wedge}}}{}^{\wedge}\breve{)} \right).$$

- De même, si $SVAR(n,x)$ signifie que x est un symbole de variable de type n, nous pouvons définir les symboles de formules élémentaires, le prédicat $SFORMÉ(x)$, comme suit :

$$\exists n \exists y \exists z \left(SVAR(n,y) \wedge SVAR(n+1,z) \wedge x = z{}^{\wedge}\bar{(}{}^{\wedge}y{}^{\wedge}\bar{)} \right).$$

Il y a à l'intérieur de $\mathfrak{P}$ à la fois des descriptions et des descriptions de descriptions. Par exemple, le théorème A une fois formalisé serait une description de théorème à l'intérieur duquel on retrouverait à la fois des descriptions de formules (celles de Π) et des descriptions de descriptions de formules (celles de P). Remarquons, par ailleurs, qu'en formalisant Π, nous avons formalisé la fonction de Gödel Γ, par γ disons. La fonction de Gödel Γ_1 définie sur termes descriptifs de $\mathfrak{P}$ doit satisfaire une seule contrainte : accorder les mêmes valeurs que Γ aux descriptions de P (et non pas les descriptions de descriptions de P).

Une fois que nous avons le langage $\mathfrak{P}$, nous pouvons procéder à la preuve. Fixons certaines conventions d'écriture pour faciliter le passage d'un niveau de langage à un autre. D'abord, les symboles $x, y, z, \ldots$ seront les variables de $\mathfrak{P}$, $\bar{x}, \bar{y}, \bar{z}, \ldots$ des descriptions de variables (ils dénotent donc les variables de Π), et $\breve{x}, \breve{y}, \breve{z}, \ldots$ des descriptions de descriptions de variables (ils dénotent les variables de P). Par ailleurs, si $R(x)$ est une relation $\mathfrak{P}$ et si elle peut être décrite, alors $\overline{R(x)}(\bar{x})$

dénotera cette description. De même, si une description de description de $R(x)$ existe, on la notera par $\overline{\overline{R(x)}}(\breve{x})$. On remarquera que $\overline{\overline{R(x)}(\bar{x})}(\breve{x}) = \overline{\overline{R(x)}}(\breve{x})$. Rappelons au lecteur que la preuve de ce théorème utilise à nouveau la formule $Q(x,y)$, laquelle est représentable selon le théorème A par la description de formule $\mathcal{Q}(\bar{x},\bar{y})$ (c'est-à-dire que $\widehat{Q(x,y)}(\bar{x},\bar{y}) = \mathcal{Q}(\bar{x},\bar{y})$). Ainsi, la description de formule $\overline{Q(x,y)}(\bar{x},\bar{y})$ est « représentable » par la description de description de formule $\overline{\mathcal{Q}(\bar{x},\bar{y})}(\breve{x},\breve{y})$. Avec ces quelques indications préliminaires, la démonstration peut se poursuivre ainsi :

- La preuve qui mène jusqu'à $Coh(\kappa) \supset \forall x Q(x,p)$ se formalise dans Π. Il s'agit plus précisément d'une preuve de la description $\overline{Coh(\kappa) \supset \forall x Q(x,p)}$ de $Coh(\kappa) \supset \forall x Q(x,p)$.
- À partir de la description $\mathcal{Q}(\bar{x},\bar{y})$, nous construisons la description $\bar{\forall} \wedge \bar{x} \wedge \mathcal{Q}(\bar{x},\bar{y})$ et nous posons $p = \Gamma\left(\bar{\forall} \wedge \bar{x} \wedge \mathcal{Q}(\bar{x},\bar{y})\right)$ et $r = \Gamma\left(\mathcal{Q}(\bar{x},\bar{p})\right)$ pour obtenir la description $\overline{Q(x,y)}(\bar{x},\bar{p})$.
- Nous construisons $\bar{\forall} \wedge \bar{x} \wedge \overline{Q(x,y)}(\bar{x},\bar{p})$ et remarquons que $\overline{Coh(\kappa) \supset \forall x Q(x,p)}$ est en fait $\overline{Coh(\kappa)} \wedge \bar{\supset} \wedge \bar{\forall} \wedge \bar{x} \wedge \overline{Q(x,y)}(\bar{x},\bar{p})$.
- À partir de ce moment, la preuve de $\overline{Coh(\kappa) \supset \forall x Q(x,p)}$ se traduit sans trop de difficultés.
- On peut démontrer que la description $\overline{Q(x,y)}$ est équivalente à la relation $\mathcal{Q}(\bar{x},\bar{y})$ qui la représente. Nous avons donc que $\overline{Q(x,y)}(\bar{x},\bar{y}) \wedge \bar{\supset} \wedge \mathcal{Q}(\bar{x},\bar{y})$ est une description de théorème, et par conséquent $\overline{Coh(\kappa)} \wedge \bar{\supset} \wedge \bar{\forall} \wedge \bar{x} \wedge \mathcal{Q}(\bar{x},\bar{p})$ est une description de théorème également.

- Dans la théorie $\mathfrak{P}$, il est possible d'arithmétiser le fait qu'une description de formule soit une description de théorème (c'est-à-dire un théorème de Π). Ceci est possible grâce au fait que les axiomes de Π sont récursivement énumérables. Soit $Dém_{\Pi}(x,y)$ cette relation, c'est-à-dire la relation affirmant que x est le nombre de Gödel d'une démonstration dans Π de la formule dont le nombre de Gödel est y; et soit $D_{\Pi}(x)$ la relation qui affirme que x est le nombre de Gödel d'une formule démontrable dans Π.
- Si $c=\Gamma(Coh(\kappa))$, nous avons alors que $\Gamma\left(\overline{Coh(\kappa)}\wedge\bar{\supset}\wedge\bar{\forall}\wedge\bar{x}\wedge Q(\bar{x},\bar{p})\right)$ est $impl(c,Gén(17,r))$ et donc $Dém_{\Pi}(n,impl(c,Gén(17,r)))$, où n est le nombre de Gödel de la démonstration dans Π de $\overline{Coh(\kappa)}\wedge\bar{\supset}\wedge\bar{\forall}\wedge\bar{x}\wedge Q(\bar{x},\bar{p})$. Autrement dit, $D_{\Pi}(impl(c,Gén(17,r)))$.
- De la formule $D_{\Pi}(impl(c,Gén(17,r)))$, nous pouvons aisément montrer maintenant que $D_{\Pi}(c)\supset D_{\Pi}(Gén(17,r))$.

Voici un tableau qui schématise les différents niveaux de description des formules :

MÉTA-MÉTALANGAGE $\mathfrak{P}$		
NIVEAU $\mathfrak{P}$	NIVEAU Π	NIVEAU P
	$\overline{Q(x,y)}(\bar{x},\bar{y})$	
	$Q(\bar{x},\bar{y})$	$\overline{Q(\bar{x},\bar{y})}(\ddot{x},\ddot{y})$
	$\bar{\forall}\wedge\bar{x}\wedge\overline{Q(x,y)}(\bar{x},\bar{p})$	
	Une preuve de $Coh(\kappa)\supset\bar{\forall}\wedge\bar{x}\wedge Q(\bar{x},\bar{p})$	
	$Coh(\kappa)\supset\bar{\forall}\wedge\bar{x}\wedge Q(\bar{x},\bar{p})$	
$D_{\Pi}(impl(c,Gén(17,r)))$		
$D_{\Pi}(c)\supset D_{\Pi}(Gén(17,r))$		

Cette modification de la preuve nous montre donc que si la cohérence est démontrable (dans P ou dans Π) alors l'énoncé $\overline{\forall} \wedge \overline{x} \wedge Q(\overline{x}, \overline{p})$ (dont le nombre de Gödel est $Gén(17,r)$) est démontrable dans Π (et non dans P). Il n'est donc pas possible d'obtenir la contradiction voulue, à savoir $D_{\kappa}\left(Gén(17,r)\right)$, telle que le nécessite la preuve. Si nous pouvions démontrer que $D_{\Pi}\left(Gén(17,r)\right)$ était contradictoire également, nous pourrions compléter la preuve, mais une telle contradiction n'est pas immédiate à mes yeux. La preuve ne peut donc pas être « complétée » de cette manière.

4. Exposition moderne des théorèmes de Gödel

La critique donnée à la section précédente ne porte que sur la démonstration effectuée par Gödel dans son article original. Elle n'exclut pas la possibilité qu'il existe d'autres moyens – valides cette fois-ci – pour arriver à la démonstration des théorèmes d'incomplétude. Je propose de donner une démonstration des théorèmes d'incomplétude par l'entremise d'une exposition donnée quelque cinquante ans après Gödel, celle de Mendelson.[32] Mendelson démontre les théorèmes de Gödel très élégamment en capitalisant sur les améliorations et les raccourcis de leurs preuves, dont le lemme de diagonalisation (dû à une observation de Carnap) et le théorème de Löb. Naturellement, dans le but d'éviter des longueurs inutiles au texte, je tenterai de minimiser au plus la durée de cette nouvelle incursion théorique.

---

[32] Mendelson 1987 : pp. 116-175.

4.1. Le langage du premier ordre sous examen

De nos jours, il n'est plus coutume de cadrer les théories arithmétiques dans les *Principia Mathematica*. Malgré l'immense popularité qu'ils ont connu dans la première moitié du vingtième siècle, ils se retrouvent aujourd'hui occultés par les théories du premier ordre qui sont beaucoup plus simples et faciles à manipuler. C'est donc pour une certaine arithmétique du premier ordre[33] que nous démontrerons les théorèmes de Gödel à présent. Commençons par la description générale d'une théorie du premier ordre.[34] D'abord, une *théorie du premier ordre* $\mathcal{L}$ comporte les symboles primitifs « , », « (», «) », « $\forall$ », « $\neg$ » et « $\Rightarrow$ » de même que des symboles pour les variables : $x_1, x_2, ..., x_n, ...$. De plus, il contient des symboles de prédicats « A_k^n » (n et k sont des entiers positifs) en nombre fini non nul ou dénombrable, des symboles de constantes « a_k » en nombre fini (peut être nul) ou dénombrable, et des symboles de fonctions « f_k^n » en nombre fini (peut être nul) ou dénombrable. Ainsi, l'alphabet d'un langage du premier ordre $\mathcal{L}$ est fixé dès que nous avons déterminé ses symboles de prédicats, de constantes et de fonctions.[35]

Parmi les expressions remarquables de $\mathcal{L}$, il y a les *termes* et les *formules*. Les termes sont définis récursivement de la manière suivante :

(T1) Les variables et les constantes (s'il y en a) de $\mathcal{L}$ sont des termes;

(T2) Si $t_1, t_2, ..., t_n$ sont des termes et si f_k^n est un symbole de fonction, alors $f_k^n(t_1, t_2, ..., t_n)$ est également un terme;

---

[33] Deux remarques s'imposent. D'abord, il y a plusieurs arithmétiques du premier ordre selon la complexité de la théorie. Ensuite, la démonstration donnée plus bas s'appliquera à un ensemble de théories arithmétiques, et non pas seulement à une seule spécifiquement.

[34] Le « premier ordre » fait ici référence à la nature des variables et des quantificateurs.

[35] On remarquera qu'un langage a toujours au moins un symbole de prédicat. Toutefois, certains langages peuvent, selon notre définition, ne pas avoir de symboles de constantes ou de fonctions. Par exemple, un langage qui formalise une structure d'ordre partielle n'a besoin que d'un symbole de relation binaire.

(T3) Tous les termes sont donnés par (1) et (2).

Avec les termes, il nous est possible de définir les expressions qui feront figure de propositions élémentaires dans le système : les formules atomiques. Si $t_1, t_2, \ldots, t_n$ sont des termes et A_k^n est un symbole de prédicat, alors $A_k^n(t_1, t_2, \ldots, t_n)$ est une *formule atomique*. Les formules sont ensuite définies récursivement de la manière suivante :

(F1) Les formules atomiques sont des formules;
(F2) Si $\mathcal{A}$ et $\mathcal{B}$ sont des formules et si x est une variable alors $\neg\mathcal{A}$, $(\mathcal{A} \Rightarrow \mathcal{B})$ et $\forall x\mathcal{A}$ sont des formules;
(F3) Les formules sont uniquement données par les règles (1) et (2).[36]

Si $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ sont des formules, alors formules suivantes sont des axiomes en commun à toute théorie du premier ordre $\mathcal{L}$:

(A1) $\mathcal{A} \Rightarrow (\mathcal{B} \Rightarrow \mathcal{A})$;
(A2) $(\mathcal{A} \Rightarrow (\mathcal{B} \Rightarrow \mathcal{C})) \Rightarrow ((\mathcal{A} \Rightarrow \mathcal{B}) \Rightarrow (\mathcal{A} \Rightarrow \mathcal{C}))$;
(A3) $(\neg\mathcal{B} \Rightarrow \neg\mathcal{A}) \Rightarrow ((\neg\mathcal{B} \Rightarrow \mathcal{A}) \Rightarrow \mathcal{B})$;
(A4) $\forall x_i \mathcal{A}(x_i) \Rightarrow \mathcal{A}(t)$ si t est un terme qui est libre pour x dans $\mathcal{A}(x_i)$;[37]
(A5) $\forall x_i (\mathcal{A} \Rightarrow \mathcal{B}) \Rightarrow (\mathcal{A} \Rightarrow \forall x_i \mathcal{B})$ si F ne contient pas d'occurrences libres de x_i.

Avec l'alphabet, les axiomes supplémentaires constituent ce qui caractérise véritablement une théorie du premier ordre. Nous verrons plus loin ceux qui

[36] On remarquera que si un langage n'a pas de symboles de prédicat, selon ces définitions, il n'aura pas de formules.
[37] Un terme t est libre pour une variable x dans une formule $F(x)$ (trois paramètres figurent dans la définition) si x n'apparaît pas libre dans $F(x)$ dans la portée d'un quantificateur liant une variable de t.

définiront l'arithmétique de base S pour laquelle nous démontrerons les théorèmes d'incomplétude.

La dimension inférentielle d'une théorie du premier ordre ressemble beaucoup à celle des *Principia*, les règles sont :

(RI1) Le *modus ponens* : si $\mathcal{A}$ et $(\mathcal{A} \Rightarrow \mathcal{B})$ alors $\mathcal{B}$;

(RI2) La généralisation : de $\mathcal{A}$, il s'ensuit $\forall x \mathcal{A}$.

Enfin, la notion de preuve dans un système du premier ordre est la même qu'ailleurs. Une suite $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ de formules est une *preuve* si pour tout $1 \leq i \leq n$: 1) $\mathcal{A}_i$ est un axiome de $\mathcal{L}$ ou 2) il existe $\mathcal{A}_j$ et $\mathcal{A}_k$ avec $j,k < i$ telles que $\mathcal{A}_k$ est $(\mathcal{A}_j \Rightarrow \mathcal{A}_i)$ ou $\mathcal{A}_i$ est $\forall x \mathcal{A}_k$ (c'est-à-dire que $\mathcal{A}_i$ est obtenue de formules précédentes par *modus ponens* ou par généralisation). Une formule $\mathcal{A}$ est un *théorème* de $\mathcal{L}$, noté $\vdash_{\mathcal{L}} \mathcal{A}$, s'il existe une preuve qui termine par $\mathcal{A}$.

Le langage du premier ordre qui nous intéressera par la suite sera le langage S. Il contient : 1) un symbole de constante a_1 que nous renommons « 0 »; 2) trois symboles de fonctions f_1^1, f_1^2 et f_2^2 que nous renommons « ' », « + » et « · » respectivement (successeur, addition et multiplication); et 3) un symbole de prédicat A_1^2 que nous renommons « = » (le symbole d'égalité). Les axiomes spécifiques à S sont au nombre de neuf :

(S1) $x_1 = x_2 \Rightarrow (x_1 = x_3 \Rightarrow x_2 = x_3)$

(S2) $x_1 = x_2 \Rightarrow x_1' = x_2'$

(S3) $0 \neq x_1'$

(S4) $x_1' = x_2' \Rightarrow x_1 = x_2$

(S5) $x_1 + 0 = x_1$

(S6) $x_1 + x_2' = (x_1 + x_2)'$

(S7) $x_1 \cdot 0 = 0$

(S8) $x_1 \cdot (x_2') = x_1 \cdot x_2 + x_1$

(S9) Pour toute formule $\mathcal{A}(x)$ de S, $\mathcal{A}(0) \Rightarrow (\forall x(\mathcal{A}(x) \Rightarrow \mathcal{A}(x')) \Rightarrow \forall x \mathcal{A}(x))$

4.2. Les fonctions et relations récursives

Nous avons vu qu'une des idées fondamentales à la démonstration des théorèmes de Gödel est la représentabilité des fonctions et relations récursives. Soit K une extension de S. Rappelons qu'une relation $R(X_1,...,X_n)$[38] est représentable dans K s'il existe une formule $\mathcal{R}(x_1,...,x_n)$ de K telle que pour tous nombres $k_1,...,k_n$:

(1) Si $R(k_1,...,k_n)$ alors $\vdash_K \mathcal{R}(\overline{k}_1,...,\overline{k}_n)$;

(2) Si $\neg R(k_1,...,k_n)$ alors $\vdash_K \neg\mathcal{R}(\overline{k}_1,...,\overline{k}_n)$.

