La vérité

Henri Galinon

Version longue et non corrigée d'un texte paru dans Poggiolesi, Wagner (éd.), Précis de philosophie de la logique et des mathématiques, vol. 1, chap. 3, pp. 109-166, Editions de la Sorbonne, Paris, 2021.

Note : La section 4.3 de ce manuscrit ne figure pas dans la version publiée.

Pour toute référence, merci de se reporter au texte publié.

1 La vérité

Henri Galinon

1. Introduction

Si la notion de vérité occupe une place importante dans les recherches logico-philosophiques contemporaines, c'est en grande partie à l'influence du logicien polonais Alfred Tarski que nous le devons[1]. Pour en prendre la mesure, il faut rappeler brièvement quel était le climat philosophique des années vingt et trente et les obstacles que rencontrait alors l'usage philosophique de la notion de vérité. Le premier trouvait sans doute sa source dans une hostilité diffuse à l'égard d'un réalisme métaphysique naïf que semblait devoir véhiculer l'usage de la notion classique de vérité comme correspondance du discours à son objet. Ni les nouvelles conceptions épistémiques de la vérité comme cohérence ou limite d'une enquête idéale n'étaient venues concurrencer la notion classique dans les cercles pragmatistes, ni les mots d'ordre vérificationnistes du cercle de Vienne ne semblaient devoir laisser de place à un usage philosophique légitime de la notion classique de vérité[2]. De façon concomitante, le tournant épistémique était également bien engagé en philosophie de la logique et des mathématiques sous l'influence des propositions intuitionnistes et formalistes. Mais aux doutes sur la légitimité philosophique d'un appel à la notion de vérité pour décrire les relations du langage au monde – *a fortiori* pour le langage des mathématiques – s'ajoutait une interrogation peut-être plus fondamentale encore sur la possibilité d'un usage simplement cohérent de la notion de vérité,

1. Leon Horsten parle à juste titre de « tournant tarskien » (Horsten [2011]).

2. Sur la variété des conceptions de la vérité ayant cours dans la première moitié du vingtième siècle, on pourra consulter Damnjanovic and Candlish [2007]. Sur les réticences que suscita le travail de Tarski au sein du cercle de Vienne, et notamment chez Neurath, on pourra consulter Mancosu [2008].

interrogation d'autant plus aiguë que l'époque n'en avait pas fini avec les répliques de la crise des fondements des mathématiques. Dans ce contexte, la parution des théorèmes d'incomplétude de Gödel en 1931 et, dans son aspiration pour ainsi dire, la parution en 1935 du travail de Tarski sur la vérité[3] ouvrirent une nouvelle époque. Faisant œuvre à la fois philosophique et logique, Tarski montrait ce que devait être une définition formalisée de la vérité et donnait la démonstration de plusieurs résultats fondamentaux touchant la possibilité et l'impossibilité d'une définition de ce genre sous différentes conditions. En montrant qu'un usage cohérent de la notion de vérité était possible, les résultats positifs établissaient la possibilité d'une sémantique scientifique et ouvraient une nouvelle ère pour la logique, mais aussi pour la philosophie des sciences et du langage, dans laquelle la vérité jouerait désormais un rôle central tout au long du vingtième siècle.

Les questions logiques et philosophiques liées à la notion de vérité sont trop nombreuses et trop éparses pour pouvoir être traitées ici, même en s'en tenant aux seuls thèmes de la tradition tarskienne. Il y a beaucoup à dire, et beaucoup a déjà été dit, sur le développement du point du vue modèle-théorique en logique mathématique et l'utilisation du travail en linguistique et en philosophie du langage[4]. L'objectif de ce chapitre est beaucoup plus limité : introduire le lecteur au travaux logiques et philosophique de Tarski sur la vérité « tout court » et à quelques-uns de leurs prolongements contemporains. La première section présente les principaux résultats de Tarski dans ce domaine. Les deuxième et troisième sections sont consacrées aux travaux ultérieurs sur les paradoxes de la vérité. Quarante ans après Tarski, le travail de Kripke marque une percée majeure dans le traitement de paradoxes et son influence contemporaine justifie que la deuxième section lui soit entièrement consacrée, tandis que la troisième section présente un tour d'horizon, non exhaustif, des développements plus récents. La quatrième et dernière section revient sur l'interprétation philosophique des travaux de Tarski en portant l'accent sur la discussion des positions dites « déflationnistes », lesquelles ont, depuis les remarques de Quine sur la vérité, occupé une place grandissante, jusqu'à devenir prépondérante, dans les débats.

3. Tarski [1935]. Pour un exposé de son travail par Tarski à l'intention des philosophes, on consultera le classique Tarski [1944]. Le regard que pose Tarski [1969] sur la portée philosophique et scientifique de son travail sur la vérité est également éclairant.

4. Sur le développement de la notion de modèle on pourra consulter Hodges [1985].

2. Tarski

2. 1 Définir la vérité : pourquoi et comment

Lorsque l'emploi d'une notion est fragilisé par les doutes que l'on nourrit sur sa légitimité, une façon incontestable de dissiper ces doutes est de produire une définition explicite de cette notion sur une base indubitable. C'est exactement la voie choisie par Tarski pour rétablir la notion de vérité dans ses droits. Sans doute faut-il préciser de quelle notion de vérité il s'agit et quelles sont les ambitions exactes du projet. La notion de vérité que Tarski souhaite définir est une notion qui s'applique à des énoncés et non à ces objets plus fuyants que sont les propositions. Le prédicat de vérité sera en conséquence défini relativement à un langage donné, et il n'entre pas dans le projet de Tarski de donner une définition plus générale de la vérité pour un langage $\mathcal{L}$ où $\mathcal{L}$ a la valeur d'une variable, même si la méthode utilisée doit être applicable pour produire différentes définitions pour différentes valeurs de $\mathcal{L}$. Les langages pour lesquels le prédicat de vérité est défini seront des langages formalisés, c'est-à-dire interprétés et munis d'une structure exactement spécifiée[5]. Il ne peut donc s'agir ni d'un langage formel[6] ni du langage ordinaire, mais plutôt de quelque chose qui s'apparente à un fragment quelque peu idéalisé de ce dernier. La théorie dans laquelle sera conduite la définition de la notion de vérité pour un langage donné devra être vierge de toute notion sémantique, et plus généralement de toute notion primitive qui pourrait nourrir le scepticisme au regard des exigences scientifiques les plus sévères : des notions permettant de décrire la syntaxe du langage et un appareil logique suffisamment puissant doivent suffire. Enfin, il faut préciser que la notion de vérité qu'il s'agit de définir est la notion classique de vérité, et non une notion concurrente telle par exemple la notion pragmatique évoquée plus haut de vérité-cohérence.

Mais à quoi reconnaîtrons-nous qu'il s'agit du prédicat *de vérité* pour le langage considéré ? À ce point, le projet de définition ne peut pas faire l'économie d'une analyse conceptuelle de la notion de vérité. Celle que propose Tarski pour la notion classique de vérité[7] a fait date par l'économie de ses

5. Voir Tarski [1935], section ?

6. La définition classique de la notion de vérité d'un énoncé d'un langage formel dans une structure d'interprétation, telle que nous la connaissons aujourd'hui en logique, émergera lentement quelque années après les travaux sur la vérité. Sur l'émergence de la notion de modèle et de son rapport à la notion de vérité telle qu'elle est élaborée en 1935, on pourra consulter Hodges [1985] et Milne [1999].

7. Cette analyse est présentée en détail dans Tarski [1944].

moyens. Tarski suggère en effet qu'un prédicat Vr est un prédicat de vérité dans un langage $\mathcal{L}_1$ pour un langage $\mathcal{L}_2$ s'il satisfait le schéma-T suivant :

Vr(s) si et seulement si p

où « s » doit être remplacé par un nom (dans $\mathcal{L}_1$) d'un énoncé de $\mathcal{L}_2$ et « p » par sa traduction (dans $\mathcal{L}_1$)[8].
Chacune des instances de ce schéma, observe Tarski, articule en effet sobrement, sans recours aux notions plus ou moins obscures et inutiles de « fait » ou de « correspondance », pour un énoncé donné, la condition qui doit être réalisée dans les choses mêmes pour qu'il soit vrai. Ainsi par exemple le célèbre biconditionnel

« La neige est blanche » est vrai si, et seulement si, la neige est blanche.

manifeste-t-il aussi exactement que possible la correspondance des mots et des choses qui est réalisée si la phrase mentionnée dans le membre gauche du biconditionnel est vraie. Cette analyse sert de base à l'élaboration d'un critère d'adéquation, que Tarski appelle la convention-T, qui doit permettre de garantir que le prédicat défini selon le cahier des charges présenté plus haut est bien un prédicat de vérité au sens classique de la notion. En substance le critère d'adéquation peut être formulé de la façon suivante :

Adéquation (au sens de Tarski) Soit T une théorie contenant un prédicat de vérité Vr pour un langage $\mathcal{L}$. Vr est adéquat pour $\mathcal{L}$ si pour tout énoncé ϕ de $\mathcal{L}$:

$$T \vdash \mathbf{Vr}(s) \leftrightarrow p$$

où « s » est un nom du langage de T qui désigne ϕ et p est une traduction de ϕ.

Quelle est la portée exacte de cette analyse de la vérité et de la convention-T ? Personne ne doute que, moyennant certaines réserves dont il sera longuement question bientôt, ces biconditionnels-T soient vrais ; ni que le critère d'adéquation formulé par Tarski permette en effet de garantir qu'un prédicat qui le satisfait s'applique toujours exactement aux énoncés vrais d'un langage et soit donc en ce sens « extensionnellement » correct. Mais avons-nous là le début d'une élucidation philosophiquement intéressante

8. Si $\mathcal{L}_2$ est une extension de $\mathcal{L}_1$, on admettra que « p » est l'énoncé s lui-même.

de la notion de vérité ? Ici les avis divergent. Dummett [1958-1959] et Putnam [1985] l'ont vivement contesté, Quine [1970] l'a fermement défendu. Nous reviendrons à cette question dans la dernière section de ce chapitre.

Les conditions de succès du projet de définition de Tarski étant désormais suffisamment précisées, ce projet peut-il être mené à bien ? Tarski a le double mérite d'avoir montré que la vérité était indéfinissable[9] et d'avoir montré comment la définir[10] – la marque d'un grand homme. On peut résumer les résultats de Tarski de la façon suivante. Une théorie formulée dans un langage $\mathcal{L}$ ne peut pas contenir de prédicat de vérité adéquat pour $\mathcal{L}$ – *a fortiori* elle ne peut en contenir de définition explicite. Si $\mathcal{L}$ est un langage dont les variables parcourent tout ce qui est[11], il n'existe pas de langage $\mathcal{L}'$ et de théorie de ce langage dans laquelle il serait possible de définir explicitement le prédicat de vérité pour $\mathcal{L}$ de façon adéquate. Dans le cas d'un tel langage $\mathcal{L}$ pour lequel le projet de définition explicite d'un prédicat de vérité ne peut être mené à bien, il est néanmoins possible d'introduire axiomatiquement un prédicat de vérité primitif qui satisfasse la condition d'adéquation sur le fragment de langage $\mathcal{L}$ ne contenant pas le prédicat de vérité. Pour tous les autres langages, Tarski montre que la vérité est définissable explicitement. Dans la suite de cette section, nous revenons plus en détail sur ces résultats.

2. 2 Du paradoxe au théorème

Sitôt établi le caractère central des biconditionnels-T pour notre compréhension du concept de vérité, il nous faut prendre la mesure de la difficulté que représentent pour le projet de définition les célèbres paradoxes de la vérité. Nous nous bornerons ici à rappeler celui du Menteur. Considérons l'énoncé suivant :

(λ) λ n'est pas vrai.

Supposons (λ) vrai. Alors (λ) n'est pas vrai (d'après le biconditionnel-T cor-

9. Tarski [1935], Th. I.
10. Tarski [1935], Th. II. et Th. III.
11. Si un tel langage existe. L'idée d'un tel langage est liée à la notion de langage universel. La position de Tarski a varié sur l'existence d'un langage logique universel. Voir le *postscriptum* à Tarski [1935], et l'analyse de Rouilhan [1998]. Dans ce volume, voir le chapitre 18. Une partie de la discussion contemporaine est formulée comme une interrogation sur la possibilité d'exprimer la généralité absolue – voir le recueil de Rayo and Uzquiano [2006].

respondant[12], lu gauche à droite). Donc l'hypothèse était fausse et (λ) n'est pas vrai. Mais c'est ce que dit (λ), donc (λ) est vrai (d'après le biconditionnel-T, de droite à gauche). Absurde.

L'énoncé du Menteur est un énoncé du langage naturel qui a la particularité d'être auto-référentiel : l'énoncé (λ) parle de l'énoncé (λ) lui-même. Cette particularité peut sembler étrange, mais il est facile de voir à la réflexion que l'autoréférence est par elle-même un phénomène relativement courant et tout à fait bénin du langage ordinaire. Ainsi, il n'y a aucune difficulté à signaler ici que le chapitre d'ouvrage que vous êtes en train de lire traite de la notion de vérité, ou que cette phrase commence par un « A ». Cette auto-référence peut être produite au moyen d'un vocabulaire spécifique (déictiques, verbes pronominaux, etc.), mais elle peut aussi être un fait empirique contingent, ce qui arrive par exemple si un énoncé affirme que tout énoncé tombant sous une certaine description est faux et qu'il se trouve être l'unique énoncé tombant sous la description en question. Le phénomène n'est pas cantonné au langage ordinaire. On peut aussi construire dans le langage formalisé de l'arithmétique des énoncés qui fonctionnent à la façon d'énoncés autoréférentiels s'attribuant à eux-même un prédicat P quelconque. Un énoncé ϕ de ce genre est prouvablement équivalent dans le système formel de l'arithmétique de Peano à l'énoncé $P(\phi)$. Ce résultat dû à Gödel est connu sous le nom de lemme de diagonalisation[13] :

Lemme de diagonalisation Pour toute formule $\Phi(x)$ de $\mathcal{L}_A$, il existe une formule ψ de $\mathcal{L}_A$ telle que : $PA \vdash \psi \leftrightarrow \Phi(\ulcorner\psi\urcorner)$

Mais s'il est possible de formuler, pour tout prédicat arithmétiquement définissable P un énoncé s'auto-attribuant P, alors le paradoxe du Menteur devient une preuve de l'indéfinissabilité d'un prédicat de vérité pour $\mathcal{L}_A$ dans PA et ses extensions[14]. En généralisant un peu on obtient :

Théorème d'indéfinissabilité de la vérité Soit T une théorie consis-

12. Le biconditionnel-T correspondant est :

$$Vr(\lambda) \leftrightarrow \lambda$$

autrement dit, en réécrivant λ explicitement :

$$Vr(\lambda) \leftrightarrow \neg Vr(\lambda)$$

13. Les axiomes de l'arithmétique de Peano et le lemme de diagonalisation sont présentés au chapitre 20.

14. Pour une introduction légère mais précise aux relations historiques entre paradoxes et développement de la logique on pourra consulter Quine [2011].

tante formulée dans un langage $\mathcal{L}$, et telle que PA est interprétable dans T. Alors il n'existe aucune formule Φ de $\mathcal{L}$ telle que, pour toute formule ψ de $\mathcal{L}$:

$$T \vdash \psi \leftrightarrow \Phi(\ulcorner \psi \urcorner)$$

Preuve : D'après le lemme de diagonalisation, il existe un énoncé λ tel que : $T \vdash \lambda \leftrightarrow \neg\Phi(\ulcorner \lambda \urcorner)$. Si le prédicat de vérité Φ était définissable dans T, on aurait également : $T \vdash \lambda \leftrightarrow \Phi(\ulcorner \lambda \urcorner)$. Donc T serait inconsistante[15].

Le paradoxe informel du menteur est transformé en un théorème d'impossibilité. On peut aller un peu plus loin en montrant que le lemme de diagonalisation vaut également pour des prédicats quelconques que l'on pourrait introduire à titre de prédicats primitifs dans le langage de l'arithmétique. On montrerait alors, en reprenant le raisonnement précédent, que non seulement le prédicat de vérité n'est pas arithmétiquement définissable, mais en outre qu'il ne peut même pas être introduit à titre primitif de façon adéquate dans une théorie consistante :

Théorème d'impossibilité de Tarski Soit T une extension de PA formulée dans un langage L_{Vr} contenant un prédicat de vérité. Si T est consistante alors le prédicat de vérité n'est pas adéquate pour L_{Vr} – i.e. T ne contient pas à titre de théorèmes tous les biconditionnels-T de la forme $Vr(\phi) \leftrightarrow \phi$ pour $\phi \in \mathcal{L}_{\mathrm{Vr}}$.

2. 3 Définition et usages cohérents de la vérité

Reprenons le raisonnement du Menteur de façon plus détaillée. En voici une dérivation possible en déduction naturelle (format séquent) :

15. Voir également chapitre 20, section 2.

$$\dfrac{\dfrac{\dfrac{\dfrac{Vr(\lambda) \vdash Vr(\lambda)}{Vr(\lambda) \vdash \lambda}\,\text{(Vr-Elim)}}{Vr(\lambda) \vdash \neg Vr(\lambda)} \qquad Vr(\lambda) \vdash Vr(\lambda)}{\dfrac{Vr(\lambda), Vr(\lambda) \vdash \bot}{Vr(\lambda) \vdash \bot}\,\text{(Contraction)}}\,\neg\text{-Elim}}{\dfrac{\dfrac{\dfrac{\vdash \neg Vr(\lambda)}{\vdash \lambda}\,\text{(Vr-Intro)}}{\vdash Vr\lambda}\,(\neg\text{-Elim})}{\vdash \bot}}$$

Cette dérivation montre que si l'on veut établir un usage consistant de la notion de vérité dans un langage, il n'y a que deux options. La première est de limiter l'application des règles qui gouvernent le prédicat de vérité, c'est-à-dire de la règle d'élimination :

$$\dfrac{\Gamma \vdash Vr(\phi)}{\Gamma \vdash \phi}\,\text{(Vr-Elim)}$$

ou de la règle d'introduction :

$$\dfrac{\Gamma \vdash \phi}{\Gamma \vdash Vr(\phi)}\,\text{(Vr-Intro)}$$

Cette restriction revient, *dans le cadre de la logique classique*, à restreindre la classe de biconditionnels-T dont on admet la validité. La seconde option est de renoncer à certaines des autres règles d'inférence mobilisées dans la preuve, et donc de réviser d'une manière ou d'une autre la logique classique[16].

16. Noter que pour mesurer l'étendue des révisions il faut tenir compte de l'existence d'autres dérivations du même paradoxe mais aussi de paradoxes voisins. Le paradoxe de Curry est instructif à cet égard car il ne mobilise pas directement la négation mais l'implication matérielle (et en particulier la règle d'introduction de l'implication). On le présente informellement sans s'y attarder. Soit (c) l'énoncé :

(*c*) Si (*c*) est vrai, alors 0=1.

Supposons l'énoncé que l'énoncé (c) n'est pas vrai, alors l'antécédent du conditionnel est faux,

Si cette seconde option a retenu l'attention de nombre de ses successeurs, Tarski pour sa part n'est pas prêt à renoncer à la logique classique. Or d'un autre côté, amputer la classe des biconditonnels-T valides n'a pas davantage de sens d'un point de vue méthodologique si on fait de la validité des biconditonnels-T la condition d'adéquation du prédicat de vérité ! La porte est donc étroite. Avec ces attendus, nous avons vu que la conclusion inévitable du paradoxe du menteur est que le langage $\mathcal{L}'$ dans lequel nous définissons ou introduisons un prédicat de vérité adéquat pour $\mathcal{L}$ ne peut pas être $\mathcal{L}$ lui-même. Mais avec ces mêmes attendus il est encore possible de se demander dans quelles relations doivent se trouver le premier, appelons-le le métalangage, et le second, appelons-le le langage-objet, pour qu'un prédicat vérité adéquat pour le langage-objet puisse être défini explicitement dans le métalangage. La réponse à cette question est le résultat positif principal de la monographie de 1935[17] :

Définissabilité de la vérité Soit $\mathcal{L}$ un langage formalisé. S'il existe un métalangage $M\mathcal{L}$ *essentiellement plus riche* que $\mathcal{L}$ alors il est possible de définir explicitement dans une théorie de langage $M\mathcal{L}$ un prédicat $\mathrm{Vr}_{\mathcal{L}}$ de vérité adéquat pour $\mathcal{L}$.

La condition de plus grande richesse essentielle revient en substance à exiger que le métalangage contienne, outre les variables parcourant le domaine que parcourent les variables du langage-objet lui-même (par exemple de domaine des entiers dans le cas où le langage-objet est celui de l'arithmétique), des variables dont le domaine contient à titre d'éléments les domaines que parcourent les variables du langage-objet. Par conséquent, si le langage-objet contient des variables dont le domaine est universel[18], la définition explicite est impossible. Dans le cas contraire, la définition, dont le détail diffère au cas par cas selon les langages-objets pour lesquels on souhaite définir le prédicat de vérité, est obtenue selon une méthode constante[19].

---

donc le conditionnel doit être vrai. Mais le conditionnel est justement l'énoncé (c). Absurde. Donc (c) est vrai. Mais si (c) est vrai, alors l'antécédent du conditionnel est vrai, et puisque si un conditionnel vrai et que son antécédent est vrai son conséquent doit l'être aussi, il faut donc que 0=1. Pour aller plus loin, voir Beall and Murzi [2013].

17. Le résultat s'entend pour les langages formalisés dont la structure satisfait certaines conditions satisfaites par les langages formalisés ordinaires. Sur ces conditions, voir Tarski [1935], section ?.

18. Pour être plus précis il faudrait distinguer les langages selon qu'ils contiennent des types syntaxiques ou non. Nous renvoyons à nouveau le lecteur au chapitre 18.

