



HAL
open science

Des preuves par la vérité

Henri Galinon

► **To cite this version:**

| Henri Galinon. Des preuves par la vérité. Figures de la vérité, éditions ISTE, 2019. hal-02323379

HAL Id: hal-02323379

<https://hal.science/hal-02323379v1>

Submitted on 21 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pénultième version.

A paraître dans Rivenc F. (éd.), *Figures de la vérité*, éditions ISTE, 2019.

Des preuves par la vérité.

Henri Galinon

I. Introduction

Il fait peu de doute que donner la preuve d'une proposition c'est prouver qu'elle est vraie et, réciproquement, que prouver la vérité d'une proposition n'est rien d'autre que de donner une preuve de la proposition en question. Cette équivalence a quelque chose de troublant. Pour établir que la neige est blanche il faut établir la blancheur de la neige, et pour établir qu'un homme est généreux il faut établir sa générosité. Le cas du prédicat de vérité est tout à fait singulier si l'on admet que pour établir qu'il est vrai que la neige est blanche il faut et il suffit d'établir... la blancheur de la neige. Car alors, même à admettre que la vérité est théoriquement distincte de la prouvabilité¹, le problème demeure de savoir ce que le recours à la notion de vérité, à sa définition, à ses lois, permettrait *d'établir* ou, pour le dire autrement, quelle différence la différence entre établir une proposition et établir sa vérité pourrait bien faire *en pratique*, donc pour des questions ou des tâches qui ne sont pas dictées par notre seul intérêt pour l'éclaircissement de la nature de la vérité. C'est cette interrogation qui nourrit le soupçon pragmatiste formulé par Richard Rorty:

« Les pragmatistes pensent que si quelque chose ne fait pas de différence en pratique, elle ne doit pas faire de différence en philosophie. Cette conviction leur rend suspecte l'insistance des philosophes sur la distinction de la justification et de la vérité » [ROR 95 p. 281]

Il y a donc un problème d'*existence de preuves par la vérité*. Sans preuve *par* la vérité, on pourra légitimement s'interroger sur l'intérêt pratique de cette notion de vérité. A l'inverse, pourtant, l'idée même de preuve par la vérité ne va pas sans susciter une certaine perplexité : comment le détour par des attributions de vérité à des propositions, plutôt qu'à ces propositions elles-mêmes, pourrait-il nous donner quelque prise épistémique que ce soit ?

Ce que l'on peut faire de la vérité, observera-t-on, dépend sans doute de la notion de vérité que l'on a vue : la théorie pragmatique de vérité-cohérence n'est pas équivalente à la théorie métaphysique des vérifacteurs (*truth makers*), qui est elle-même distincte de la théorie habermasienne de la vérité comme consensus ou de la théorie aristotélicienne de la vérité

¹ Contre une certaine forme d'anti-réalisme, en pointant par exemple qu'est vraie la loi selon laquelle "si p, alors 'p' est vrai", tandis que son double étrange "si p, alors 'p' est prouvable" ne peut être vraie que d'une certaine notion idéalisée, et peut-être douteuse, de prouvabilité.

comme correspondance, chacune portant avec elle, sous réserve qu'elles puissent être formulées de façon satisfaisante, ses usages propres et ses éventuelles conséquences pratiques. Mais l'existence et l'intérêt de notre problème implique simplement ceci : que la notion de vérité obéisse à l'équivalence fondamentale que nous avons formulée entre une proposition p et la proposition qu'il est vrai que p ou, pour parler en termes d'énoncés, entre un énoncé p et l'énoncé « p est vrai ». Sans doute de nombreuses théories de la vérité prennent-elles cette équivalence au sérieux, et peu nombreux sont les philosophes ouvertement disposés à la récuser entièrement.² Elle occupe toutefois une place centrale dans le « déflationnisme » contemporain, lequel présente pour cette raison un intérêt spécial pour notre étude. C'est que le déflationniste tient que la vérité ce n'est en effet *que* cela : une notion qui obéit à une forme de ce que j'ai appelé ci-dessus le principe d'équivalence (et qu'il faudra préciser le temps venu). L'évaluation des usages de la vérité permis par le seul principe d'équivalence - le point dont nous sommes parti -, est donc solidairement une évaluation du déflationnisme comme théorie de la vérité. Or on a généralement, je crois et c'est ce dont j'espère donner une idée dans ce qui suit, sous-estimé la valeur que confère à la notion de vérité ce seul principe d'équivalence et, par conséquent, telle sera ma conclusion paradoxale, sous-estimé la valeur que le déflationnisme – ou peut-être devrais-je dire un certain déflationnisme – confère à la vérité. Mais c'est sans doute par là, par le déflationnisme, qu'il faut commencer.