Dans le traitement que donne Mendelson à la représentabilité, nous serons surtout intéressés par la notion correspondante pour les fonctions. Une fonction arithmétique $f(X_1,...,X_n)$ est dite *représentable* dans K si et seulement s'il existe une formule $\mathcal{R}(x_1,...,x_n,x_{n+1})$ de K telle que pour tous nombres naturels $k_1,...,k_n$ et m :

(1) Si $f(k_1,...,k_n) = m$ alors $\vdash_K \mathcal{R}(\overline{k}_1,...,\overline{k}_n,\overline{m})$;

(2) $\vdash_K \exists_1 x_{n+1} \mathcal{R}(\overline{k}_1,...,\overline{k}_n, x_{n+1})$.

[38] J'emploie des majuscules pour ces variables pour les distinguer des variables de K.

Nous verrons que les fonctions récursives sont précisément les fonctions représentables dans S.

La classe des fonctions récursives est définie de manière analogue à celle de Gödel. D'abord, il y a les fonctions initiales que sont :

(FI1) La fonction nulle : $Z(x)=0$ pour tout x;

(FI2) La fonction successeur : $N(x)=x+1$;

(FI3) Les fonctions de projection : $U_i^n(x_1,...,x_n)=x_i$ pour tous $x_1,...,x_n$.

Ensuite les règles :

(Réc1) La substitution : la fonction $f(x_1,...x_n)$ est obtenue par substitution de la fonction $g(x_1,...,x_m)$ (de m arguments) et des fonctions $h_1(x_1,...,x_n),...,h_m(x_1,...,x_n)$ (de n arguments chaque) si

$$f(x_1,...x_n)=g(h_1(x_1,...,x_n),...,h_m(x_1,...,x_n)).$$

(Réc2) La récursion : la fonction $f(x_1,...x_n,x_{n+1})$ est obtenue par récursion des fonctions $g(x_1,...,x_n)$ (de n arguments) et $h(x_1,...,x_n,x_{n+1},x_{n+2})$ si

$$f(x_1,...,x_n,0)=g(x_1,...,x_n)$$
$$f(x_1,...,x_n,y+1)=h(x_1,...,x_n,y,f(x_1,...,x_n,y)).$$

(Réc3) L'opérateur μ. Soit $g(x_1,...,x_n,y)$ une fonction telle que pour tous $x_1,...,x_n$, il existe au moins un y tel que $g(x_1,...,x_n,y)=0$. Pour une telle fonction, nous signifions par $\mu y(g(x_1,...,x_n,y)=0)$ le plus petit nombre y qui satisfait $g(x_1,...,x_n,y)=0$. Si

$$f(x_1,...,x_n)=\mu y(g(x_1,...,x_n,y)=0)$$

alors on dit que $f(x_1,...,x_n)$ est obtenue de $g(x_1,...,x_n,y)$ par l'opérateur μ.

Une fonction f est dite *récursive* si et seulement si elle peut être obtenue des fonctions initiales (FI1), (FI2) ou (FI2) par un nombre fini d'applications des règles (Réc1), (Réc2) et (Réc3), c'est-à-dire s'il existe une suite de fonctions $f_1,\ldots,f_n$ telles que, pour tout i, f_i est soit une fonction initiale soit une fonction obtenue de fonctions $f_{j_1},\ldots,f_{j_r}$ avec $j_1,\ldots,j_r < i$ par l'application des règles (Réc1), (Réc2) ou (Réc3). Une fonction est dite *primitivement récursive* si elle est récursive et il est possible de l'obtenir sans l'application de la règle (Réc3).

Il est possible de démontrer ensuite que :

THÉORÈME. Toutes les fonctions récursives sont représentables dans K.
COROLLAIRE. Toutes les relations récursives sont représentables dans K.[39]

4.3. L'arithmétisation de la métamathématique

Soit K une théorie du premier ordre qui possède les mêmes symboles que S.[40] L'arithmétisation de K se fait de la même manière que celle de la théorie P des *Principia*. On attribue d'abord à chaque symbole primitif de K un nombre par une fonction g :

(	)	,	$\neg$	$\Rightarrow$	$\forall$
3	5	7	9	11	13

x_k	a_k	f_k^n	A_k^n
$13+8k$	$7+8k$	$1+8(2^n3^k)$	$3+8(2^n3^k)$

[39] Le passage des fonctions aux relations se fait par la fonction caractéristique d'une relation.
[40] On remarquera que K n'est pas nécessairement une extension de S.

On définit ensuite une fonction Γ qui attribue un nombre à toute expression $u = u_0 u_1 \cdots u_r$ de K (les u_i sont des symboles primitifs) de la manière suivante :

$$\Gamma(u_0 u_1 \cdots u_r) = 2^{g(u_0)} 3^{g(u_1)} \cdots p_r^{g(u_r)}$$

où p_i désigne le i-ième nombre premier ($p_0 = 2, p_1 = 3, p_2 = 5, \ldots$).[41] Si $U_0, U_1, \ldots, U_r$ sont des expressions (c'est-à-dire une suite d'expressions), le nombre associé à la suite sera

$$\Gamma(U_0, U_1, \ldots, U_r) = 2^{\Gamma(U_0)} 3^{\Gamma(U_1)} \cdots p_r^{\Gamma(U_r)}.$$

La fonction Γ associe donc à chaque expression et à chaque suite d'expressions de K un nombre, et le théorème fondamental de l'algèbre (le fait que tout nombre ait une factorisation en nombres premiers unique à l'ordre près) nous donne que 1) la fonction Γ est injective (deux expressions/suites d'expressions ont des nombres associés différents), et 2) les propriétés syntaxiques des expressions/suites d'expressions sont lisibles à même leurs nombres de Gödel respectifs.

L'arithmétisation que nous avons accompli pour la théorie P se laisse faire également pour la théorie S. Notamment, il existe une fonction récursive $\delta(x)$ et une relation récursive $Dém_S(x, y)$ telles que :

- $\delta(n)$ = le nombre de Gödel de la formule obtenue d'une formule de nombre de Gödel n ayant une seule variable libre en substituant la variable libre pour le nombre n. C'est-à-dire : si $\mathcal{A}(x_1)$ est une formule qui a une seule variable

[41] On remarquera la différence entre le nombre associé à un symbole et le nombre associé à une expression composée de ce seul symbole.

libre x_1 et dont le nombre de Gödel est n, $\delta(n)$ est le nombre de Gödel de $\mathcal{A}(\overline{n})$.

- $Dém_S(x, y)$ si et seulement si x est le nombre de Gödel d'une preuve dans S de la formule dont le nombre de Gödel est y.

Puisque les fonctions et relations récursives dans S sont représentables, il existe des formules $\mathfrak{D}(x_1, x_2)$ et $\mathfrak{D}\acute{e}m_S(x_1, x_2)$ qui les représentent.

4.4. Le premier théorème d'incomplétude

Carnap avait remarqué dès 1934 (dans *La syntaxe logique du langage*) qu'une des idées fondamentales à la démonstration des théorèmes d'incomplétude était la notion de point fixe ou de diagonalisation. Cette idée prend la forme du théorème suivant :

THÉORÈME DU POINT FIXE. Pour toute formule $\mathcal{A}(x_1)$ de S d'une seule variable libre x_1, il existe une formule close $\mathcal{B}$ telle que

$$\vdash_S \mathcal{B} \Leftrightarrow \mathcal{A}(\ulcorner \mathcal{B} \urcorner).$$

(Notons que $\ulcorner \mathcal{B} \urcorner$ est une abréviation pour $\overline{\Gamma(\mathcal{B})}$, c'est-à-dire le symbole qui représente le nombre de Gödel de $\mathcal{B}$.)[42]

[42] Le symbole qui représente le nombre n (non nul) est évidemment $0\underbrace{'' \cdots '}_{n \text{ fois}}$.

Sans entrer dans la preuve du théorème, mentionnons seulement la formule $\mathcal{B}$ qui remplit les exigences du théorème. Nous définissons d'abord la formule $\mathcal{C}$ comme $\forall x_2(\mathfrak{D}(x_1,x_2) \Rightarrow \mathcal{A}(x_2))$ où $\mathfrak{D}(x_1,x_2)$ est la formule qui représente la fonction récursive $\delta(x)$. Nous posons ensuite $\mathcal{B}$ comme étant

$$\forall x_2(\mathfrak{D}(\ulcorner \mathcal{C} \urcorner, x_2) \Rightarrow \mathcal{A}(x_2)).$$

Intuitivement, nous pouvons voir $\mathcal{B}$ comme $\mathcal{A}(\overline{\delta(a)})$ où $a = \Gamma(\mathcal{A}(x_1))$.

Le premier théorème d'incomplétude découle du théorème du point fixe appliqué à la formule $\forall x_2 \neg \mathfrak{D}\acute{e}m_S(x_2,x_1)$. Soit $\mathcal{G}$ le point fixe de $\forall x_2 \neg \mathfrak{D}\acute{e}m_S(x_2,x_1)$. Nous avons alors :

PREMIER THÉORÈME D'INCOMPLÉTUDE. Si S est ω-cohérente alors $\nvdash_S \mathcal{G}$ et $\nvdash_S \neg\mathcal{G}$.[43]

PREUVE. D'abord supposons que $\vdash_S \mathcal{G}$. Il existe donc un nombre n tel que $D\acute{e}m_S(n,g)$ où $g = \Gamma(\mathcal{G})$. Puisque $D\acute{e}m_S(x,y)$ est représentable, nous avons $\vdash_S \mathfrak{D}\acute{e}m_S(\overline{n},\overline{g})$, c'est-à-dire $\vdash_S \mathfrak{D}\acute{e}m_S(\overline{n},\ulcorner \mathcal{G} \urcorner)$. D'un autre côté, puisque $\mathcal{G}$ est le point fixe de $\forall x_2 \neg \mathfrak{D}\acute{e}m_S(x_2,x_1)$, nous avons que $\vdash_S \mathcal{G}$ entraîne $\vdash_S \forall x_2 \neg \mathfrak{D}\acute{e}m_S(x_2,\ulcorner \mathcal{G} \urcorner)$. Mais dans ce cas, nous avons $\vdash_S \neg \mathfrak{D}\acute{e}m_S(\overline{n},\overline{g})$ aussi, contredisant la (ω-) cohérence de S. Nous avons donc $\nvdash_S \mathcal{G}$.

Supposons ensuite que $\vdash_S \neg\mathcal{G}$. Par la première partie de cette preuve, nous avons que $\neg D\acute{e}m_S(n,g)$ pour tout n. Puisque $D\acute{e}m_S(x,y)$ est représentable, nous avons que $\vdash_S \neg \mathfrak{D}\acute{e}m_S(\overline{n},\overline{g})$ pour tout n. Mais de $\vdash_S \neg\mathcal{G}$, il s'ensuit que

[43] L'ω-cohérence est définie ici de la même manière que dans P.

$\vdash_S \neg\forall x_2 \neg \mathcal{D}\acute{e}m_S(x_2,\overline{g})$, ce qui contredit l'$\omega$-cohérence de S. Nous avons donc $\nvdash_S \neg\mathcal{G}$.

Il est bien connu que Rosser a amélioré les hypothèses de ce théorème, réduisant l'ω-cohérence à la cohérence seule :

THÉORÈME DE GÖDEL-ROSSER. Si S est cohérente, alors il existe une formule close $\mathcal{R}$ telle que $\nvdash_S \mathcal{R}$ et $\nvdash_S \neg\mathcal{R}$.

La formule $\mathcal{R}$ est obtenue en appliquant le théorème du point fixe à une certaine formule. Pour la définir, il faut d'abord nous rappeler que la fonction $N\acute{e}g(x)$ – qui retourne comme valeur le nombre de Gödel de la négation de l'expression dont le nombre de Gödel est x – est récursive. Il existe donc une formule $\mathcal{N}\acute{e}g(x_1,x_2)$ de S qui la représente. La formule que nous cherchons est

$$\forall x_2(\mathcal{D}\acute{e}m_S(x_2,x_1) \Rightarrow \forall x_3(\mathcal{N}\acute{e}g(x_1,x_3) \Rightarrow \exists x_4(x_4 \leq x_2 \wedge \mathcal{D}\acute{e}m_S(x_4,x_3)))\,.$$

Quelques remarques s'imposent sur ces deux théorèmes. Certaines généralisations peuvent être données. Soit K une théorie avec égalité qui possède les mêmes symboles que S. Si la fonction $\delta(x)$ est représentable dans K, alors le théorème du point fixe est valide pour K. De plus, si 1) $D\acute{e}m_K(x,y)$ est une relation récursive, 2) $\vdash_K 0 \neq 1$ et 3) toute fonction récursive est représentable dans K, alors le premier théorème d'incomplétude tient également pour K. Pour démontrer la version améliorée de Rosser, il faut exiger en plus que, pour tout nombre n, $\vdash_K x_1 \leq \overline{n} \Rightarrow (x_1 = 0 \vee x_1 = \overline{1} \vee \cdots \vee x_1 = \overline{n})$ et $\vdash_K x_1 \leq \overline{n} \vee \overline{n} \leq x_1$.

Le système arithmétique du premier ordre « minimal » dans lequel nous pouvons démontrer le premier théorème d'incomplétude est le système de Robinson, lequel peut être donné par treize axiomes seulement. C'est donc un exemple de système incomplet axiomatisable par un nombre fini d'axiomes, contrairement à S qui, à cause de l'axiome d'induction, possède un nombre infini d'axiomes. Il est donc possible de démontrer ce théorème d'incomplétude dans toute extension récursivement axiomatisable[44] de R. S est évidemment une telle extension de R.

Un exemple de système arithmétique complet cette fois-ci qui possède néanmoins l'axiome d'induction est celui de Presburger. L'arithmétique de Presburger est l'arithmétique additive. Elle est obtenue de S en supprimant les axiomes (S7) et (S8) de même que le symbole de fonction « $\cdot$ ».

4.4. Le deuxième théorème de Gödel

Rappelons qu'une théorie K est cohérente si, pour toute formule $\mathcal{A}$ de K, on n'a pas $\vdash_K \mathcal{A}$ et $\vdash_K \neg\mathcal{A}$, c'est-à-dire $\nvdash_K \mathcal{A}$ ou $\nvdash_K \neg\mathcal{A}$. Nous pouvons considérablement réduire cette condition en observant que toutes les formules de K sont démontrables du moment qu'une seule formule et sa négation le sont. En effet, si $\mathcal{A}$ est la formule qui est telle que $\vdash_K \mathcal{A}$ et $\vdash_K \neg\mathcal{A}$, alors du fait que $\vdash_K \neg\mathcal{A} \Rightarrow (\mathcal{A} \Rightarrow \mathcal{B})$ il s'ensuit que $\vdash_K \mathcal{B}$ et ce, quelle que soit la formule $\mathcal{B}$. La condition de cohérence de K se réduit donc à « non $\vdash_K \mathcal{A}$ ou non $\vdash_K \neg\mathcal{A}$ » pour une formule $\mathcal{A}$ de notre choix. Prenons, pour simplifier les choses, la formule $\neg 0 = 1$. Puisque nous savons que $\vdash_K 0 \neq 1$, la cohérence de K se réduit à la simple condition $\nvdash_K 0 = 1$. Le programme de Hilbert vise précisément à démontrer $\nvdash_S 0 = 1$ avec des moyens arithmétiques. En arithmétisant la métamathématique, le problème se formule

[44] Par une théorie récursivement axiomatisable, on entend une théorie pour laquelle les nombres de Gödel de ses axiomes sont donnés par une fonction récursive.

de la manière suivante : démontrer $\forall x_1 \neg \mathcal{D}\acute{e}m_S(x_1, \ulcorner 0 = 1 \urcorner)$ dans S, c'est-à-dire arriver à $\vdash_S \forall x_1 \neg \mathcal{D}\acute{e}m_S(x_1, \ulcorner 0 = 1 \urcorner)$. Le deuxième théorème de Gödel montre que ceci est impossible, que $\forall x_1 \neg \mathcal{D}\acute{e}m_S(x_1, \ulcorner 0 = 1 \urcorner)$ est indémontrable dans S (indécidable pour être précis). Par commodité, nous utiliserons $\mathcal{C}oh_S$ pour désigner la formule $\forall x_1 \neg \mathcal{D}\acute{e}m_S(x_1, \ulcorner 0 = 1 \urcorner)$.