19. On peut incidemment noter qu'il est aisé de définir un prédicat de vérité qui soit adéquat pour un langage qui ne comporterait qu'un ensemble fini d'énoncés, si une telle idée a un sens : un biconditionnel-T isolé est lui-même un cas limite d'une telle définition pour un prédicat de

Pour illustrer la méthode, nous donnerons ici une définition de la vérité pour le langage de l'arithmétique[20]. Le langage de l'arithmétique contient les symboles primitifs suivants $\mathcal{L}_A = \{0, S, +, .\}$[21]. On suppose que dans la théorie formulée dans le métalangage il est possible de définir la morphologie de $\mathcal{L}_A$ ainsi que la structure de n'importe quelle théorie formalisée particulière de ce langage[22].

vérité s'appliquant à un unique énoncé. Pour un nombre fini d'énoncés $\phi_1, \cdots, \phi_n$, la définition suivante serait adéquate : Pour tout énoncé x, $Vrai(x)$ si et seulement si $(x = \ulcorner\phi_1\urcorner \wedge \phi_1) \vee \cdots \vee (x = \ulcorner\phi_n\urcorner \wedge \phi_n)$, où $\ulcorner\phi_i\urcorner$ est un nom de l'énoncé ϕ_i. On pourrait montrer également qu'il est possible de définir arithmétiquement un prédicat de vérité adéquat pour l'ensemble des énoncés atomiques du langage de l'arithmétique, ou bien l'ensemble des énoncés de $\mathcal{L}_{PA}$ de complexité donnée – *cf.* par exemple Boolos et al. [2002]. On peut cependant douter qu'il s'agisse là de « langages », ces fragments n'étant pas clos pour les règles de composition des énoncés ni pour les règles de déduction, et l'application d'un prédicat de vérité à ces fragments présente un intérêt limité.

20. Tarski [1935] présente l'exemple d'une définition pour le calcul des classes.

21. Pour des exemples de théories arithmétiques formulées dans ce langage on se reportera aux axiomes de l'arithmétique de Robinson et de Peano présentés au chapitre 20.

22. Selon la méthode familière des manuels de logique. Les ressources nécessaires se résument d'abord à la possibilité de donner des noms pour chacune des expressions primitives de $\mathcal{L}_A$ puis de spécifier quelles suites de signes primitifs constituent des termes, des formules ou encore de énoncés du langage $\mathcal{L}_A$. Pour les besoins de notre exemple, $\overline{x}$ désignera dans le métalangage la variable « x » de notre langage-objet, $\overline{S}$ le symbole de fonction successeur « S », $\overline{\neg}$ le symbole de la négation etc. On se donnera ensuite les axiomes gouvernant l'opération syntaxique de concaténation des signes et suites de signes, que l'on notera de la façon suivante : « $\ulcorner xy\urcorner$ » désigne la suite de signes obtenue en apposant la suite y immédiatement à droite de la suite x. Ces axiomes énoncent les lois de l'opération de concaténation, par exemple : si x et y sont des expressions alors $\ulcorner xy\urcorner$ est une expression ; ou encore que la classe des expressions est la plus petite classe qui contient les variables $\overline{x}$, $\overline{y}$, etc., les parenthèses $\overline{(}$ et $\overline{)}$, les constantes logiques $\overline{\equiv}$, $\overline{\neg}$, $\overline{\wedge}$,$\overline{\forall}$,..., les constantes non-logiques $\overline{S}$, $\overline{+}$, $\overline{\cdot}$, $\overline{0}$ et close par l'opération de concaténation. Sur la base de ces axiomes, on peut alors définir l'ensemble des termes de $\mathcal{L}_A$:

Définition de l'ensemble des termes de $\mathcal{L}_A$

1. $\overline{x}, \overline{y}$...sont des termes
2. si t_1, t_2 sont des termes alors $\ulcorner\overline{S}t_1\urcorner$, $\ulcorner t_1\overline{+}t_2\urcorner$, $\ulcorner t_1\overline{\cdot}t_2\urcorner$ sont aussi des termes
3. Rien d'autre n'est un terme

On définirait de façon analogue l'ensemble des formules de $\mathcal{L}_A$, l'ensemble des variables libres d'une formule ainsi que l'ensemble des énoncés $\mathcal{L}_A$. Il est à noter ici que la restriction de ces définitions à la morphologie de $\mathcal{L}_A$ est inessentielle, et que l'on pourrait décrire de surcroît dans le métalangage la morphologie du métalangage lui-même.

Pour définir une théorie axiomatisée spécifique du langage-objet étudié on procèderait en commençant par définir l'ensemble des axiomes, puis la clôture par les règles de déduction. Par exemple la définition de l'ensemble Ax_Q des axiomes de Q commencerait simplement ainsi : $Ax_Q(x)$ ssi x est un énoncé et $(x = \ulcorner\overline{\forall}\overline{x}\overline{\neg}\overline{\neg}(\overline{x}\overline{\equiv}\overline{0})\overline{\vee}\urcorner$ etc. La description d'une théorie particulière du langage $\mathcal{L}_A$ n'est nullement nécessaire à la définition de la vérité pour $\mathcal{L}_A$, qui n'en dépend pas, mais elle a son intérêt pour étudier les conséquences de la définition de la vérité, comme nous le verrons plus loin. Notons pour terminer que le fragment de la théorie du métalangage permettant de conduire les définitions des notions morphologiques et preuve-théoriques du langage-objet est structurellement identique à une théorie arithmétique élémentaire. Par un procédé de codage on peut donc si on le souhaite identifier théorie arithmétique élémentaire et théorie syntaxique. Dans la suite de ce chapitre on supposera

Sur cette base, on définit la notion de dénotation d'un terme du langage $\mathcal{L}_A$ relativement à une assignation de la façon suivante :

Assignation, dénotation d'un terme Une assignation σ est une fonction qui à toute variable de $\mathcal{L}$ associe un (unique) objet de l'univers de discours de $\mathcal{L}$, ici les entiers.

La dénotation d'un terme de $\mathcal{L}$ pour une assignation σ se définit par induction :

— $Den(\overline{x}, \sigma) = \sigma(\overline{x})$
— si t_1, t_2 sont des termes de $\mathcal{L}_A$, si $Den(t_1, \sigma) = x_1$ et $Den(t_2, \sigma) = x_2$ alors
  — $Den(\ulcorner \overline{S} t_1 \urcorner, \sigma) = S(x_1)$
  — $Den(\ulcorner t_1 \overline{+} t_2 \urcorner, \sigma) = x_1 + x_2$
  — $Den(\ulcorner t_1 \overline{.} t_2 \urcorner, \sigma) = x_1 . x_2$

On définit ensuite la satisfaction pour les formules atomiques du langage. Dans le cas du langage de l'arithmétique le seul symbole prédicatif primitif est le symbole binaire de l'égalité, les formules atomiques sont donc des équations.

Définition de la satisfaction pour les formules atomiques (val^+) : Pour toute formule atomique x et fonction d'assignation σ, $val^+(x, f)$ si et seulement s'il existe deux termes t_1 et t_2 de $\mathcal{L}_A$ tels que $x = \ulcorner t_1 = t_2 \urcorner$ et $Den(t_1, \sigma) = Den(t_2, \sigma)$. Inversement, $val^-(x, \sigma)$ si et seulement s'il existe deux termes t_1 et t_2 de $\mathcal{L}_A$ tels que $x = \ulcorner t_1 = t_2 \urcorner$ et $Den(t_1, \sigma) \neq Den(t_2, \sigma)$.

On peut alors définir inductivement la notion de satisfaction d'une formule x de $\mathcal{L}_A$ par une assignation σ.

Définition de la satisfaction $Sat(x, \sigma)$ est le plus petit ensemble tel que :
$Sat(x, \sigma)$ si et seulement si x est atomique et $val^+(x, \sigma)$
ou il existe y tel que $x = \ulcorner \overline{\neg} y \urcorner$ et $\neg Sat(y, \sigma)$
ou il existe y, z, tels que $x = \ulcorner y \overline{\wedge} z \urcorner$ et $Sat(y, f)$ et $Sat(z, \sigma)$
ou il existe y, z, tels que $x = \ulcorner y \overline{\vee} z \urcorner$ et $Sat(y, \sigma)$ ou $Sat(z, \sigma)$
ou il existe y, z, tels que $x = \ulcorner y \overline{\Rightarrow} z \urcorner$ et si $Sat(y, \sigma)$ alors $Sat(z, \sigma)$
ou il existe une formule y et une variable v, $x = \ulcorner \overline{\forall} v y \urcorner$ et pour toute v-variante σ' de σ, $Sat(y, \sigma')$.[23]

tacitement partout que les théories considérées contiennent ces ressources minimales. Afin de ne pas alourdir les notations et obscurcir le propos, nous nous permettrons de nombreux abus de notation en ne distinguant pas, sauf lorsque c'est nécessaire, entre les énoncés, noms d'énoncés, et codes d'énoncés.

23. Une fonction d'assignation σ' est une v-variante d'une fonction d'assignation σ si elle est

La relation de satisfaction Sat est définie comme la plus petite relation satisfaisant les clauses ci-dessus : cette relation existe-t-elle ? On touche ici à la raison pour laquelle la notion de satisfaction n'est pas définissable explicitement pour tout langage-objet. Dans le cas en effet où les quantificateurs du langage-objet parcourent tout ce qui existe, alors cette relation ne peut pas être définie dans le métalangage. Il suffit pour s'en convaincre d'observer que la classe des couples dont le premier élément est $\overline{x = x}$ et qui appartiennent à Sat n'est pas un ensemble, et que par conséquent Sat n'est pas définie dans un tel cas.

La définition de la vérité pour les énoncés de $\mathcal{L}_A$ suit de la définition de la satisfaction de la même façon que, dans la définition standard de la vérité dans une structure d'interprétation, cette dernière suit de la définition de la satisfaction dans une structure :

Définition de la vérité pour le langage de l'arithmétique x est un énoncé vrai de $\mathcal{L}_A$ si et seulement si il est satisfait par toute assignation σ.

Observons que dans le cas d'un langage comme celui de l'arithmétique qui contient des noms de chacun des objets de l'univers du discours, il est possible de définir la vérité d'un énoncé de $\mathcal{L}_A$ sans faire de détour par la notion de satisfaction relativement à une assignation. On définit d'abord la dénotation des termes clos en posant que $Den(\overline{0}) = 0$, que si t est un terme clos alors $Den(S(t)) = S(Den(t))$, etc. Puis on définit la vérité pour les énoncés atomiques : $val^+(\phi)$ si et seulement si il existe deux termes clos t_1 et t_2 tels que $\phi = \ulcorner t_1 = t_2 \urcorner$ et $Den(t_1) = Den(t_2)$. Un abus de notation – partout où cela est possible – permet d'alléger la lecture sans créer d'ambiguïté préjudiciable à la compréhension. On peut alors définir inductivement le prédicat de vérité comme le plus petit ensemble tel que :

1. Si ϕ est un énoncé atomique de L_A : $Vr(\phi) \leftrightarrow val^+(\phi)$
2. Si ϕ est un énoncé de L_A : $Vr(\neg\phi) \leftrightarrow \neg Vr(\phi)$
3. Si ϕ et ψ sont des énoncés de L_A : $Vr(\phi \wedge \psi) \leftrightarrow Vr(\phi)$ et $Vr(\psi)$
4. Si ϕ et ψ sont des énoncés de L_A : $Vr(\phi \vee \psi) \leftrightarrow Vr(\phi)$ ou $Vr(\psi)$
5. Si ϕ et ψ sont des énoncés de L_A : $Vr(\phi \rightarrow \psi) \leftrightarrow$ si $Vr(\phi)$ alors $Vr(\psi)$
6. Si $\phi(x)$ est une formule de L_A contenant au plus une variable libre x : $Vr(\forall x \phi(x)) \leftrightarrow$ pour tout terme clos t, $Vr(\phi(t/x))$ (où $\phi(t/x)$ est

identique à σ sauf éventuellement en la variable v, à laquelle elle peut ou non assigner une valeur différente de celle que lui assigne σ.

l'énoncé obtenu par substitution dans $\phi(x)$ du terme clos t à toutes les instances de la variable libre x.

La définition ainsi obtenue répond-elle au cahier des charges présenté pour commencer ? Oui : on pourrait montrer qu'elle est bien adéquate au sens de la convention-T[24] ; elle ne repose en outre sur aucune notion sémantique non-définie, uniquement sur des notions arithmético-syntaxiques et ensemblistes[25].

La question reste ouverte de savoir si un usage cohérent d'un prédicat de vérité adéquat est possible pour un langage pour lequel la définition explicite est, elle, impossible. Sur cette question Tarski observe qu'un prédicat primitif adéquat de vérité pour un langage quelconque peut toujours être introduit de façon cohérente dans le langage si cette introduction respecte la distinction entre langage-objet (le langage de départ) et métalangage (contenant le prédicat de vérité) en limitant l'application du prédicat vérité à des énoncés ne contenant pas le prédicat de vérité lui-même :

Théorème
Soit T une théorie dans un langage $\mathcal{L}$ ne contenant pas de prédicat de vérité. Si T est consistante, alors $T \bigcup \{Vr(\phi) \leftrightarrow \phi : \phi \in \mathcal{L}\}$ aussi.

L'ensemble $\{Vr(\phi) \leftrightarrow \phi : \phi \in \mathcal{L}\}$ des biconditionnels-T pour un prédicat de vérité constitue une théorie axiomatique de la vérité, que l'on note en général DT dans la littérature contemporaine[26], à la fois adéquat au sens de Tarski et consistante. Tarski montre en fait un peu plus que la consistance de $T \bigcup DT$ relativement à T. Il montre en effet que DT étend conservativement T ce qui signifie essentiellement que l'ajout de DT à T non seulement n'entre pas en conflit avec T, mais que cette ajout ne peut jamais entrer en conflit avec aucune extension ultérieure de T qui serait formulée dans le langage de T lui-même. Plus précisément :

Définition des extensions conservatives
Soit T une théorie d'un langage $\mathcal{L}$ et T' une extension de T dans un langage $\mathcal{L}'$. T' est une extension conservative T si et seulement si, pour tout énoncé

24. Par une induction dans la méta-métathéorie sur la complexité des formules. Voir par exemple Tarski [1983], p. 195.

25. On notera que la notion d'assignation d'un objet à une variable n'est pas une notion sémantique mais une notion purement mathématique, et que la notion de traduction d'un langage dans un autre n'est pas mobilisée dans la définition mais seulement dans le critère d'adéquation formulé dans un tiers langage. Voir Milne [1997] pour une discussion plus approfondie des ressources mobilisées par la définition tarskienne.

26. En référence à la « théorie décitationnelle » (*disquotational*) de la vérité parfois attribuée à Quine. Voir la dernière section du chapitre pour quelques précisions.

$\phi \in \mathcal{L}$:

$$T' \vdash \phi \quad \Rightarrow T \vdash \phi$$

Le résultat positif de Tarski peut donc être formulé de la façon suivante :

Conservativité de DT
Soit T une théorie dans un langage $\mathcal{L}$. Alors $T \bigcup DT$ est conservative sur T.

La preuve repose sur le fait que toute dérivation dans $T \bigcup DT$ mobilise au plus un nombre fini de biconditionnels-T parmi ses prémisses. Or, comme nous l'avons noté plus haut[27] en passant il est possible de définir explicitement dans T un prédicat de vérité partiel ne s'appliquant qu'à un nombre fini d'énoncés de $\mathcal{L}$; on peut donc réécrire la dérivation dans T en remplaçant partout le prédicat de vérité par sa définition.[28]

2. 4 Fécondité de la définition

Que permet la définition de la vérité ? Il est à noter que, même lorsque la définition explicite de la vérité pour un langage est possible, cette définition ne donne pas de critère de vérité, au sens il n'est pas possible dans la métathéorie où se tient cette définition de décider, pour chaque énoncé du langage-objet, s'il est vrai ou non[29]. La définition explicite de la vérité pour un langage $\mathcal{L}$ dans une théorie formulée dans un métalangage permet néanmoins de démontrer un certain nombre de propositions[30].

1. *Principe de contradiction dans le langage-objet*. Pour tout énoncé ϕ du langage-objet, $\neg Vr(\phi)$ ou $\neg Vr(\neg\phi)$
2. *Principe du tiers exclu dans le langage-objet*. Pour tout énoncé ϕ du langage-objet, $Vr(\phi)$ ou $Vr(\neg\phi)$

Si la théorie dans laquelle est définie le prédicat de vérité pour le langage-objet $\mathcal{L}$ est une extension d'une théorie T de $\mathcal{L}$, on peut en outre prouver les propositions suivantes dans la théorie étendue[31] :

27. Note 18.
28. Voir Tarski [1935] théorème III.
29. Sur la distinction entre définition et critère de vérité on pourra consulter Rivenc [2001].
30. Voir Tarski [1983], section ?
31. Ici Pr_T est un prédicat de $\mathcal{L}$ qui représente la prouvabilité dans T. On suppose que ce prédicat satisfait les conditions de prouvabilité mentionnées chap. 20 section 4.2.

Principe de réflexion global
Tous les théorèmes de T sont vrais :

$$\forall \phi \in \mathcal{L}\ (Pr_T(\phi) \rightarrow Vr(\phi)).$$

On se souvient en outre que les recherches fondationelles des années trente avait été marquées par les théorèmes d'incomplétude de Gödel, que l'on peut formuler ici de la façon suivante[32] :

Premier théorème d'incomplétude Si T est consistante et contient les ressources pour décrire sa propre syntaxe[33], alors il existe un énoncé ϕ de $\mathcal{L}$ tel que :

$$T \nvdash \phi \quad \text{et} \quad T \nvdash \neg\phi$$

Second théorème d'incomplétude Si T est consistante, $T \nvdash CONS_T$.

où $CONS_T$ est l'énoncé $\neg Pr_T(0 = 1)$ qui affirme que l'énoncé $0 = 1$ n'est pas prouvable dans T, autrement dit que T est consistante[34].

Dans ce contexte, la méthode de définition de la vérité développée par Tarski permet d'ajouter à ces résultats un codicille intéressant. Dans la (méta-)théorie MT qui étend T et qui contient la définition de la vérité pour le langage de T, la proposition suivante devient en effet un théorème :

Théorème de consistance T est consistante.

Ce résultat est une conséquence directe du principe de réflexion. On peut en effet raisonner dans MT de la façon suivante : tous les théorèmes de T sont vrais. De plus, s'il est vrai que $0 = 1$ alors $0 = 1$ (adéquation du prédicat de vérité). Or $0 \neq 1$ (on suppose que MT est une extension de T et contient l'arithmétique élémentaire). Donc il n'est pas vrai que $0 = 1$. Donc $0 = 1$ n'est pas un théorème de T. Donc T est consistante[35].

32. Voir chap. 20 pour une présentation rigoureuse des résultats d'incomplétude.
33. éventuellement *via* codage, comme c'est le cas de l'arithmétique de Robinson Q ou de Peano PA – voir chapitre 20.
34. Voir chapitre 20.
35. Soit, plus explicitement :
    1. $MT \vdash \forall \phi \in \mathcal{L}(Pr_T(\phi) \rightarrow Vr(\phi)$ [par hypothèse]
    2. $MT \vdash Vr(0 = 1) \rightarrow 0 = 1$ [adéquation de Vr]
    3. $MT \vdash \neg(0 = 1)$ [par hypothèse MT est une extension de T et contient l'arithmétique élémentaire]
    4. $MT \vdash \neg Vr(0 = 1)$ [par 2 et 3]

Les preuves de consistance de T données dans une extension de T ne peuvent renforcer notre conviction dans la consistance de T[36] – le risque d'inconsistance est plus grand dans l'extension que dans T elle-même – mais le résultat explicite le lien conceptuel entre vérité et consistance. Dans la même veine, il est possible d'établir dans la métathéorie que la vérité (pour un langage suffisamment riche) ne peut être réduite à la prouvabilité dans un système formel donné.

Irréductibilité 1 Soit $\mathcal{L}$ aussi riche que le langage de l'arithmétique, et T une théorie dans $\mathcal{L}$. Il y a un énoncé vrai de $\mathcal{L}$ qui n'est ni prouvable ni réfutable dans T.

Ou, formulé un peu différemment :

Irréductibilité 2 Si $\mathcal{L}$ est suffisamment riche, l'ensemble des énoncés vrais de $\mathcal{L}$ n'est pas récursivement axiomatisable[37].

Nous avons observé[38] que certaines théories – essentiellement celles dans lesquelles il est possible d'interpréter l'arithmétique de Robinson – peuvent contenir à titre de partie une description de leur propre syntaxe, et en particulier prouver ou réfuter des énoncés portant sur la possibilité de prouver ou de réfuter en leur sein certains énoncés, et la question de savoir si une théorie de ce genre est ou non consistante est une question formulable dans le langage de la théorie elle-même. Or lorsque nous sommes en situation de définir explicitement la vérité pour le langage $\mathcal{L}$ d'une telle théorie T dans une extension de cette théorie formulée dans un métalangage, le théorème de consistance montre, en conjonction avec le second théorème d'incomplétude de Gödel, que l'extension est non-conservative.

Que reste-t-il de ces usages théoriques de la notion de vérité dans le cas où la définition explicite de la vérité n'est pas possible pour le langage-objet ? Nous avons vu alors que Tarski suggérait l'introduction d'un prédicat primitif de vérité gouverné par les axiomes DT dont l'innocuité pouvait être garantie. Las, l'innocuité de DT n'est que l'autre face de sa faiblesse : aucun des résultats précédents ne peut être dérivé dans l'extension d'une théorie T par les seuls axiomes aléthiques de DT : ni les formulations sémantiques du tiers exclu et du principe de contradiction, ni le principe de

5. $MT \vdash \neg Pr_T(0 = 1)$ [4 et 1]
6. $MT \vdash T$ est consistante [reformulation de 5]

36. Tarski [1983], p. 236-237 section ?
37. Sur la notion d'ensemble récursivement énumérable, voir le chapitre 20.
38. Voir note 21.

réflexion sur T, ni le théorème de consistance de T, ni même la thèse d'irréductibilité de la vérité à la prouvabilité dans T[39]. Pour les mêmes raisons, DT ne permet pas d'établir les lois de compositionnalité qui régissent l'interaction du prédicat de vérité et des constantes logiques et qui figuraient en bonne place dans la définition explicite de la vérité vue plus haut.