II. Déflationnisme

En quoi une réponse à la question « qu'est-ce que la vérité ? » doit-elle consister ? On peut distinguer sur cette question deux attitudes possibles. J'appellerai la première l'attitude « classique », la seconde l'attitude « critique ». L'attitude classique exige une réponse à la question de la nature de la relation du langage au monde et, sur cette base, une explication de ce en quoi cela consiste pour un discours ou une pensée, de satisfaire à la norme du vrai. Selon ce point de vue, la discussion sur la nature de la vérité est intimement liée au débat entre réalisme et anti-réalisme et, surtout, ne peut pas véritablement en faire l'économie. Ainsi, une réponse « réaliste » à la question de la nature de la vérité tentera de défendre l'idée que le discours/la pensée articule des représentations, et de définir la vérité comme correspondance de ces représentations à la « réalité ». On sait que ce programme n'a rien d'une promenade de santé. Il faudra ainsi expliquer ce que sont les représentations, en donnant un compte-rendu plausible de la nature de ce qu'elles représentent, de ce en quoi consistent leurs propriétés représentationnelles et de la façon dont nous les acquérons. Une réponse « antiréaliste » tendra à défendre l'idée qu'il n'existe au contraire rien de tel qu'une représentation du monde tel qu'il est, et cherchera à construire l'idée d'adéquation de nos pensées (croyances, assertions) comme une forme de convenance à certaines dimensions de notre expérience, celle-ci pouvant être définie de façon plus ou moins large en relation avec nos perceptions sensibles, à l'expérience objective ou intersubjective, voire à nos buts, à un ensemble de normes qui la structure et à notre inscription dans une communauté. Si "vérité" est le nom de cette adéquation alors,

² Encore que la possibilité d'aménagements permettant de l'accommoder ne soit pas toujours évidente. Faut-il accepter que p si, et seulement si il est *maximalement utile sur le long terme* d'accepter que p ? Pour accepter cette équivalence il faudrait admettre, ce qui n'a rien d'évident, qu'il est *maximalement utile sur le long terme* de l'accepter.

toujours selon cette attitude classique exigeante, il échoit à l'antiréaliste qui souhaite présenter sa conception de la vérité d'articuler les médiations de cette relation complexe de notre pensée à notre expérience du monde. Selon l'attitude classique, et dans des versions réalistes ou contraire antiréalistes, la théorie platonicienne des Formes et de la participation, les *Méditations métaphysiques*, les *Essais sur l'entendement humain* ou la *Critique de la raison pure* sont autant de théories de la vérité bien que, pourrait-on soutenir, la notion de vérité n'y soit que marginalement travaillée pour elle-même; et réciproquement une théorie de la vérité peut difficilement être autre chose qu'une grande théorie métaphysique ou épistémologique.

Au contraire ce que j'ai appelé, faute de mieux, l'approche critique, abandonne ces présupposés touchant 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 concluante? De quel point de vue l'entreprend-on? Comment l'intègre-t-on au vaste édifice du savoir et en particulier aux sciences naturelles? Au lieu de présumer de ce que doit accomplir une théorie de la vérité, l'approche critique recommande de s'arrêter d'abord plus longuement à l'analyse de l'idée de vérité elle-même, à ses usages, à son champ d'application. C'est cette perspective qui est au travail lorsque Frege, dans un passage souvent cité, fait observer 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 je sens une odeur de violette*. Il semblerait que rien n'est ajouté 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 [FRE 71, p. 174])

Ramsey fait preuve de la même attention lorsqu'il note que, dans ses emplois les plus simples le prédicat de vérité semble être *redondant*, et qu'il ajoute que, même dans les emplois où il semble 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.³

Un tel examen de la notion de vérité entrepris dans cette perspective critique pourrait *in fine* être compatible avec, et se voir complété par, le développement de ce que nous avons appelé une approche classique. C'est par exemple certainement le projet de Ramsey lui-même⁴. Mais tel n'est pas nécessairement le cas et deux autres issues sont possibles. La première : l'approche critique pourrait tout à fait nous découvrir que la notion de vérité est profondément incohérente et qu'elle doit être abandonnée – quelque chose comme une conclusion nietschéenne s'en suivrait, une forme de nihilisme aléthique. La seconde nous retiendra davantage ici : l'analyse critique pourrait révéler à la fois que la notion de vérité est parfaitement légitime *et* que ce que nous avons appelé l'approche classique est une impasse, provisoire ou définitive. C'est

³ Voir [RIV 98] pour une présentation et une discussion plus approfondie de la position de [RAM 27] et [RAM 91].

⁴ Voir [RIV 98] sur ce point.

précisément dans une affirmation de ce genre que je vois l'un des traits caractéristiques du déflationnisme.

Le déflationnisme comporte en effet trois ingrédients fondamentaux, et c'est sans doute chez Quine qu'on les discerne le mieux. Quine juge d'abord, c'est le premier ingrédient, la notion de vérité suffisamment expliquée au cas par cas par chaque biconditionnel-T, c'est-à-dire par les énoncés du type⁵ :

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

On aura reconnu dans ce biconditionnel une version du principe d'équivalence par lequel nous avons commencé. Or, comme le note Quine, l'intuition classique de la vérité comme correspondance est bien saisie par les biconditionnels-T, chacun énonçant clairement et sobrement quelles conditions doivent être réalisées pour qu'un énoncé soit vrai. Pour avoir une définition philosophiquement satisfaisante du prédicat de vérité, il *suffit* donc de suivre la voie ouverte par Tarski et de donner une définition qui implique les équivalences-T.