Pour arriver à cette démonstration, il faut d'abord faire un détour par les propriétés générales d'un prédicat de prouvabilité. Soit $\mathcal{P}r(x_1)$ la formule $\exists x_2 \mathcal{D}\acute{e}m_S(x_2, x_1)$. Nous pouvons démontrer que le prédicat $\mathcal{P}r(x_1)$ satisfait les trois conditions suivantes (appelées les conditions de dérivabilité de Hilbert-Bernays) :

(HB1) Si $\vdash_S \mathcal{A}$ alors $\vdash_S \mathcal{P}r(\ulcorner \mathcal{A} \urcorner)$;

(HB2) $\vdash_S \mathcal{P}r(\ulcorner (\mathcal{A} \Rightarrow \mathcal{B}) \urcorner) \Rightarrow (\mathcal{P}r(\ulcorner \mathcal{A} \urcorner) \Rightarrow \mathcal{P}r(\ulcorner \mathcal{B} \urcorner))$;

(HB3) $\vdash_S \mathcal{P}r(\ulcorner \mathcal{A} \urcorner) \Rightarrow \mathcal{P}r(\ulcorner \mathcal{P}r(\ulcorner \mathcal{A} \urcorner) \urcorner)$.

Démontrer que $\mathcal{P}r(x_1)$ satisfait (HB1) et (HB2) est assez simple, démontrer qu'il satisfait (HB3) – ce que nous ne ferons pas ici – est plus complexe.[45] La démonstration du deuxième théorème d'incomplétude est grandement facilitée par l'utilisation de ces propriétés. Il nous faut cependant démontrer au préalable le théorème de Löb.

D'une certaine manière, le théorème de Löb se trouve à répondre à une question concernant le point fixe de $\mathcal{P}r(x_1)$. Si $\mathcal{H}$ est un point fixe de $\mathcal{P}r(x_1)$, il est équivalent à une formule qui affirme sa prouvabilité.[46] Est-il prouvable? Löb répond par l'affirmative, plus précisément :

---

[45] La démonstration de (HB3) peut être trouvée dans (Hilbert & Bernays 2001 : Tome II, pp. 348 et suivantes).

[46] Cette formule est souvent appelée formule de Henkin.

THÉORÈME DE LÖB. Si $\mathcal{H}$ est une formule close telle que $\vdash_S \mathcal{P}r(\ulcorner\mathcal{H}\urcorner) \Rightarrow \mathcal{H}$, alors $\vdash_S \mathcal{H}$.

La preuve utilise essentiellement les trois conditions de dérivabilité.

Le deuxième théorème de Gödel se démontre facilement à partir du théorème de Löb :

DEUXIÈME THÉORÈME D'INCOMPLÉTUDE. Si *S* est cohérente alors $\nvdash_S \mathcal{C}oh_S$.

PREUVE. Nous savons déjà que $\vdash 0 \neq 1$. Par la cohérence de *S*, nous avons que $\nvdash 0 = 1$. Par le théorème de Löb, nous savons que $\vdash_S \mathcal{P}r(0=1) \Rightarrow 0=1$ entraîne que $\vdash_S 0=1$; donc $\nvdash_S \mathcal{P}r(0=1) \Rightarrow 0=1$. Si $\vdash_S \neg\mathcal{P}r(0=1)$, par le fait que $\vdash_S \neg\mathcal{P}r(0=1) \Rightarrow (\mathcal{P}r(0=1) \Rightarrow 0=1)$ (une instance de tautologie), nous aurions $\vdash_S 0=1$; donc $\nvdash_S \neg\mathcal{P}r(0=1)$ aussi. Mais $\neg\mathcal{P}r(0=1)$ est précisément $\mathcal{C}oh_S$. Ainsi $\nvdash_S \mathcal{C}oh_S$.

On peut montrer que la formule $\mathcal{C}oh_S$ est indécidable dans *S*.

5. La postérité de l'incomplétude

Les théorèmes d'incomplétude de Gödel ont fait l'objet de nombreuses interprétations, certaines modestes et légitimes d'autres carrément douteuses. Arrêtons-nous ici le temps d'exposer quelques mécompréhensions communément rencontrées sur le sujet. Pour plusieurs, les théorèmes d'incomplétudes de Gödel démontrent une sorte de limitation essentielle à la connaissance humaine, voire à l'esprit. En ce sens, l'incomplétude de l'arithmétique viendrait appuyer une forme d'agnosticisme, l'idée que certaines vérités échapperaient perpétuellement à l'entendement humain. De telles interprétations ne sont pas le seul apanage des théosophes, même un éminent scientifique comme Stephen Hawking répand de pareilles confusions en la matière.

Dans une conférence intitulée *Gödel and the end of physics* donnée en 2002,[47] il étend l'incomplétude de l'arithmétique à la physique, arguant que toute théorie physique contient (au moins) l'arithmétique et se trouve ainsi touchée par les théorèmes d'incomplétude. Sans nous lancer dans des considérations expansives sur le sujet, mentionnons seulement que la notion arithmétique de vérité ou de satisfaction qui est au cœur de l'incomplétude n'a pas son correspondant dans les théories physiques. Toutes les tentatives (infructueuses) pour préciser une notion de vérification satisfaisante pour les énoncés empiriques témoignent déjà par elles-mêmes de l'abîme existant entre une théorie scientifique et la réalité. Prétendre que l'incomplétude gödelienne contribue de manière significative (voire même tout court) à l'incomplétude de la physique relève un peu de la paranoïa épistémologique.

La plus singulière interprétation de l'incomplétude est, à mon sens, celle de Roger Penrose.[48] Celui-ci trouve dans le théorème de Gödel une justification à l'effet que nos processus cognitifs sont non algorithmiques, car nous sommes effectivement en mesure de connaître l'énoncé qui est vrai mais non prouvable dans P (l'énoncé dont le *ndg* est $Gén(17, r)$). À son tour, la nature non algorithmique de la pensée est une indication claire, selon lui, de ce que nos processus cognitifs sont des processus quantiques.

L'interprétation canonique et moins exaltée des résultats de Gödel veut que l'incomplétude soit un résultat d'abord et avant tout de logique mathématique, un résultat métalogique stipulant que tout système formel suffisamment puissant pour contenir l'arithmétique (ou une copie isomorphe à celle-ci) est (1) incomplet (ne peut

---

[47] On peut trouver le texte de cette conférence à http://www.damtp.cam.ac.uk/strings2002/dirac/hawking/.

[48] *The Emperor's New Mind* (1989) et *Shadows of the Mind* (1994).

pas démontrer toutes les propositions vraies[49] dans ce système) et (2) inapte à démontrer sa propre cohérence (à moins d'être contradictoire, auquel cas tout est démontrable). Par ailleurs, comme le fit remarquer Hatcher,[50] il faut ajouter aux hypothèses de ces théorèmes la condition tacite suivante : ... *en utilisant la méthode d'arithmétisation de Gödel (ou une méthode analogue) pour représenter la théorie dans l'arithmétique*... C'est-à-dire que les théorèmes d'incomplétude de Gödel porte aussi sur une certaine façon de représenter une théorie dans l'arithmétique. Si une autre méthode de « projection » est utilisée, il se pourrait que les théorèmes ne s'appliquent pas.[51] Enfin, il faut insister sur le fait que l'impossibilité de démontrer la cohérence de *P* n'est pas absolue, elle est relative (encore une fois) aux moyens utilisés. L'impossibilité en question est de démontrer *avec l'arithmétique*[52] que l'arithmétique est non contradictoire. Ainsi, le fait que Gentzen[53] ait démontré la cohérence de l'arithmétique avec l'induction transfinie (des moyens qui dépassent largement le cadre de l'arithmétique) n'entre pas en contradiction avec le second théorème d'incomplétude.

Les théorèmes de Gödel portent donc un coup dur au programme de Hilbert : d'une part, la méthode axiomatique en sort écorchée par les limitations « essentielles » qui lui sont posées par le premier théorème d'incomplétude; d'autre part, la métamathématique s'avère insuffisante afin de réaliser l'objectif de démontrer la non contradiction d'une théorie aussi simple que l'arithmétique de Peano. Par ailleurs, le premier théorème a aussi révélé un fait pour le moins inattendu : la vérité est un concept considérablement plus fort que la prouvabilité. La morale semble donc être que le métalangage est là pour rester, en ce sens que le langage formel lui-même n'est pas suffisant pour rendre compte de certaines intuitions que nous entretenons à l'endroit de la vérité. Aussi, si nous souhaitons que le métalangage nous soit d'une

---

[49] Il s'agit évidemment de la vérité telle que conçue en logique.
[50] Hatcher 1982 : p. 207.
[51] Hatcher fait références aux travaux de Feferman et Takeuti qui vont en ce sens.
[52] Et la méthode d'arithmétisation de Gödel.
[53] Gentzen 1969 : p. 132-213.

certaine utilité, il faut augmenter considérablement ses capacités. Exit le langage universel, exit le finitisme.

CHAPITRE VI

Le concept de métalangage de Tarski

0. Introduction

Avec Tarski nous arrivons à la formulation moderne du métalangage, c'est-à-dire à la définition rigoureuse d'une théorie formelle qui peut étudier non seulement les propriétés syntaxiques d'un langage donné mais également les rapports sémantiques qu'il entretient avec sa référence. La principale différence entre le métalangage tarskien et la métamathématique hilbertienne est la capacité chez le premier à définir un concept de vérité. La cohérence d'une théorie étant une affaire symbolique (ou syntaxique), Hilbert s'était donné des moyens métathéoriques plutôt modestes. De son côté, Tarski s'intéresse à définir un concept formel de vérité applicable à une théorie quelconque. Ce changement de cap apporte des difficultés supplémentaires par son cadre général et par le fait que la vérité soit une notion plus complexe que la non-contradiction. En effet, la vérité n'est pas une affaire de syntaxe seulement, elle apparaît dans une relation entre le langage et ce qu'il représente. Afin de donner une définition correcte et adéquate du concept de vérité, Tarski devra pousser l'analyse métathéorique jusqu'à en faire une discipline. Par ailleurs, nous verrons que Tarski limite son concept de vérité aux langages formels seulement. Cette contrainte lui est imposée par un résultat célèbre portant sur l'impossibilité de définir un concept de vérité dans un langage sémantiquement clos, un résultat qui aura un rayonnement important dans la philosophie analytique du vingtième siècle. L'essentiel de ce chapitre sera consacré à l'exposition du concept de vérité tarskien, lequel concept inclut une notion détaillée de métalangage.[1]

1. La vérité comme prédicat pour les langages formels

Pour aborder un problème aussi vaste et central, un problème que Tarski qualifie lui-même comme étant le plus important en philosophie de la connaissance, il importe de préciser ce que nous entendrons de manière préliminaire par « vérité ». À l'époque où Tarski écrit, un bon nombre de conceptions alternatives de la vérité circulent déjà, qu'il s'agisse de la vérité comme correspondance, comme vertu pragmatique, comme cohérence, etc. Selon ses dires, Tarski s'emploiera à formaliser une notion de vérité correspondance, celle qui lui semble la plus primitive et qui remonte à Aristote.[2] Définir le concept de vérité pour Tarski ne consiste pas à exposer l'ensemble du vrai. Par ailleurs, définir le concept de vérité, il suffit seulement de préciser la signification de la locution « p est vrai » (où p est une proposition), ce qui ne permet d'aucune manière de savoir que p est vrai ou pas de manière générale. En gros, le concept de vérité qui sera formalisé possède la forme suivante :

(T) x est vrai si et seulement si p,

[1] Le concept de vérité tarskien fait essentiellement l'objet de *Le concept de vérité dans les langages formels* dans (Tarski 1972 : p. 157-276) ou, en anglais, dans (Tarski 1956 : p. 152-278).

[2] « Dire de ce qui est qu'il est, et dire de ce qui n'est pas qu'il n'est pas, voilà le vrai; dire de ce qui est qu'il n'est pas, et de ce qui n'est pas qu'il est, voilà le faux. » *Métaphysique*, Livre Γ, chapitre VII.

où « x » est un nom de la proposition « p ».

La première section de son article fondateur sur la question de la vérité, *Le concept de vérité dans les langages formalisés*, est consacrée à un résultat négatif concernant le langage ordinaire ou tout autre langage qui renferme son défaut caractéristique. Par sa capacité à exprimer tout ce qui est dicible, le langage ordinaire jouit d'une qualité remarquable : l'universalité. En particulier, le langage ordinaire est en mesure de se décrire lui-même, c'est-à-dire, pour chaque expression de ce langage, il existe une expression qui a pour dénotation cette dernière. Un stratagème (parmi tant d'autres) existant dans le langage ordinaire consiste à former un nom pour une expression en plaçant l'expression entre guillemets : le mot « chien » désigne l'animal, mais le mot « « chien » » désigne le mot qui désigne l'animal. Un langage qui possède cette propriété est dit sémantiquement clos. Or il s'avère que de tels langages souffrent d'un terrible défaut : ils sont contradictoires.

La contradiction vient de la capacité de parler du langage dans le langage et formuler ainsi une variante de l'antinomie du menteur. Tarski offre une construction qui va comme suit : utilisons le symbole « c » pour désigner la proposition

(1) « c » n'est pas vrai.

Par notre définition opératoire du concept de vérité (T), nous avons que

(2) « « c » n'est pas vrai » est vrai si et seulement si « c » n'est pas vrai.

D'autre part, par l'identité entre « c » et (1), nous avons aussi que

(3) « c » est vrai si et seulement si « « c » n'est pas vrai ».

En substituant (2) dans (3), il en résulte la contradiction :

« c » est vrai si et seulement si « c » n'est pas vrai.

Cette construction s'appuie abondamment sur les capacités expressives du langage : nous avons exploité le fait qu'un symbole puisse désigner une proposition dans laquelle sa propre vérité est niée. Le langage naturel permet donc des acrobaties autoréférentielles qui mènent à des contradictions. Un autre exemple de construction autoréférentielle pathologique est la suivante. Soit « $Hm(x)$ » et « $Ht(x)$ » les prédicats définis pour chaque prédicat x (« x » est un nom de prédicat) :

(a) $Hm(x)$ ssi x satisfait le prédicat x (le nom du prédicat satisfait le prédicat),

(b) $Ht(x)$ ssi x ne satisfait pas x (le nom du prédicat ne satisfait pas le prédicat).[3]

(« Hm » est mis pour « homologique » et « Ht » pour « hétérologique ».) Par exemple, nous avons que Hm(« fini »), Hm(« est une description de six mots ») et Hm(« est écrit en français »), et d'autre part que Ht(« infini »), Ht(« animal ») et Ht(« a trois lettres »). Puisque le langage ordinaire contient le nom de prédicat « Ht », la question se pose à savoir si Hm(« Ht ») ou bien Ht(« Ht »). Ces deux cas sont exhaustifs et exclusifs : l'un d'entre eux doit se produire, et ils sont la négation l'un de l'autre. Supposons qu'il s'agisse du premier : si Hm(« Ht »), c'est que « Ht » satisfait le prédicat « $Ht(x)$ », c'est-à-dire que Ht(« Ht ») et donc $\neg Hm$(« Ht »). Il faut donc que nous ayons Ht(« Ht »). Mais dans ce cas, « Ht » satisfait le prédicat « $Ht(x)$ » et il s'ensuit que Hm(« Ht »), c'est-à-dire $\neg Ht$(« Ht »). Dans les deux cas, nous aboutissons à une contradiction.

Cette situation mène Tarski à la conclusion énoncée plus haut à l'effet qu'il ne saurait y avoir un concept de vérité dans le langage naturel compatible avec (T). Il

[3] L'exemple est une adaptation de l'antinomie Grelling-Nelson.

faut donc renoncer au langage ordinaire si nous désirons obtenir des résultats moindrement concluants. Par ailleurs, il faut renoncer à tout autre langage ayant des capacités expressives similaires, un langage suffisamment riche pour contenir une copie de lui-même en lui-même.[4] Un langage pour lequel une définition de la vérité pourra être donnée n'aurait donc pas les ressources expressives propres à un langage sémantiquement clos, il ne pourra pas se décrire lui-même. C'est ainsi qu'il sera nécessaire, afin d'étudier convenablement ce langage, de faire appel à un autre langage, un métalangage, à l'intérieur duquel seront étudiées les propriétés syntaxiques et sémantiques du premier langage. Tarski expose cette contrainte comme suit :

> [C]haque fois que nous étudions le langage d'une science déductive formalisée, nous devons distinguer nettement le langage dont nous parlons du langage que nous parlons, la science, objet de l'étude, de la science étudiant la précédente. Les noms des expressions du premier langage, ainsi que ceux des relations existant entre ces expressions, appartiennent déjà à cet autre langage dit *métalangage* (dont d'ailleurs le premier langage peut être un fragment). La description de ses expressions, la définition de notions plus complexes, surtout des notions liées à la construction d'une science déductive [...], la détermination des propriétés de ses notions rentrent dans la tâche de cette autre science dite *métascience*. (Tarski 1972 : p. 173).