Cette dernière remarque suggère une alternative à DT pour gouverner un prédicat primitif de vérité. Pourquoi en effet ne pas reprendre à titre d'axiomes les lois de composition de la vérité et des connecteurs logiques, ceux-là mêmes qui figurent à titre de sous-formules dans la définition explicite de la vérité ? En tirant parti des particularités du langage de l'arithmétique notées plus haut, cette théorie axiomatique compositionnelle (désormais CT) peut être présentée de façon ramassée :

La théorie compositionnelle CT

1. Pour toute formule atomique $\phi \in \mathcal{L}_{PA} : Vr(\phi) \leftrightarrow val^{+}(\phi)$, où val^{+} est un prédicat de vérité défini pour les formules atomiques de T via la notion de dénotation (voir plus haut).
2. Pour toute formule $\phi \in \mathcal{L}_{PA} : Vr(\neg\phi) \leftrightarrow \neg Vr(\phi)$
3. Pour toute formule $\phi \in \mathcal{L}_{PA} : Vr(\phi \wedge \psi) \leftrightarrow Vr(\phi) \wedge Vr(\psi)$
4. Pour toute formule $\phi \in \mathcal{L}_{PA} : Vr(\phi \vee \psi) \leftrightarrow Vr(\phi) \vee Vr(\psi)$
5. Pour toute formule $\phi \in \mathcal{L}_{PA} : Vr(\forall x\phi) \leftrightarrow \forall t Vr(\phi(t/x))$[40]

Cette théorie présente des propriétés intéressantes. Dans $T \bigcup CT$ on peut en effet prouver, non seulement les lois de composition de la vérité et des constantes logiques, mais également le principe général de contradiction et du tiers exclu ou encore, en raisonnant par récurrence sur la préservation de la vérité dans l'application des règles de déduction, le principe de réflexion sur T et le théorème de consistance de T ainsi que la thèse

39. Nous avons vu que l'extension de T par DT est conservative, ce qui exclut la preuve du théorème de consistance de T. Le théorème de consistance est une conséquence du principe de réflexion, ce dernier n'est donc pas davantage prouvable que le premier. Les principes de contradiction et du tiers exclu sont des énoncés généraux sur le prédicat de vérité, mais aucun énoncé général portant sur la notion de vérité ne peut être dérivé de la seule donnée des biconditonnels-T – intuitivement, toute dérivation dans $T + DT$ ne mobilise qu'un nombre fini de biconditionnels, et donc ne contraint l'application du prédicat de vérité que sur un nombre fini d'énoncés.

40. Cette dernière clause doit s'entendre de la façon suivante : l'énoncé $\ulcorner \forall x\phi \urcorner$ est vrai si et seulement si pour tout terme t du langage, l'énoncé obtenu à partir de ϕ en substituant t à ϕ est vrai. Pour que cette définition soit correcte, il faut supposer que le langage contient des termes pour chaque élément du domaine du discours du langage. Voir plus haut la définition de la vérité pour le langage de l'arithmétique.

d'irréductibilité, c'est-à-dire les résultats majeurs que nous attendons de la possibilité de raisonner sur la vérité[41].

Faut-il pour conclure se réjouir et déclarer que Tarski a dompté les paradoxes de la vérité ? Nous avons vu que la solution tarskienne des paradoxes était fondée sur la distinction entre langage-objet et métalangage. Ne peut-on donc rendre raison d'un usage du prédicat de vérité appliqué à des énoncés mobilisant la notion de vérité ? Il ne faut pas se méprendre sur les limites que la discipline tarskienne des niveaux de langage impose à l'usage des prédicats de vérité, en particulier quant à la possibilité d'user de façon consistante d'un prédicat de vérité pour un langage contenant lui-même un prédicat de vérité. En effet la même discipline qui permet d'introduire ou de définir un prédicat de vérité pour un langage-objet qui n'en contient pas peut être reconduite au besoin pour définir ou introduire axiomatiquement dans un nouveau (méta-)métalangage un prédicat de vérité pour les énoncés du premier métalangage devenu le langage-objet de ce nouveau (méta-)métalangage. La hiérarchie des métalangages $M_1\mathcal{L}$, $M_2\mathcal{L}$, ..., $M_\alpha\mathcal{L}$ et de leurs prédicats de vérité Vr_1, Vr_2, ..., Vr_α,... ne s'appliquant qu'aux énoncés des langages d'indices inférieurs strictement à leur propre indice, peut ainsi se poursuivre à l'infini[42]. Ce n'est pas dire toutefois que la discipline tarskienne est la bonne : la camisole fonctionne, mais est-elle pertinente ? En s'interdisant tout écart à la logique classique et en insistant sur l'inaltérabilité de la condition d'adéquation (que Tarski maintient même dans le cas d'une introduction d'un prédicat primitif de vérité dans un langage en faisant vivre ce qu'on pourrait appeler le fantôme des « niveaux de langage »), Tarski s'interdit de rendre raison d'une affirmation aussi bénigne que

« "la neige est blanche" est vraie » est vrai

41. La précision concernant l'emploi d'un raisonnement par récurrence sur la vérité a son importance. Si l'on ne tolère les applications du schéma d'induction qu'à des formules du langage-objet ou du langage de la syntaxe, on ne peut pas dériver le principe de réflexion sur T, ni le théorème de consistance. En effet, si cette restriction est forcée, alors $T \bigcup CT$ devient en fait une extension conservative de T. Dans cette extension, les principes généraux de contradiction et du tiers exclu sont en revanche toujours des théorèmes. Le revers de la fécondité démonstrative de cette théorie compositionnelle est sa non-conservativité, et la conséquence de cette non-conservativité est que la consistance de la théorie de base ne garantit plus la consistance de son extension aléthique : montrer la fécondité de l'emploi du prédicat de vérité ou montrer son innocuité, il faut choisir son projet. Pour examen logiquement approfondi et systématique de la des différentes théories de la vérité, nous renvoyons le lecteur à l'ouvrage de référence Halbach [2014].

42. Plus précisément dans les ordinaux transfinis. Tarski ne présente pas cette idée dans sa monographie sur la vérité. Pour une présentation détaillée des hiérarchies de prédicat de vérité *à la Tarski* on pourra se reporter à Halbach [1995].

dont le prédicat de vérité est, selon toute vraisemblance, appliqué à un énoncé comportant le même prédicat de vérité et qui, au contraire de l'énoncé du Menteur, est logiquement parfaitement innocent. La discipline tarskienne est insatisfaisante au moins en ceci qu'elle fait l'économie d'une analyse qui permettrait de distinguer les énoncés pathologiques d'autres applications non pathologiques du prédicat de vérité à des énoncés contenant ce prédicat de vérité. Cette analyse en retour ne peut être purement syntaxique et doit mobiliser des concepts sémantiques. Montrer que l'on peut définir la notion de vérité sans s'appuyer sur des notions sémantiques primitives, ou caractériser sémantiquement les énoncés paradoxaux, fallait-il donc choisir ? Nous avons vu que la réponse au paradoxe par la distinction des niveaux de langages était commandée par le choix de ne pas toucher à la logique classique ni à la condition d'adéquation. Voyons à présent si, en relâchant l'une ou l'autre de ces exigences, il est possible de rendre compte de la possibilité d'une application cohérente d'un prédicat de vérité aux énoncés du langage auquel il appartient.

3. Vrai de vrai : Kripke

Le travail de Kripke peut recevoir deux interprétations. Selon la première, il appartient à la famille des solutions non classiques aux paradoxes ; selon la seconde c'est la théorie naïve de la vérité, et non la logique classique, qui y est amendée. Cette double possibilité témoigne en fait de la fécondité d'outils logiques et conceptuels qui demeurent fondamentaux pour comprendre tous les développements ultérieurs des solutions aux paradoxes de la vérité. Pour cette raison nous les présentons en détail.

3. 1 La théorie de Kripke

Kripke remarque que le bon diagnostic des problèmes issus des applications itérées du prédicat de vérité ne peut pas être syntaxique, mais seulement sémantique. Dans une interprétation des applications itérées de la vérité en termes de hiérarchie de langages, certains énoncés parfaitement non problématiques ne peuvent pas recevoir leur conditions de vérité intuitives. Considérons l'exemple suivant :

— Abel dit :
  — a1) 2+2=4
  — a2) Tout ce que dit Béatrice est faux.
— Béatrice dit :
  — b1) Tout ce que dit Abel est faux.

a1), a2) et b1) semblent exprimer des propositions déterminées. Intuitivement, a1) est vrai, donc b1) doit être faux, et par conséquent a2) doit être vrai. Si les prédicats de vérité étaient implicitement typés de sorte que Vr_n ne s'applique avec sens qu'à des énoncés ne contenant pas de prédicat de vérité Vr_p pour $n \leq p$, on ne pourrait pas assigner de « type » aux prédicats de vérité qui rende justice aux conditions de vérité intuitives de ces énoncés.

En fait, remarque Kripke, la plupart des énoncés paradoxaux ne sont tels qu'en vertu de certaines circonstances empiriques. La pathologie des énoncés paradoxaux n'est pas un trait intrinsèque de ces énoncés et « Tous les crétois sont des menteurs » ou « Tout ce que Nixon dit à propos du Watergate est faux », peuvent ou non être paradoxaux selon les circonstances : tout dépend ce que disent le crétois ou de ce que dit Nixon à propos du Watergate. En fait, observe Kripke, nos assertions utilisant le prédicat de vérité sont « risquées »[43], parvenant ou échouant à exprimer une proposition selon les circonstances. Se garantir de tout risque de paradoxes en employant des prédicats de type ou de niveaux fixés par avance, c'est donc prévenir l'entorse par l'amputation. Ne peut-on rendre compte de la possibilité qu'un langage $\mathcal{L}$ contiennent son propre un prédicat de vérité pour $\mathcal{L}$ pour peu qu'on admette que dans certaines circonstances certains énoncés n'ont pas de valeur de vérité (ou n'expriment pas de propositions si c'est la même chose) ?

Modèles partiels

Pour montrer qu'un usage consistant d'une notion de vérité non-typée est possible, Kripke construit un modèle d'un langage $\mathcal{L}_{\mathrm{Vr}}$ contenant son propre prédicat de vérité Vr. Dans cette structure d'interprétation, on admettra que les énoncés peuvent recevoir non pas deux, mais trois valeurs sémantiques, que l'on notera conventionnellement $1, 0, \frac{1}{2}$. Intuitivement ces trois valeurs sémantiques reflètent respectivement le fait qu'il est permis

43. Kripke [1975], p. 692.

d'asserter un énoncé, qu'il est permis d'asserter sa négation, ou qu'il n'est permis ni de l'asserter ni d'asserter sa négation étant donné les circonstances décrites dans le modèle. Les énoncés « ordinaires » recevront ainsi une valeur sémantique classique correspondant à l'idée que, selon les énoncés considérés, on peut les asserter ou asserter leur négation ; mais certains énoncés contenant le prédicat de vérité pourront ne pas recevoir de valeur sémantique ordinaire, reflétant l'intuition qu'on ne peut ni les asserter ni asserter leur négation. Dans ce nouveau cadre sémantique, à quoi reconnaîtra-t-on que Vr est bien un prédicat de vérité pour le langage ? L'analyse tarskienne mettait en avant l'importance des biconditionnels-T : mais faut-il comprendre que ce qui est essentiel est que ϕ et $Vr(\phi)$ aient même valeur sémantique (soient « équi-assertables »), ou bien que le biconditionnel lui-même soit assertable ? Dans une sémantique classique, le fait qu'un énoncé ϕ reçoive la même valeur sémantique que l'énoncé $Vr(\phi)$ est équivalent au fait que le biconditionnel $Vr(\phi) \rightarrow \phi$ reçoive la valeur sémantique 1. Mais dans le nouveau cadre que nous envisageons, ce n'est plus forcément le cas, tout dépendra de la façon dont nous définirons le comportement sémantique du biconditionnel matériel. Le critère d'adéquation que Kripke choisit de retenir est le plus direct :

Adéquation (au sens de Kripke)
Le prédicat de vérité $\mathrm{Vr}_{\mathcal{L}}$ est adéquat si et seulement si tout énoncé ϕ de $\mathcal{L}$ a la même valeur sémantique que l'énoncé $\mathrm{Vr}(\phi)$.

Nous décrivons l'interprétation de $\mathcal{L}_{\mathrm{Vr}}$ en suivant deux étapes. Dans la première nous réglons la question générale de la définition de la satisfaction des formules dans une sémantique à trois valeurs. Dans la seconde nous examinons comment produire une sémantique de ce genre adéquate pour le langage $\mathcal{L}_{\mathrm{Vr}}$. Relativement à la première question, un premier problème concerne la compositionnalité du langage : comment la valeur d'une formule quelconque est-elle déterminée à partir de la valeur sémantique de ses sous-formules ? D'un point de vue purement formel les options sont nombreuses, mais certaines contraintes s'imposent naturellement. En particulier on exigera que la nouvelle sémantique soit une généralisation de la sémantique classique en ce sens que, lorsque des formules ayant des valeurs sémantiques classiques 1 ou 0 sont composées avec les connecteurs et quantificateurs habituels, la formule composée reçoive alors sa valeur sémantique habituelle[44]. Les deux schèmes de composition fort et faible de

44. Dans ce cas, selon Kripke, nous ne changeons pas de logique mais mettons en œuvre un moyen de traiter les énoncés sémantiquement défectueux. Voir la remarque de Kripke [1975],

Kleene répondent à cette contrainte (voir tableau).

		Kleene faible K_3^-			Kleene fort K_3^+		
p	q	$\neg p$	$p \vee q$	$p \wedge q$	$\neg p$	$p \vee q$	$p \wedge q$
1	1	0	1	1	0	1	1
1	0	0	1	0	0	1	0
0	1	1	1	0	1	1	0
0	0	1	0	0	1	0	0
1	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	1	$\frac{1}{2}$
$\frac{1}{2}$	1	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$
0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	0
$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

TABLE 1.1 – Modes de composition

On peut interpréter ces différentes valeurs de vérité comme reflétant un certain type d'information disponible sur les énoncés. Dans cette perspective, la valeur sémantique $\frac{1}{2}$ signifie que les conditions d'assertabilité de l'énoncé ne sont pas totalement déterminées, à la différence de celles des énoncés recevant les valeurs classiques, pour lesquels il est parfaitement déterminé s'ils peuvent être affirmés ou niés. L'ordre partiel « de détermination » (on parle aussi parfois d'ordre « informationnel ») engendré sur les valeurs sémantiques peut être représenté par l'arbre suivant :

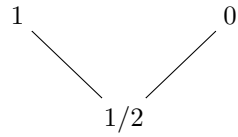

FIGURE 1.1 – L'ordre informationnel sur les valeurs sémantiques

Chacun des deux schèmes fort et faible correspond à une interprétation différente de l'idée qu'un énoncé recevant la valeur sémantique $\frac{1}{2}$ ne peut être asserté faute d' « information disponible » ou de « détermination ». Le schème faible est naturel par exemple si la valeur $\frac{1}{2}$ assignée à un énoncé est interprétée comme absence de signification de l'énoncé, tandis que le schème fort s'impose si la valeur $\frac{1}{2}$ représente le fait qu'un énoncé doué de signification possède néanmoins un statut sémantique « indéterminé » pour

p. 700, note 18.

une raison ou une autre[45]. Ces options n'épuisent pas les possibilités[46], ce qui est crucial cependant pour la construction à venir est que les schèmes de compositions retenus jouissent de la propriété suivante de monotonie :

Monotonie des schèmes de composition de Kleene Si la valeur sémantique d'une sous-formule augmente (selon l'ordre informationnel), alors la valeur sémantique de l'énoncé global augmente (selon l'ordre informationnel).

Venons-en aux énoncés atomiques et à la façon dont l'interprétation des prédicats détermine leur valeur sémantique. Dans la sémantique classique, les prédicats sont habituellement interprétés par la donnée de leur extension – un sous-ensemble du domaine d'une structure d'interprétation – et cette donnée détermine à la fois les objets dont on peut affirmer le prédicat et, complémentairement, les objets du domaine dont on peut nier le prédicat (tous ceux qui ne sont pas dans l'extension du prédicat). Mais puisque nous souhaitons faire une place à des objets et des prédicats tels qu'un tel prédicat ne peut être ni affirmé ni nié de tels objets, il faut aménager le cadre habituel. La proposition de Kripke est d'interpréter chaque prédicat par un couple (E, A) de sous-ensembles disjoints du domaine de la structure. E est l'extension du prédicat et contient les objets dont on peut affirmer le prédicat, et A son anti-extension, contenant les objets du domaine dont on peut nier que le prédicat s'y applique.

Dans l'application qui nous intéresse, ce cadre sémantique est mis au service d'une preuve de consistance d'un langage contenant son propre prédicat de vérité. On supposera donc donné un langage de base $\mathcal{L}$ non problématique, ne contenant pas de notions sémantiques (par exemple le langage

45. Mentionnons également un autre schème d'évaluation sémantique des formules qui a un intérêt propre, – en fait il s'agit plutôt d'un raffinement des schèmes précédents – le schème de supervaluation. Ce dernier n'est pas compositionnel (la valeur sémantique du composé ne dépend pas seulement de la valeur sémantique des composants) mais il a la vertu de sauver les schémas classiques de tautologie – $\phi \vee \neg\phi$ etc. On peut se représenter l'intention qui la motive de la façon suivante. Si un énoncé qui reçoit la valeur sémantique $\frac{1}{2}$ est en attente d'un supplément d'information qui déterminera s'il peut être asserté ou nié alors, pour certains énoncés comportant des sous-énoncés dont les conditions d'assertabilité sont encore indéterminées, on peut néanmoins déjà voir qu'il ne fera aucune différence pour leur assertabilité d'apprendre au bout du compte si le composant doit être asserté ou nié. En effet si toute résolution de l'indétermination de l'assertabilité des composants converge sur le caractère assertable de l'énoncé global alors les conditions d'assertabilité de ce dernier sont en fait déjà déterminées. Ainsi, bien qu'un énoncé comme le menteur (λ) puissent se révéler ne pouvoir recevoir que la valeur sémantique $\frac{1}{2}$, le schème de supervaluation fait droit à l'idée que cela ne remet pas en cause la validité de toutes les instances du tiers exclu.

46. Encore faut-il qu'elles correspondent à une interprétation philosophiquement intéressante des valeurs sémantiques.

de l'arithmétique vu plus haut). On considérera une structure d'interprétation classique de ce langage, structure que l'on notera $\mathcal{M}$ et dont on suppose que le domaine contient, entre autres objets, tous les énoncés du langage $\mathcal{L}$ ainsi que du langage obtenu en ajoutant au vocabulaire de $\mathcal{L}$ un prédicat de vérité Vr. Ce que veut montrer Kripke, c'est qu'il est possible d'étendre cette structure classique $\mathcal{M}$, en une structure contenant en plus l'interprétation (non-classique) d'un prédicat de vérité adéquat au sens précisé plus haut. Le langage L_{Vr}[47] va donc être interprété dans une structure $\langle \mathcal{M}, (E, A)\rangle$ où A et E sont respectivement l'extension et l'anti-extension de Vr. Plus formellement, on peut définir la satisfaction dans une structure d'interprétation non-classique pour L_{Vr} par une assignation σ de la façon suivante[48].

Définition Satisfaction dans une structure partielle (schème fort de Kleene). Pour les formules ne contenant pas le prédicat Vr, la définition de la satisfaction est la définition habituelle. Pour les autres formules de L_{Vr} on procède par récurrence[49] :

— Formules atomiques :
$|Vr(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$ ssi $\sigma(x) \in E$
$|Vr(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$ ssi $\sigma(x) \in A$
$|Vr(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = \frac{1}{2}$ sinon.

— Négation :
$|\neg\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$ ssi $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$
$|\neg\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$ ssi $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$

— Disjonction :
$|\phi(x) \vee \psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$ ssi $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$ ou $|\psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$
$|\phi(x) \vee \psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$ ssi $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$ et $|\psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$

— Conjonction :
$|\phi(x) \wedge \psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$ ssi $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$ et $|\psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 1$
$|\phi(x) \wedge \psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$ ssi $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$ ou $|\psi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle} = 0$

— Quantification Universelle :

47. C'est-à-dire le langage $\mathcal{L}$ étendu avec un prédicat de vérité Vr.
48. La définition de la satisfaction que nous avons choisie utilise le schème fort de Kleene car c'est celle que Kripke semble juger la plus naturelle, mais on adapterait naturellement la définition à des schèmes de composition différents.
49. Dans la définition qui suit, la notation $|\phi(x)|^{\sigma}_{\langle \mathcal{M},(E,A)\rangle}$ désigne la valeur sémantique de la formule ϕ pour l'assignation σ dans la structure d'interprétation $\langle \mathcal{M}, (E, A)\rangle$.

$|\forall x\phi(x)|^{\sigma}_{\langle\mathcal{M},(E,A)\rangle} = 1$ ssi pour toute x-variante σ' de σ,$|\phi(x)|^{\sigma'}_{\langle\mathcal{M},(E,A)\rangle} = 1$

$|\forall x\phi(x)|^{\sigma}_{\langle\mathcal{M},(E,A)\rangle} = 0$ ssi il existe une x-variante σ' de σ telle que $|\phi(x)|^{\sigma'}_{\langle\mathcal{M},(E,A)\rangle} = 0$

— $\phi \rightarrow \psi$ et $\exists x\phi(x)$ sont introduits à titre d'abréviations de $\neg\phi \vee \psi$ et $\neg\forall x\neg\phi(x)$ respectivement.