Cependant, de ce que l'on peut rendre compte de l'intuition de la correspondance, et c'est là le mouvement critique et le deuxième ingrédient, il ne suit pas qu'on puisse élaborer une théorie générale de la vérité comme correspondance du langage au monde, générale au sens d'une théorie qui expliquerait dans notre langage à quelles conditions un énoncé est vrai pour L avec L *variable*. Les raisons de Quine pour juger cette théorie générale impossible sont ultimement reliées à ses considérations sur la notion de référence : on pourrait certes utiliser le travail de Tarski pour réduire la vérité pour un langage L quelconque à une théorie de la référence dans L⁶, mais il faudrait en rester là car il ne peut en général exister, selon Quine, de théorie empiriquement fondée permettant de déterminer (dans notre langage donc) la référence d'une expression référentielle de L (lorsque L n'est pas notre langage). La célèbre expérience de pensée de la traduction radicale est supposée montrer que l'affirmation « L'expression « gavagai » réfère aux lapins » n'a pas de valeur de vérité déterminée, ou plus généralement que de nombreux schèmes de références sont compatibles avec l'ensemble des comportements linguistiques observables d'un locuteur. Pour Quine les seules attributions de référence qui ont des conditions d'assertion *déterminées* doivent donc être arrimées d'une façon ou d'une autre à celles qui ont des conditions d'assertion *logiquement déterminées*, c'est-à-dire celles que je fais dans mon propre langage sur le modèle de « 'Lapin' réfère aux lapins ». De même, les équivalences-T dont les conditions d'assertion sont objectivement déterminées sont celles qui sont dérivables à partir de celles que chacun formule à propos d'énoncés de son propre langage,

⁵ On renverra le lecteur à [QUI 93] et [QUI 70 chap. 2] pour des expressions nettes de ce point de vue.

⁶ Faute de place nous passons vite sur ce point. En gros : l'énoncé « Pa » est vrai si et seulement si ce qui est désigné par « a » possède la propriété désignée par « P ». Tarski montre que la vérité des énoncés complexes est une construction logique à partir de la satisfaction d'énoncés simples conçue sur ce modèle. Pour une clarification plus complète de ce qu'il faut entendre ici par la réduction de la vérité à la référence dans le cas où L est une variable, et plus généralement sur le lien entre cette question et celle des conditions de succès d'une réduction empirique de la notion de vérité à partir du travail de Tarski, je renvoie le lecteur au bel article de Field [FIE 72].

sur le modèle de l'affirmation que « la neige est blanche » est vrai si, et seulement si, la neige est blanche.⁷

Le dernier ingrédient concerne l'usage du prédicat vérité. L'apport et l'originalité de Quine est de le fonder entièrement sur ce qu'il qualifie de nécessité logique. Car comment affirmer autrement qu'en ces mots que la dernière phrase formulée par Platon était vraie, ou que tous les énoncés de la forme « p ou non p » sont vrais ? C'est un problème de limites épistémiques, de finitude : je ne peux pas affirmer la dernière phrase énoncée par Platon, pas plus que je ne peux affirmer tous les énoncés de la forme « p ou non p ». Je pourrais certes dire que je crois (le contenu exprimé par) la dernière phrase de Platon, ou que je suis disposé à endosser tous les énoncés de la forme « p ou non p », mais je formulerais alors des observations sur mes états mentaux, et non un contenu en rapport logique direct avec celui de la dernière phrase prononcée par Platon ni la généralisation attendue des instances du schéma « p ou non p ». Parce que nous sommes des êtres finis, il nous faut donc parfois nous résoudre à parler du langage, à faire des détours par la *désignation* ou la *description* de contenus que nous ne pouvons exprimer ou penser directement, pour continuer à parler de réalités non linguistiques, et c'est à cela que sert le prédicat de vérité.⁸ En quel sens exactement, en parlant de la vérité d'énoncés ou de propositions, arithmétiques par exemple, continuai-je à parler d'arithmétique ? Au sens où, pour continuer avec cet exemple, mon affirmation que la théorie arithmétique axiomatisée par Peano est vraie est une affirmation qui a des conséquences *arithmétiques* (par exemple $2+2=4$!), conséquences que mes affirmations que cette théorie est intéressante, ou élégante, ou que Kurt l'admet, n'ont pas par elles-mêmes. D'autres prédicats, demandera-t-on, ne jouissent-ils pas de cette propriété remarquable ? Il est clair que si l'application d'un prédicat P à un ensemble d'énoncés doit avoir pour conséquence logique⁹ ces énoncés eux-mêmes, il faut que ce prédicat satisfasse le principe de décitation suivant : *si P (« p ») alors p* . Le prédicat logique de vérité est simplement le plus faible possible satisfaisant ce principe, puisqu'il satisfait aussi la réciproque : *si p alors « p » est vrai*. Autrement dit, si un autre prédicat, P' , satisfait le schéma de décitation alors, si P' s'applique à un énoncé, cet énoncé est également vrai. C'est par exemple le cas de prédicats comme « être connu » (la connaissance est la croyance vraie justifiée) ou « être prouvable » (qui contient également pour ainsi dire comme ingrédient le