La distinction entre dire et montrer de Wittgenstein est donc reprise ici intégralement avec les précautions de Russell : distinguer le langage *à* l'étude du langage *de* l'étude. Tarski, au contraire de Frege et Russell, pose le métalangage comme étant nécessaire à la réussite de l'entreprise théorique; il lui revient donc d'avoir compris son importance et d'avoir élaboré les propriétés générales de la relation entre langage objet et métalangage. Hilbert a bien sûr formulé une relation semblable, mais nous avons vu qu'il a limité son étude au langage de l'arithmétique et son métalangage à l'arithmétique intuitive. Par ailleurs, le théorème sur le caractère antinomique du

---

[4] Autrement dit, un langage sémantiquement clos.

langage ordinaire a amené Tarski à faire du métalangage un langage formalisé aussi, contrairement à la métamathématique hilbertienne qui est un chapitre de l'arithmétique intuitive.

2. Le métalangage tarskien

Si L est le langage formel à étudier, le métalangage ML qui servira à définir un concept de vérité sur L devra contenir : 1) L ou une traduction L' de L, 2) des noms pour les expressions de L et 3) quelques ressources logiques (aussi peu que possible) comme le calcul propositionnel de même que certains éléments du calcul des prédicats, de la théorie des ensembles et de l'arithmétique. Par exemple, si le langage pour lequel nous voulons donner une définition de la vérité est l'arithmétique élémentaire, c'est-à-dire les nombres naturels munis des opérations d'addition et de multiplication ainsi que les relations d'égalité et d'ordre, le métalangage devra inclure *au moins* l'arithmétique élémentaire (ou une traduction de celle-ci) et un nom pour chaque expression de cette dernière ainsi qu'un certain éventail de théorie logique pour manipuler les noms (dont une relation d'égalité entre expressions et, éventuellement, un principe d'induction pour ces expressions).[5] Suivant la construction tarskienne, le métalangage est nécessairement plus riche que le langage objet.

Comme nous l'avons remarqué plus haut, cette situation contraste avec celle de la métamathématique hilbertienne. Dans celle-ci, les outils métamathématiques sont toujours les mêmes et ils sont considérés plus faibles que l'arithmétique ou l'analyse. Cette disparité s'explique d'abord par la complexité théorique de la notion de vérité. Définir la vérité exige un arsenal plus important que celui nécessaire à une preuve de non contradiction : le métalangage doit non seulement parler du symbole, il

[5] Ces ressources logiques varient d'un contexte à l'autre. Pour plus de détails, voir l'exemple du calcul des classes qui suit à la section 2.2.

doit aussi traiter de sa dénotation. Si Hilbert a réussi à éviter cette difficulté, c'est qu'il a restreint son concept de vérité correspondance aux propositions réelles pour ne laisser qu'une vérité cohérence aux propositions idéelles. Tarski ne s'engage pas dans le débat qui consiste à délimiter l'ensemble des propositions mathématiques signifiantes, il part de l'hypothèse qu'une théorie est signifiante et définit un concept de vérité en conséquence.

2.1. Le calcul des classes

Tarski illustre d'abord son concept de vérité sur le calcul des classes, non pas pour des raisons inhérentes à celui-ci mais uniquement pour des questions de simplicité. Afin de rendre justice aux détails de la thèse tarskienne, qui sont trop souvent approximés dans des expositions générales, arrêtons-nous un instant sur le calcul des classes. Celui-ci est constitué à la base des quatre types de symboles :

1) un répertoire dénombrable de variables : $x_0, x_1, x_2, \ldots$;
2) une relation binaire, l'inclusion : $\subset$;
3) deux opérateurs logiques, la négation et la somme logique : $\neg$ et $\vee$;
4) un quantificateur universel : $\forall$.

Les variables sont les seuls termes de ce langage et elles représentent des classes. L'ensemble des fonctions propositionnelles est défini récursivement comme suit :

a) $x_n \subset x_m$ est une fonction propositionnelle quels que soient n et m;

b) si F est une fonction propositionnelle, alors $\forall x_n F$ est une fonction propositionnelle quel que soit n;

c) si F et G sont des fonctions propositionnelles, alors $\neg F$ et $F \vee G$ en sont également.

Un énoncé de ce langage est une fonction propositionnelle ne possédant aucune variable libre. Le concept de vérité s'appliquera uniquement aux énoncés. Une fois les expressions bien définies, il faut ajouter les axiomes et les règles inférentielles. Je laisse tomber la description détaillée de ces axiomes. Mentionnons seulement qu'ils regroupent les axiomes du calcul propositionnel ainsi que les axiomes spécifiques au calcul des classes. Enfin, les règles inférentielles sont au nombre de quatre, elles sont : la substitution, le détachement (*modus ponens*), l'introduction et l'élimination du quantificateur universel.

On observera qu'un langage formel est introduit à l'aide d'une série de consignes qui peuvent être regroupées sous quatre catégories différentes selon qu'elles servent à l'introduction (1) des règles sur la formation des expressions admissibles dans le langage, (2) des règles précisant la classe des énoncés (les entités qui seront susceptibles d'être vraies), (3) des d'axiomes (les vérités primitives du système), et (4) des règles d'inférence permettant la déduction dans le système. Puisque le langage ordinaire conduit à des antinomies, le métalangage qui servira à définir le prédicat de vérité pour le calcul des classes devra lui aussi être un langage formel. Il faudra donc préciser sa structure suivant chacun de ces quatre points.

2.2. Le métalangage du calcul des classes

Nous savons déjà que le métalangage devra contenir le calcul des classes (ou une traduction) de même qu'un nom pour chaque expression de celui-ci. Ainsi, l'ensemble des expressions admissibles sera considérablement augmenté. D'abord au niveau des termes, en plus des variables de classes, il y a aura des noms pour les expressions du calcul des classes et des variables pour ces noms. De même, en plus de la relation d'inclusion, il y aura une relation d'égalité « = » définie exclusivement pour les termes et une forme élémentaire d'appartenance « $\in$ » (qui servira notamment dans l'axiome d'induction). Pour définir l'ensemble des termes, des fonctions propositionnelles et des énoncés, il faudra le faire à l'aide de définitions récursives,[6] comme celles que nous avons données pour les termes et fonctions propositionnelles du calcul des classes.

Le changement le plus important, du langage objet au métalangage, se situe au niveau des axiomes. L'introduction de nouveaux termes pour décrire le langage objet nous contraint d'introduire de nouveaux axiomes pour traduire le fait qu'ils sont des noms des expressions du langage objet.

Avant de parler des axiomes, un mot sur les noms descriptifs s'impose. Pour que les noms de nos expressions soient utilisables sur le plan descriptif, ils doivent « révéler » les propriétés structurales des expressions qu'ils décrivent. Par exemple, on pourrait employer les noms « Pierre », « Jean » et « Jacques » pour décrire les expressions « $\forall x_2 \neg x_3 \subset x_2$ », « $x_3 \subset x_2$ » et « x_2 ». Toutefois, cette terminologie aurait le fâcheux défaut de ne rien dire sur les expressions constituantes de Pierre, Jean et Jacques et, en particulier, de rester muette sur les relations entre ces expressions, notamment sur le fait que Jacques soit un terme qui figure dans Jean et que Jean soit une sous formule de Pierre. La solution est donc de prendre une terminologie qui épouse davantage la structure des expressions, ce que Tarski nomme des descriptions structurales.[7] Par conséquent, il nous faut des noms descriptifs qui imitent les expressions, c'est-à-dire des noms que nous engendrons par exactement les même définitions récursives qui servent à engendrer les expressions dont ils sont les noms. Les descriptions structurales sont par ailleurs d'autant plus nécessaires étant donnée la nature indécomposable des noms propres. L'utilisation des guillemets (les

[6] Certaines de ces définitions seront seulement « transfiniment » récursives.

[7] Dans un premier un premier temps, dans la première section de l'article, Tarski emploie le terme « description structurale » dans un sens très large. Ce n'est que plus loin qu'il peaufine l'usage de l'expression dans le sens que j'emploie présentement.

noms citationnels) pour générer des noms d'expressions, bien que fort simple et efficace, n'est pas suffisante à la tâche qu'on se donne : « Pierre » est autant partie de « Pierre est l'ami de Jacques » que le mot « rat » est une partie du nom « Socrate »[8] (c'est-à-dire pas du tout).

Revenons donc aux axiomes. Les axiomes du métalangage se divisent en deux groupes. Le premier groupe concerne les axiomes logiques de nature générale, nécessaires à la logique mathématique. Il est donc important d'inclure le calcul des propositions, quelques rudiments de la théorie des ensembles (dont la relation d'appartenance) et certains éléments du calcul des prédicats, voire même de l'arithmétique.[9] (J'omettrai de donner ici même une description de ces axiomes. Je me contenterai seulement d'indiquer que, selon la conception tarskienne du métalangage, il est important que ces axiomes soient exprimés dans le formalisme.) Le deuxième groupe d'axiomes est relatif à la nature spécifique du langage objet à l'étude. Les axiomes de ce groupe se subdivisent à leur tour en deux parties : d'une part, il y a les axiomes du langage objet (qu'il faut inclure dans le métalangage) et d'autre part, il y a les axiomes qui portent sur les noms des expressions du langage objet.

Les expressions et les axiomes du calcul des classes que nous importons dans le métalangage nous étant connus (ou supposés tels), notre tâche principale sera de formuler les axiomes qui concernent la construction des descriptions structurales. La première étape consiste à définir la classe des entités qui nous serviront de nom d'expressions :

(n1) Les symboles « $\overline{\vee}$ », « $\overline{\neg}$ », « $\overline{x_n}$ » (quel que soit n), « $\overline{\forall}$ », et « $\overline{\subset}$ » sont des noms (informellement : « $\overline{\vee}$ » est le nom de $\vee$, « $\overline{\neg}$ » le nom de $\neg$, « $\overline{x_n}$ » de x_n, « $\overline{\forall}$ » de $\forall$ et « $\overline{\subset}$ » de $\subset$);

[8] Pour reprendre un exemple de Quine.

[9] Pour les détails, voir (Tarski 1956 : p. 171).

(n2) Si « α » et « β » sont des noms alors « $\alpha \wedge \beta$ » est un nom aussi (éventuellement, il s'agira de la concaténation des noms « α » et « β »).

Le comportement de la relation d'égalité avec ces expressions est donné par les axiomes suivants :

(é1) Les noms « $\overline{\vee}$ », « $\overline{\neg}$ », « $\overline{x_n}$ » (quel que soit n), « $\overline{\forall}$ », et « $\overline{\subset}$ » sont distincts, c'est-à-dire que l'égalité ne tient pas entre toute paire d'éléments distincts.

(é2) Si « x » et « y » sont des expressions, $x \wedge y$ est distinct des noms primitifs qui figurent dans é1.

(é3) Si « x », « y », « s » et « t » sont des expressions, $x \wedge y = s \wedge t$ si et seulement si 1) $x = s$ et $y = t$ ou 2) il existe une expression « r » telle que $x = s \wedge r$ et $t = r \wedge y$ ou 3) il existe une expression « r » telle que $s = x \wedge r$ et $y = r \wedge t$.[10]

Avec les axiomes (n1), (n2) et (é1) à (e3), nous avons circonscrit la classe des expressions ainsi que la relation d'égalité sur ceux-ci.[11] En fait, pour raisonner par induction sur la classe des expressions, il nous faut un peu plus :

(Ind) Soit X un classe qui satisfait les conditions suivantes : 1) $\overline{\vee}$, $\overline{\neg}$, $\overline{x_n}$, $\overline{\forall}$ et $\overline{\subset} \in X$; 2) si $x \in X$ et $y \in X$ alors $x \wedge y \in X$. Alors toute expression appartient à X.[12]

Il faut remarquer que ce principe d'induction ne s'applique qu'aux expressions et non pas à tous les objets du métalangage.

[10] Il faudrait idéalement que ces axiomes soient rendus dans le formalisme du métalangage.

[11] On remarquera que la présentation du langage est faite de manière récursive. Ceci répond à une exigence d'effectivité à laquelle Tarski se soumet.

[12] La relation d'appartenance est nécessaire ici.

Une fois que la classe des expressions est déterminée, on la subdivise en différentes parties par des définitions récursives pour éventuellement aboutir à la classe des énoncés. Sans entrer trop dans les détails :

1) x est une inclusion de $\overline{x_n}$ dans $\overline{x_m}$: $x = \overline{\subset} \wedge \overline{x_n} \wedge \overline{x_m}$;[13]

2) x est la négation de l'expression y : $x = \overline{\neg} \wedge y$;

3) x est la disjonction de y et z : $x = \overline{\vee} \wedge y \wedge z$;[14]

4) x est une quantification universelle de l'expression y par la variable $\overline{x_n}$: $x = \overline{\forall} \wedge \overline{x_n} \wedge y$; etc.

De cette manière, on peut formuler récursivement la définition d'une fonction propositionnelle :

5) x est une fonction propositionnelle, $FP(x)$: a) il existe n et m avec $x = \overline{\subset} \wedge \overline{x_n} \wedge \overline{x_m}$; ou b) il existe y telle que $FP(y)$ et $x = \overline{\neg} \wedge y$; ou c) il existe y et z telles que $FP(y)$, $FP(z)$ et $x = \overline{\vee} \wedge y \wedge z$; ou d) il existe y avec $FP(y)$ et une variable $\overline{x_n}$ telles que $x = \overline{\forall} \wedge \overline{x_n} \wedge y$.

Une fois que la notion de variable libre est précisée, il est aisé définir ce qu'on entend par énoncé : x est un énoncé si et seulement si x est une fonction propositionnelle et x n'a pas de variables libres, ce que nous représenterons par $x \in S$.

Après avoir formulé structurellement la classe des axiomes (une sous classe de la classe des énoncés), il est possible de préciser récursivement une notion de conséquence au niveau des descriptions qui nous permet en dernier lieu de délimiter

[13] Les descriptions sont écrites en notation polonaise, c'est-à-dire sans parenthèses. Dans cette notation, le symbole d'opération ou de relation est placé devant les arguments.
[14] Même remarque. La disjonction est en notation polonaise.

la classe des énoncés prouvables.[15] Il est donc possible de caractériser structurellement le nom d'un énoncé prouvable, ce qui n'est guère étonnant puisque nous avons essentiellement imité, au niveau de l'écriture, les règles du langage objet. L'utilité de la partie descriptive du métalangage deviendra apparente lorsque nous considérerons la définition de la vérité.

3. Le concept de vérité

La définition de la vérité sera construite à partir de la convention (T) :

x est vrai si et seulement p,

où « p » est une proposition et « x » un nom pour cette proposition. Plus précisément, Tarski exige d'un concept de vérité $V(x)$ qu'il satisfasse les deux conditions suivantes :

a) Quelque soit la description structurale « x » d'une proposition « p », $V(x)$ si et seulement si p;
b) Si $V(x)$ alors x est le nom d'un énoncé (la classe des noms d'expressions vraies est une sous classe des noms d'énoncés).[16]

Pour arriver à définir un concept qui ait pour conséquences a) et b) et qui rencontre certaines contraintes d'effectivités, il faut passer par la notion intermédiaire de satisfaction. La notion de satisfaction est l'ancêtre de la valuation dans l'approche modèle-théorétique moderne. Elle est une manière de parler de validité pour une

[15] Naturellement, pour que la classe des expressions prouvables soit récursive, il faudrait que la classe des axiomes le soit.
[16] Ces deux conditions constituent ce que Tarski appelle la convention T.

fonction propositionnelle ayant des variables libres. Son rôle est de faire en sorte que la définition de la vérité soit *récursive.*[17] Puisque la notion de satisfaction est complexe et risque d'occulter la forme générale du concept de vérité tarskien, commençons par illustrer ce concept pour un cas simple où elle n'est pas nécessaire : le calcul propositionnel.

Supposons d'abord que : 1) le calcul des propositions est composé d'une famille infini de variables propositionnelles $p_0, p_1, p_2, \ldots$ avec deux opérations « $\neg$ » et « $\vee$ »; 2) les noms descriptifs du calcul propositionnel sont $P_0, P_1, P_2, \ldots$ pour les propositions, $\overline{\neg}$ pour la négation et $\overline{\vee}$ pour la disjonction; et 3) les règles sur la concaténation ainsi que pour la formation des noms structuraux sont essentiellement les mêmes que (n1), (n2), (é1), (é2) et (é3) (à isomorphisme près). Ainsi, « $\overline{\vee} \wedge \overline{\neg} \wedge P_3 \wedge P_2$ » serait le nom de la proposition « $(\neg p_3 \vee p_2)$ ». Par ailleurs, supposons que nous ayons défini récursivement un prédicat $P(x)$ tel que x est un nom de proposition si et seulement si $P(x)$. Le concept de vérité tarskien pour le calcul des propositions sera donc

$$\begin{aligned} V(x) \equiv\; & \exists n(x = P_n \wedge p_n) \\ & \vee\; \exists y\big(P(y) \wedge x = \overline{\neg} \wedge y \wedge \neg V(y)\big) \\ & \vee\; \exists y \exists z\big(P(y) \wedge P(z) \wedge x = \overline{\vee} \wedge y \wedge z \wedge \big(V(y) \vee V(z)\big)\big), \end{aligned}$$

c'est-à-dire x est une proposition vraie si et seulement si : 1) il existe un entier n tel que x est la proposition « p_n » et il est le cas que p_n ; ou 2) il existe un nom « y » tel que $x = \overline{\neg} \wedge y$ et y n'est pas vrai; ou 3) il existe des noms « y » et « z » tels que $x = \overline{\vee} \wedge y \wedge z$ et y ou z sont vrais.