Interpréter $L_{\mathbf{Vr}}$

Nous avons défini le cadre sémantique qui doit rendre possible l'interprétation d'un langage contenant un prédicat de vérité non-typé. Il reste à montrer que ce langage peut contenir un prédicat de vérité qui soit adéquat, c'est-à-dire tel que, pour tout énoncé ϕ de $\mathcal{L}_{\mathrm{Vr}}$:

$$|Vr(\phi)|^{\sigma}_{\langle\mathcal{M},(E,A)\rangle} = |\phi|^{\sigma}_{\langle\mathcal{M},(E,A)\rangle}$$

Pour y parvenir, la clé est de remarquer qu'une structure avec un prédicat de vérité adéquat peut être vue comme un point fixe d'une certaine opération qui consiste à corriger itérativement l'extension d'un prédicat Vr en l'identifiant à chaque étape à l'ensemble des énoncés qui ont reçu la valeur 1 à l'étape antérieure (et symétriquement son antiextension à l'ensemble des énoncés qui ont reçu la valeur 0) : quand l'itération de l'opération ne change rien à l'extension de Vr (c'est l'idée de « point fixe »), c'est que tout énoncé ϕ qui reçoit la valeur 1 dans la structure est dans l'extension de Vr, et donc que l'énoncé atomique $Vr(\phi)$ reçoit lui aussi la valeur 1 ; et réciproquement si $Vr(\phi)$ reçoit la valeur 1 c'est que ϕ a reçu la valeur 1 à l'étape antérieure, et qu'elle la possède encore puisque cette étape est un point fixe ! (*Idem* pour l'anti-extension.)

Cet opérateur, notons-le Φ, peut être défini formellement sur la classe $\mathcal{C}$ des L_{Vr}-structures qui étendent notre structure de base[50] de la façon suivante :

Opérateur de Kripke (*Kripke Jump*)
Pour toute structure $\langle\mathcal{M},(E,A)\rangle \in \mathcal{C}$:

$$\Phi(\langle\mathcal{M},(E,A)\rangle) = \langle\mathcal{M},(E',A')\rangle$$

avec : $A' = \{\phi : \langle\mathcal{M},(E,A)\rangle \vDash_1 \phi\}$ et $E' = \{\phi : \langle\mathcal{M},(E,A)\rangle \vDash_0 \phi\}$.

50. I.e $\mathcal{C}$ = la classe des structures de la forme $\langle\mathcal{M},(E,A)\rangle$, pour $\mathcal{M}$ fixé.

Mais comment montrer que cet opérateur a un point fixe ? Commençons d'abord par ordonner (partiellement) la classe $\mathcal{C}$ des extensions de $\mathcal{M}$, selon que ces structures permettent d'affirmer ou de nier de façon déterminée le prédicat de vérité d'un plus petit ou plus grand nombre d'énoncés du langage :

Définition de $(\mathcal{C}, \leq)$
Sur la classe $\mathcal{C}$ on définit l'ordre partiel $(\mathcal{C}, \leq)$ de la façon suivante :
$\langle \mathcal{M}, (E, A) \rangle \leq \langle \mathcal{M}, (E', A') \rangle$ ssi $E \subseteq E'$et $A \subseteq A'$.

Dans cet ordre partiel, qui étend à $\mathcal{C}$ l'intuition gouvernant l'ordre informationnel sur les valeurs de vérité, on dira que deux structures u, v, sont cohérentes entre elles s'il existe une structure $w \in (\mathcal{C}, \leq)$ qui les majore toutes les deux – i.e. telle que $u \leq w$ et $v \leq w$. Par exemple, la structure $\langle \mathcal{M}, (E = \{1 + 1 = 2\}, A = \{2 + 1 = 8\}) \rangle$ et la structure $\langle \mathcal{M}, (E = \{3 + 2 = 5, \}, A = \{2 + 1 = 8, 2 = 4\}) \rangle$ sont cohérentes, car elles sont majorées par (entre autres) la structure $\langle \mathcal{M}, (E = \{1 + 1 = 2, 3 + 2 = 5\}, A = \{2 + 1 = 8, 2 = 4\}) \rangle$, qui appartient également à $(\mathcal{C}, \leq)$. Tandis que les structures $\langle \mathcal{M}, (E = \{1 + 1 = 2, Vr(\ulcorner 1 + 1 = 2 \urcorner)\}, A = \{2 + 1 = 8\}) \rangle$ et $\langle \mathcal{M}, (E = \{3 + 2 = 5, \}, A = \{2 + 1 = 8, 2 = 4, Vr(\ulcorner 1 + 1 = 2 \urcorner)\}) \rangle$ ne sont pas cohérentes car aucune structure ne peut les majorer les deux à la fois. La cohérence de deux structures en ce sens reflète l'idée qu'on peut voir les deux structures comme des approximations d'une même structure plus déterminée (« informationnellement »). Prolongeant cette idée, on dit qu'un sous-ensemble quelconque de $(\mathcal{C}, \leq)$ est cohérent dans $(\mathcal{C}, \leq)$ si tous ses éléments sont cohérents deux à deux. On pourrait ainsi se convaincre facilement que, par exemple, l'ensemble infini des éléments de $(\mathcal{C}, \leq)$ pour lesquels l'extension de Vr est vide est un sous-ensemble cohérent de $(\mathcal{C}, \leq)$. Avec ces définitions on peut remarquer à présent la chose suivante :

Proposition
Tout sous-ensemble cohérent Y de $(\mathcal{C}, \leq)$ a un plus petit majorant dans $(\mathcal{C}, \leq)$.

Preuve : On vérifie facilement que la structure obtenue en prenant pour extension de Vr la réunion des extensions assignées à Vr dans les structures appartenant à Y, et en procédant de façon analogue pour l'anti-extension, est le plus petit majorant de Y.

Définition (Ordre cohérent-complet)
On dit que $(\mathcal{C}, \leq)$ est un ordre partiel cohérent-complet si tous ses sous-ensembles cohérents ont un plus petit majorant dans C.

La dernière observation fondamentale est que l'opérateur Φ de Kripke a la propriété d'être monotone pour l'ordre considéré :

L'opérateur Φ est monotone sur $(\mathcal{C}, \leq)$
$\langle \mathcal{M}, (E, A)\rangle \leq \langle \mathcal{M}, (E', A')\rangle \Rightarrow \Phi(\langle \mathcal{M}, (E, A)\rangle) \leq \Phi(\langle \mathcal{M}, (E', A')\rangle)$

La preuve procède par induction sur la complexité des formules et repose fondamentalement sur le comportement des connecteurs et quantificateurs choisis, et plus précisément sur le fait qu'ils sont eux-mêmes monotones sur l'ordre informationnel (la propriété de monotonie du schème de Kleene mentionnée plus haut)[51].

Nous sommes maintenant en mesure d'affirmer le théorème suivant, qui généralise un théorème de Tarski sur les treillis complets[52], et garantit l'existence d'un point fixe pour l'opérateur de Kripke sur $(\mathcal{C}, \leq)$:

Théorème du point fixe
Tout opérateur monotone sur un ordre partiel cohérent-complet possède un point point fixe.[53]

Ce résultat assure donc l'existence d'une interprétation adéquate du langage $\mathcal{L}$ contenant son prédicat de vérité. Cependant cette construction révèle bien plus. On pourrait montrer effet qu'il existe de nombreux points fixes de la construction. L'un d'entre eux est le plus petit et il est intrinsèque

51. Esquisse de preuve : Pour abréger, on note $\Phi(A)$ l'extension de Vr dans la structure $\Phi(\langle \mathcal{M}, (E, A)\rangle)$ (idem pour A, E, E').
- Soit ϕ une formule atomique. Si ϕ ne contient pas de prédicat partiel (i.e. « Vr »), il n'y a rien à dire. Supposons $\phi = Vr(x)$. Si $\phi \in \Phi(A)$ alors $|Vr(x)|\langle \mathcal{M}, (E, A)\rangle) \vDash_1 Vr(x)$. Donc $x \in A$. Donc $x \in A'$. Donc $(\langle \mathcal{M}, (E', A')\rangle) \vDash_1 Vr(x)$. Donc $\phi \in \Phi(A')$. On raisonnerait de même avec E et E'.
- Supposons le résultat vrai pour les formules ψ de complexité au plus n, *i.e.* $(\psi \in \Phi(A) \rightarrow \psi \in \Phi(A')) \wedge (\psi \in \Phi(E) \rightarrow \psi \in \Phi(E'))$.
  Soit $\phi = \neg\psi$. Par définition $\phi \in \Phi(A)$ ssi $(\langle \mathcal{M}, (E, A)\rangle) \vDash_1 \phi$ ssi $(\langle \mathcal{M}, (E, A)\rangle) \vDash_0 \psi$. Par hypothèse de récurrence, si $\psi \in \Phi(E)$ alors $\psi \in \Phi(E')$, i.e. $(\langle \mathcal{M}, (E', A')\rangle) \vDash_0 \psi$. Mais alors $(\langle \mathcal{M}, (E', A')\rangle) \vDash_1 \phi$, i.e. $\phi \in \Phi(A')$.On raisonnerait de même sur les antiextensions.
- Analogue pour les autres connecteurs et les quantificateurs.

52. Théorème dit de Tarski-Knaster. Les treillis complets sont les ordres partiels dont tous les sous-ensembles ont une borne supérieure et une borne inférieure – c'est-à-dire un plus petit majorant et un plus grand minorant.

53. Esquisse de preuve : Soit $(X, \leq)$ un opcc quelconque, et f un opérateur monotone sur $(X, \leq)$. Soit x_0 un élément de X et Y le sous-ensemble de X défini par $Y = \{y : x_0 \leq y \wedge y \leq f(y)\}$. Supposons que Y possède un élément maximal m. Par définition de Y on a $m \leq f(\lambda)$, donc par monotonie $f(\lambda) \leq ff(\lambda)$. Donc $f(\lambda)$ est dans Y aussi et, puisque $m \leq f(\lambda)$ et que m est maximal, $m = f(\lambda)$. Il reste donc à montrer que Y possède bien un élément maximal. Mais ceci est une conséquence du fait que Y est lui-même un opcc (à vérifier) et du lemme de Zorn[54]. Pour une preuve détaillée nous renvoyons le lecteur à Visser [2004] ou à Gupta and Belnap [1993] Théorème 2C :3 p.64 et 2C :5 p. 66.

au sens où il est cohérent avec tout autre point fixe. L'ensemble des points fixes a lui-même une structure d'ordre cohérent-complet et l'ensemble des points fixes intrinsèques une structure de treillis complet, si bien que ce dernier possède donc un plus petit mais également un plus grand élément ! Ce dernier est donc le plus grand point fixe « non-arbitraire » : ne mériterait-il pas d'être lui aussi candidat à l'extension de Vr[55] ?

La résolution du problème de départ révèle que la condition d'adéquation sous-détermine l'extension du prédicat de vérité : si Φ admet en fait de nombreux points fixes, quel est le « bon » candidat ? Kripke suggère la réponse suivante : le plus petit point fixe, car il a une interprétation naturelle comme le résultat d'un processus idéalisé d'apprentissage de la signification du prédicat de vérité. Supposons en effet qu'un locuteur ne sache rien de la sémantique du prédicat de vérité, et que lui soit seulement donné l'explication suivante : on peut asserter $Vr(\phi)$ exactement quand on peut asserter ϕ. Au départ, il ne sait pas appliquer le prédicat de vérité. L'interprétation de son langage est donc représentée par la structure partielle $\langle \mathcal{M}, (\emptyset, \emptyset)\rangle$. Dans ce langage certains énoncés sont assertables : tous ceux qui ne contiennent pas le prédicat de vérité et qui sont vrais en vertu de faits non sémantiques. Appliquant l'instruction qui lui a été fournie pour appliquer le prédicat de vérité, l'agent sait donc maintenant que, puisque l'énoncé « la neige est blanche » est assertable, l'énoncé « "la neige est blanche" est vrai » l'est aussi. Le modèle partiel qui représente le nouvel état de son appréhension du langage après cette première étape n'est autre que $\Phi(\langle \mathcal{M}, (\emptyset, \emptyset)\rangle)$. Dans ce langage où l'extension et l'anti-extension du prédicat de vérité ont été mises à jour, de nouvelles assertions deviennent possibles, dont la reconnaissance entraîne à son tour une nouvelle mise à jour des conditions d'assertabilité de Vr représentée par $\Phi(\Phi(\langle \mathcal{M}, (\emptyset, \emptyset)\rangle))$, etc. Cette suite de modèles partiels, que l'on peut prolonger dans le transfini (en prenant la réunion des interprétations précédentes aux ordinaux limites), est croissante (non strictement)[56] et doit atteindre un point fixe où se stabiliser[57]. Or ce point fixe est le plus petit point fixe de Φ. Il est le candidat

55. L'existence d'un plus petit point fixe peut être obtenue de différentes manières. Nous donnons une autre démonstration ci-dessous, qui reprend celle de Kripke [1975]. La structure de treillis des points fixes intrinsèques est démontrée dans Visser [2004], Th. 20 p. 188 ou Gupta and Belnap [1993], Th. 2C.12 p. 71.

56. Il est clair que $\mathcal{M}, (\emptyset, \emptyset)\rangle \leq \Phi(\langle \mathcal{M}, (\emptyset, \emptyset)\rangle)$, puisque $\langle \mathcal{M}, (\emptyset, \emptyset)\rangle$ est le plus petit élément de la classe des structures partielles. Or Φ est monotone. Donc $\Phi(\mathcal{M}, (\emptyset, \emptyset)\rangle) \leq \Phi(\Phi(\langle \mathcal{M}, (\emptyset, \emptyset)\rangle))$, etc. Une induction transfinie transforme ceci en preuve.

57. Sinon il y aurait une bijection de la classe des ordinaux dans l'ensemble des couples disjoints d'ensemble d'énoncés de L_{Vr}. Cette preuve de l'existence d'un plus petit point fixe de l'opérateur de Kripke ne dépend pas de l'analyse des ordres consistant-complet donnée plus

le plus « naturel » à l'extension du prédicat de vérité au sens donc où c'est celui qui est « engendré » par les instructions données plus haut : la sémantique du plus petit point fixe est celle qui est apprise par approximations successives[58].

Enoncés fondés, énoncés pathologiques

Quel que soit le choix du point fixe, Kripke a donc montré en un sens qu'un langage pouvait contenir un prédicat de vérité pour lui-même. Mais le cadre sémantique proposée offre également un bel outillage pour clarifier nos jugements sur différents types d'énoncés intuitivement bizarres. Les énoncés qui prennent la valeur 0 ou 1 dans le plus petit point fixe, Kripke les appelle « fondés » (« grounded »). Leur valeur de vérité est entièrement déterminée par l'ensemble des faits non-sémantiques et la règle de l'équivalence de l'assertion d'un énoncé et de l'attribution de la vérité à cet énoncé (une forme d'intuition déflationniste relativement à la nature de la vérité). Les énoncés comme le Menteur ou le Véridique

$$(\tau) \qquad \mathbf{Vr}(\tau)$$

n'ont pas de valeur de vérité déterminée dans le plus petit point fixe[59], mais on peut les distinguer. En fait le Menteur n'a de valeur de vérité déterminée dans aucun point fixe[60], tandis que le véridique est vrai dans certains points fixes, et faux dans d'autres, sa valeur de vérité pouvant être assignée arbitrairement[61].

Les énoncés paradoxaux sont ceux qui, parmi les énoncés non-fondés, n'ont de valeur déterminée dans aucun point fixe. Ainsi le menteur est-il paradoxal, mais le véridique, qui peut recevoir une valeur sémantique en-

haut.

58. L'idée que L est la sémantique déterminée par la procédure d'apprentissage par approximations successives devrait être précisée pour trouver un sens cognitif. Le fait que le processus se prolonge dans le transfini n'est pas nécessairement un problème si l'agent peut, sinon parcourir, du moins représenter distinctement chaque ordinal transfini et avec lui l'étape correspondante de ce processus. Kripke note que plus petit point fixe est atteint à l'étape ω_1^{CK} du processus. Or ω_1^{CK} est l'ordinal approché par la suite des ordinaux récursifs, c'est-à-dire des types de bons-ordres décidables.

59. Pour le voir, raisonner par induction sur les ordinaux : le Véridique n'a pas de valeur déterminé dans $\langle \mathcal{M}, (\emptyset, \emptyset) \rangle$, et s'il n'a pas de valeur déterminé dans $\langle \mathcal{M}, (E, A) \rangle$, il n'en a pas non plus dans $\Phi(\langle \mathcal{M}, (E, A) \rangle)$.

60. L'application du « Kripke Jump » Φ envoie le menteur dans l'anti-extension de la structure image s'il était dans l'extension de la structure argument, et vice et versa.

61. Pour le voir, montrer que : si $\tau \in A$ alors (v)$\in \Phi(A)$ et si $\tau \in A$ alors $\tau \in \Phi(A)$

tière dans certains points fixes, n'est pas paradoxal bien qu'il ne soit pas fondé. Nous avons vu en outre que certains énoncés ont toujours la même valeur sémantique dans les points fixes où ils reçoivent une valeur sémantique entière. Ce sont des énoncés qui ont une valeur sémantique « intrinsèque ». À la différence du véridique un énoncé comme « Cette phrase ou sa négation est vraie » reçoit la valeur sémantique $\frac{1}{2}$ ou 1, mais jamais 0, quel que soit le point fixe considéré. On le voit, le cadre sémantique développé par Kripke permet une classification des énoncés en fonction de leur comportement dans différentes circonstances sémantiques et empiriques décrites inscrites dans les modèles de point-fixe, et les distinctions opérées, loin d'être de purs artefacts, ont une correspondance intuitive claire. C'est l'une des vertus de la construction.

Interprétation et limites de la construction de Kripke

Dans le plus petit point fixe (par exemple) qui doit interpréter le prédicat de vérité, le menteur reçoit la valeur sémantique $\frac{1}{2}$. Prenant appui sur notre construction, il est alors tentant de conclure que par conséquent, finalement, le menteur n'est pas vrai. Cette stratégie de revanche du Menteur est toutefois discutable, car elle repose sur l'identification du prédicat « Vrai » et du prédicat « avoir la valeur sémantique 1 », identification à laquelle il faut résister[62]. Dans le langage $\mathcal{L}_{\mathrm{Vr}}$ interprété dans le plus petit point fixe, on ne peut pas dire que le Menteur n'est pas vrai, pas plus qu'on ne dire que le Menteur est vrai, pas plus qu'on ne peut dire que le Menteur est soit vrai soit non vrai, ce qui est représenté dans la construction par le fait que ces énoncés reçoivent tous la valeur sémantique $\frac{1}{2}$.

Si cette question de la revanche peut être réglée de cette façon, il convient de souligner plusieurs limites de la construction. La première limite était d'emblée intégrée dans l'ambition du projet, mais mérite d'être rappelée. La construction de Kripke montre qu'un langage peut contenir un prédicat de vérité non typé et adéquat, mais ne montre pas qu'il peut exister un langage dans lequel il serait possible de définir explicitement la vérité pour ce langage : le langage dans lequel est conduite la construction de Kripke

62. En ce sens la stratégie du « closing-off » suggérée en passant par Kripke n'est peut-être pas judicieuse. Elle permet néanmoins de voir facilement que la construction peut alors être comprise comme une preuve de consistance classique des biconditionnels-T pour la classe des énoncés qui sont dans l'extension du prédicat de vérité Vr dans le point fixe. Voir sur ce point les remarques conclusives de cette section, et dans la section suivante la présentation du point de vue méthodologique développé par Hartry Field.

est un métalangage classique et essentiellement plus riche que le langage-objet contenant son propre prédicat de vérité. En ce sens « le fantôme de la hiérarchie de Tarski est toujours avec nous » (Kripke [1975], p. 714). Mentionnons que d'autres tentatives ont été faites pour s'affranchir un peu plus de cette limite, notamment par Hintikka[63].

Une autre limite de la construction, elle aussi intégrée dans l'ambition du projet kripkéen, réside en ceci qu'elle se borne à montrer qu'un langage donné peut contenir un prédicat de vérité adéquat sans s'aventurer dans la formulation des lois aléthiques qui devraient être valides dans tout langage contenant son propre prédicat de vérité[64]. Certains auteurs ont ultérieurement complété le travail de Kripke sur ce point en explorant différentes possibilités naturellement suggérées par la construction sémantique[65].

Mais au-delà de ces premières limitations, c'est la faiblesse expressive du langage obtenu qui laisse un sentiment d'inachevé. Non seulement la plupart des concepts sémantiques définis dans la théorie kripkéenne appartiennent au métalangage et ne peuvent être exprimés dans le langage contenant son prédicat de vérité (ainsi de « fondé », « paradoxal », « intrinsèque »), mais le langage ne contient pas de ressources expressives propres permettant de caractériser sémantiquement le comportement spécifique d'un énoncé comme le Menteur, par exemple, qui met en échec le tiers exclu. D'autre part la façon dont le résultat de possibilité est obtenu, parce que l'obtention d'un point fixe repose essentiellement sur la monotonie des opérateurs logiques, est consubstantielle à une faiblesse expressive du langage et de la logique associée. Ainsi le langage de Kleene ne contient pas de négation forte, une négation qui prendrait la valeur 1 quand l'affirmation a une valeur sémantique différente de 1 (l'introduction d'une telle négation ruinerait la monotonie de l'opérateur de Kripke). Passe encore pour la négation forte, dont on pourrait peut-être arguer que son besoin est une projection du point du vue du métalangage classique sur le langage-objet. Mais les relations logiques fondamentales entres énoncés validées par la sémantique de Kleene-Kripke ne permettent pas de rendre compte de nos raisonnements les plus ordinaires. Il n'y a ainsi aucun schéma de tautologie dans la logique associée à la sémantique partielle *via* la définition ordinaire

63. Voir Hintikka [1998] et le bilan dans Rouilhan and Bozon [2006].

64. On a vu dans la première section que Tarski exigeait pour sa part la validité des biconditionnels-T.

65. Par exemple Kremer [1988] pour une version possible dans une logique non classique, ou Feferman [1991] dans un cadre classique. Voir les remarques sur les théories axiomatiques dans la section suivante.

de la vérité logique[66]. L'absence d'un conditionnel matériel permettant de valider des règles d'inférence attendues a également concentré l'attention. On peut bien sûr introduire $\phi \rightarrow \psi$ selon la définition habituelle $\neg\phi \vee \psi$, mais cette implication n'est pas l'outil dont nous avons besoin. On peut par exemple vérifier que, pour la définition naturelle de la conséquence logique comme préservation de la valeur sémantique 1, la règle d'inférence $(A \rightarrow B \vDash (C \rightarrow A) \rightarrow (C \rightarrow B))$ n'est pas valide[67] ; on peut aussi vérifier par exemple que les deux affirmations suivantes sont correctes : $\phi \vDash \phi$ et $\nvDash \phi \rightarrow \phi$. La question immédiatement ouverte par le travail de Kripke est donc : un langage peut-il contenir son propre prédicat de vérité et être suffisamment riche pour soutenir le raisonnement ordinaire et permettre l'expression du caractère sémantiquement problématique de certains énoncés ?