⁷ Cet arrimage à notre langage des attributions de propriétés sémantiques est le sens de la remarque de Quine selon laquelle la vérité est une notion « immanente » à notre langage. Voir [QUI 61, §6]. Il est vrai que Quine semble parfois prêt à aller plus loin encore et à admettre une forme d'indétermination des attributions de référence à l'intérieur même de notre langage, renvoyant ainsi notre traduction de l'expression « lapin » de notre langage pris comme objet d'investigation, par le mot « lapin » dans le langage que nous employons pour parler de notre langage, à un simple choix parmi d'autres possibles (voir [QUI 69]).

⁸ Sous cette lecture, « La dernière phrase prononcée par Platon est vraie » est l'expression adéquate de la généralisation des instances du schéma « si la dernière phrase prononcée par Platon est « p », alors p » (on laisse ici de côté la difficulté que fait surgir le fait que Platon s'exprimait en grec).

⁹ Modulo l'identification des énoncés figurant dans l'ensemble, nous reviendrons sur ce point.

prédicat *logique* de vérité)¹⁰. En un sens fort, il n'y a donc que le prédicat logique de vérité qui puisse accomplir la fonction logique visée par Quine.¹¹

Les contempteurs du déflationnisme ne nient pas que le prédicat de vérité contribue à l'expressivité du langage ni qu'il permette d'accomplir la fonction logique dont il vient d'être question. Mais ils insistent sur le fait que la 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 objectivable 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 (y compris « l'effet de correspondance »), leurs opposants que la vérité est la description d'une relation objectivable du langage au monde (à nous d'en rendre raison) et que c'est l'usage logique qui est dérivé. Le tableau que nous venons de brosser à grands traits d'une *conception logique de la vérité*, est non seulement celle de Quine mais également, à quelques nuances près, celle de Paul Horwich ou encore de Hartry Field. Cette conception affirme que l'usage du prédicat de vérité est légitime, que la notion de vérité est une notion logique, et que la vérité elle-même n'est pas une propriété naturalisable ; elle n'implique pas le relativisme pas plus qu'elle n'implique par elle-même de renoncer à des lois logiques telles que le tiers exclu ou le principe de non contradiction.¹² Que faut-il retenir de ce tableau philosophique de la vérité ? Le prédicat de vérité ne serait-il qu'un auxiliaire utile à l'occasion mais qu'un peu de continence dans nos ambitions expressives permettrait d'oublier tout à fait ? A quel point les usages logiques de la vérité sont-ils philosophiquement significatifs ? Je voudrais montrer à présent que l'on se tromperait en sous-estimant leur portée. Je commence dans la prochaine section par faire un crochet par la philosophie de la logique, et cela pour deux raisons. D'abord pour lever une hypothèque qui pèse sur la conception déflationniste de la vérité sous le nom d'argument de la conservativité. Ensuite parce que les preuves dites « sémantiques », qui sont au cœur de cet argument, sont en elle-même instructives pour comprendre la portée des usages du concept

¹⁰ « Prouvable » au sens ordinaire semble en effet satisfaire la loi « Si 'p' est prouvable alors p ». En ce sens de « prouvable », un énoncé faux ne peut pas être prouvable. Il faut distinguer ce prédicat du prédicat de prouvabilité, parfois utilisé par les logiciens, signifiant « formellement dérivable dans X » et qui est toujours relatif à un certain système X d'axiomes et de règles d'inférences - Gödel a montré que si X est cohérent et contient un nombre fini de règles et d'axiomes alors il n'est pas possible d'affirmer *dans X* « Si p est prouvable dans X alors p » pour tout énoncé p. Il faut distinguer également ce prédicat d'un prédicat plus faible de justification qui *ne* satisfait pas la loi « Si p est justifié alors p ». Savoir qu'un énoncé est justifié en ce dernier sens ne suffit pas à affirmer l'énoncé en question.

¹¹ Cela ne signifie pas que les équivalences-T suffisent à *définir* un prédicat de vérité extensionnellement correct. Ce qu'il faudra mobiliser à cette fin dépendra de ce que l'on attend d'une définition ainsi que de considérations annexes (en particulier l'arrière-plan logique dans lequel nous souhaitons conduire la définition de ce prédicat.) Si l'on demande une définition axiomatique et que l'on se place dans une logique qui satisfait le théorème de compacité, on admet généralement que la bonne théorie de cette notion « logique » de vérité est une théorie récursivement axiomatisée qui « généralise », pour parler comme Tarski, les équivalences-T. Pour les précisions techniques, on pourra se reporter à [ARA à paraître].