[17] La contrainte d'effectivité est beaucoup plus faible ici que chez des finitistes. Comme nous le verrons plus loin, le concept de vérité tarskien devra faire appel à une forme de récursion transfinie.

On remarquera que le concept de vérité pour le calcul des propositions a nécessité un calcul des prédicats ayant des variables et des quantificateurs pour les nombres naturels (la variable dans l'expression ci-dessus est « n »), des variables et des quantificateurs pour les descriptions structurales (les variables « x », « y » et « z ») ainsi qu'une relation d'égalité pour les descriptions. Dans le calcul des classes (d'ordre fini), il faudra (au besoin) ajouter à cette liste des variables et des quantificateurs pour les classes ainsi que des quantificateurs et des variables pour les suites de classes. Les suites de classes joueront un rôle prépondérant dans la notion de satisfaction. Selon le langage objet L, il est possible que nous n'ayons pas à ajouter ces ressources supplémentaires si nous pouvons définir les nombres et les suites de classes dans L ou la traduction de L qui apparaît dans le métalangage. Dans ce cas, il sera moins laborieux de définir l'extension de L qui constituera son métalangage.

Pour traiter de manière uniforme les notions de quantification dans des fonctions propositionnelles, Tarski utilise un stratagème ingénieux qui utilise les suites de classes. À la base, son concept de satisfaction est plutôt simple et se formule ainsi : la classe « a » satisfait la FP « $F(x_0)$ » (d'une seule et unique variable « x_0 ») ssi $F(a)$. Les difficultés apparaissent lorsqu'une fonction propositionnelle possède plus d'une variable libre (ou pas du tout), comme la fonction « $F(x_{i_0}, x_{i_1}, ..., x_{i_n})$ ». Pour remédier à ce problème, les classes ne seront pas prises individuellement mais regroupées en suites. On dira donc que la suite $f = (f_0, f_1, f_2 ...)$ composée des classes $f_0, f_1, f_2 ...$ satisfait « $F(x_{i_0}, x_{i_1}, ..., x_{i_n})$ » ssi $F(f_{i_0}, f_{i_1}, ..., f_{i_n})$. Il reste cependant un détail important à clarifier : nous devons préciser la notion de satisfaction d'une fonction propositionnelle à l'intérieur de laquelle apparaît une (ou plusieurs) variable(s) liée(s) par un (ou des) quantificateur(s). Le cas le plus simple est celui de l'énoncé $\forall x_0 F(x_0)$ où « x_0 » est la seule variable de $F(x_0)$. Tarski propose la définition suivante : f satisfait $\forall x_0 F(x_0)$ si et seulement si, pour tout g avec $f_i = g_i$ sauf possiblement en $i = 0$, g satisfait $F(x_0)$. Avec cette précision, en tenant compte de ce qui a été dit sur le calcul des propositions, nous pouvons enfin formuler le concept de satisfaction : si x est une fonction propositionnelle et si f est une suite propositionnelle alors

$$\begin{aligned} sat(f,x) \equiv\ & \exists n \exists m \left(x = \overline{\subset} \wedge \overline{x_n} \wedge \overline{x_m} \wedge f_n \subset f_m \right) \\ & \vee\ \exists y \left(FP(y) \wedge x = \overline{\neg} \wedge y \wedge \neg sat(f,y) \right) \\ & \vee\ \exists y \exists z \left(FP(y) \wedge FP(z) \wedge x = \overline{\vee} \wedge y \wedge z \wedge \left(sat(f,y) \vee sat(f,z) \right) \right) \\ & \vee\ \exists n \exists y \left(FP(y) \wedge x = \overline{\forall} \wedge \overline{x_n} \wedge y \wedge \forall g \left(\neg \forall m \left(m = n \vee g_m = f_m \right) \vee sat(g,y) \right) \right), \end{aligned}$$

c'est-à-dire que f satisfait x ssi : ou 1) il existe des entiers (non négatifs) n et m tels que x est le nom de $x_n \subset x_m$ et $f_n \subset f_m$; ou 2) il existe un nom de FP « y » tel que $x = \overline{\neg} \wedge y$ et y n'est pas satisfait par f; ou 3) il existe des noms de FPs « y » et « z » tels que $x = \overline{\vee} \wedge y \wedge z$ et f satisfait y ou z; ou 4) il existe un entier n et un nom de FP « y » tel que $x = \overline{\forall} \wedge \overline{x_n} \wedge y$ et pour toute suite g égale à f sauf peut-être à la n-ième place, g satisfait y.

Nous pouvons enfin définir la notion tarskienne de vérité :

$$V(x) \equiv x \in S \wedge \forall f \left(sat(f,x) \right),$$ [18]

c'est-à-dire : x est vrai si et seulement si x est un énoncé et x est satisfait par toutes les suites de classes.[19] L'ensemble des énoncés vrais sera dénoté par Vr. À l'aide des

[18] « f » est une variable de suite de classes. Sinon, si f est une variable comme une autre, il faudrait remplacer la formule dans le champ du quantificateur $\forall f$ par : $SC(f) \supset sat(f, x)$, où $SC(f)$ si et seulement si f est une suite de classes.

[19] Il aura fallu utiliser pour cette définition des variables et des quantificateurs pour les nombres naturels, les classes, les noms et les suites de classes.

axiomes du métalangage, nous pouvons démontrer sans difficulté que le prédicat se comporte adéquatement :

THÉORÈME 1 (Le principe de contradiction). $\forall x\left(x \in S \supset \neg\left(x \in Vr \wedge \overline{\neg}{}^\frown x \in Vr\right)\right)$.

THÉORÈME 2 (Le principe du tiers exclu). $\forall x\left(x \in S \supset \left(x \in Vr \vee \overline{\neg}{}^\frown x \in Vr\right)\right)$.[20]

Une dernière remarque. Il est possible de définir la notion de satisfaction ainsi que la notion de vérité de manière à ce qu'elles soient relatives à un certain domaine, plutôt que générales comme c'est le cas des définitions données ci-haut.

4. Vérité vs prouvabilité

Avec cette définition en main, la comparaison entre la notion formelle de preuve, laquelle est présente depuis les premiers balbutiements de la logique mathématique, et la notion de vérité devient formellement tangible. Nous avons vu que les résultats les plus énigmatiques de la logique mathématique procédaient d'ailleurs du décalage entre ces deux notions.[21] Arrêtons-nous donc pour formuler la notion de preuve dans le métalangage du calcul des classes.

Les méthodes déductives de ce langage se réduisent à 1) la substitution, 2) le *modus ponens* (appelé aussi « détachement »)[22], 3) l'introduction de quantificateur universel, et 4) l'élimination du quantificateur universel. La règle de substitution exige que nous ayons formalisé au préalable la substitution des variables, ce que nous

[20] Tarski 1956 : p. 197.

[21] On pensera au premier théorème d'incomplétude de Gödel, évidemment.

[22] La raison de cette appellation est simple. Si les opérations logiques primitives sont « ~ » et « ∨ », le *modus ponens* s'énonce alors comme : de « *p* » et « *~p∨q* » on peut inférer « *q* ». Ainsi, « *q* » est détachée de « *~p∨q* ».

prendrons pour acquis[23] : $Sub(x,n,m)$ sera la formule obtenue de x en substituant la variable $\overline{x_n}$ pour la variable $\overline{x_m}$. La formalisation des règles (1) à (4) exigera d'autre part la notion intermédiaire de conséquence du n-ième dégré. Si X est un ensemble de fonctions propositionnelles ($x \in X \supset FP(x)$) une conséquence du n-ième dégré de X sera :

$$
\begin{aligned}
x \in Con_n(X) \equiv\ & (n = 0 \wedge x \in X) \\
& \vee (n > 0 \wedge x \in Con_{n-1}(X)) \\
& \vee (n > 0 \wedge \exists y \exists k \exists l (FP(y) \wedge y \in Con_{n-1}(X) \wedge x = Sub(y,k,l))) \\
& \vee \left(n > 0 \wedge \exists y \exists z \left(FP(y) \wedge FP(z) \wedge \overline{\neg} {}^\wedge y \in Con_{n-1}(X) \wedge \overline{\vee} {}^\wedge \overline{\neg} {}^\wedge y {}^\wedge x \in Con_{n-1}(X)\right)\right) \\
& \vee \left(n > 0 \wedge \exists y \exists k \left(FP(y) \wedge y \in Con_{n-1}(X) \wedge x = \overline{\forall} {}^\wedge \overline{x_k} {}^\wedge y\right)\right) \\
& \vee \left(n > 0 \wedge \exists y \exists k \exists l \left(FP(y) \wedge \overline{\forall} {}^\wedge \overline{x_k} {}^\wedge y \in Con_{n-1}(X) \wedge x = Sub(y,k,l)\right)\right)
\end{aligned}
$$

Chaque terme de la disjonction traite un cas : a) tous les éléments de X sont des conséquences; b) les conséquence du $n-1$-ième degré sont aussi des conséquences du n-ième degré; c) une formule obtenue d'une conséquence par substitution est une conséquence (règle 1); d) une formule obtenue de conséquences par détachement est une conséquence (règle 2); e) une formule obtenue d'une conséquence par généralisation est une conséquence (règle 3); et f) une formule obtenue d'une conséquence par élimination d'un quantificateur est une conséquence (règle 4). Ainsi, x est une conséquence logique des formules propositionnelles dans X, $x \in Con(X)$, si et seulement s'il existe un entier naturel n tel que x est la n-ième conséquence de X, $\exists n(x \in Con_n(X))$. Si X est vide, on dira que x est démontrable.

[23] Bien qu'elle soit simple, la notion de substitution pose certaines difficultés techniques qui ne sont d'aucun intérêt pour nous. Pour les détails de la définition, voir (Tarski 1956 : p. 180).

Puisque la vérité est un prédicat s'appliquant seulement aux énoncés, restreignons-nous à ne considérer que les énoncés prouvables du calcul des classes, les x tels que $x \in S \wedge x \in Con(\varnothing)$. L'ensemble des énoncés démontrables sera nommé $Dém$. Dans ce contexte, le théorème de Gödel se traduit par un décalage entre les ensembles Vr et $Dém$:

THÉORÈME 3. Si $X \subseteq Vr$ alors $Con(X) \subseteq Vr$. En particulier, $Con(Vr) \subseteq Vr$.

THÉORÈME 4. Vr est un ensemble non contradictoire et déductivement clos.

THÉORÈME 5. Tous les énoncés démontrables sont vrais : $Dém \subseteq Vr$.

THÉORÈME 6. Il existe des énoncés vrais indémontrables : $Dém \neq Vr$.[24]

Le dernier théorème est la conséquence d'un lemme stipulant qu'un certain énoncé et sa négation, en l'occurrence l'énoncé « $\overline{\forall} ^\wedge \overline{x_1} ^\wedge \overline{\forall} ^\wedge \overline{x_2} ^\wedge \overline{\subset} ^\wedge \overline{x_1} ^\wedge \overline{x_2}$ » et sa négation « $\overline{\neg} ^\wedge \overline{\forall} ^\wedge \overline{x_1} ^\wedge \overline{\forall} ^\wedge \overline{x_2} ^\wedge \overline{\subset} ^\wedge \overline{x_1} ^\wedge \overline{x_2}$ », n'appartiennent ni l'un ni l'autre à $Dém$.

Comme nous avons vu au dernier chapitre, la pathologie décrite par le théorème 6 n'est pas particulière au calcul élémentaire des classes,[25] elle affecte également le calcul des classes d'ordre infini, une théorie qui diffère du calcul élémentaire des classes seulement par la présence de variables et de quantificateurs d'ordres quelconques.[26] Tarski démontre l'analogue des théorèmes de Gödel avec son concept de vérité :

(I) Si le langage objet *L* avec métalangage *ML* a les ressources nécessaires pour définir une interprétation de *ML* dans lui-même, il est impossible de définir un prédicat de vérité pour *L* dans *ML*;

---

[24] Ces théorèmes sont des conséquences plus ou moins directes du métalangage, sauf pour le dernier qui est un peu plus technique. Voir (Tarski 1956 : p. 198-9).

[25] Dans le cas du calcul élémentaire des classes, il est possible de modifier le système d'axiomes pour qu'il soit complet. Ce qui ne sera pas possible en général.

[26] Le calcul des classes d'ordre infini est essentiellement le système de base des *Principia*.

(II) S'il est possible de définir un prédicat de vérité de L dans ML et si le prédicat de démontrabilité *Dém* est interprétable dans L, il existe un énoncé vrai mais qui est indémontrable.

La démonstration de ces méta-métathéorèmes repose sur l'existence d'une correspondance biunivoque entre les expressions du langage L et les nombres naturels de sorte qu'une proposition portant sur des expressions de L (qui sont typiquement les propositions du métalangage) devient une proposition portant sur des nombres. Dans les hypothèses de (I), on suppose que toutes les propositions portant sur des expressions de L (donc toute les propositions de ML) sont traduisibles dans L; dans les hypothèses de (II), on suppose seulement que la notion de démontrabilité est traduisible dans L.[27] Pour les fins de la preuve, la correspondance entre les expressions de L et les nombres naturels sera donnée par la suite $\phi = (\phi_1, \phi_2, \phi_3, \ldots)$, c'est-à-dire que ϕ_n sera corrélé au nombre n. Étant donné qu'il est possible de définir l'arithmétique dans le calcul des classes d'ordre infini, ceci revient à dire que les énoncés de L sont corrélés à des classes.

La preuve de (I) est essentiellement une adaptation de l'antinomie du menteur et procède comme suit.[28] Supposons qu'un prédicat de vérité pour L soit définissable dans ML. Pour chaque nombre naturel x (x étant une variable de classe dans le calcul des classes), l'expression « $\phi_x \notin Vr$ » est donc une expression du métalangage. Par les hypothèses de (I), « $\phi_x \notin Vr$ » est traduisible dans L par une certaine expression arithmétique $\psi(x)$ pour chaque (nombre naturel) x de sorte que $\phi_x \notin Vr \equiv \psi(x)$ (*).[29] Par ailleurs, puisque $\psi(x)$ est une expression du calcul des classes, il existe un

[27] On supposera également qu'un prédicat de vérité est définissable.
[28] Il va sans dire que la preuve donnée ici n'est seulement qu'une esquisse. Pour qu'elle soit complète, il faudrait spécifier la nature exacte de la correspondance et formaliser les notions du méta-métalangage qui lui sont nécessaires.
[29] Pour arriver à (*), il faut un peu plus que l'hypothèse que V est traduisible dans L. Il faut que ϕ soit telle que $\phi_x \notin V$ est traduisible dans L.

nombre k tel que ϕ_k est le nom de $\psi(x)$. En remplaçant x par k dans l'équivalence (*), nous obtenons $\phi_k \notin Vr \equiv \psi(k)$ (**). Mais « ϕ_k » est le nom d'une proposition équivalente à $\psi(k)$, par la convention (T), nous avons

$$\phi_k \in Vr \equiv \psi(k).$$

Ce qui est en contradiction avec (**). Ainsi, si le langage objet L d'un métalangage ML est suffisamment puissant pour exprimer (une interprétation de) ML, ML ne peut définir un prédicat de vérité pour L sans être contradictoire.

L'argument peut être modifié sans problèmes pour montrer (II) en remplaçant Vr par $Dém$.[30] En effet, soit l'expression $\phi_x \notin Dém$. Selon les hypothèses de (II), il est possible de traduire $\phi_x \notin Dém$ par une expression numérique $\theta(x)$ de sorte que $\phi_x \notin Dém \equiv \theta(x)$ (#). Par la correspondance ϕ et le fait que $\theta(x)$ soit une expression du calcul des classes, il existe l tel que ϕ_l est de nom de $\theta(x)$. En substituant x par l dans (#), nous obtenons $\phi_l \notin Dém \equiv \theta(l)$ (##).[31] Puisqu'il existe un prédicat de vérité pour ce langage,[32] nous avons que $\phi_l \in Vr \equiv \theta(l)$ et donc $\phi_l \notin Dém \equiv \phi_l \in Vr$ par (##). Or $Dém \subseteq Vr$ dans ce langage (par le théorème 5 de cette section), donc l'énoncé ϕ_l est vrai mais indémontrable (ϕ_l est également indécidable parce que $\overline{\neg}\,{}^\wedge \phi_l \notin Dém$).