Revenons pour conclure sur l'interprétation classique de la construction de Kripke à laquelle nous avons fait allusion en commençant. Du point de vue classique la preuve d'existence des modèles de point fixe montre qu'on peut sans risque admettre dans un langage classique la validité des biconditionnels-T pour l'ensemble des énoncés fondés[68]. Par conséquent, étant donnée une théorie-objet T consistante (langage $\mathcal{L}$), l'extension obtenue (langage L_{Vr}) en ajoutant à titre d'axiomes à T tous les biconditionnels-T de la forme $Vr(x) \leftrightarrow \phi$ où il faut remplacer ϕ par un énoncé *fondé* de L_{Vr} et « x » par un nom de cet énoncé, est consistante. Un langage classique peut donc contenir un prédicat de vérité adéquat pour son fragment contenant tous les énoncés fondés[69].

66. $\vDash \phi$ si et seulement si $|\phi|_{\mathcal{M}} = 1$ dans toute structure $\mathcal{M}$. Lorsque dans un schéma de formule, par exemple $A \vee \neg A$, on substitue l'énoncé du Menteur aux lettres d'énoncés, on obtient toujours une formule dont la valeur sémantique est $\frac{1}{2}$ d'après les règles de composition de Kleene.

67. Pour la mettre en échec, considérer une structure dans laquelle A et B reçoivent la valeur 1 et C est l'énoncé du menteur.

68. Prendre une interprétation de $\mathcal{L}_{\mathrm{Vr}}$ où dans un plus petit point fixe, la transformer structure d'interprétation classique en supprimant l'anti-extension du prédicat de vérité. La structure $\mathcal{M}'$ ainsi obtenue est un modèle classique des biconditionnels-T pour les énoncés fondés. En revanche $\mathcal{M}' \nvDash Vr(\lambda) \leftrightarrow \lambda$: en effet λ n'étant pas dans l'extension de Vr, le membre gauche est vrai ($\mathcal{M}' \nvDash Vr(\lambda)$) et le membre droit faux ($\mathcal{M}' \vDash \neg Vr(\lambda)$).

69. Sur ce point, voir aussi les remarques concernant les théories axiomatiques classiques dans la section 4.4.

4. Après Kripke

4. 1 La théorie révisionnelle de la vérité

Si la construction de Kripke établit qu'un langage peut contenir son propre prédicat de vérité (dans les limites évoquées dans la section précédente), on peut s'interroger sur sa valeur en tant que *description* de la sémantique du prédicat ordinaire de vérité. Anil Gupta a insisté sur le fait que pour résoudre de façon satisfaisante le problème des paradoxes il était d'abord nécessaire de posséder une description correcte de la sémantique du prédicat de vérité ordinaire. Or à cet égard, l'approche par la sémantique de points fixes de Kripke est insatisfaisante. Tout d'abord le prédicat ordinaire de vérité semble être réellement paradoxal, ce que le prédicat de vérité n'est pas dans la sémantique de Kripke. Dans une perspective descriptive, l'abandon de la sémantique bivalente pour rendre compte de la signification du prédicat de vérité ne va pas non plus de soi, et elle ne peut en tout cas être justifiée par le problème normatif de la cohérence des attributions de vérité. Or indépendemment de la question du caractère paradoxal de certains usages de la vérité, l'idée selon laquelle le prédicat de vérité ordinaire possède une extension déterminée comme point fixe de l'opérateur de Kripke dans une sémantique à trois valeurs permet-elle de rendre compte de la façon dont nous assignons la vérité à des énoncés ? Nous avons déjà observé que la multiplicité des points fixes fragilisait cette valeur descriptive. Mais Anil Gupta va plus loin dans la critique en faisant remarquer que certaines attributions de vérité que nous jugeons naturelles ne peuvent tout simplement pas être expliquées par des sémantiques de point fixe pour le prédicat de vérité. Considérez la situation suivante[70] :

- Alice dit :
  - a1) 2+2=3
  - a2) La neige est toujours noire
  - a3) Tout ce que dit Benoît est vrai
  - a4) 10 est un nombre premier
  - a5) Une chose que dit Benoît n'est pas vraie
- Benoît dit :
  - b1) 1+1=2
  - b2) 2 est premier
  - b3) La neige est parfois blanche

70. Gupta [1982].

— b4) Au plus une chose dite par Alice est vraie

Dans le plus petit point fixe, a3), a5) et b4) ont pour valeur $\frac{1}{2}$, l'évaluation des uns dépendant de l'évaluation des autres. Mais intuitivement il semble clair que b4) est vrai, donc a3) aussi et a5) est fausse. De façon similaire, on peut modifier l'exemple pour mettre en échec différents points fixes et différents schèmes[71].

D'un point de vue descriptif la sémantique de Kripke se révèle donc plutôt insatisfaisante : pour des fragments de langages qui n'exposent à aucun risque d'antinomie la sémantique kripkéenne livre des verdicts contraires à l'usage[72]. N'est-il pas possible de faire mieux ?

La théorie révisionnelle de la vérité est une tentative originale et élégante dans cette direction qui a émergé d'abord de façon indépendante dans Herzberger [1982] et Gupta [1982], et à laquelle Gupta a donné par la suite beaucoup plus d'ampleur en collaboration avec Belnap dans Gupta and Belnap [1993]. L'idée centrale de Gupta est qu'une sémantique descriptivement adéquate du prédicat de vérité doit rendre compte de ce que la signification du prédicat de vérité est celle d'une *règle de révision*. C'est dans cet esprit que Gupta ré-interprète l'idée tarskienne selon laquelle les biconditionnels-T « définissent » la notion de vérité : correctement compris, les biconditionnels-T sont en fait l'expression de cette règle de révision de l'extension du prédicat de vérité. Ainsi l'importance du biconditionnel-T

« la neige est blanche » est vrai si et seulement si la neige est blanche

pour la notion de vérité n'est pas tant qu'un tel biconditionnel est assertable (même s'il se trouve l'être), mais qu'il est l'expression de cette norme de révision de l'extension du prédicat de vérité que nous cherchons à satisfaire si nous comprenons la signification du prédicat de vérité. Ainsi au sens où l'entend Gupta, le biconditionnel-T mobilisant l'énoncé M du Menteur :

M est vrai si et seulement si M n'est pas vrai

s'il n'est certainement pas assertable (c'est une contradiction logique), n'en est cependant pas moins une expression correcte de cette règle de révision

71. Y compris les schèmes supervaluationnels : monter sémantiquement en a3) et a5) pour éviter que Alice ne se contredise classiquement.

72. Gupta étend sa critique à la définition kripkéenne des énoncés paradoxaux. C'est la pertinence de l'approche des phénomènes sémantiques pathologiques par la sémantique de point fixe qui est remise en question de façon plus globale. Voir Gupta and Belnap [1993].

qui commande la sémantique du prédicat de vérité. Il y a paradoxe de la vérité précisément parce que la signification du prédicat de vérité est déterminée par une règle de révision et que cette règle de révision ne converge pas quand elle est appliquée au menteur : sous l'hypothèse que le Menteur n'est pas dans l'extension du prédicat de vérité, on peut donc affirmer que le Menteur n'est pas vrai, mais alors d'après la règle de révision il faut intégrer le Menteur à l'extension du prédicat de vérité ; dans cette nouvelle hypothèse où le Menteur est dans l'extension du prédicat de vérité, on peut donc affirmer que le Menteur est vrai ; mais alors la règle de révision ordonne que le Menteur soit retiré de l'extension du prédicat de vérité ; et ainsi de suite. Cette instabilité de l'énoncé du Menteur dans la suite des révisions du prédicat de vérité tranche avec celui de l'énoncé « La neige est blanche ». Supposons en effet pour commencer que l'énoncé « la neige est blanche » ne soit pas dans l'extension du prédicat de vérité. Comme on peut affirmer que la neige est blanche (ce sont des faits non sémantiques qui déterminent cela, pas la sémantique de la notion de vérité), il faut d'après la règle de révision revoir l'hypothèse de départ et intégrer l'énoncé « la neige est blanche » à l'extension du prédicat de vérité. Une nouvelle révision ne changera rien : l'extension de « neige » et « blanche » n'ayant pas changé on peut toujours affirmer que la neige est blanche, et la règle de révision de l'extension du prédicat de vérité a pour effet de maintenir l'énoncé en question dans l'extension du prédicat de vérité, et ce dans toute la suite des révisions du prédicat de vérité. Cette convergence s'observe également – une étape plus tôt –, si l'hypothèse faite au départ sur l'extension du prédicat de vérité est qu'elle contient l'énoncé « la neige est blanche ». Avant d'entrer un peu plus dans le détail, il convient de dissiper une confusion à laquelle peut donner lieu la mise en regard de cette description et de l'évocation dans la section précédente, à propos de la sémantique de Kripke, d'un processus itératif d'apprentissage de la sémantique du prédicat de vérité. Dans l'argument de Kripke en faveur de la sémantique du plus petit point fixe pour le prédicat de vérité, il s'agissait de suggérer que la sémantique correcte pouvait s'apprendre par approximations successives : en commençant par un langage avec une sémantique défaillante pour le prédicat de vérité, et en corrigeant l'interprétation petit à petit l'interprétation du langage, on pouvait espérer atteindre l'horizon du langage sémantiquement adéquat – ce langage adéquate étant non classique. Ici il s'agit au contraire d'intégrer l'idée de règle de révision à la sémantique du prédicat de vérité elle-même : langage classique ou non, comprendre le prédicat de vérité c'est chercher à rendre les jugements de vérité en fidélité à cette règle.

Pour fixer les idées de façon un peu plus précise et faciliter la comparaison avec le cadre kripkéen, on représentera le processus de révision de l'extension du prédicat de vérité comme une succession de structures d'interprétation dans laquelle seule change l'extension du prédicat de vérité. Le point de départ est une structure classique d'interprétation $\mathcal{M}$ d'un langage de base $\mathcal{L}$ augmenté d'un prédicat de vérité Vr. Au point initial du processus, Vr est interprété de façon arbitraire par un ensemble $\overline{Vr}^0 = h_0$ d'énoncés. On définit alors de la façon suivante une suite ordinale de révisions de l'interpétation du prédicat de vérité :

— $\overline{Vr}^0 = h_0$.
— $\overline{Vr}^{\alpha+1} = \{\phi : \langle \mathcal{M}, \overline{Vr}^{\alpha} \rangle \vDash \phi\}$
— Pour un ordinal limite β, on intègre un énoncé dans l'extension s'il y figure dans toutes les étapes antérieures à partir d'un certain point, autrement dit $\overline{Vr}^{\beta} = \{\phi : \exists\alpha\forall\gamma(\alpha < \gamma < \beta \rightarrow \langle \mathcal{M}, \overline{Vr}^{\gamma} \rangle \vDash \phi\}$

À titre d'illustration, considérons le cas où $\overline{Vr}^0 = \emptyset$. Pour cette extension du prédicat de vérité, la structure associée rend un certain nombre de verdicts, par exemple :

$$\langle \mathcal{M}, \overline{Vr}^0 \rangle \vDash \neg Vr(\text{la neige est blanche})$$
$$\langle \mathcal{M}, \overline{Vr}^0 \rangle \vDash \neg Vr(\tau) \text{ ou encore}$$
$$\langle \mathcal{M}, \overline{Vr}^0 \rangle \vDash \neg Vr(\lambda)$$

Si l'on révise maintenant l'extension de Vr à la lumière de la valeur sémantique que reçoivent les énoncés dans la structure précédente alors, dans la nouvelle structure, les verdicts sémantiques peuvent avoir changés. Ainsi :

$$\langle \mathcal{M}, \overline{Vr}^1 \rangle \vDash Vr(\text{la neige est blanche})$$
$$\langle \mathcal{M}, \overline{Vr}^1 \rangle \vDash \neg Vr(\tau)$$
$$\langle \mathcal{M}, \overline{Vr}^1 \rangle \vDash Vr(\lambda)$$

En effet l'énoncé du menteur, qui était assertable dans la première structure, passe dans l'extension du prédicat de vérité après une première révision. Quant au véridique, l'énoncé qui dit de lui-même qu'il est vrai, il n'est toujours pas vrai après cette étape de révision. On se souvient que Kripke distinguait le véridique du Menteur comme étant à la fois non-fondé et non paradoxal. Ici les ressources de la sémantique révisionnelle ne cèdent rien à la sémantique des points fixes et permet également de rendre compte de

la spécificité de tels énoncés : nous laissons au lecteur le soin de vérifier que si le véridique est dans l'hypothèse de départ, il y reste, tandis que s'il n'y est pas, il n'y entre pas – il y a bien quelque chose d'arbitraire dans l'affirmation du véridique.

Dans la terminologie de Gupta et Belnap, un énoncé est *stablement vrai* dans une suite de révisions s'il entre dans l'extension du prédicat de vérité à partir d'un certain ordinal pour ne plus en sortir lors des révisions ultérieures. Un énoncé stablement faux dans une suite de révisions n'est dans aucune des extensions à partir d'un certain ordinal, et un énoncé est instable s'il n'est ni stablement vrai ni stablement faux. Ce ne sont là que quelques propriétés remarquables dont la combinaison ouvre déjà une large palette de classifications sémantiques, un énoncé pouvant être stablement vrai, stablement faux, instable dans une, dans toute ou dans aucune suite de révisions[73]. Ainsi le menteur est il instable dans toute les suites, l'énoncé « 1+1=2 » stablement vrai dans toutes les suites[74], le véridique est stablement vrai dans certaines suites et stablement faux dans d'autres.

Sans aller plus avant dans l'exploration des possibilités offertes, quel bilan tirer de ce travail sur les paradoxes ? À certains égards, la sémantique révisionnelle représente un progrès relativement à la sémantique de Kripke. D'un point de vue descriptif, elle ne renonce pas à rendre compte de ce prédicat de vérité qui est le nôtre et dont la signification est à la manœuvre quand nous conduisons notre raisonnement paradoxal sur l'énoncé du Menteur. Elle rend compte également de façon fine, le lecteur pourra le vérifier par lui-même, de nos intuitions concernant les attributions de vérité à Alice et Bernard dans la situation présentée plus haut. D'un point de vue méthodologique, le cadre révisionnel ne rend pas nécessaire de supposer que notre langage fonctionne sur une sémantique à trois valeurs pour rendre compte de la signification du prédicat de vérité, allégeant sur ce point la charge de la justification. Il offre en outre, nous l'avons vu, un cadre d'analyse sémantique puissant.

En dépit de ces qualités descriptives, il n'est pas certain que la théorie révisonnelle constitue un progrès décisif relativement à question *normative* soulevée par les paradoxes – comment les résoudre, comment un langage satisfaisant peut-il contenir son propre prédicat de vérité, quels principes

73. Rappel : les différentes suites de révisions sont déterminées par les différentes hypothèses de départ.

74. On suppose que la structure d'interprétation du langage de base est un modèle de PA par exemple.

aléthiques sont vrais, quels raisonnements sont permis dans un langage contenant son propre prédicat de vérité ? En effet ce que la théorie révisionnelle a à offrir sur ces questions est moins enthousiasmant. La règle de révision indique comment réviser des hypothèses concernant l'extension du prédicat de vérité, mais, en définitive, elle ne dit pas encore quelles affirmations catégoriques sont permises sur ce qui est vrai et ce qui ne l'est pas dans une structure dont l'interprétation des termes non sémantiques a été fixée. La suggestion la plus simple est sans doute d'identifier le catégoriquement vrai, entendu comme ce qui est assertable dans les circonstances décrites par le modèle de base, au stablement vrai dans toutes les suites de révision. Ainsi, la sémantique révisionnelle de la vérité permet de rendre compte de l'assertabilité de l'énoncé $Vr(1+1=2)$, tandis que ni le Menteur, ni le véridique ne pourraient être affirmés, pas plus que leurs négations. À cet égard, la situation n'est donc pas si différente de qu'elle était dans la proposition de Kripke, tandis qu'à d'autres égards elle est moins bonne : si le biconditionnel-T du Menteur[75] ne peut pas être affirmé – et il ne peut pas l'être en logique classique – sa négation, $\neg(\mathrm{Vr}(\lambda) \leftrightarrow \neg\mathrm{Vr}(\lambda))$, étant stablement vraie dans toutes les suites de révision, peut l'être. Que la négation de certains biconditionnels-T puisse être catégoriquement vraie dans une sémantique qui leur fait jouer un rôle définitionnel, c'est une fatalité du choix de la logique classique, mais cela a de quoi étonner. On concluera peut-être que la construction kripkeénne n'est au bout du compte pas moins fidèle à l'intuition de l'équivalence en préservant l'intersubstituabilité *salva veritate* de ϕ à $Vr(\phi)$ en tout contexte extensionnel. Il faut noter enfin que la promesse de la préservation de la logique classique, qui est un marqueur de l'approche révisionnelle, est à prendre avec précaution[76]. Pour s'en convaincre, considérons un énoncé tel que $Vr(\lambda) \vee \neg Vr(\lambda)$. Cet énoncé, étant une instance de tautologie, est stablement vrai dans toutes les suites de révisions, donc catégoriquement vrai dans quelque structure d'interprétation du langage que ce soit. Mais ni $Vr(\lambda)$ ni $\neg Vr(\lambda)$ ne le sont. La règle de raisonnement par cas[77] n'est donc pas valide. On peut inférer une contradiction de l'assertion de $Vr(\lambda)$, et de même de l'assertion de $\neg Vr(\lambda)$, mais non de l'assertion de leur disjonction.

75. $\mathrm{Vr}(\lambda) \leftrightarrow \neg\mathrm{Vr}(\lambda)$.

76. Comme le fait remarquer Field [2008], chap. 6, les supervaluations mobilisées pour restaurer la logique classique dans l'approche kripkéenne ne permettent pas de faire mieux sur ce point.

77. En déduction naturelle : la règle d'élimination de la disjonction.

4. 2 Enrichir le langage

Nous avons noté que les limites de la construction kripkéenne étaient pour une part des limites d'expressivité d'un langage qui contient, certes, son propre prédicat de vérité, mais dans lequel, faute d'implication matérielle digne de ce nom, il est à peine possible de conduire un raisonnement logique ordinaire. Ainsi, nous avons observé que $\phi \vDash \phi$ alors que le conditionnel défini par $\neg\phi \vee \phi$, dont la validité reflète classiquement cette relation dans le langage lui-même[78], n'est pas valide dans la sémantique de Kripke. Confronté à cette difficulté, pourquoi ne pas essayer d'enrichir le langage en introduisant les notions logiques qui lui font défaut, et en particulier un conditionnel digne de ce nom ? C'est la tâche à laquelle s'est attelé Hartry Field dans une série de travaux qui culmine avec la parution de Field [2008]. Le cahier des charges de la sémantique de Field pour un langage contenant son propre prédicat de vérité exige d'une part un prédicat de vérité non-typé qui satisfasse le principe d'intersubstituabilité de ϕ et $Vr(\phi)$ dans tout contexte extensionnel et d'autre part un conditionnel qui satisfasse une bonne part des lois logiques attendues[79], et dans laquelle tous les biconditionnels-T (formulés à l'aide du nouveau conditionnel) doivent être valides (adéquation au sens de Tarski). Techniquement parlant, ce tour de force permettant de sauver la théorie naïve de la vérité[80] est réalisé en combinant les idées de Kripke sur l'existence de structures multivaluées dans lesquelles le prédicat de vérité est interprété par un point fixe[81] et les outils de la théorie révisionnelle pour fixer la sémantique des conditionnels[82].

Dans la sémantique ainsi obtenue, la restriction de la validité du tiers exclu permet de tenir en échec le raisonnement paradoxal sur l'énoncé du menteur, en dépit de la validité du biconditionnel $Vr(\lambda) \leftrightarrow \neg Vr(\lambda)$. La sémantique de Field, comme celle de Kripke, garantit que l'on ne peut affirmer l'énoncé du Menteur ni affirmer sa négation en ne donnant ni à l'un

78. C'est le théorème de déduction en classique.

79. Quelques exemples : $\vDash \phi \rightarrow \phi$, $\vDash (\phi \wedge \psi) \rightarrow \phi$, $\vDash \phi \rightarrow \neg\neg\phi$, le modus ponens, la transitivité du conditionnel, etc. Le conditionnel ne peut satisfaire toutes les règles classiques, sous peine de paradoxe de Curry, il s'agit donc non pas de valider toutes les lois classiques mais de contenir au maximum l'écart tout en préservant la consistance. À titre d'exemple, la validité du principe d'importation est perdue : $\phi \rightarrow (\psi \rightarrow \chi) \nvDash (\phi \wedge \psi) \rightarrow \chi$. Pour plus de détails, voir Field [2008], chap. 17.