¹² Je partage l'avis de Dummett [DUM 91] et de Field [FIE 01] selon qui l'expression adéquate de l'indétermination factuelle, ou de l'anti-réalisme relatif à un domaine du discours, est à chercher dans la réforme de la logique elle-même. Il est vrai que l'on mobilise généralement la notion de vérité pour exprimer ces lois, comme lorsque l'on dit « Tout énoncé est soit vrai soit faux », ou bien « Il existe des énoncés ni vrais ni faux ». Mais il est cohérent de se représenter l'usage de la notion de vérité dans la formulation de ces lois comme dérivé de l'emploi « logique » présenté par Quine. La question est alors celle de savoir quel genre de justifications sont pertinentes et véritablement éclairantes à l'appui ou à la charge d'une loi logique.

logique de vérité.¹³ Dans la dernière section, je dégage des enjeux épistémologiques plus généraux de cet usage logique pour ce que j'appellerai l'*attitude réflexive*.

III. Du problème de la conservativité à la nécessité de la réflexion.

Le déflationniste, on l'a dit, fait fond sur le principe d'équivalence entre la proposition p et la proposition qu'il est vrai que p , ou dans la version quiniennienne entre l'énoncé « ' p ' est vrai » et l'énoncé p lui-même. Ce principe semble pourtant se heurter à l'existence de ce que les logiciens appellent parfois les preuves sémantiques, preuves que Tarski, dans sa célèbre monographie sur la vérité [TAR 35], mettait d'ailleurs en avant pour illustrer la fécondité de sa définition de la vérité. Selon Stewart Shapiro et Jeffrey Ketland¹⁴ l'existence de ces preuves constitue en fait une réfutation de la conception déflationniste de la vérité. Leur argument est le suivant:

1. (Conservativité) Si le déflationnisme est correct, on ne doit pouvoir déduire de la vérité d'une théorie aucun énoncé formulable dans le vocabulaire de cette théorie¹⁵ qu'on ne puisse déjà déduire de la théorie elle-même.
2. Il existe une théorie telle que, de l'affirmation de sa vérité on peut déduire un certain énoncé formulé dans son vocabulaire et qui n'est pas déductible de la théorie elle-même¹⁶ (fait logique)
3. Donc le déflationnisme est faux.

Commençons par clarifier la deuxième prémisse. Le fait logique allégué contre le déflationniste est fondé sur les célèbres phénomènes d'incomplétude. Une conséquence bien connue du second théorème d'incomplétude de Gödel est en effet que la théorie arithmétique de Peano en premier ordre (PA) ne permet pas de prouver un certain énoncé, formulable dans le vocabulaire arithmétique, qui exprime la cohérence de la théorie. Or cet énoncé de la cohérence de PA *déviante dérivable* sous l'hypothèse supplémentaire de la vérité de PA, c'est-à-dire sous l'hypothèse que tous les théorèmes de PA sont vrais. L'affirmation de la vérité de PA permet donc de dériver un énoncé arithmétique formulable dans le langage de PA que PA elle-même ne permet pas de prouver.¹⁷

¹³ Je me suis efforcé dans la section suivante comme dans l'ensemble du texte de limiter le jargon technique de la logique et d'être juste *en substance*, quitte à sacrifier la correction de la lettre et des notations.

¹⁴ Shapiro [SHA 98], Ketland [KET 99]. Sur les aspects logiques du débat et pour bibliographie on pourra se reporter à Cieslinski [CIE 17]. Ce débat sur la portée des arguments de conservativité contre le déflationnisme fait largement écho à un débat antérieur portant sur les relations entre les théorèmes d'incomplétude et les programmes nominalistes en mathématiques.

¹⁵ On suppose tacitement que la théorie dont il est question ne touche pas, de près ou de loin, à des questions sémantiques et ne contient donc pas le prédicat de vérité.

¹⁶ Si toutefois la théorie est cohérente. Le fait logique sous-jacent forme le contenu même des célèbres théorèmes d'incomplétude de Gödel parus en 1931. Les détails logiques sont laissés de côté. Ils sont classiques et exposés par exemple dans Tarski [TAR 35]. Pour une introduction précise avec des notations logiques accessibles, on pourra consulter [ARA à paraître].

¹⁷ La cohérence d'une théorie formalisée est en effet essentiellement un fait arithmétique (touchant l'existence ou non de certaines suites finies de signes), si bien que l'on peut interpréter les énoncés de l'arithmétique comme portant sur la structure d'une théorie, et pourquoi pas, donc, une théorie de l'arithmétique. Pour une introduction techniquement accessible à la façon dont ce codage peut être opéré, voir à nouveau [ARA à paraître]. Notons que rien d'essentiel ne se joue dans ce codage. Nous aurions tout aussi bien pu éviter ici le détour par le codage en choisissant en lieu et place de PA une théorie S syntaxique dont l'objet serait la description de la structure formelle de S . Cela ne change rien sur le fond et le choix de PA est devenu standard.