Ces deux résultats exposent des conséquences importantes pour la notion de métalangage. Une façon d'interpréter (I) serait de dire que la métathéorie doit être *essentiellement* plus riche que sa théorie objet pour éviter la contradiction. Mais

[30] (II) est un équivalent du premier théorème d'incomplétude de Gödel.

[31] La même remarque qu'à (*) s'applique ici : il faut que ϕ soit telle que $\phi_x \notin Dém$ est traduisible dans L.

[32] Pour éviter que ML soit contradictoire par (I), on supposera que ML est suffisamment riche pour que le prédicat Vr ne soit pas traduisible dans L.

même si le métalangage est plus riche que le langage objet, assez riche pour ne pas être intégralement reproduit dans le langage objet, si le prédicat de démontrabilité est traduisible dans le langage objet, il existe selon (II) un énoncé vrai qui est indémontrable. Malheureusement pour la métamathématique, le prédicat de démontrabilité est traduisible dans le langage objet lorsque celui-ci contient l'arithmétique.

5. Exposition moderne du concept de vérité

Puisque nous avons déjà introduit les conceptions syntaxiques modernes pour un langage du premier ordre au chapitre V, profitons-en pour introduire les conceptions sémantiques modernes pour ce même type de langage. Le lecteur constatera que la définition moderne du concept de vérité suit de très près la définition tarskienne donnée plus haut. Je donne la définition générale de même qu'un exemple détaillée de ce prédicat pour le langage S (chapitre V) :

(I) Une *interprétation* M d'un langage $\mathcal{L}$ (du premier ordre) est la donnée d'un ensemble D (un domaine) et d'une application qui associe à chaque symbole de prédicat A_k^n une relation $(A_k^n)^M \subset D^n$, à chaque symbole de fonction f_k^n (s'il y en a) une opération $(f_k^n)^M : D^n \to D$ et à chaque symbole de constante a_n (s'il y en a) un élément $(a_n)^M \in D$.

L'interprétation $\mathcal{N}$ de S possède l'ensemble $\mathbb{N}$ des nombres naturels comme domaine; associe aux symbole de constante « 0 » le nombre 0; associe aux symboles de fonctions « ' », « + » et « · » la fonction successeur, l'addition et la multiplication respectivement; et associe au symbole d'égalité « = » la relation d'égalité. Puisque notre métalangage est cadré dans la théorie des ensembles, remarquons qu'il est possible de donner une caractérisation de ce modèle dans des termes purement ensemblistes ($\mathbb{N}$ serait $\bigcap\{x \mid \varnothing \in x \wedge \forall y(y \in x \Rightarrow y \cup \{y\} \in x)\}$, 0 serait $\varnothing$, la fonction successeur serait $x \mapsto x \cup \{x\}$, etc.).

(II) Il faut ensuite donner une notion de *satisfaction*. De manière générale, soit Σ l'ensemble de toutes les suites dénombrables d'éléments dans le domaine D d'une interprétation de $\mathcal{L}$ et soit $s = (s_1, s_2, \ldots) \in \Sigma$. On définit une fonction $s(t)$ pour chaque terme t de $\mathcal{L}$ de la manière suivante :

(s1) $s(x_i) = s_i$;

(s2) $s(a_i) = (a_i)^M$;

(s3) $s(f_k^n(t_1, \ldots, t_n)) = (f_k^n)^M(s(t_1), \ldots, s(t_n))$.

On dira que s est i-différente de s' si pour tout $j \neq i$ $s_j = s'_j$. Dans l'interprétation $\mathcal{N}$, pour (s2) il suffit de mentionner que $s(0) = \varnothing$, et pour (s3) que $s(t_1') = s(t_1) \cup \{s(t_1)\}$, que $s(t_1 + t_2)$ vaut la somme de $s(t_1)$ et $s(t_2)$, et que $s(t_1 \cdot t_2)$ vaut le produit de $s(t_1)$ et $s(t_2)$. La satisfaction d'une formule par $s = (s_1, s_2, \ldots)$, notée $\vDash_s$, est ensuite définie comme :

(s4) $\vDash_s A_k^n(t_1, t_2, \ldots, t_n)$ ssi $(A_k^n)^M(s(t_1), \ldots, s(t_n))$, c'est-à-dire ssi $(s(t_1), \ldots, s(t_n)) \in (A_k^n)^M$;

(s5) $\vDash_s \neg\mathcal{A}$ ssi $\nvDash_s \mathcal{A}$;

(s6) $\nvDash_s (\mathcal{A} \Rightarrow \mathcal{B})$ ssi s satisfait $\vDash_s \mathcal{A}$ mais $\nvDash_s \mathcal{B}$;

(s7) $\vDash_s \forall x_i \mathcal{A}$ ssi $\vDash_{s'} \mathcal{A}$, pour tout $s' \in \Sigma$ i-différente de s.

Pour la théorie *S*, seule (s4) est à spécifier : $\models_s t_1 = t_2$ ssi $s(t_1) = s(t_2)$, où on remarquera que dans la deuxième expression « = » dénote la relation d'égalité ensembliste.

(III) La notion de vérité en découle immédiatement : une formule $\mathcal{A}$ est *vraie* dans l'interprétation *M*, noté $\models_M \mathcal{A}$, si et seulement si pour toute suite $s \in \Sigma$, $\models_s \mathcal{A}$. $\mathcal{A}$ est *fausse* dans l'interprétation *M*, noté $\not\models_M \mathcal{A}$, ssi pour toute suite *s*, $\not\models_s \mathcal{A}$.

Nous avons donc notre concept de vérité pour *S*, le prédicat « $\models_{\mathcal{N}}$ ».

Les deux théorèmes de Tarski de la section précédente se démontrent aisément à l'aide du théorème du point fixe. Si *K* est une théorie, T_K sera l'ensemble des nombres de Gödel de formules vraies. Le théorème (I) se traduit alors par :

THÉORÈME IA. Soit *K* une théorie qui a les mêmes symboles que *S* et à l'intérieur de la laquelle la fonction diagonale est représentable.[33] Si *K* est cohérente, alors la propriété $x \in T_K$ n'est pas représentable dans *K*.

PREUVE. Supposons le contraire. Soit $\mathcal{T}(x_1)$ une formule qui représente $x \in T_K$ dans *K*, c'est-à-dire que pour tout *n* :

(i) $n \in T_K$ entraîne $\vdash_K \mathcal{T}(\overline{n})$;

(ii) $n \notin T_K$ entraîne $\vdash_K \neg\mathcal{T}(\overline{n})$.

Appliquons le théorème du point fixe à la formule $\neg\mathcal{T}(x_1)$. Nous obtenons une formule $\mathcal{B}$ telle que $\vdash_K \mathcal{B} \Leftrightarrow \neg\mathcal{T}(\ulcorner\mathcal{B}\urcorner)$ (*). Il y a deux cas à considérer. (1) Si $\vdash_K \mathcal{B}$

[33] Ces hypothèses sont nécessaires pour que le théorème du point fixe soit applicable. *S* remplit ces conditions.

alors $b \in T_K$ où $b = \Gamma(\mathcal{B})$. Par (i), nous avons $\vdash_K \mathcal{T}(\overline{b})$. Mais $\vdash_K \mathcal{B}$ entraîne par (*) que $\vdash_K \neg\mathcal{T}(\ulcorner\mathcal{B}\urcorner)$, c'est-à-dire que $\vdash_K \neg\mathcal{T}(\overline{b})$, ce qui contredit cohérence de *K*. (2) Si $\nvdash_K \mathcal{B}$ alors $b \notin T_K$. Par (ii), nous avons $\vdash_K \neg\mathcal{T}(\overline{b})$. Par (*), nous avons donc que $\vdash_K \mathcal{B}$, contredisant l'hypothèse $\nvdash_K \mathcal{B}$.

Le second théorème est un corollaire du premier théorème d'incomplétude. Soit D_S l'ensemble des nombres de Gödel des formules démontrables de *S*. Nous avons :

THÉORÈME IIA. Si *S* est cohérente, $D_S \subsetneq T_S$.

PREUVE. Dans un premier temps, il est facile de démontrer que $D_S \subset T_S$. Pour montrer l'inégalité, il suffit d'appliquer le premier théorème d'incomplétude pour obtenir une formule close $\mathcal{G}$ telle que $\nvdash_S \mathcal{G}$ et $\nvdash_S \neg\mathcal{G}$. Soit *g* le nombre de Gödel de $\mathcal{G}$. Nous avons alors que

(i) $g \notin D_S$ et $N\acute{e}g(g) \notin D_S$;

(ii) $g \in T_S$ ou bien $N\acute{e}g(g) \in T_S$.

Ce qui implique que $D_S \neq T_S$. (Pour les curieux, c'est $g \in T_S$.)

6. Tarski et la philosophie

Reprenons du début. Tarski définit un prédicat de vérité qui s'inspire de la vérité correspondance, un prédicat qui s'applique au nom d'une proposition (et non pas à la proposition elle-même). Selon les critères tarskiens formulés dans la convention T, une définition de la vérité doit avoir pour conséquences les propositions suivantes : « x » est vrai si et seulement si p, où « x » est un nom de la proposition « p ». Ces équivalences font en sorte que la définition est *matériellement* correcte. Une définition de la vérité dans le langage ordinaire est impossible parce que celui-ci est sémantiquement clos. Trois conséquences s'ensuivent : i) un prédicat de vérité n'est définissable que pour les langages formels à expressivité limitée; ii) puisque le langage formel ne saurait contenir les noms de ses propres expressions, il faut faire appel à un autre langage, en l'occurrence un métalangage, qui contient les descriptions structurales de ce langage; et iii) le métalangage est forcément un langage formel (à expressivité limitée) sans quoi on risque de tomber dans les mêmes antinomies sémantiques. Une définition de la vérité est *formellement* correcte si elle est donnée seulement *pour* un langage formel (langage objet) *dans* un langage formel (métalangage) en respectant certaines conditions finitistes. Il est possible de construire une définition de la vérité pour le calcul des classes (d'ordre fini) qui est formellement et matériellement correcte. Le métalangage nécessaire à cette définition comprend : 1) le calcul des classes, 2) des descriptions structurales pour le calcul des classes, et 3) des ressources logiques supplémentaires pour définir le prédicat de vérité. Dans ce cas précis, le métalangage est plus riche que le langage objet. Nous avons ensuite considéré la possibilité qu'un langage contenant l'arithmétique puisse interpréter en totalité ou en partie son métalangage, ce qui nous a conduit aux méta-métathéorèmes (I) et (II). Enfin, retenons que la distinction entre langage et métalangage que Tarski développe est une notion relative. Contrairement à Hilbert, il ne considère pas la métathéorie une théorie fixe et invariable. Un langage est un métalangage relatif à autre et non pas de manière absolue, il pourrait lui-même être langage objet.[34]

[34] Tarski 1944 : p. 48.

Parmi les conséquences philosophiques des considérations de Tarski sur le métalangage et la notion de vérité – tirées par Tarski lui-même et entretenues par d'autres par la suite – nous retrouvons notamment l'idée que le métalangage est essentiellement plus riche que le langage objet, du moins lorsque celui-ci se trouve à définir un prédicat de vérité pour le langage. Puisque le métalangage est plus riche que le langage objet, il nécessite un métalangage pour avoir lui-même un concept de vérité, ce qui nous amène à l'autre idée communément répandue sur la notion de métalangage : le fait que le métalangage présuppose un méta-métalangage et que le méta-métalangage présuppose un méta-méta-métalangage, et ainsi de suite. Sur le plan philosophique, il semblerait donc que s'aventurer dans la voie du métalangage implique une forme de régression à l'infini. Dans *Tarski On "Essentially Richer" Metalanguages*, Dave DeVidi et Graham Solomon réagissent à cette affirmation et contestent les conséquences que semblent en tirer Tarski et sa postérité philosophique.[35]

DeVidi et Solomon précisent d'abord que les métalangages ne viennent pas en famille, qu'un métalangage est nécessaire lorsque nous voulons étudier un autre langage, qu'il n'a pas à être défini sinon. Mais l'objectif de DeVidi et Solomon est surtout de montrer que la condition qu'un métalangage soit essentiellement plus riche que son métalangage n'est pas nécessaire du tout, que Tarski n'a pas de justification pour cette affirmation. En fait, Tarski n'a même pas précisé ce qu'il entendait par un langage essentiellement plus fort qu'un autre. Il est clair que si nous construisons un métalangage suivant les indications données plus haut, le métalangage est plus fort que le langage objet dans un certain sens. Mais dans un autre sens, il se peut très bien que le métalangage ne soit au bout du compte qu'une extension conservatrice du langage objet.[36] Si le métalangage de la construction de Tarski est une extension

[35] DeVidi & Solomon 1997.

[36] Un théorie T_1 exprimée dans un langage L_1 est une extension conservatrice d'une théorie T exprimée dans le langage L si : 1) le langage L_1 contient tout le vocabulaire de L, et 2) la classe des théorèmes de T_1 qui s'écrivent avec le vocabulaire L seulement est précisément la classe de théorèmes de T.

conservatrice du langage objet, les seuls éléments qui contribuent au fait qu'il soit essentiellement plus riche sont les termes structurellement descriptifs et les quelques axiomes qui les régissent. Ainsi, l'extension n'est pas très généreuse.

Une des affirmations sur lesquelles DeVidi et Solomon reviennent est qu'une interprétation du métalangage dans le langage objet existe du moment que le langage objet est aussi « riche » que le métalangage. Les théorèmes (I) et (II) exposent les conséquences d'une telle situation. Mais comme j'ai fait remarquer en bas de page lors de leur démonstration, ces théorèmes reposent sur une hypothèse plus forte que celle de la représentabilité du métalangage dans le langage objet seulement. Il faut que la corrélation donnée entre le métalangage et le langage objet (décrite par ϕ dans la preuve) soit elle-même donnée par le métalangage, pour ensuite appliquer la corrélation à elle-même et obtenir les résultats souhaités. DeVidi et Solomon exposent des résultats intéressants en ce sens, dont celui de Gupta (1982) qui établit qu'un certain langage classique non contradictoire peut formuler en lui-même sa propre définition de vérité. Le point important de la définition est que la corrélation code juste assez pour représenter ce langage sans pour autant coder le codage.[37] Ainsi, dans ce cas, le langage serait son propre métalangage. Pire encore (pour la thèse tarskienne bien entendu), il existe des théories pour lesquelles un concept de vérité peut être défini dans une métathéorie plus faible que la théorie elle-même. Notamment, un concept de vérité pour la théorie des ensembles de Zermelo peut être défini dans une extension P de la théorie des ensembles finis.[38] Ces résultats nous invitent donc à considérer les affirmations de Tarski avec prudence.

---

[37] Dans cette discussion, la définition de la vérité peut prendre des aspects différents que celle de Tarski. DeVidi et Solomon se permettent cet écart.

[38] Reste à voir si ces résultats sont démontrés dans le métalangage lui-même ou dans un méta-métalangage de ces deux théories. Dans ce le dernier cas, les conclusions de DeVidi et Solomon me sembleraient non fondées.

CONCLUSION

La condition métathéorique

Le portrait du métalangage tel qu'il s'est dessiné au cours de cette analyse pourrait être résumé de la manière suivante : si $\mathcal{S}$ est le système que nous souhaitons étudier, la métathéorie de $\mathcal{S}$ est essentiellement la donnée d'un couple $(\Delta_{\mathcal{S}}, \mathcal{T})$, où $\Delta_{\mathcal{S}}$ est une description de $\mathcal{S}$, c'est-à-dire un ensemble de noms structurellement générés pour la théorie $\mathcal{S}$ accompagné d'une relation d'égalité et d'un principe d'induction (s'appliquant aux noms seulement), et où $\mathcal{T}$ est une théorie dont la nature varie en fonction de l'étude effectuée sur $\mathcal{S}$. La métamathématique, dans sa forme la plus élémentaire, exigerait seulement $\Delta_{\mathcal{S}}$, c'est-à-dire que $\mathcal{T}$ pourrait être vide. Une version plus forte ajouterait à $\mathcal{T}$ quelques éléments de la théorie des ensembles. De son côté, Gödel exige que $\mathcal{T}$ contienne (au moins) l'arithmétique récursive. Celui de Tarski demande que $\mathcal{T}$ contienne suffisamment de notions ensemblistes de même qu'une copie de $\mathcal{S}$ ou, à tout le moins, une théorie qui lui est isomorphe. Enfin, la démonstration de cohérence de Gentzen pour l'arithmétique demande que $\mathcal{T}$ dispose de l'induction transfinie. Il n'y a donc pas lieu de parler d'*une* métathéorie pour $\mathcal{S}$, mais de plusieurs.