80. On désigne par là l'ensemble des biconditionnels-T pour un langage contenant son prédicat de vérité.

81. Voir la section consacrée au travail de Kripke.

82. Nous renvoyons le lecteur à Field [2008] pour un exposé détaillé, en particulier à l'exposé simplifié du chapitre 16.

ni à l'autre la valeur 1 dans quelque interpétation du langage. Quant à la question de savoir s'il faut admettre, à la lumière du métalangage, qu' *in fine* le menteur n'est pas vrai, et en tirer la conclusion revancharde que l'on devrait pouvoir affirmer le Menteur s'il était formulé avec un prédicat de vérité conforme à sa vocation, la réponse de Field est que le métalangage n'est qu'un instrument permettant d'établir la cohérence d'un certain nombre de lois du langage-objet et qu'il n'y a aucune raison philosophique de vouloir faire coïncider extensionnellement le prédicat du métalangage « avoir la valeur sémantique 1 » et le prédicat « Vr » du langage-objet – le premier n'est pas la norme du second[83]. Reste alors la question de l'expressivité du langage. Le statut sémantique du Menteur est distinct de celui d'un énoncé comme $1 + 1 = 2$, cette spécificité est manifestée de différentes façons dans le langage lui-même[84], mais peut-on la *dire* ? Ce n'est pas le moindre des mérites de la construction de Field que de montrer comment introduire dans le langage-objet un opérateur D qui permette l'expression *interne au langage-objet* du statut sémantique du Menteur et de ce qui le distingue de celui de l'énoncé $1 + 1 = 2$. L'opérateur D s'applique à des énoncés de sorte que $D\phi$ peut s'interpréter à peu près comme « De façon déterminée : ϕ » – une affirmation plus forte que la simple affirmation de ϕ. À l'aide de cet opérateur de détermination on peut alors affirmer dans le langage que le menteur n'est pas vrai de façon déterminée, ou dit autrement, qu'il n'est pas déterminé que le menteur n'est pas vrai : $\vDash \neg D(\neg Vr(\lambda))$. Bien sûr ce progrès dans la possibilité d'exprimer le statut sémantique du Menteur ne va pas pas sans un progrès parallèle dans la possibilité de formuler de nouveaux énoncés paradoxaux, et le langage contient désormais un nouveau Menteur λ_1 qui dit de lui-même $\neg DVr(\lambda_1)$ et dont on peut se demander s'il est vrai de façon déterminée ou non. La sémantique permet d'apporter à *ce* paradoxe une réponse analogue à celle apportée au menteur ordinaire, en garantissant que λ_1 est sémantiquement défectueux au sens où il ne reçoit pas la valeur sémantique 1, pas plus que ne reçoit la valeur 1 l'affirmation de sa détermination, de sorte que l'instance du tiers exclu $D\neg DVr(\lambda_1) \vee \neg D\neg DVr(\lambda_1)$ n'est pas valide. Le statut sémantique défectueux de λ_1 ne peut donc pas être exprimé à l'intérieur du langage par l'affirmation que λ_1 n'est pas vrai de façon déterminée – puisque c'est précisément ce que dit λ_1 ! Devons-nous alors demeurer muets sur le statut sémantique de λ_1 ? Non, car nous pouvons à présent itérer l'opérateur de

83. Cette perspective instrumentale permet également de justifier le fait que le métalangage employé dans la construction soit classique. Sur ces questions, voir Field [2008], chap. 2.

84. Par exemple, la mise en échec de l'instance correspondante du tiers exclu.

détermination et affirmer sans paradoxe : $\models \neg DD\neg D(\lambda_1)$. Et ainsi de suite à l'infini[85].

L'approche de Field se présente donc comme un développement particulièrement abouti de certaines des idées en germe dans la construction de Kripke. Au point de départ de ce projet il y a la décision théorique de chercher à sauver l'intersubstituabilité de $Vr(\phi)$ et de ϕ au prix d'un renoncement à la logique classique. La façon dont Field choisit de s'écarter de la logique classique – ou de la généraliser, selon le point de vue adopté – n'est cependant pas la seule possible, comme nous allons le montrer à présent.

4. 3 Redefinir la notion de conséquence

Les valeurs sémantiques attribuées dans les structures d'interprétation aux énoncés du langage-objet étudié jouent un double rôle : rendre compte de la façon dont les conditions d'assertion de ces énoncés dépend des conditions d'assertion de leur composants, et rendre compte des relations logiques entre énoncés. La première tâche est celle qui est accomplie par la définition de la satisfaction dans une structure (ou par les tables de vérité) ; la seconde par la définition de la conséquence logique. Or dans une sémantique multivalente cette dernière peut être définie de différentes façons, donnant lieu à autant de logiques différentes. Si le nombre de valeurs sémantiques mobilisées dans la structure d'interprétation d'un langage ne détermine pas en lui-même la logique du langage, à quelles conditions faut-il tenir deux logiques L_1 et L_2 pour identiques ? Sur cette question également il convient d'être prudent. Sans doute l'identité de leur ensemble respectif de tautologies, $\models_{L_1} \phi \Leftrightarrow \models_{L_2} \phi$, n'y suffit pas, et l'on sera enclins à exiger l'identité extensionnelle des relations de conséquence, $\Gamma \models_{L_1} \phi \Leftrightarrow \Gamma \models_{L_2} \phi$. Y faut-il de surcroît identité des méta-règles d'inférence ?[86], ou des méta-métarègles ? Peu importe ici où l'on situe la frontière exacte de la logique classique, ce qui compte est d'être au clair sur ce que

85. Ou plutôt dans le transfini. Voir Field [2008] chap. 22 pour plus de précisions sur ce point. On notera que pour représenter la gradation dans la détermination des énoncés la sémantique fait usage non pas de trois valeur de vérité, comme dans la sémantique de Kripke, mais d'une infinité. Sur les sémantiques qui comportent une infinité de valeurs de vérité, voir le chapitre ?.

86. Les métarègles d'inférence sont les règles de clôture pour la relation de conséquence. $\phi \rightarrow \phi, \phi \models \phi$ est une règle d'inférence. En revanche :

l'on sacrifie.

Nous avons vu dans la section 2 la manière dont on peut tirer parti du passage de deux à trois valeurs sémantiques pour modifier la logique, sans toucher nominalement à la définition classique de la conséquence : ϕ est conséquence de $\psi_1, \psi_2, ...$ si ϕ reçoit la valeur 1 dans les structures où $\psi_1, \psi_2, ...$ reçoivent la valeur 1. Avec cette définition et l'emploi des connecteurs forts de Kleene on obtient la logique de Kleene. Nous avons déjà noté que cette logique ne contenait aucun schéma de tautologie. On pourrait pourtant s'interroger sur le mérite de cette définition de la conséquence dans un langage dont la sémantique est structurée autour, non pas de deux, mais de trois valeurs de vérité. Si pour construire un langage qui ressemble au nôtre nous avons besoin de trois valeurs sémantiques, ne faut-il pas reprendre l'examen de la définition de la conséquence logique ? Une autre considération indépendante appuie la pertinence d'une remise à plat : la notion de conséquence logique classique n'est pas exempte de tout reproche[87]. Nombre d'auteurs ne se sont-ils pas déjà interrogés, en particulier, sur la validité de la règle classique qui veut que n'importe quel énoncé soit conséquence logique d'une contradiction ? L'examen de conscience auquel nous contraignent les paradoxes est peut-être l'occasion de reconsidérer cette règle.

C'est le chemin emprunté par Graham Priest. Selon Priest, le raisonnement du menteur est correct et la véritable leçon en est que le Menteur et sa négation sont tous deux vrais.[88] Il y a, défend Priest, de vraies contradictions[89]. Cependant, et ici Priest se fait en outre l'avocat des logiques *paraconsistantes* contre la logique classique, tout énoncé n'est pas conséquence logique d'une contradiction. Une sémantique correcte devrait donc à la fois rendre compte de la possibilité qu'un langage contienne son propre prédicat de vérité, et de l'invalidité de la règle d'explosion : $\psi, \neg\psi \not\models \phi$. La version la plus simple de cette idée est développée par Priest sous le nom de *Logique*

$$\frac{\Gamma \models \phi \rightarrow \psi \qquad \Gamma \models \phi}{\Gamma \models \psi}$$

est une méta-règle d'inférence. De même que deux logiques peuvent avoir le même ensemble de tautologies mais avoir des relations de conséquence logique distinctes, deux logiques peuvent avoir des relations conséquences coextensionnelles mais ne pas satisfaire le même ensemble de métarègles.

87. Voir le chapitre 11.

88. Si l'on définit la fausseté d'un énoncé comme la vérité de sa négation, alors il y a des énoncés qui sont à la fois vrai et faux – c'est la marque du *dialéthisme*.

89. Au sens propre d'un énoncé d'un certain genre qui est *vrai*, pas au sens d'un énoncé qui serait authentiquement d'un certain genre !

des paradoxes (LP)[90] . On définit pour un langage contenant son prédicat de vérité une structure d'interprétation à la Kripke, en utilisant le schème fort de Kleene et en assignant au prédicat de vérité l'interprétation qui est la sienne dans le plus petit point fixe. Mais tandis que la conséquence logique classique est définie par la préservation de la vérité (au sens de la valeur sémantique 1), Priest la définit dualement comme préservation de la « non fausseté » :

Conséquence logique $\vDash_{LP}$
$\psi_1, \psi_2, ... \vDash_{LP} \phi$ si et seulement si il n'existe pas de structure $\mathcal{M}$ dans laquelle les prémisses $\psi_1, \psi_2, ...$ reçoivent toutes des valeurs sémantiques distinctes de 0 tandis que le conséquent ϕ reçoit la valeur sémantique 0.

Dans le plus petit point fixe construit sur un modèle standard $\mathcal{M}$ de l'arithmétique, il est clair que $|0 \neq 0|_{\mathcal{M}} = 0$, que $|\neg Vr(\lambda)|\mathcal{M} = \frac{1}{2}$, que $|Vr(\lambda)|_{\mathcal{M}} = \frac{1}{2}$ et donc que $0 \neq 0$ n'est pas conséquence de $\{Vr(\lambda), \neg Vr(\lambda)\}$: c'est l'illustration centrale de l'invalidité de la règle classique d'explosion dans LP. Un moment de réflexion permet de se convaincre que toutes les tautologies classiques sont en revanche LP-valides, car toutes les instances de schéma de tautologie reçoivent dans les structures de point fixe soit la valeur sémantique 1 soit la valeur sémantique $\frac{1}{2}$. LP souffre cependant des mêmes défauts fondamentaux que la logique de Kleene. Comme le langage de Kleene, le langage de LP est faiblement expressif et ne contient pas de conditionnel digne de ce nom – il faudra donc l'ajouter. Si LP est en outre assez accueillante pour valider entièrement tous les biconditionnels-T – y compris donc pour les énoncés pathologiques comme le Menteur[91], LP est peut-être un peu trop accueillante puisque la négation des biconditionnels-T problématiques n'est pas moins LP-valide qu'eux[92] ! Ces défauts, toutefois, si réels qu'ils soient, ne sont pas incorrigibles, et Priest lui-même a montré comment à enrichir le langage de LP et sophistiquer sa sémantique d'une façon analogue à celle de Field.[93] Reste l'intention d'ensemble de la solution, qui heurte nos habitudes sémantiques sur deux plans : sur le plan de la théorie de la vérité, avec l'idée que des contradictions puissent être vraies ; sur le plan de la logique, avec leur nouvelle innocuité. Priest

90. Voir Priest [1979] pour un premier exposé
91. $Vr(\lambda) \leftrightarrow \neg Vr(\lambda)$ reçoit la valeur sémantique $\frac{1}{2}$ dans toutes les structures de point fixe, donc ne reçoit jamais la valeur 0, donc est valide au sens de LP.
92. $\neg(Vr(\lambda) \leftrightarrow \neg Vr(\lambda))$ reçoit également la valeur sémantique $\frac{1}{2}$ dans toutes les structures de point fixe.
93. On trouvera une discussion synthétique des approches paraconsistantes aux paradoxes dans la cinquième partie de Field [2008].

soutient que ces habitudes sémantiques sont de peu de poids au regard des bénéfices intellectuels de la réforme dont il se fait l'avocat. La question est alors de savoir si d'autres réformes concurrentes ne pas font mieux, et à moindre coût.

L'espace logique des solutions aux paradoxes ouvert par les sémantiques multivalentes est donc plus vaste qu'il ne paraît au premier regard, nous allons en donner un dernier exemple. La sémantique de LP invalide la règle d'explosion, mais elle laisse inchangée les règles structurelles de la logique, c'est-à-dire les règles les plus générales dont la formulation n'engage aucun connecteur particulier : on peut citer parmi elles la réflexivité[94], affaiblissement[95], la contraction[96] ou encore la transitivité[97]. Plutôt que de modifier des règles opérationnelles de la logique qui règlent les rapports des connecteurs à la relation de conséquence, ne faut-il pas prioritairement enquêter sur la pertinence d'une réforme des règles structurelles elle-mêmes ? C'est une approche de ce genre qui est proposée dans Cobreros et al. [2014]. Sur le fond désormais habituel d'une sémantique kripkéenne pour un langage contenant son propre prédicat de vérité, les auteurs introduisent l'idée qu'il faut faire droit à une gradation des standards d'assertion. Les énoncés qui reçoivent la valeur 1 peuvent être assertés selon un standard strict[98], tandis que les énoncés qui reçoivent la valeur $\frac{1}{2}$ ne peuvent l'être que selon un standard plus tolérant. Si la conséquence logique dit quelque chose de l'assertabilité de la conclusion étant donnée l'assertabilité des prémisses, la multiplicité des standards d'assertion appelle naturellement la distinction de différentes notions de conséquence logique. Les auteurs définissent alors un argument valide dans la logique $S3T$[99] comme un argument tel que si ses prémisses sont assertables strictement, alors sa conclusion est assertable :

Conséquence logique $\vDash_{S3T}$
$\psi_1, \psi_2, ... \vDash_{S3T} \phi$ si et seulement si il n'existe pas de structure $\mathcal{M}$ telle que $|\psi_1| = |\psi_2| = ... = 1$ et $|\phi|_{\mathcal{M}} = 0$.

Dans $S3T$ les règles structurelles de monotonie et contraction sont respectées, ainsi que de nombreuses propriétés de la logique classique telles que le théorème de déduction ou la possibilité de la preuve par cas. $S3T$

94. $\phi \vDash \phi$
95. Si $\Gamma \vDash \phi$ alors $\Gamma, \Gamma' \vDash \phi$
96. Si $\Gamma, \psi, \psi \vDash \phi$ alors $\Gamma, \psi \vDash \phi$.
97. Si $\Gamma \vDash \phi$ et $\Gamma', \phi \vDash \psi$ alors $\Gamma, \Gamma' \vDash \psi$
98. Dans les conditions idéalisées décrites par la structure
99. Pour « Strict Tolerant Transparent Truth ».

étend même conservativement la logique classique : une inférence ne mobilisant pas le prédicat de vérité est valide classiquement si et seulement si elle est $S3T$-valide, ce qui implique que $S3T$ est consistante. Les schéma d'inférences classiquement valides sont $S3T$-valides dans le langage étendu contenant le prédicat de vérité. $S3T$ préserve donc de nombreuses propriétés de la logique classique. Pour maintenir la cohérence, l'écart le plus significatif réside dans l'abandon contrôlé de la transitivité de la conséquence[100]. En effet un argument Π_1 qui serait $S3T$-valide peut conduire d'une prémisse ϕ dont la valeur sémantique est 1 à une conclusion ψ dont la valeur sémantique est $\frac{1}{2}$ et un argument Π_2 être $S3T$-valide conduire de la prémisse ψ dont la valeur sémantique est $\frac{1}{2}$ à une conclusion χ dont la valeur sémantique est 0, le tout sans que l'argument obtenu en enchaînant Π_1 et Π_2 soit lui-même valide au sens de $S3T$, puisque que conduisant d'une prémisse ϕ dont la valeur sémantique est 1 à une conclusion χ dont la valeur sémantique est 0.

$S3T$ est ainsi une une logique que l'on pourrait qualifier de non cartésienne, au sens où elle met en échec la méthode analytique d'épreuve de la validité des arguments : certaines inférences sont globalement incorrectes quoique localement correctes partout. Cette idée n'est pas si extravagante qu'elle en a l'air : les paradoxes soritiques ne sont-ils des exemples naturels de tels raisonnements ? Peut-être faut-il compter la vérité au rang des ces prédicats « non-analytiques » qui, comme les prédicats vagues, mettent en échec la transitivité de la conséquence. Une autre difficulté surgit toutefois avec le fait que $S3T$ ne répond pas à l'idée naturelle de conséquence valide comme préservation, et en particulier à l'idée que la conséquence logique devrait *a minima* préserver la validité elle-même. Or les inférences $S3T$-valides ne préservent pas la $S3T$-validité : $Vr(\lambda) \wedge \neg Vr(\lambda)$ est $S3T$-valide[101], et l'inférence de $Vr(\lambda) \wedge \neg Vr(\lambda)$ à $0 = 1$ est $S3T$-valide[102], cependant $0 = 1$ n'est pas $S3T$-valide.

Signalons pour clore sur la relation de conséquence que d'autres approches substructurelles ont été explorées. On peut par exemple bloquer les paradoxes en renonçant à la règle de contraction. Le logicien russe Grisin avait été un pionnier de cette approche pour la résolution des paradoxes de la théorie des ensemble au début des années quatre-vingt.[103] Dans les

100. $\phi \vDash_{S3T} \psi$ et $\psi \vDash_{S3T} \chi$ n'impliquent pas $\phi \vDash_{S3T} \chi$
101. Puisque l'énoncé reçoit la valeur $\frac{1}{2}$ dans toutes les structures de point-fixe
102. Aucune structure n'assigne la valeur 0 à la conclusion qui assignerait la valeur 1 à la prémisse.
103. Grisin [1982]. Cantini [2003] fait remonter la première logique sans contraction à Fitch.

années quatre-vingt dix Greg Restall[104] a suivie une voie analogue. Plus récemment Elia Zardini s'en est fait un avocat énergique Zardini [2011]. Pour aller plus loin dans la présentation et l'évaluation des possibilités offertes en matière de révision de la logique pour traiter des paradoxes, on pourra consulter Murzi and Carrara [2014].

Dans l'espace logiquement contraint des solutions aux paradoxes, c'est comparativement que nous évaluons les possibilités existantes. Mais comment cette comparaison doit-elle être conduite ? S'agit-il de trouver une représentation plus fidèle de notre pratique réelle ou de la corriger ? Et s'il s'agit de changer de logique, que signifie un tel changement et quels en seront les critères ? Sur le chemin non-classique vers la résolution des paradoxes, le voyageur doit donc affronter l'énigme de la pluralité des logiques, qui n'est pas moins profonde celle de la vérité.[105] C'est l'un des mérites de ces approches que d'en explorer les solidarités.

4. 4 Approches axiomatiques classiques

Nous avons déjà observé que pour certains philosophes, Tarski ou Gupta par exemple, le prix logique à payer pour le maintient de la théorie naïve de la vérité était trop élevé. Mais réaffirmer la logique classique ne suffit pas, et s'il l'on veut sauver l'usage d'un prédicat de vérité non-typé il est souhaitable d'en proposer une théorie, c'est-à-dire de coucher sur le papier un ensemble consistant de principes qui gouvernent cet usage de la notion dans tout langage. Or les approches classiques doivent renoncer à l'idée qu'une théorie de la vérité puisse contenir tous les biconditionnels-T et cette impossibilité même est source d'une certaine profusion – comment choisir parmi tant de possibilités également étrangères à l'idéal auquel on a renoncé[106] ?

104. Dans sa thèse Restall [1994].

105. Sur ces questions nous renvoyons le lecteur au chapitre 19.

106. Une approche peut être classique et non-axiomatique, comme dans Gupta and Belnap [1993], mais également être axiomatique et non classique, comme par exemple dans l'article déjà cité de Kremer [1988] ou dans Horsten [2011]. Classique ou non, l'approche axiomatique prétend parfois à deux avantages méthodologiques relativement à l'approche modèle-théorétique. Le premier est de fournir d'emblée et pour ainsi dire automatiquement une « logique de la vérité » c'est-à-dire l'ensemble des énoncés aléthiques valides. Une approche modèle-théorétique peut demeurer incomplète en omettant de préciser relativement à quelle classe de structures il faut définir la validité. Même sans cette omission, on peut définir la validité sans donner le moindre indice de ce à quoi peuvent ressembler les énoncés aléthiques valides pour la définition considérée – il se pourrait même que la classe des validités alé-

Le premier travail post-kripkéen influent conduit dans cet esprit classique et axiomatique est présenté dans Feferman [1991][107]. La théorie de Feferman axiomatise dans un langage classique les intuitions sémantiques arrêtées dans les structures de point-fixe de Kripke[108], et telle est la raison pour laquelle on nomme généralement KF cette théorie pour Kripke-Feferman. En écrivant $F(\phi)$ pour $Vr(\neg\phi)$, et en utilisant les conventions de notations introduites plus haut, on peut formuler les axiomes de KF de la façon suivante :

Axiomes de KF

1. Pour toute formule atomique ϕ de L_{PA} : $Vr(\phi) \leftrightarrow val^{+}(\phi)$
2. Pour toute formule atomique ϕ de L_{PA} : $F(\phi) \leftrightarrow val^{-}(\neg\phi)$
3. $\forall\phi$ de L_{Vr} : $F(\neg\phi) \leftrightarrow Vr(\phi)$
4. $\forall\phi, \psi$ de L_{Vr} : $Vr(\phi \wedge \psi) \leftrightarrow (Vr(\phi) \wedge Vr(\psi))$
5. $\forall\phi, \psi$ de L_{Vr} : $F(\phi \wedge \psi) \leftrightarrow (F(\phi) \vee F(\psi))$
6. $\forall\phi(x)$ de L_{Vr} : $Vr(\forall x\phi(x)) \leftrightarrow \forall y Vr(\phi(y))$
7. $\forall\phi(x)$ de L_{Vr} : $F(\forall x\phi(x)) \leftrightarrow \exists y F(\phi(y))$
8. $\forall\phi$ de L_{Vr} : $Vr(Vr(\phi)) \leftrightarrow Vr(\phi)$
9. $\forall\phi$ de L_{Vr} : $F(Vr(\phi)) \leftrightarrow Vr(\neg\phi)$
10. $\forall\phi$ de L_{Vr} : $\neg[Vr(\phi) \wedge Vr(\neg\phi)]$

Feferman montre que cette théorie est consistante en prouvant que les structures de points fixes de Kripke, une fois rendues à la logique classique en y supprimant l'anti-extension de Vr, sont des modèles de KF. Il est clair que KF renonce à la théorie naïve de la vérité : certains biconditionnels-T,

thiques ne soit pas récursivement énumérable. Le second avantage auquel peut prétendre l'approche axiomatique est qu'elle n'est pas sujette à certaines limitations intrinsèques de l'approche sémantique. En adoptant un cadre modèle-théorique pour interpréter les langages, l'approche sémantique fait implicitement l'hypothèse que le domaine des quantificateurs du langage étudié est un ensemble, et à strictement parler les preuves de consistance qu'elle apporte ne peuvent donc jamais concerner un langage permettant d'exprimer la généralité absolue (sur cette question voir la note). En réponse à cette difficulté, Field [2008], s'inspirant d'un argument célèbre de Kreisel, invoque le caractère instrumental de la théorie des modèles dont l'emploi est essentiellement justifié par le théorème de complétude pour la logique du premier ordre. Notons symétriquement que l'avantage de la plus grande universalité revendiquée à cet égard par les approches axiomatiques est annulé si les preuves de consistance des principes aléthiques considérés est administrée par des méthodes sémantiques.