Voyons à présent la première prémisse de l'argument. Lorsque T est une théorie formulée dans un langage L et T' une extension de T formulée dans un langage L' , les logiciens disent que la théorie T' est conservative sur la théorie T si et seulement si tout énoncé formulé dans le langage L et prouvable dans T' est aussi prouvable dans T . La première prémisse de l'argument revient donc à exiger que l'affirmation de la vérité d'une théorie étende conservativement la théorie en question, exige que la seconde prémisse montre précisément ne pas être satisfaite dans le cas de PA. Mais pourquoi cette demande de conservativité faite au déflationniste ? C'est que, de prime abord, la conservativité semble n'être qu'une version *affaiblie* du principe d'équivalence : ce dernier n'implique-t-il pas qu'on ne peut déduire de l'hypothèse de la vérité davantage qu'il n'est possible de déduire de ce à quoi est attribuée la vérité ? L'enjeu de la question tient à ceci que, pour éviter une conclusion désastreuse pour le déflationniste il n'y a qu'une issue : récuser l'exigence de conservativité.

Pour récuser la demande de conservativité, il faut d'abord rappeler que le principe d'équivalence a toujours été entendu comme s'appliquant aux propositions individuelles ou aux énoncés individuels donnés explicitement entre guillemets. En effet, dans le cas où l'on applique le prédicat de vérité à un ensemble *infini* d'énoncés, il n'a jamais pu être question de soutenir l'équivalence logique entre l'ensemble infini des énoncés et l'énoncé de leur vérité. C'est *au contraire* parce qu'il n'est pas possible d'affirmer l'ensemble infini des énoncés, que Quine soutient qu'il faut avoir recours au prédicat de vérité : non pour affirmer alors un énoncé qui leur serait logiquement équivalent, mais bien pour affirmer, dans un métalangage, une généralisation qui, moyennant l'identification des énoncés visés, les impliquera tous grâce au principe de décitation (appliqué dans son domaine de validité, c'est-à-dire à des énoncés nommés entre guillemets).

Ce point étant rétabli, revenons alors à la dérivation de la cohérence de PA, celle-là même qui met échec la demande de conservativité formulée par Shapiro et Ketland, et voyons si elle menace la position déflationniste. Le point de départ est le suivant :

(Principe de réflexion global [Ref_G]) Tous les théorèmes de la théorie PA sont vrais

Ce principe implique toutes les instances du schéma suivant :

(Schéma de réflexion local [Ref_L]) Si « p » est un théorème de PA alors p

Il implique donc aussi l'énoncé suivant qui en est une instance particulière :

Si « $0=1$ » est un théorème de PA alors $0=1$

Lui-même est équivalent à sa contraposée :

(C) Si $0 \neq 1$ alors « $0=1$ » n'est pas un théorème de PA

Mais dans PA on peut prouver que $0 \neq 1$. Donc dans PA augmenté du schéma de *réflexion local*, on peut dériver que « $0=1$ » n'est pas un théorème de PA (par *modus ponens* à partir de (C) et de la preuve que $0 \neq 1$), c'est-à-dire que PA est cohérente.

Cette preuve de cohérence fait-elle problème pour le déflationnisme ? On peut distinguer dans le déroulé précédent *deux* effets logiques de l'affirmation de la vérité de PA. Le premier, en remontant de la conclusion aux prémisses, est en fait imputable aux seules instances du schéma de réflexion local¹⁸ : nous avons vu que l'on pouvait dériver de ces instances la cohérence de PA. Mais la vérité ne joue aucun rôle à ce niveau puisqu'elle ne figure pas dans les instances du schéma et que l'on dérive la cohérence de PA à partir d'un principe logiquement moins fort que l'affirmation de la vérité de PA. Le prédicat de vérité ne figurant pas dans les instances du schéma de réflexion local, l'échec de la conservativité sur PA des instances de ce schéma ne contredit pas la conception logique de la vérité. Le second effet est « l'effet de généralité », en vertu duquel Ref_G est lui-même logiquement plus fort que l'ensemble infini des instances du schéma Ref_i . Si l'effet de généralité devait aggraver l'échec de la conservativité en permettant de dériver d'autres énoncés arithmétiques que ceux qu'il est possible de dériver des seules instances de Ref_i , ce surplus logique, quoique intimement lié, lui, aux conditions du déploiement du prédicat de vérité, n'en resterait pas moins philosophiquement bénin : nous avons déjà relevé que cet effet de généralité n'était pas en conflit avec les thèses déflationnistes, en étant même un ingrédient essentiel. Il semble donc au bout du compte que l'échec de la conservativité n'ait pas grand-chose à voir avec un échec du déflationnisme.¹⁹ Elle permet en revanche d'entrevoir quelques réponses à la question posée au départ de notre enquête : quelle différence la vérité fait-elle en pratique ?