Aujourd'hui, la théorie de prédilection $\mathcal{T}$, telle qu'elle se manifeste dans l'usage, est la théorie des ensembles de Zermelo-Fraenkel. On n'a qu'à ouvrir un manuel standard en logique mathématique pour constater l'omniprésence de cette théorie comme « médium » métathéorique, la seule exception à ma connaissance étant *Proof Theory and Logical Complexity* de Jean-Yves Girard. Pourtant, le métalangage du calcul des classes que nous avons défini au chapitre précédent ne s'appuie pas au niveau métathéorique sur la théorie des ensembles mais sur le calcul des classes lui-même. Bien entendu, n'importe quelle théorie isomorphe au calcul des classes peut être employé pour ce faire, et la théorie des ensembles contient très certainement une partie qui lui est isomorphe. Quoi qu'il en soit, ce que je veux faire remarquer ici est que la définition moderne d'un modèle et de la vérité dans ce modèle, telle qu'elle est exposée à la section cinq du précédent chapitre, ne laisse pas la possibilité à d'autres théories : un modèle ne peut être qu'une certaine structure ensembliste. Ce virage ensembliste n'est pas étonnant après tout, considérant l'énorme succès qu'a connu la théorie des ensembles au vingtième siècle comme langage mathématique « universel ». Mais la conséquence en est que le métalangage est maintenant invariablement associé à la théorie des ensembles, une association qui est loin d'être nécessaire. Un métalangage peut épouser d'autres tendances théoriques, les exemples donnés plus haut en témoignent.

Si une théorie admet plusieurs métathéories, la question se pose à savoir si les métathéories s'entendront toujours sur les propriétés de la théorie objet. Naturellement, la réponse est non. Nous savons déjà que les métalangages arithmétiques de Gödel et de Gentzen diffèrent par au moins un théorème : la cohérence de l'arithmétique. Il est toutefois inutile d'aller chercher des exemples aussi complexes pour mettre en évidence cette « relativité métathéorique ». Prenons le calcul propositionnel $\mathcal{P}$ et les deux métathéories $(\Delta_{\mathcal{P}}, \mathcal{T}_1)$ et $(\Delta_{\mathcal{P}}, \mathcal{T}_2)$, où $\mathcal{T}_1$ est le calcul des prédicats classique et $\mathcal{T}_2$ le calcul des prédicats intuitionniste. De toute évidence, l'énoncé métathéorique « pour toute proposition P, $\vdash_{\mathcal{P}} P$ ou $\nvdash_{\mathcal{P}} P$ » est un théorème de $(\Delta_{\mathcal{P}}, \mathcal{T}_1)$ mais pas de $(\Delta_{\mathcal{P}}, \mathcal{T}_2)$, à cause de la validité et de la non validité du tiers exclu dans $\mathcal{T}_1$ et $\mathcal{T}_2$ respectivement. Pour prendre un autre exemple dans la lignée de la théorie des ensembles, on pourrait se demander si la démonstration du lemme de Lindenbaum dépend de l'axiome de choix.[1]

Le fait que les propriétés d'un langage soient relatives à une certaine métathéorie compromet-il l'analyse métathéorique dans son ensemble? L'analyse métathéorique a-t-elle une valeur, ou est-ce seulement un exercice formel vain? Pour répondre à ces questions, il serait peut-être utile de reprendre rapidement les étapes qui nous ont amené cette forme d'étude. D'abord, il eut la transition du langage ordinaire au langage formel. Ensuite, il eut le constat de l'insuffisance de l'intuition et du raisonnement philosophique *a priori* pour garantir que le langage formel possédait certains attributs fondamentaux. Ici, nous avions deux choix : être stoïques (c'est-à-dire suspendre notre jugement)[2] ou continuer à connaître. Si nous choisissons de connaître, l'insuffisance de l'intuition a pour conséquence la mise en valeur de la métathéorie. Mais après l'intuition, c'était au tour de la métathéorie de connaître certaines limitations. Devant ces limitations, une certaine expansion du métalangage s'imposa et, par conséquent, une pluralité au niveau métathéorique. Était-ce une erreur de s'obstiner à connaître, aurait-il fallu suivre la prescription stoïque? La question soulève immédiatement les accusations de circularité ou de régression à l'infini qui sont souvent formulées à l'endroit des approches métathéoriques. Ainsi, la valeur de l'approche métathéorique dépend en quelque sorte de sa capacité à répondre à ces critiques.

---

[1] La démonstration du lemme de Lindenbaum utilise l'axiome du choix, mais je ne sais pas s'il en dépend intrinsèquement.

[2] C'est la prescription de Wittgenstein dans le *Tractatus*.

Attardons nous d'abord sur le caractère spécifique de la circularité et de la régression à l'infini telles que nous les retrouvons dans la métathéorie. Normalement, un raisonnement qui cherche à établir $\mathcal{A}$ est dit circulaire s'il présuppose $\mathcal{A}$. Mais la circularité métathéorique n'est pas de cet ordre. Poincaré accuse Hilbert d'être circulaire parce que la métamathématique utilise le principe d'induction comme *méthode* pour démontrer la validité du principe d'induction comme *énoncé*. La même remarque s'impose pour la régression à l'infini. Il y a donc une distinction importante à faire ici entre l'usage et la mention : le fait d'utiliser un principe au niveau métathéorique n'entraîne pas de manière immédiate que l'énoncé de ce principe dans le langage objet sera démontrable.[3] Un contre exemple suffit pour nous en convaincre. Le fait d'étudier le calcul des prédicats intuitionniste avec un métalangage classique ne rendra pas le tiers exclu démontrable dans ce calcul. Nous pouvons également citer le second théorème de Gödel : la cohérence du langage objet est supposée au niveau du métalangage mais elle reste néanmoins indémontrable dans le langage objet.

Cette nuance est importante, mais elle ne règle pas toute la question de la circularité. La critique de Poincaré demeure. L'ambition de Hilbert était de démontrer que l'arithmétique était non contradictoire, et pour ce faire il a construit une arithmétique formelle sur laquelle la raison métamathématique pouvait opérer. Abstraction faite des arguments kantiens de Hilbert, sur le plan purement logique, la métathéorie hilbertienne se donne l'arithmétique pour démontrer que l'arithmétique est cohérente. Si l'objectif final est éventuellement de déduire de la cohérence de l'arithmétique formelle que l'arithmétique informelle est « saine », il y a effectivement un cercle car l'arithmétique informelle est considérée saine au départ. À la question de savoir ce que la méthode métathéorique vaut, nous pouvons donc donnée une réponse provisoire. Rien, si l'objectif est de montrer de manière métathéorique et anhypothétique qu'une théorie dans son ensemble est valide. La

[3] Cette observation a d'abord été faite par Husserl au §19 de la troisième recherche dans *Recherches* (Husserl 1970).

validité d'un théorème affirmant la cohérence d'une théorie ne sera jamais supérieure à celle des méthodes qui ont permis de le démontrer. Une preuve de cohérence a donc une valeur éminemment relative. Cette remarque vaut autant pour une preuve de cohérence par modèle (exhiber une structure qui satisfait un système) que les preuves par démonstration (démontrer dans une métathéorie l'impossibilité d'obtenir une contradiction); dans le premier cas, c'est le modèle qui est garant de la cohérence de la théorie et dans le second, ce sont les méthodes déductives utilisées dans la démonstration. À l'inverse, l'incohérence est une chose qui se laisse démontrer directement dans la théorie, et sans la moindre métathéorie. En ce sens, Russell avait raison de dire que la validité d'un système était une affaire essentiellement inductive : la validité de l'arithmétique dans un sens absolu tient à ce qu'elle n'a pas été démentie à ce jour.

Cette situation n'empêche pas toutefois que l'analyse métathéorique soit fertile lorsque l'objectif est autre que de démontrer ce qui est déjà explicitement supposée. Si la cohérence de la théorie est déjà supposée de toute manière, aussi bien s'en servir. Si le fait de déduire des conséquences de cette hypothèse est considéré comme étant circulaire, aussi bien considérer l'ensemble des mathématiques comme circulaires, car en mathématiques non plus on ne trouvera jamais de conclusions plus fortes que les hypothèses utilisées. Syntaxiquement parlant, il est vrai que l'activité mathématique se réduit à une immense tautologie, puisque tous les théorèmes d'une théorie sont fixés dès que le système d'axiomes est donné. Mais pour les êtres finis et limités que nous sommes, des êtres incapables de percevoir spontanément et immédiatement les conséquences les plus reculées d'une famille hypothèses, la déduction a pour rôle d'exposer ces conséquences inattendues. Ainsi, l'étude métathéorique peut jouer exactement ce rôle : exposer les propriétés d'une théorie, des propriétés qui pour plusieurs seront inattendues.

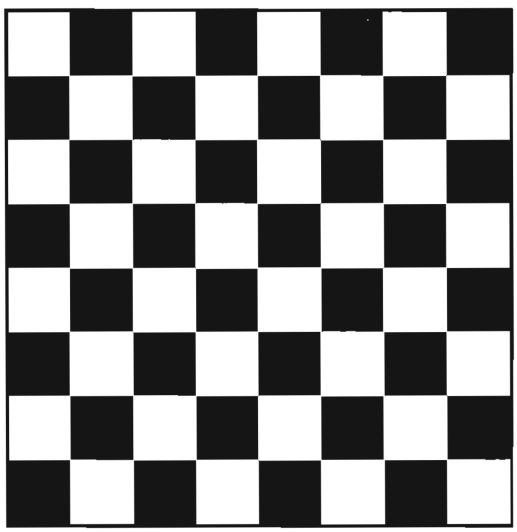
Une application : l'incomplétude revue

Appliquons maintenant notre perspective métathéorique au problème de la complétude et illustrons les retombées positives d'une telle approche. Partons de la question suivante : pourquoi est-il décevant d'apprendre qu'une théorie est incomplète (dans le sens du premier théorème)? Autrement dit, pour quelle raison espérions-nous qu'une théorie soit complète? Bien que la complétude soit définie de manière purement syntaxique, aucune explication syntaxique pourra dire en quoi consiste cette déception. Dans un premier temps, nous pourrions expliquer cette déception par les attentes formulées à l'endroit de la notion de prouvabilité, des attentes de nature sémantique : nous voudrions que la prouvabilité s'érige en concept de vérité. L'incomplétude se comprend donc mieux en référence à une interprétation et le concept de vérité qui s'ensuit. Comme nous le savons, le concept de vérité d'une théorie est précisément défini pour que le tiers exclu soit valide, c'est-à-dire que toute formule close est ou bien vraie ou bien fausse. Selon le premier théorème d'incomplétude de Gödel, cette propriété du tiers exclu fait défaut à la prouvabilité dans la plupart des systèmes arithmétiques. C'est donc dire qu'il existe des formules vraies qui sont indémontrables (et des formules fausses qui sont irréfutables). Voilà le principal problème avec cette incomplétude.

Pourquoi devrions-nous être étonnés par cette situation? En premier lieu, et de façon générale, l'idée de distinguer la prouvabilité de la vérité est venue très tard dans l'histoire de la logique et, jusqu'à lors, il y avait cette prémisse tacite dans les sciences formelles à l'effet qu'une formule vraie est forcément démontrable, que sa vérité consiste précisément en sa prouvabilité. Avec le premier théorème de Gödel et le théorème de Tarski, ce mythe tombe mais l'étonnement persiste.

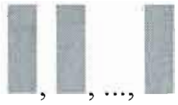
L'étonnement persiste parce que le système d'axiomes qui définit l'arithmétique semble motivé en partie par une ontologie réaliste où le tiers exclu est valide. En effet, quelle que soit la formule $\mathcal{F}$, nous avons $\vdash \mathcal{F} \vee \neg\mathcal{F}$, mais ceci n'entraîne pas que $\vdash \mathcal{F}$ ou $\vdash \neg\mathcal{F}$. De la même manière, bien qu'il soit possible de montrer que $\vdash \neg\neg\mathcal{F} \Rightarrow \mathcal{F}$, nous n'avons pas de façon générale que non-$\vdash \neg\mathcal{F}$ entraîne $\vdash \mathcal{F}$. Enfin, pour prendre un exemple avec des quantificateurs, de $\vdash \mathcal{F}(\overline{n})$ pour tout *n*, il ne s'ensuit pas toujours que $\vdash \forall x \mathcal{F}(x)$ (*S* n'est pas ω-complète). En effet, si je nous rapporte à la démonstration du premier théorème d'incomplétude, nous avions $\neg Dém(n, g)$ pour tout *n* (où $g = \Gamma(\mathcal{G})$) et donc (par représentabilité de $Dém(x, y)$) que $\vdash_S \neg\mathfrak{D}ém(\overline{n}, \overline{g})$ pour tout *n*. Mais nous ne pouvons pas avoir $\vdash_S \forall x_2 \neg\mathfrak{D}ém(x_2, \overline{g})$ sans quoi nous aurions $\vdash_S \mathcal{G}$.[4] On en conclut donc que le prédicat de prouvabilité « $\vdash_S$ » ne traduit pas fidèlement la signification des connecteurs du langage objet au métalangage, il manifeste une forme d'opacité. Rappelons qu'il en est tout autrement pour le prédicat de vérité, celui-ci reste fidèle à la signification de « $\forall$ », « $\neg$ » et « $\Rightarrow$ ».

Pour mieux saisir l'enjeu de l'incomplétude, je vais construire une analogie avec un certain jeu. Considérons un jeu plus simple, joué avec un échiquier :

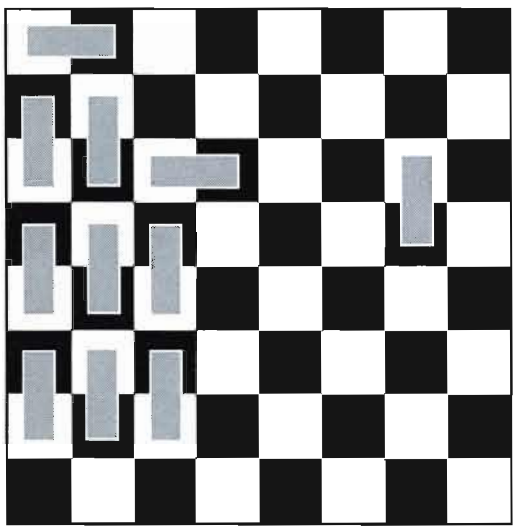


[4] Cette forme d'incomplétude se nomme l'ω-incomplétude.

et des dominos :



L'objectif du jeu consiste à faire de recouvrements de l'échiquier avec des dominos, un recouvrement étant la disposition d'un certain nombre de dominos sur l'échiquier telle que chaque domino recouvre exactement deux cases juxtaposées (horizontalement ou verticalement) et aucune case n'est recouverte par plus d'un domino. Par exemple, la disposition suivante serait un recouvrement :



Nous nous demandons maintenant si certains recouvrements sont possibles. En particulier, existe-t-il un recouvrement qui laisse à découvert seulement la case inférieure gauche et la case supérieure droite (finalement, les cases extrêmes de la diagonale noire)? La réponse est non, cette position est inatteignable par recouvrements. La démonstration est facile. Il suffit de remarquer qu'un domino recouvre toujours une case noire et une case blanche. Le nombre total de cases blanches recouvertes (resp. découvertes) est donc égal au nombre total de cases noires recouvertes (resp. découvertes). Un recouvrement qui laisse les deux cases extrêmes de la diagonale noire à découvert est donc impossible.

L'analogie a pour but de donner une certaine intuition au premier théorème d'incomplétude de Gödel. Si nous considérons les formules de S comme des positions et les prédicats métathéoriques de prouvabilité et de vérité comme des manières d'atteindre ces positions, il devrait être possible de comprendre l'enjeu de l'incomplétude avec les jeux (fictifs) suivants :

(J1) Atteindre des positions par démonstration : Un joueur de J1 atteint une position $\mathcal{A}$ si et seulement si $\vdash_S \mathcal{A}$, c'est-à-dire si et seulement s'il trouve une démonstration de $\mathcal{A}$ dans S.

(J2) Atteindre des positions par vérification : Un joueur de J2 atteint une position $\mathcal{A}$ si et seulement si $\vDash_{\mathcal{N}} \mathcal{A}$, c'est-à-dire si et seulement s'il est vrai dans l'interprétation $\mathcal{N}$.

Il est assez simple de montrer que le jeu (J2) est une extension de (J1). On peut vérifier sans trop de difficultés que :

(V1) Si $\mathcal{A}$ est un axiome de S, alors $\vDash_{\mathcal{N}} \mathcal{A}$;

(V2) Si $\vDash_{\mathcal{N}} \mathcal{A}$ et $\vDash_{\mathcal{N}} \mathcal{A} \Rightarrow \mathcal{B}$ alors $\vDash_{\mathcal{N}} \mathcal{B}$;

(V3) Si $\vDash_{\mathcal{N}} \mathcal{A}(x_1)$ alors $\vDash_{\mathcal{N}} \forall x_1 \mathcal{A}(x_1)$.