107. Les premières présentations de KF datent de la fin des années soixante-dix. Voir l'historique de Feferman [2008].

108. Sur l'interprétation classique de la construction de Kripke, voir plus haut section 2.

par exemple $Vr(\lambda) \leftrightarrow \neg Vr(\lambda)$ ne sont pas des théorèmes de KF. KF présente néanmoins l'intérêt d'être *compositionnelle* : elle prouve les grands principes de composition des connecteurs logiques avec le prédicat de vérité. Cette théorie n'est toutefois pas sans défaut en tant que théorie de la notion de vérité. Le plus dirimant est sans doute l'écart engendré par KF entre la logique externe et la logique interne au prédicat de vérité. En effet, l'ensemble des énoncés prouvables dans KF et l'ensemble des énoncés prouvablement vrais dans KF sont distincts, le second étant strictement inclus dans le premier. Or cet état de fait a pour conséquence contre-intuitive que les ressources de KF permettront d'affirmer certains énoncés tout en ne permettant pas de les affirmer vrais. Ainsi par exemple, la logique de KF étant classique, $KF \vdash Vr(\lambda) \vee \neg Vr(\lambda)$, mais $KF \nvdash Vr(Vr(\lambda) \vee \neg Vr(\lambda))$[109]. Autrement dit, dans KF, la règle de preuve

$$\frac{\vdash \phi}{\vdash Vr(\phi)} \text{ (NEC)}$$

n'est pas valide.

Peut-on remédier au problème ? L'axiome 8 nous montre que ce problème n'affecte pas la logique interne du prédicat de vérité de KF, c'est-à-dire la théorie de la vérité que l'on obtiendrait en retenant au titre d'axiomes, non pas les axiomes de KF, mais les énoncés déclarés vrai dans KF. Pourquoi dès lors ne pas renoncer à KF, et déclarer que notre théorie de la vérité est en fait la logique interne de KF, c'est-à-dire constituée exactement de l'ensemble des énoncés déclarés Vr par KF ? Cette suggestion formulée par Reinhardt [1986] et souvent discutée depuis, n'est pas sans intérêt, mais elle présente aussi ses difficultés. D'abord, cette logique n'est pas classique[110]. Cela n'est pas forcément un problème mais nous éloigne des options théoriques de cette section – c'est à d'autres théories non-classiques qu'il faudrait comparer ses mérites. Ensuite elle n'est pas axiomatisée, et même si l'ensemble de ses théorèmes est récursivement énumérable, tant que nous n'en avons pas une axiomatisation maniable nous acceptons une théorie en ne sachant que peu de choses des grands principes qui la gouvernent – on est donc loin également de l'esprit des approches axiomatiques.

109. Pour le voir, il suffit de se souvenir que l'énoncé $Vr(\lambda) \vee \neg Vr(\lambda)$ n'est dans l'extension de Vr dans aucune structure de point-fixe kripkéenne.
110. Comme le montre le fait déjà relevé que $KF \nvdash Vr(Vr(\lambda) \vee \neg Vr(\lambda))$.

Au fond, ce qui apparaît à l'analyse est qu'avec KF Feferman a choisi de privilégier certains principes aléthiques (compositionalité par exemple) au détriment d'autres principes (la règle NEC) et que ces choix méritent d'être discutés plus à fond à la lumière d'autres possibilités consistantes. Dans un article qui a fait date, Friedman and Sheard [1987] prennent le problème à bras le corps en formulant un ensemble de huit axiomes[111] et quatre règles[112] rassemblant les principes les plus en vue parmi ceux qu'une théorie de la vérité devrait *a priori* devoir satisfaire. Si ces axiomes et ces règles ne sont pas conjointement satisfaisables, l'ensemble de leurs combinaisons consistantes est un espace riche qu'il convient d'explorer. L'état des lieux exhaustif établi par les auteurs est la carte du tendre de l'amateur classique[113] de paradoxes aléthiques. Il n'est pas possible ici de revenir en détail sur les résultats de cet article, mais nous relèverons, parmi les nombreux systèmes dont la consistance est démontrée, le système FS dont les axiomes ont été depuis reformulés de la façon suivante :

Axiomes de FS

1. Pour tout énoncé atomique ϕ de $\mathcal{L}_{PA}$: $Vr(\phi) \leftrightarrow val^{+}(\phi)$
2. Pour tout énoncé ϕ de $\mathcal{L}_{\text{Vr}}$: $Vr(\neg\phi) \leftrightarrow \neg Vr(\phi)$
3. Pour tout énoncé ϕ de $\mathcal{L}_{\text{Vr}}$: $Vr(\phi \wedge \psi) \leftrightarrow Vr(\phi) \wedge Vr(\phi(x))$
4. Pour tout énoncé ϕ de $\mathcal{L}_{\text{Vr}}$: $Vr(\phi \vee \psi) \leftrightarrow Vr(\phi) \vee Vr(\phi(x))$
5. Pour tout énoncé $\phi(x)$ de $\mathcal{L}_{\text{Vr}}$: $Vr(\forall x\phi(x)) \leftrightarrow \forall x Vr(\phi(x))$
6. Enfin deux règles de preuve[114] :
   — NEC : d'un preuve de ϕ inférer $Vr(\phi)$
   — CONEC : d'une preuve de $Vr(\phi)$ inférer ϕ

Comme KF, le système FS peut retenir l'attention à plus d'un titre. FS est un système d'axiomes pour un prédicat non-typé de vérité qui présente la plupart des vertus dont jouit la théorie TC pour les prédicats typés : elle rend compte de la composition du prédicat de vérité avec les constantes

111.

T-In	$\phi \rightarrow Vr(\phi)$	T-Out	$Vr(\phi) \rightarrow \phi$
T-Cons	$\neg(Vr(\phi) \wedge Vr(\neg\phi)$	T-Comp	$Vr(\phi) \vee Vr(\neg\phi)$
T-Rep	$Vr(\phi) \rightarrow Vr(Vr(\phi))$	T-Del	$Vr(Vr(\phi)) \rightarrow Vr(\phi)$
U-inf	$\forall n Vr(\phi(n)) \rightarrow Vr(\forall x\phi x)$	E-inf	$Vr(\exists x\phi(x) \rightarrow \exists n Vr(\phi(n))$

112.

T-intro	$\phi / Vr(\phi)$	T-Elim	$Vr(\phi)/\phi$
¬T-Intro	$\neg\phi/\neg Vr(\phi)$	¬T-Elim	$\neg Vr(\phi)/\neg\phi$

113. Comprendre le logicien attaché à la logique classique.

114. Une règle de preuve est une règle d'inférence d'application limitée en ce sens qu'elle ne permet de tirer de conclusion qu'à partir d'un énoncé ou d'un ensemble d'énoncés eux-mêmes établis par une preuve, c'est-à-dire au terme d'une dérivation *dont toutes les hypothèses ont été déchargées*. C'est cette restriction qui permet d'éviter de prêter le flanc au raisonnement du menteur dans l'application des règles ci-dessous.

logiques, permet de prouver le principe général de contradiction et celui du tiers exclu pour le langage $\mathcal{L}_{\text{Vr}}$. Il ne permet pas de dériver tous les biconditonnels-T, mais les règles NEC et CONEC permettent du moins de dériver les biconditionnels-T pour les énoncés prouvables, d'asserter les énoncés dont la vérité a été prouvée, aussi bien que d'asserter la vérité des énoncés assertables. En outre FS est consistant et la preuve sémantique standard de sa consistance, qui fait appel à des séquences de révisions, montre qu'elle entretient des liens naturels avec la théorie révisonnelle de la vérité[115]. En dépit de ses qualités, FS est toutefois grevée d'un défaut dont on laissera chacun apprécier la gravité : FS est ω-inconsistante, c'est-à-dire que les modèles de $PA + FS$ sont des modèles non-standards de l'arithmétique[116].

Les deux théories que nous venons d'évoquer, KF et FS ont en commun de procéder à une axiomatisation *via* des principes de « compositionalité » du prédicat de vérité. Mais puisque la condition naïve d'adéquation d'un prédicat de vérité réside dans l'assertabilité des biconditionnels-T, pourquoi ne procéder de façon directe en proposant à titre d'axiomes un ensemble bien choisi d'équivalences-T, un peu à la façon dont Tarski lui-même avait procédé dans le cas du prédicat de vérité typé en proposant à titre d'axiomes – non sans émettre quelques réserves critiques – les bicontionnels-T pour les énoncés ne contenant pas de prédicat de vérité. En dépit de son attrayante apparence de simplicité, cette approche présente aussi ses propres difficultés, car si l'on s'épargne sur ce chemin la peine d'un choix entre des principes généraux également attrayants, le choix des bicondionnels à retenir ne va pas de soi une fois rappelé qu'on ne peut les garder tous sous peine d'inconsistance. Ne pourrait-on se fixer sur un ensemble de biconditonnels maximalement consistants[117] ? Cette suggestion se heurte à plusieurs difficultés. La première est qu'il existe une infinité des tels ensembles. La seconde est plus profonde et vient de la découverte de ce que, si les biconditonnels-T de la théorie $DT_{\mathcal{L}}$ discutée par Tarski étaient déductivement innocents, voire faibles, cette faiblesse n'a plus cours lorsque l'on parle de biconditonnels-T mobilisant des prédicats de vérité non-typés. En effet McGee [1992] a montré que pour toute théorie T dans un langage $\mathcal{L}_{\text{Vr}}$, il existe un ensemble E de biconditonnels-T tel que pour tout énoncé de ϕ de $\mathcal{L}_T$:

115. Pour une esquisse de la preuve, voir Horsten [2011], p. 108.
116. Pour une discussion techniquement accessible, on pourra consulter Horsten [2011].
117. C'est-à-dire un ensemble consistant dont toute extension est inconsistant.

$T \vdash \phi$ si et seulement si $E \vdash \phi$[118].

À la lumière de ce résultat, le choix arbitraire d'un ensemble maximalement consistant de biconditionnels-T apparaît ni plus ni moins innocent que le choix arbitraire d'une théorie de la vérité parmi toutes les théories possibles, et le repli sur les biconditionnels-T une impasse méthodologique.

Au moment de conclure cet aperçu des tentatives logico-philosophiques de résolution des paradoxes de la vérité, tentatives tant sémantiques qu'axiomatiques, tant classiques que paracomplètes[119] ou paraconsistantes, le lecteur restera peut-être sur l'impression qu'en matière de réponse aux paradoxes c'est aujourd'hui moins l'imagination logique qui manque que des lignes philosophiques directrices qui permettraient d'affermir notre jugement. Il n'est donc pas étonnant qu'un certain nombre d'articles aient paru depuis une dizaine d'années pour tenter de clarifier ce que nous devons prioritairement attendre d'une solution des paradoxes de la vérité[120]. Cette réflexion est cependant loin d'être achevée tant il est vrai qu'elle appelle une ré-évaluation philosophique plus large de nos idées sur la nature de la logique, de la vérité, et même de l'objectif que nous poursuivons quand nous prétendons vouloir résoudre les paradoxes. Dans la dernière section nous reprenons l'autre fil que nous avions dégagé au départ dans le travail de Tarski, le fil de l'analyse conceptuel de la vérité, et nous discutons de sa postérité contemporaine. N'est-ce pas justement d'une analyse de genre que nous attendons un guide sur les voies du paradoxe ?

118. Soit ϕ un énoncé prouvable dans T. On peut prendre par diagonalisation le point fixe de $Vr(x) \leftrightarrow \phi$. On a : $PA \vdash \psi \leftrightarrow (Vr(\psi) \leftrightarrow \phi)$, donc $PA \vdash (\psi \leftrightarrow Vr(\psi)) \leftrightarrow \phi$ et l'on a un biconditonnel-T équivalent à ϕ. En collectant les biconditonnels obtenus de cette façon pour chaque théorème de T, on obtient une collection de biconditonnels-T qui prouve tous les théorèmes de T.

119. Les approches paracomplètes mettent en œuvre des logiques qui invalident le tiers exclu.

120. Leitgeb [2007] en est un exemple souvent cité, Horsten and Halbach [2015] en est un autre. Le travail sur les paradoxes dans Field [2008] est un des meilleurs exemples contemporain de recherche logique articulée à des considérations épistémologiques générales et systématiques.

5. Théories de la vérité et nature de la vérité

5. 1 Déflationnisme

Tarski a défendu et promu la fécondité de l'usage de la notion de vérité dans les sciences formalisées, mais la relativisation du projet de définition de la vérité à un langage donné[121], la neutralité revendiquée par Tarski à l'égard de toute analyse métaphysique, et la place conceptuellement centrale qu'il a donnée à ces platitudes que sont les biconditionnels-T laissent ouvertes nombre de questions traditionnelles sur la vérité. Si certains philosophes ont vu là une limitation majeure de la portée philosophique de l'entreprise tarskienne, d'autres ont au contraire tenté de saisir la positivité philosophique de cette économie en réarticulant sur cette base même un discours philosophique qui réponde aux interrogations traditionnelles sur la nature de la vérité.

On peut en effet distinguer deux attitudes philosophiques et avec elles deux types d'attentes, vis-à-vis de la notion de vérité. Partant de l'observation que la notion de vérité articule une norme centrale de la relation de notre langage au monde, une analyse de la vérité doit, selon un premier point de vue, d'une part expliciter la nature de la relation du langage au monde et d'autre part, à la lumière de cette explicitation, nous dire en quoi consiste la satisfaction de cette norme. Selon ce point de vue, la discussion sur la nature de la vérité est intimement liée au débat entre réalisme et anti-réalisme et ne peut pas en faire l'économie. Si le réalisme est vrai, le discours articule des représentations d'états du monde, et la vérité pourrait être définie comme correspondance de ces représentations à la réalité[122]. S'il n'existe au contraire rien de tel qu'une représentation du monde tel qu'il est, alors l'idée centrale d'adéquation de nos « représentations » doit être définie autrement en lien, peut-on présumer, avec certaines dimensions de notre expérience, quelque soit la manière dont celle-ci est définie[123]. Si « vé-

121. Tarski ne définit pas « Vrai dans $\mathcal{L}$ » pour $\mathcal{L}$ variable. Voir Black [1949] pour une critique sur cette base de la portée du travail de Tarski en tant qu'élucidation de la notion de vérité. Ce type de critique est repris par Putnam [1985] ou Putnam [1991]. Field [1972]) voit dans cette limitation un échec du projet de réduction de la notion de vérité à des notions non sémantiques, projet qui demeure suspendu à une réduction ultérieure de la notion générale de dénotation des noms et des prédicats d'un langage.

122. En dépit des apparences le programme n'a rien d'une promenade de santé. Il faut expliquer ce que sont les représentations, ce que sont les choses susceptibles de leur correspondre ou non – des faits ?, mais que sont des faits sinon ce qui correspond à ces représentations ? –, et en quoi consiste cette mystérieuse correspondance.

123. En tenant compte ou non de sa dimension normative existentielle – de nos buts, de ce

rité » est le nom de cette adéquation alors, selon cette attitude exigeante, la théorie de la vérité aura la lourde charge de déployer toutes les médiations de cette relation complexe de notre pensée au monde. Que nos inclinations soient réalistes ou anti-réalistes, la théorie platonicienne des Formes et de la participation, le *Discours de la méthode* de Descartes, la *Critique de la raison pure* de Kant, ou *Raison, Vérité et Histoire* de Putnam sont alors essentiellement autant de théories de la vérité, bien que la notion de vérité n'y soit que marginalement travaillée pour elle-même ; et réciproquement une théorie de la vérité peut difficilement être d'emblée autre chose qu'une grande théorie métaphysique ou épistémologique[124].

Une autre approche abandonne ces présupposés sur la charge qui échoit à une théorie de la vérité – une telle charge, demandera-t-on, peut-elle seulement être honorée de façon rationnelle et conclusive ? Au lieu d'attaquer bille en tête la description des relations du langage au monde on s'interroge sur la notion de vérité elle-même, ses usages et son champ d'application légitime. Après tout, nous avons observé qu'une telle notion n'était peut-être même pas cohérente – quel sens aurait alors l'idée que le réalisme métaphysique est une certaine conception de la vérité ? C'est dans cette seconde perspective – que l'on pourrait qualifier de perspective critique sur la notion de vérité – que le travail de Tarski prend un relief philosophique particulier et que peut se développer ce que l'on a appelé le déflationnisme en matière de vérité.

Dans cette seconde perspective, la remarque fondatrice serait plutôt celle de l'évanescence des attributions de vérité. Frege [1971] observe déjà que le prédicat de vérité semble ne rien ajouter à l'expression d'une pensée :

> Il vaut aussi de remarquer que la proposition *je sens une odeur de violette* a même contenu que la proposition *il est vrai que que je sens une odeur de violette*. Il semblerait que rien n'est ajoutée à la pensée quand je lui attribue la propriété d'être vraie. [...] Serait-ce que nous ayons à faire à quelque chose qui ne peut nullement être appelé propriété dans le sens usuel ? (La pensée, *in* Frege [1971], p. 174)

Ramsey[125] observe également que dans ses emplois les plus simples le pré-

que nous regardons comme nos succès, ou seulement de ce qui nous serait donné dans la perception comme étant la perception du monde et non de nous-même par exemple, ou ce qui survit de cette expérience à l'épreuve de l'intersubjectivité etc.

124. Pour une introduction plus large aux théories de la vérité, on pourra consulter Ludwig [2016], Engel [1989] et en anglais Kirkham [1995].

125. Ramsey [1927].

dicat de vérité semble être *redondant*, et ajoute[126] que dans les emplois même où il est inéliminable – comme dans l'énoncé « La dernière phrase prononcée par Platon est vraie » – le prédicat de vérité pourrait être éliminé sans perte si nous disposions d'outils de quantification sur les positions d'énoncés – « $\forall p$ (Platon a dit que $p \rightarrow p$) » dit passablement la même chose que l'énoncé mentionné précédemment[127]. C'est à Quine toutefois que l'on fait généralement remonter le déflationnisme contemporain. Quine juge la notion de vérité suffisamment expliquée au cas par cas par chaque biconditionnel-T[128]. mais l'équi-assertabilité d'un énoncé et de l'attribution de la vérité à cet énoncé est à mettre en perspective avec ses thèses sceptiques quant à la possibilité d'objectiver la signification d'un énoncé[129]. La vérité est donc une notion immanente à un langage donné[130] et il ne peut y avoir de théorie générale de la vérité pour un langage quelconque en général. L'intuition classique de la vérité comme correspondance est saisie par les biconditionnels-T mais ne peut-être consolidée dans une théorie générale de la correspondance du langage au monde. Ce que l'on pourrait appeler l' « effet de correspondance » est plutôt une projection induite par notre disposition à parler du monde en faisant un détour par un discours sur notre propre langage, comme lorsque je dis que « Barcelone » est le nom d'une capitale européenne. Lors de cette « montée sémantique » je me mets à parler du langage et j'ai besoin d'un outil pour restaurer la référence au réel et continuer à parler du monde tout en parlant du langage – c'est précisément le rôle des notions sémantiques comme la dénotation et la vérité. Ce rôle est d'autant plus crucial, ajoute le déflationniste, que la montée sémantique est parfois inévitable : comment affirmer autrement qu'en ces mots que la dernière phrase formulée par Platon était certainement vraie, ou que tous les théorèmes de PA sont vrais ? Mes limites épistémiques sont telles que je ne peux ni affirmer la dernière phrase énoncée par Platon ni affirmer tous les théorèmes de PA. Du point de vue de Quine la notion de vérité répond donc essentiellement à un besoin logico-linguistique. En face, les contempteurs du déflationnisme ne nient pas que le prédicat de vérité contribue à l'expressivité du langage, mais ils insistent sur le fait que la

126. Dans Ramsey [1991].
127. Voir Rivenc [1998] pour une présentation et une discussion plus approfondie de la position de Ramsey.
128. On renverra le lecteur aux passages de Quine [1993], ou Quine [2008] chap. 2 pour des expressions nettes de ce point de vue.
129. Sauf de façon triviale pour les énoncés de son propre langage : « La neige est blanche » signifie très exactement que la neige est blanche.
130. L'idée selon laquelle la notion de vérité est « immanente » au langage est formulée au paragraphe 6 de Quine [2010].

contribution du prédicat de vérité (en plus de celle des quantificateurs et autres auxiliaires logiques tout à fait classiques mobilisés) réside justement en ceci qu'il signifie l'existence d'un certain genre de relation objective entre le langage et le monde. En somme, les déflationnistes soutiennent que l'usage logique est premier et que les autres usages se comprennent à cette lumière[131], leurs opposants que l'usage descriptif et explicatif de la vérité est premier et que l'usage logique est dérivé[132]. Cette opposition, on le pressent, n'est pas sans conséquence sur le problème des paradoxes qui nous a occupé dans les sections précédentes. En effet si le prédicat de vérité est une notion purement logique dont le rôle est l'expression de la généralité alors, dans la mesure où ce rôle est un effet direct de l'intersubstituabilité de $Vr(\phi)$ et ϕ en tout contexte extensionnel, cette intersubstituabilité s'impose au déflationniste plus qu'à tout autre comme l'horizon de toute solution des paradoxes de la vérité[133].