Le phénomène logique mis en avant par Shapiro et Ketland n'en demeure en effet pas moins instructif sur les usages de la notion de vérité. Nous avons dit que notre finitude nous interdisait l'expression directe d'un ensemble infini d'énoncés ou d'énoncés dont nous ignorons le contenu et, pour cette raison, nous faisait parfois obligation d'affirmer, non pas directement l'ensemble des énoncés exprimant le contenu auquel nous souhaitons nous lier logiquement, mais la vérité de ces énoncés identifiés sous une certaine description. L'échec de la conservativité nous rappelle que la médiation de ces descriptions auxquelles nous avons recours pour identifier les énoncés n'est pas logiquement triviale, et que la situation logique et épistémologique qui est la nôtre lorsque le prédicat de vérité est logiquement inéliminable de nos assertions est fondamentalement distincte de celle dans laquelle nous attribuons la vérité à des énoncés de notre propre langage désignés de façon transparente entre guillemets. C'est

¹⁸ Il est vrai que ces énoncés sont équivalents aux instances du schéma: si « p » est un théorème de PA alors « p » est vrai. Cette équivalence n'entre pas en contradiction avec la conception logique de la vérité qui, au contraire, l'implique.

¹⁹ Ce type d'argumentation est déjà utilisé dans [TEN 02]. Pour qu'elle soit tout à fait convaincante, il faudrait montrer en outre que le schéma de réflexion global lui-même n'est pas dérivable dans PA de ce qui pourrait se présenter de façon crédible comme une pure définition du prédicat de vérité pour le langage de PA. Si tel était le cas, la réponse du berger à la bergère serait toute trouvée : notre justification des principes de réflexion serait en fait elle-même un effet de notre saisie de la nature de la vérité. Mais tel n'est pas le cas. Le principe de réflexion global et ses instances ne deviennent dérivables en utilisant les axiomes récursifs définissant le prédicat de vérité dans PA qu'à condition de *renforcer simultanément les axiomes de PA*, renforcement permis par les possibilités expressives nouvelles offertes par l'usage du prédicat de vérité défini. Sur ces questions voir [FIE 99], et [GAL 2015] pour un prolongement.

dans la possibilité de faire travailler ce contraste à notre avantage que réside justement notre intérêt pratique pour la notion logique de vérité.

Ce n'est donc pas tant l'échec de la conservativité qui importe au déflationniste que, d'abord, l'observation duale que la formulation de nos engagements logiques vis-à-vis d'un ensemble infini d'énoncés *ne peut* parfois se faire *qu'*au moyen d'une affirmation qui étend non conservativement l'ensemble des énoncés visés. Celui qui ne connaît l'ensemble des théorèmes de PA que sous cette description (« Les théorèmes de PA ») et souhaite formuler un énoncé qui les implique tous, n'a d'autre choix que d'affirmer en ces mots que tous les théorèmes de PA sont vrais. De façon apparemment paradoxale, il peut donc toujours dériver la cohérence de PA à partir du contenu qu'il a nécessairement formulé, alors même que la cohérence de PA n'est pas dérivable de PA elle-même²⁰. « Tous les théorèmes de la théorie T sont vrais » n'est pas une conséquence logique de T, mais l'accepter est parfois une condition de possibilité de l'engagement logique de notre pensée avec T.

Ce que ce paradoxe révèle de l'usage du prédicat de vérité apparaît plus clairement encore à l'analyse de la situation d'un sujet qui, tout en n'étant pas en position de décrire un certain ensemble infini d'énoncés, endosserait néanmoins tous ces énoncés de façon pour ainsi dire immanente, en ceci qu'il se trouverait *suivre une règle* dont il ne connaît pas la formule mais qui engendre l'ensemble des énoncés en question. C'est la situation d'un sujet qui aurait asserté directement les *axiomes* d'une théorie et dont la *pratique* inférentielle serait conforme aux règles logiques. Dans pareil cas on peut dire que le sujet accepte la théorie, c'est-à-dire l'ensemble infini de ses énoncés, de façon *dispositionnelle*. Le principe de réflexion globale a alors pour un tel sujet le sens d'une *explicitation* des engagements implicites dans sa pratique inférentielle. La formulation de ce principe est en fait le produit exact de la double exigence de *connaissance réflexive* et de *solidarité de notre intelligence vis-à-vis de notre pratique* : quand bien même il est logiquement cohérent d'accepter dispositionnellement une théorie tout en acceptant qu'elle n'est pas cohérente (c'est logiquement possible puisque la négation de la cohérence de T est logiquement compatible avec T), le faire n'est pas « réflexivement cohérent ». ²¹ La notion logique de vérité, *via* la formulation du principe de réflexion, permet de donner un sens précis à cette notion de cohérence réflexive d'un ensemble d'énoncés ou de croyances comme cohérence logique de ces énoncés ou de ces croyances avec les principes de réflexion que nous acceptons. ²²

²⁰ Si du moins PA est cohérente.

²¹ Cette perspective est proche de celle tacitement adoptée par des logiciens tels que Solomon Feferman qui étudient les aspects logiques de ces extensions. Voir [FEF 62] et [FEF 99]. Pour Feferman, le sujet qui accepte les axiomes de PA et suit la pratique inférentielle classique accepte déjà PA - nous préférons dire *dispositionnellement* -, est rationnellement contraint d'accepter la cohérence de PA, quoiqu'il n'y soit pas logiquement contraint. Il faudrait nuancer ce point en rappelant, dans un esprit Kripkeo-Wittgensteinien, que notre pratique inférentielle passée est toujours finie et qu'elle peut donc être décrite diversement comme obéissant à des règles générales incompatibles entre elles. Dans le passage de ma pratique inférentielle à la formulation réflexive de la règle à laquelle obéit cette pratique, il y a donc un acte *d'interprétation* de ma propre pratique. En conversation, J.M. Salanskis parlait à ce sujet, de façon suggestive, d'une extension herméneutique de la connaissance.