Ceci montre donc que, pour tout $\mathcal{A}$, si $\vdash_S \mathcal{A}$ alors $\vDash_{\mathcal{N}} \mathcal{A}$, c'est-à-dire que toute position atteignable dans le jeu (J1) l'est aussi dans le jeu (J2). Le théorème d'incomplétude nous montre que cette inclusion est stricte, qu'il existe des positions atteignables dans le jeu (J2) qui ne le sont pas dans (J1). Notamment, nous avons que :

(V4) Si, pour tout n, $\vDash_{\mathcal{N}} \mathcal{A}(\overline{n})$, alors $\vDash_{\mathcal{N}} \forall x_1 \mathcal{A}(x_1)$

et nous savons que ce n'est pas le cas pour le prédicat « $\vdash_S$ ».

D'une certaine manière, (V4) nous indique que celui qui joue à (J2) dispose de pouvoirs transfinis : il peut déduire $\vDash_{\mathcal{N}} \forall x_1 \mathcal{A}(x_1)$ de la suite infinie d'hypothèses $\vDash_{\mathcal{N}} \mathcal{A}(0)$, $\vDash_{\mathcal{N}} \mathcal{A}(\overline{1})$, $\vDash_{\mathcal{N}} \mathcal{A}(\overline{2})$, ... Nous pourrions modifier la donne du jeu (J1) et permettre au joueur de faire des déductions transfinies. Il faudrait simplement admettre les preuves de longueur infinie dans S et ajouter la règle :

(ω) De $\vdash_S \mathcal{A}(0)$, $\vdash_S \mathcal{A}(\overline{1})$, $\vdash_S \mathcal{A}(\overline{2})$,... on peut déduire $\vdash_S \forall x_1 \mathcal{A}(x_1)$.

Cette règle est parfois appelée la règle de Tarski ou de Carnap.[5] Chose étonnante s'il en est une, il s'avère que ce nouveau jeu est complet : si $\mathcal{A}$ est une formule close, une (et une seule) des positions $\mathcal{A}$ ou $\neg\mathcal{A}$ est atteignable. Ainsi, le prédicat de prouvabilité, augmenté de la règle oméga de Tarski, est complet; il se comporte comme un prédicat de vérité.

Quelle est donc la morale de cette analogie? L'analyse métathéorique nous a permis de constater que : 1) le prédicat de prouvabilité n'est pas fidèle à la signification « intentionnelle » des connecteurs de S; 2) le prédicat de vérité de son côté l'est; 3) le prédicat de vérité est essentiellement un prédicat de prouvabilité transfini ou, ce qui revient au même, le prédicat de prouvabilité transfini est un prédicat de vérité; donc 4) l'incomplétude exprime une discorde entre la signification intentionnelle des connecteurs et notre concept constructif de preuve. Voilà donc un résultat métathéorique qui est loin d'être trivial ou circulaire. Sans métathéorie, cette discorde n'aurait jamais pu être observée.

[5] Voir (Tarski 1974 : pp. 5 et suivantes).

Vers l'infini et plus loin encore

Le concept de métalangage tel que je le défends dans ce mémoire prend donc des distances importantes par rapport au fondationalisme qui lui a donné vie. Ce n'est pas dire pour autant que les mathématiques sont sans fondements ou que le formalisme (tel que décrit par von Neumann)[6] est la seule option philosophique sur le marché, c'est plutôt réaliser que la question même des fondements doit être revue. Si par « fonder » on entend la démonstration anhypothétique de la validité d'une théorie, il est évident alors que le projet fondationaliste est voué à l'échec. En mathématiques comme ailleurs, la certitude cartésienne est un idéal épistémique irréaliste. Mais si fonder une théorie consiste à démontrer avec une métathéorie que la théorie possède certains attributs souhaitables, qu'il s'agisse de la complétude, de la cohérence ou d'une autre propriété, il y a, me semble-t-il, de l'espoir pour une forme de fondationalisme. Fonder, dans ce nouveau sens amendé, est donc toujours relatif au méta-système employé pour étudier la théorie objet.

De cette réflexion sur la notion de métathéorie, nous pouvons tirer un certain nombre de conclusions. D'abord, la méthode métathéorique n'est pas stérile lorsqu'elle cherche à déterminer des propriétés autres que la cohérence absolue. Même un résultat de cohérence relative, qui donne une mesure de la compatibilité entre deux théories, est non trivial, en ce sens qu'il nous informe sur quelque chose que nous ignorions. L'utilisation du « patrimoine » logique pour déterminer la position spécifique d'une nouvelle théorie contribue clairement à l'accroissement de notre connaissance, malgré qu'il ne s'agisse pas d'une connaissance anhypothétique. Il semble que le problème ne soit pas tant la relativité métathéorique que la quête de la certitude cartésienne inhérente aux écrits fondateurs de la logique mathématique. Lorsque nous laissons tomber ce fondationnalisme anhypothétique, et les preuves de cohérence absolue qui lui sont propres, la relativité n'est plus un problème. En fait,

[6] Benacerraf & Putnam 1983 : pp. 61-66.

elle nous expose la véritable nature de l'analyse formelle : une mise en relation réciproque de nos théories logiques.

Le paradigme métathéorique développé ci-dessus s'insère dans une conception où la logique est une famille de sciences formelles, chacune ayant des visées différentes et des degrés de rigueur ou de formalisation différents. Une science formelle $\mathcal{S}$ peut être comparée à une autre science formelle $\mathcal{T}$, ou encore $\mathcal{S}$ peut être utilisée pour étudier $\mathcal{T}$ et vice versa. Aucun critère absolu ne peut être donné pour trancher de manière définitive sur le sort d'une de ces sciences. C'est essentiellement à la pratique historique et actuelle de la logique de nous fournir une gradation par laquelle nous faisons plus confiance à certaines sciences plus qu'à d'autres, et l'analyse métathéorique nous permet de distribuer ou non cette confiance à de nouveaux candidats. En résumé et pour faire dans la métaphore, c'est avec des outils connus et éprouvés par l'expérience que nous construisons d'autres outils plus sophistiqués et précis que les précédents. Manifestement, dans cette progression logique ou technologique, il n'y a aucune contradiction, circularité ou régression à l'infini.

BIBLIOGRPAHIE

BARWISE, J. et KEISLER, H. J. (éd.) (1978), *Handbook of Mathematical Logic*, New York, North-Holland Publishing.

BENACERRAF, P. et PUTNAM, H. (éd.) (1983), *Philosophy of Mathematics : Selected Readings*, 2ième éd., Cambridge, Cambridge University Press.

BLACK, M. (1966), *A Companion to Wittgenstein's Tractatus*, Ithaca (N. Y.), Cornell University Press.

BOOLE, G. (1992), *Les lois de la pensée*, trad. DIAGNE, S. D., Paris, Vrin.

BOOLOS, G. (1979), *The Unprovability of Consistency : An Essay in Modal Logic*, Cambridge, Cambridge University Press.

CARNAP, R. (1967), *The Logical Structure of the World & Pseudoproblems in Philosophy*, trad. par GEORGE, R. A., Berkeley, University of California Press.

CARNAP, R. (1951), *The Logical Syntaxe of Language*, trad. par SMEATON, A., London, Routledge & Kegan Paul.

CAUCHY, L. A., (1989), *Analyse algébrique*, Sceaux, Jacques Gabay.

COHEN, R. S. et MARION, M. (éd.) (1995), *Québec Studies in the Philosophy of Language, Part 1 : Logic, Mathematics, Physics and History of Science*, Boston, Kluwer Academic.

CORI, R. et LASCAR, D. (2003), *Logique mathématique : cours et exercices corrigés, vol. I et II*, Paris, Dunod.

CHURCH, A. (1956), *Introduction to Mathematical Logic*, Vol. I, Princeton, Princeton University Press.

DEVIDI, D. et SOLOMON, G. (1999), *Tarski on 'Essentially' Richer Metalanguages*, Journal of Philosophical Logic, Vol. 20, No. 1, pp. 1-28.

DOKIC, J. et ENGEL, P. (2001), *Ramsey : Vérité et succès*, Paris, PUF.

DUMMETT, M. (1973), *Frege : Philosophy of Language*, London, Duckworth.

DUMMETT, M. (1978), *Truth and Other Enigmas*, Cambridge (Mass.), Harvard University Press.

ENGEL, P. (1996), *Philosophie et psychologie*, Paris, Gallimard.

FREGE, G. (1964), *The Basic Laws of Arithmetic : Exposition of the System*, trad., éd. et intr. par FURTH, M., Berkeley, University of Calfornia Press.

FREGE, G. (1969), *Les fondements de l'arithmétique*, trad. et intr. par IMBERT, C., Paris, Seuil.

FREGE, G. (1971), *Écrits logiques et philosophiques*, trad. et éd. par IMBERT, C., Paris, Seuil.

FREGE, G. (1972), *Conceptual Notation and Related Articles*, éd., trad. et introd. par BYNUM, T. W., Oxford, Clarendon Press.

FREGE, G. (1980), *Gottlob Frege: Philosophical and Mathematical Correspondence*, trad. KAAL, H. et éd. MCGUINNESS, B., Chicago, University of Chicago Press.

FREGE, G. (1984), *Collected Papers on Mathematics, Logic, and Philosophy*, éd. par MCGUINNESS, B. et trad. par BLACK, M., Oxford, Blackwell.

GEACH, P. (1976), *Saying and showing in Frege and Wittgenstein*, dans Hintikka (dir.), *Essays in Honour G. H. von Wright*, Acta Philosophica Fennica, vol. 28, n. 1-3, p. 54-70.

GENTZEN, G. (1969), *The Collected Papers of Gerhard Gentzen*, éd. par SZABO, M., Amsterdam, North Holland.

GIAQUINTO, M. (2002), *The Search for Certainty : A Philosophical Account of the Foundations of Mathematics*, Oxford, Clarendon Press.

GIRARD, J.-Y. (1987), *Proof Theory and Logical Complexity*, Vol. I, Naples, Bibliopolis.

GIRARD, J.-Y. (1999), *On the Meaning of Logical Rules I : Syntax vs. Semantics*, Computational Logic, éd. par BERGER et SCHWICHTENBERG, pp. 215-272, SV, Heidelberg.

GIRARD, J.-Y. (2000), *On the Meaning of Logical Rules II : Multiplicatives and Additives*, Foundation of Secure Computation, éd. par BAUER et STEINBRÜGGEN, pp. 183-212, Amsterdam, IOS Press.

GOLDFARB, W. (1979), *Logic in the Twenties : The Nature of the Quantifier*, Journal of Symbolic Logic 44, pp. 351-68.

HAMILTON, A. G. (1988), *Logic for Mathematiciens*, éd. révisée, Cambridge, Cambridge University Press.

HATCHER, W. (1982), *The Logical Foundations of Mathematics*, Oxford, Pergamon Press.

HECK JNR., R. G. (1998), *Grundgesetze der Arithmetik* I §§29-32, *Notre Dame Journal of Formal Logic* 38, pp. 437-74.

HECK JNR., R. G. (à venir), *Frege and Semantics*.

HILBERT, D. (1950), *Foundations of Geometry*, trad. par TOWNSEND, E., La Salle (Ill.), Open Court.

HILBERT, D. et BERNAYS, P. (2001), *Fondements des mathématiques, Tomes I et II*, trad. par GAILLARD, F., GUILLAUME, E. ET GUILLAUME, M., Paris, L'Harmattan.

HUSSERL, E. (1970), *Logical Investigations*, vol. I et II, trad. FINDLAY, J. N., London, Routledge & Kegan Paul.

HYLTON, P. (1990), *Russell, Idealism, and the Emergence of Analytic Philosophy*, Oxford, Clarendon Press.

KLEENE, S. C. (1950), *Introduction to Metamathematics*, Princeton, Van Nostrand.

KRIPKE, S. (1982), *Wittgenstein : On Rules and Private Language*, Cambridge (Mass.), Harvard University Press.

KUSCH, M. (1995), *Psychologism : A Case Study in the Sociology of Philosophical Knowledge*, London, Routledge.

LARGEAULT, J. (éd.) (1972), *Logique mathématique*, Paris, Vrin.

LAURIER, D. (1993), *Introduction à la philosophie du langage*, Liège, Mardaga.

LARGEAULT, J. (éd.) (1992), *Intuitionisme et théorie de la démonstration*, Paris, Vrin.

MARION, M. et VOIZARD, A. (éd.) (1998), *Frege : Logique et Philosophie*, Paris, L'Harmattan.

MARION, M. (1998), *Wittgenstein, Finitism, and the Foundations of Mathematics*, Oxford, Clarendon Press.

MARTIN, R. L. (éd.) (1984), *Recent Essays on Truth and the Liar Paradox*, New York, Oxford University Press.

MARTIN-LÖF, A. (1996), *On the meanings of the logical constants and the justifications of the logical laws*, Nordic Journal of Philosophy, Vol. 1, No. 1, pp. 11-60.

MARTIN-LÖF, A. (1980), *Intuitionistic Type Theory*, Naples, Bibliopolis.

MENDELSON, E. (1987), *Introduction to Mathematical Logic*, 3ième éd., Monterey (Californie), Wadsworth & Brooks/Cole Advanced Books & Software.

POINCARÉ, H. (1906), *Les mathématiques et la logique*, Revue de métaphysique et de morale 14, pp. 294-317.

POINCARÉ, H. (1968), *La science et l'hypothèse*, Paris, Flammarion.

POINCARÉ, H. (1970), *La valeur de la science*, Paris, Flammarion.

QUINE, W. V. O. (1964), *From a Logical Point of View : 9 logico-philosophical essays*, Cambridge (Mass.), Harvard University Press.

RICKETTS, T. (1985), *Frege, the* Tractatus *and the Logicocentric Predicament*, Noûs 9, pp. 3-15.

RUSSELL, B. (1919), *Introduction to Mathematical Philosophy*, Londres, G. Allen & Unwin.

RUSSELL, B. (1964), *The Principles of Mathematics*, seconde édition, Londres, G. Allen & Unwin, 1964.

RUSSELL, B. (1986), *The Philosophy of Logical Atomism and Other Essays*, éd. par SLATER, J., Londres, G. Allen & Unwin, 1986.

RUSSELL, B. (1989), *Écrits de logique philosophique*, trad. et éd. par ROY, J.-M., Paris, PUF.

RUSSELL, B. (1997), *Essais philosophiques*, prés. et trad. par CLÉMENTZ, F. et COMETTI, J.-P., Paris, PUF.

RUSSELL, B. et WHITEHEAD, A. (1962), *Principia Mathematica, to *56*, Cambridge, Cambridge University Press.

SCHERRER, J.-B. (éd.) (1989), *Le théorème de Gödel*, trad. par SCHERRER, J.-B., Paris, Seuil.

SCHRÖDER, E. (1966), *Algebra der Logik, I-III.* 1890-1910, Chelsea.

SMULLYAN, R. M. (1992), *Gödel's Incompleteness Theorems*, Oxford, Oxford University Press.

TARSKI, A. (1956), *Logic, Semantics, Metamathematics : papers from 1923 to 1938*, éd. et trad. par WOODGER, J.H., Oxford, Hackett.

TARSKI, A. (1972), *Logique, sémantique et métamathématique*, tomes. I, éd. par GRANGER, G., Paris, Armand Colin.

TARSKI, A. (1974), *Logique, sémantique et métamathématique*, tomes. II, éd. par GRANGER, G., Paris, Armand Colin.

TARSKI, A. (1986), *Collected Papers*, vol. I-IV, éd. par GIVANT, S.R. et MCKENZIE, R. éditeurs, Birkhäuser, Basel.

VAN HEIJENOORT, J. (éd.) (1967), *From Frege to Gödel : A Sourcebook in Mathematical Logic, 1879-1931*, Cambridge (Mass.), Harvard University Press.

VAN HEIJENOORT, J. (1985), *Selected Essays*, Naples, Bibliopolis.

WITTGENSTEIN, L. (1961), *Tractatus Logicus-Philosophicus*, trad. par KLOSSOWSKI, P. intr. par RUSSELL, B., Paris, Gallimard.

WITTGENSTEIN, L. (1974), *Philosophical Investigations*, trad. par ANSCOMBE, G. E. M., Oxford, Basil Blackwell.

WITTGENSTEIN, L. (1974), *Philosophical Grammar*, éd. par RHEES, R. et trad. par KENNY, A., Oxford, Basil Blackwell.

WITTGENSTEIN, L. (1978), *Remarks on the Foundations of Mathematics*, 3ième éd., éd. par WRIGHT, G. H., RHEES, R. et ANSCOMBE, G. E. M., trad. par ANSCOMBE, G. E. M., Oxford, Basil Blackwell.

WITTGENSTEIN, L. (2004), *Tractatus Logicus-Philosophicus*, trad. par PEARS, D. F. et MCGUINNESS, B. F., intr. par RUSSELL, B., Londres, Routledge Classics, 2004.