Mais comment trancher entre ces deux perspectives sur la vérité, quel critère tangible pourrait avoir valeur de test ? Une façon d'en préciser l'opposition est de la formuler en termes de pouvoir explicatif de la notion de vérité. S'il s'avère nécessaire de faire appel à une notion substantielle de vérité pour rendre compte de certains faits alors la perspective déflationniste s'en trouve certainement affaiblie ; si à l'inverse le recours à une théorie substantielle de la vérité est inutile et que le rôle logique du prédicat de la vérité s'avère suffisant pour rendre compte de nos meilleurs usages, alors c'est certainement quelque chose qui doit compter en sa faveur. Dans cette formulation, le problème peut recevoir une interprétation logiquement bien définie en termes de conservativité, une notion que nous avons déjà rencontrée dans la première section.

5. 2 La question de la conservativité

La thèse du caractère purement expressif de nos usages du prédicat de vérité se heurte – en tout cas de prime abord – à un certain nombre de re-

131. Par exemple Field [2001], Horwich [1998]. En travaillant le lien avec les paradoxes, Leon Horsten parle de notion « purement inférentielle » dans Horsten [2009] et Horsten [2011] : une notion inférentielle au sens de Horsten est une notion dont il n'y pas de loi absolument générale, et le prédicat est (prouvablement) de ce type.

132. C'est par exemple la position défendue par Gupta [1993].

133. Ainsi le déflationnisme professé par Hartry Field est-il l'une de ses motivations déclarées pour sauver la théorie naïve de la vérité.

marques que nous avons pu faire dans la première section relativement à la fécondité de l'usage de la notion de vérité. Si l'indéfinissabilité de la vérité n'est pas frontalement incompatible avec les thèses déflationnistes[134], les usages théoriques de la notion mis en avant par Tarski lui-même ne sont-ils pas autant de pierres dans le jardin déflationniste ? Le fait logique de l'existence de preuves sémantiques, de preuves *par* la vérité ne constitue-t-il pas une réfutation pure et simple de la thèse philosophique du caractère non-explicatif de la notion de vérité ? Il y a là une difficulté potentielle pour le déflationniste qu'ont bien vue Stewart Shapiro et Jeffrey Ketland[135].

En quelques mots, leur argument est le suivant :

1. (Réflexion) Une théorie T augmentée d'une théorie de la vérité dans L_T doit permettre de prouver que tous les théorèmes de T sont vrais.
2. (Conservativité) Une théorie déflationniste de la vérité doit être conservative
3. Une théorie de la vérité qui satisfait 1, n'est pas conservative.
4. Donc le déflationnisme est faux.

La première prémisse impose une contrainte à toute théorie de la vérité, à savoir, d'être logiquement suffisamment riche pour impliquer en conjonction avec une théorie T le principe de réflexion sur T. Supposons en effet que j'accepte la théorie T et que je possède un concept de vérité et une description syntaxique de T en tant que théorie. Alors dans l'esprit déflationniste, rien ne devrait me manquer pour affirmer la vérité de T, c'est-à-dire que tous les théorèmes de T sont vrais. Observons que cette première contrainte condamne par elle-même la théorie décitationnelle simple de la vérité au moins à titre logique, théorie de l'air d'innocence de laquelle le déflationniste s'inspire pour développer ses thèses philosophiques (Horwich [1998]). Elle fait porter au déflationniste la charge de présenter des principes aléthiques plus forts – par exemple la théorie compositionnelle tarskienne CT. On se souviendra alors que DT est justement conservative sur PA – au contraire de CT qui ne l'est pas[136].

134. Le déflationniste ne prétend pas que la vérité est définissable, il défend même jusqu'à un certain point une thèse contraire en la réputant indispensable.

135. Shapiro [1998], Ketland [1999]. On pourra se reportera à Cieslinski [2017] pour une bibliographie du débat. Ce débat sur la portée des arguments de conservativité contre le déflationnisme fait largement écho au débat antérieur sur les relations entre les théorèmes d'incomplétude et les programmes nominalistes pour les mathématiques.

136. Nous avons vu qu'il n'y avait pas de consensus sur ce que doivent être ces lois dans le cas où nous souhaitons faire la théorie de Vrai-dans-L dans L lui-même. Nous avons vu en outre que les théories décitationnelles pour les prédicats de vérité non typés pouvaient réserver des

La seconde prémisse de l'argument mobilise justement la notion de conservativité pour donner une explication, au sens de Carnap, du déflationnisme. Et ce critère a le mérite d'être intuitivement correct : comment soutenir que la notion de vérité ne joue pas de rôle explicatif si la théorie de la vérité permet d'expliquer (c'est-à-dire de prouver) des faits, en particulier des faits arithmétiques, que la théorie arithmétique que nous acceptons ne permet pas par elle-même de prouver ?

La troisième étape du raisonnement est simplement un fait logique. Si en effet l'extension aléthique de T permet de prouver « Tous les théorèmes de T sont vrais », alors elle permet de prouver que T est consistante (c'est le théorème de consistance de la section 1). Or le second théorème d'incomplétude de Gödel nous garantit que l'énoncé de la consistance de T[137] n'est pas dérivable dans T elle-même. Donc l'extension aléthique de T, quels qu'en soient les détails, ne peut être conservative sur T si elle satisfait (Réflexion).

Afin d'éviter une conclusion désastreuse pour le déflationniste il n'y a donc apparemment que deux possibilités : refuser le critère de conservativité (prémisse 2) ou refuser le critère de réflexion (prémisse 1). Les réponses à cet argument sont nombreuses, nous ne mentionnons que quelques possibilités. Field [1999] accepte le critère de conservativité mais refuse la conclusion de l'argument au motif que la non-conservativité n'est pas ici un effet de la théorie de la vérité. En effet, comme nous l'avons fait remarqué plus haut[138] l'extension de PA par la théorie compositionnelle CT ne suffit pas pour dériver que tous les théorèmes de PA sont vrais : il faut en outre étendre certains principes de preuve de PA elle-même, à savoir de nouvelles instances du schéma d'induction mobilisant les énoncés du langage étendu au prédicat de vérité. PA+CT *stricto sensu* est en fait une extension conservative de PA ! Il n'y a qu'un pas faire pour conclure que la non-conservativité n'est pas un effet du caractère explicatif des principes aléthique mais du pouvoir expressif qu'ils confèrent au prédicat de vérité : le pouvoir expressif de la vérité est recruté par le schéma d'induction pour augmenter la force preuve-théorique de PA et la non-conservativité, vue

surprises. Nous nous en tiendrons ici à la discussion de l'argument de la conservativité telle qu'elle se présente dans la situation la plus simple, en relation avec les prédicats de vérité typés. Il s'agit d'une retraite provisoire, étant admis que les contraintes qui pèsent sur une réponse aux paradoxes sont en principe à prendre en compte pour ajuster notre conception de la nature de la vérité, et que la réciproque est vraie également.

137. Formulé avec un prédicat de prouvabilité qui satisfait les trois conditions mentionnées chap. 20 section 4.2. du présent volume.

138. Section 1, note 39.

sous cet angle, serait donc parfaitement en ligne avec les attentes déflationnistes. D'autres philosophes sont plus enclins à récuser le critère de conservativité. Ainsi, Shapiro lui-même suggère au déflationniste d'adopter une notion plus riche de conséquence sémantique qui permette de sécuriser une forme de conservativité sémantique de la notion de vérité – l'argument de la conservativité n'est plus tant un argument contre le déflationnisme qu'un argument en faveur de la logique du second ordre à l'usage du déflationniste[139]. *A contrario* on peut faire remarquer que l'exigeance de conservativité peut être renforcée jusqu'à faire douter de sa signification et suggérer ainsi l'arbitraire de la notion choisie. Ainsi la théorie décitationnelle de la vérité elle-même n'étend pas conservativement PA au sens où certains modèles de PA ne peuvent pas être étendus à des modèles de $PA + DT$[140]. Halbach [2001] observe également que la conservativité d'une théorie de la vérité sur la logique pure est une demande qu'il n'est pas possible de satisfaire[141] ; et l'on peut à l'inverse noter que la conservativité sur PA augmenté d'une règle d'inférence infinitaire de type règle-ω est toujours satisfaite par n'importe quelle théorie de la vérité.[142] Certains philosophes ont proposé des critères logiques autres que la conservativité avec l'ambition de préciser l'intuition déflationniste tout en assurant la cohérence de sa position[143] ; d'autres suggèrent des pistes philosophiques différentes et logiquement plus ouvertes[144].

5. 3 Les principes de réflexion

Gödel avait déjà fait observer que la non-conservativité pouvait recouvrir des réalités épistémologiques différentes. Dans son essai sur l'hypothèse du continu[145], commentant l'indécidabilité dans ZF de différentes

139. La formulation de l'arithmétique de Peano en second ordre, PA^2, est sémantiquement complète, par conséquent l'adjonction à PA^2 de principes aléthiques consistants avec PA^2 forme toujours une extension sémantiquement conservative de PA^2 : si $PA^2 + T$ est consistante alors pour tout énoncé arithmétique ϕ, $PA^2 + T \vDash \phi \Rightarrow PA^2 \vDash \phi$.
140. **?**.
141. En substance, la raison en est que le prédicat de vérité s'applique à des objets – les énoncés – dont l'existence n'est pas une vérité purement logique mais est impliquée par toute théorie de la vérité.
142. PA augmenté d'une règle infinitaire de type règle-ω est une théorie complète.
143. Par exemple Fischer [2015].
144. Horsten [2011] sur le caractère purement inférentiel de la vérité, ou Bonnay and Galinon [2019].
145. Gödel [1990]

hypothèses ensemblistes, il s'interroge sur le type de justification que nous pouvons mobiliser en faveur de nouveaux axiomes. Gödel propose alors de distinguer entre extensions *intrinsèques* et extensions *extrinsèques* de ZF. Le fondement de cette distinction, selon Gödel, est que nous disposons pour les premières de sources de justifications qui sont fondamentalement les mêmes que celles qui justifient que nous affirmions les axiomes de ZF eux-mêmes – notre compréhension de la conception itérative des ensembles. Pour les secondes, en revanche, les justifications sont hétérogènes à nos justifications des axiomes de ZF. Gödel pensait ainsi que les axiomes de grand cardinaux appartiennent à la première catégorie tandis que l'axiome du choix relevait de la seconde et introduisait quelque chose de fondamentalement nouveau dans la théorie des ensembles. On peut naturellement penser que pour Gödel les principes de réflexion qui affirment la consistance d'une théorie, ou la vérité de tous ses théorèmes, constituent des extensions du premier type[146]. En tout état de cause, ces remarques invitent à réexaminer les statuts épistémologiques des principes de réflexion. Elles invitent à se demander, en particulier, s'il est possible de donner un sens précis à l'idée selon laquelle si l'on accepte une théorie T, on est *ipso facto* justifié à accepter des principes de réflexion qui en sont logiquement indépendants, tels la consistance de T ou la vérité des théorèmes de T. Cette idée de déploiement progressif du contenu implicite dans l'acceptation d'une théorie est à la racine du programme de recherche logique développé sous différentes formes par Feferman dans la seconde moitié du vingtième siècle[147]. Resituée dans un contexte plus large, l'évaluation du statut des principes de réflexion renvoie une interrogation multiforme sur l'articulation entre croyances de premier ordre et croyances de second ordre, et fait apparaître des liens avec la question classique de l'autorité en première personne aussi bien qu'avec des questions contemporaines d'épistémologie[148], avec la question classique du statut du discours réflexif aussi bien qu'avec les recherches contemporaines en philosophie de l'esprit sur les normes et

146. Voir aussi Koellner [2009] pour une formulation explicite de cette idée.

147. Voir en particulier Feferman [1962] et Feferman [1991]. Pour une discussion des racines de ce programmes, voir Feferman [2005]. Pour une discussion des résultats de complétude par itération transfinie des principes de réflexion on pourra consulter Franzen [2004] . Dean [2015] contient une discussion du contenu arithmétique des principes de réflexion avec le débat sur la notion de vérité.

148. La question du conflit entre la règle consistant à proportionner ses croyances aux preuves disponibles et les bornes de la liberté de croire dans van Fraassen [1984] en est un exemple, la question de la justification de la formation de mes croyances sur le fondement du témoignage d'autrui dans Coady [1992] en est un autre.

fonctions métacognitives.[149] Il n'est pas impossible que sur ce chemin qui l'éloigne des questions définitionnelles et fondationnelles nées des préoccupations positivistes, le déflationniste en vienne à infléchir son discours. Après tout, si posséder la notion de vérité est une condition de la dynamique interne de notre rationalité, ne devient-il pas urgent, sans renoncer aux acquis des débats antérieurs, d'approfondir l'étude des fonctions de la notion de vérité et pourquoi pas d'insister désormais sur son importance et la grandeur de ce qu'elle nous permet d'accomplir en première personne et en société ?

Bibliographie

J. C. Beall and J. Murzi. Two flavors of curry's paradox. *The Journal of Philosophy*, 110(3) :143–165, 2013.

M. Black. *Language and Philosophy*. Cornell University Press, Ithaca, 1949.

D. Bonnay and H. Galinon. Deflationary truth is a logical notion. In M. Piazza and M. Pulcini, editors, *Truth, existence and explanation*. Springer, 2019.

G. S. Boolos, J. P. Burgess, and R. C. Jeffrey. *Computability and Logic : Fourth Edition*. Cambridge University Press, 2002.

A. Cantini. The undecidability of grisin's set theory. *Studia Logica*, 74(3) : 345–368, Aug 2003.

C. Cieslinski. *The Epistemic Lightness of Truth. Deflationism and Its Logic*. Cambridge University Press, 2017.

149. Ainsi, les pragmatistes classiques ont inspiré au psychologue Asher Koriat [2012] l'hypothèse que l'évaluation de la cohérence des croyances premier ordre était l'outil cognitif privilégié pour certifier leur vérité au sens de la correspondance, tandis que les travaux recueillis dans Proust and Fortier [2018] suggèrent une large variabilité culturelle des normes métacognitives.

C. Coady. *Testimony*. Clarendon Press, 1992.

P. Cobreros, P. Egré, D. Ripley, and R. van Rooij. Reaching transparent truth. *Mind*, 122(488) :841–866, 2014.

N. Damnjanovic and S. Candlish. A brief history of Truth. In D. Jacquette, editor, *Handbook of the Philosophy of Sciences*, volume 5 : Philosophy of logic. Elsevier, 2007.

W. Dean. Arithmetical reflection and the provability of soundness. *Philosophia Mathematica*, 23(1) :31–64, 2015.

M. Dummett. Truth. *Proceedings of the Aristotelian Society*, 59, 1958-1959.

P. Engel. *La Norme du Vrai*. Gallimard, 1989.

S. Feferman. Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, 27 :259–316, 1962.

S. Feferman. Reflecting on incompleteness. *Journal of Symbolic Logic*, 51 : 1–48, 1991.

S. Feferman. Predicativity. In S. Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590–624. Oxford University Press, 2005.

S. Feferman. Axioms for determinateness and truth. *Review of Symbolic Logic*, 1(3) :204–217, 2008.

H. Field. Tarski's theory of truth. *Journal of Philosophy*, 69 :347–375, 1972. Trad. *in* Bonnay et Cozic [2009].

H. Field. Deflating the conservativness argument. *Journal of Philosophy*, 96 :533–540, 1999.

H. Field. *Truth and the Absence of Fact*. Oxford University Press, 2001.

H. Field. *Saving truth from paradox*. Oxford University Press, 2008.

M. Fischer. Deflationism and instrumentalism. In T. Achourioti, K. Fujimoto, H. Galinon, and J. Martinez, editors, *Unifying the philosophy of truth*. Springer Verlag, 2015.

T. Franzen. *Inexhaustability : A Non-Exhaustive Treatment*. Peters, A.K., 2004.

G. Frege. *Ecrits logiques et philosophiques*. Seuil, Paris, 1971.

H. Friedman and M. Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33 :1–21, 1987.

K. Gödel. What is Cantor's continuum problem? [1964]. In S. Feferman, editor, *Gödel Collected Works*, volume 2. Oxford University Press, 1990.

V. N. Grisin. Predicate and set-theoretic calculi based on logic without contractions. *Math. USSR Izvestija (English translation)*, 18(1) :41–59, 1982.

A. Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11 :1–60, 1982.

A. Gupta. A critique of deflationism. *Philosophical Topics*, 93, 1993.

A. Gupta and N. D. Belnap. *The Revision Theory of Truth*. MIT Press, 1993.

V. Halbach. Tarski hierarchies. *Erkenntnis*, 43(3) :339–367, 1995.

V. Halbach. How innocent is deflationism? *Synthese*, 126(1-2), 2001.

V. Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge, 2014.

H. G. Herzberger. Notes on naïve semantics. *Journal of Philosophical Logic*, 11 :61–102, 1982.

J. Hintikka. *The Principles of Mathematics Revisited*. Cambridge University Press, Cambridge, 1998.

W. Hodges. Truth in a structure. *Proceedings of the Aristotelian Society*, 86 :135–151, 1985.

L. Horsten. Levity. *Mind*, 118(471) :555–581, 2009.

L. Horsten. *The Tarskian Turn. Deflationism and axiomatic truth theories.* MIT Press, 2011.

L. Horsten and V. Halbach. Norms for theories of reflexive truth. In H. Galinon, T. Achourioti, J. Martinez, and K. Fujimoto, editors, *Unifying the Philosophy of truth*, 2015.

P. Horwich. *Truth*. Oxford University Press, 2nd edition, 1998.

J. Ketland. Deflationism and Tarski's paradise. *Mind*, 108 :69–94, 1999.

R. L. Kirkham. *Theories of Truth : a Critical Introduction*. MIT Press, 1995.

P. Koellner. On reflection principles. *Annals of Pure and Applied Logic*, 157 (2-3) :206–219, 2009.

A. Koriat. The subjective confidence in one 's knowledge and judgements : some metatheoretical considerations. In M. Beran, J. Brandl, J. Perner, and J. Proust, editors, *Foundations of Metacognition*, pages 213–232. Oxford University Press, 2012.

M. Kremer. Kripke and the logic of truth. *Journal of Philsophical Logic*, 17 :225–278, 1988.

S. Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72(2) : 690–716, 1975.

H. Leitgeb. What theories of truth should be like (but cannot be). *Philosophy Compass*, 2(2), 2007.

P. Ludwig. « Vérité », version académique. In M. Kristanek, editor, *L'Encyclopédie philosophique*. http ://encyclo-philo.fr/verite-a/, 2016.

P. Mancosu. Tarski, Neurath, and Kokoszynska on the Semantic Conception of Truth. In D. Patterson, editor, *New essays on Tarski and philosophy*. Oxford University Press, 2008.

V. McGee. Maximal consistent sets of instances of tarski's schema (t). *Journal of Philosophical Logic*, 21, 1992.

P. Milne. Tarski on truth and its definition. In T. Childers, P. Kolar, and V. Svoboda, editors, *Logica'96 : Proceedings og the 10th Internation Symposium*, Prague, 1997. Philosophia.

P. Milne. Tarski, truth and model theory. *Proceedings of the Aristotelian Society*, 99 :141–167, 1999. ISSN 00667374, 14679264. URL `http://www.jstor.org/stable/4545302`.

J. Murzi and M. Carrara. Paradox and logical revision. A short introduction. *Topoi*, 34(1) :7–14, 2014.

G. Priest. The logic of paradox. *Journal of philosophical logic*, 8 :219–241, 1979.

J. Proust and M. Fortier, editors. *Metacognitive diversity*. Oxford University Press, 2018.

H. Putnam. A comparison of something with something else. *New Literary History*, 17, 1985.

H. Putnam. Does the disquotational theory solve all philosophical problems? *Metaphilosophy*, 1991.

W. V. O. Quine. *Philosophy of logic*. Harvard University Press, Cambridge, Mass, 1970.

W. V. O. Quine. *La poursuite de la vérité*. Seuil, Paris, 1993.

W. V. O. Quine. *Philosophie de la logique*. Aubier Flammarion, 2008.

W. V. O. Quine. *Le mot et la chose*. Champ Flammarion, 2010.

W. V. O. Quine. Les voies du paradoxe. In S. Bozon and S. Plaud, editors, *Les voies du paradoxes et autres essais*. Vrin, 2011.

F. P. Ramsey. Facts and propositions. *Proceedings of the Aristotelian Society*, 7 :153–170, 1927.

F. P. Ramsey. On truth. *Episteme*, 16 :1–16, 1991.

A. Rayo and G. Uzquiano, editors. *Absolute generality*. Oxford University Press, 2006.

W. N. Reinhardt. Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic*, 15(2) : 219–251, May 1986.

G. Restall. *On logics without contraction*. PhD thesis, University of Queensland, 1994.

F. Rivenc. Ce que ramsey a vraiment dit, ou la théorie prophrastique de la vérité. *Philosophie*, (57) :16–50, 1998.

F. Rivenc. Définition et critère de vérité. *Philosophie*, (65), 2001.

d. Rouilhan, Philippe. Tarski et l'universalité de la logique. In F. Nef and D. Vernant, editors, *Le formalisme en Question*, Paris, 1998. Vrin.

P. d. Rouilhan and S. Bozon. La vérité de IF : Hintikka a-t-il vraiment exorcisé la malédiction de Tarski? In R. Auxier and L. E. Hahn, editors, *The Philosophy of Jaako Hintikka*. Open Court, 2006.

S. Shapiro. Proof and truth : Through thick and thin. *Journal of Philosophy*, 95(10) :493–521, 1998.

A. Tarski. The concept of truth in formalized languages. In *Logic, Semantics and Metamathematics*. Hackett Pub., Indianapolis (1983), 2nd edition, 1935. JH Woodger (trans.) ; First published as 'Der Wahrheitsbegriff in Den Formaliserten Sprachen', *Studia Philosophica* I (1935).

A. Tarski. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3) :341–376, 1944. Traduction partielle *in* Bonnay et Cozic [2009].

A. Tarski. Truth and proof. *Scientific American*, June :63–70, 1969.

A. Tarski. *Logic, Semantics, Metamathematics*. Hackett pub., 1983.

B. van Fraassen. Belief and the will. *The Journal of Philosophy*, 81(5) : 235–256, 1984.

A. Visser. *Semantics and the Liar Paradox*, pages 149–240. Springer Netherlands, Dordrecht, 2004.

E. Zardini. Truth without contra(di)ction. *The Review of Symbolic Logic*, 4 (4) :498–535, 2011.