²² Ce point montre que la cohérence logique *stricto sensu* n'est pas le seul critère « interne » de vérité, contrairement à ce que laissent parfois entendre les pragmatistes (cf. [ROR 91 p. 281]). La réflexivité est porteuse d'une norme propre. On aperçoit que cette dernière n'est pas sans rapport, je ne peux que le suggérer ici, avec ce qu'on appelle parfois *la norme du vrai*. Sur la question de savoir quels actes sont soumis à une norme de vérité, on pourra consulter [ENG 89]. [PRO 2018] offre un éclairage complémentaire en présentant des recherches sur la variabilité culturelle des normes métacognitives.

Le déflationniste peut enfin tirer de l'argument de la conservativité une autre leçon sur la portée de l'usage logique de la vérité. Cet enseignement apparaît plus clairement en prenant cette fois du champ par rapport au cas particulier de l'attribution de vérité de PA (ou à des théories incluant PA). Dans le cas général, l'attribution de la vérité à un ensemble d'énoncés se fait au moyen d'une description de cet ensemble dans un vocabulaire, c'est-à-dire un registre conceptuel, qui n'a rien de commun avec celui mis en œuvre par ces énoncés eux-mêmes, si bien que l'affirmation de la vérité des énoncés tombant sous la description en question étend en général conservativement l'ensemble d'énoncés que cette dernière décrit: de la vérité de la systématique phylogénétique (la théorie de la classification des espèces) on ne peut pas inférer l'existence d'un taxon élémentaire qui n'aurait pu être inférée de la théorie elle-même.²³ Dans ce cas, la fonction logique du prédicat de vérité peut être mise au travail, non plus seulement en vue d'une opération d'explicitation de nos engagements mais, par la mise en balance d'engagements dans des registres conceptuels différents, en vue de ce que l'on pourrait qualifier d'opération de *pivot* intellectuel et que nous allons tout de suite préciser. Nous avons dit en effet que, grâce au principe de décitation (appliqué dans son domaine de validité, c'est-à-dire à des énoncés nommés entre guillemets), l'attribution du prédicat de vérité à la description d'un ensemble d'énoncés impliquait ces énoncés eux-mêmes *modulo les moyens d'identifier les énoncés visés par la description*. Ainsi de ce que la dernière phrase prononcée par Platon est vraie on ne peut pas déduire que l'automne est doux à Athènes, à moins d'être en situation d'établir que la dernière phrase prononcée par Platon est « L'automne est doux à Athènes ». ²⁴ Dans l'exemple que nous venons de donner, les moyens requis pour l'identification de l'énoncé auquel est attribuée la vérité sont les principes et hypothèses que mobiliserait une enquête historique permettant d'établir ce que furent les derniers mots de Platon. Ces moyens, on l'accordera, n'ont par eux-mêmes aucun lien direct avec la connaissance du climat athénien, ni les recherches historiques sur la parole de Platon avec les recherches météorologiques. Pourtant, *lorsque* nous sommes en mesure d'identifier les énoncés visés par la description, l'affirmation que la dernière phrase prononcée par Platon est vraie, si elle pouvait être justifiée, ferait peser logiquement, et pour ainsi dire de façon oblique, les premières sur les secondes, l'histoire sur la climatologie. Les enjeux épistémologiques d'une telle opération sont, on le devine, dignes qu'on y insiste, et c'est ce que nous allons faire dans la dernière section.

IV. De la réflexion à l'autonomie.

Le discours déflationniste a tendu à sous-estimer la portée épistémologique de l'emploi du prédicat logique de vérité, et cela pour deux raisons. La première est qu'il lui a fallu d'abord et surtout installer dans le débat philosophique ses thèses négatives sur la *nature* de la vérité. La seconde tient à l'inscription du déflationnisme dans un champ philosophique, hérité de l'empirisme logique, dont la préoccupation dominante était la question du fondement des

²³ Cette affirmation est à prendre avec précaution et demanderait des précisions et qualifications que je ne peux faire ici. Notons que si la vérité de T permet de dériver la cohérence de T et que la cohérence de T est un fait arithmétique, il faut alors, pour échapper à l'échec de la conservativité, que T ne contienne pas d'arithmétique – une clause que l'on jugera rarement satisfaite si l'on insiste sur le fait que les théories scientifiques modernes doivent contenir les hypothèses permettant de formuler des lois, des prédictions et des explications *quantitatives*...

²⁴ Je laisse de côté le fait que Platon ne s'exprimait pas en Français. Il faudrait parler de la traduction française de la dernière phrase prononcée par Platon.

sciences naturelles dans l'expérience. Dans cette perspective, il importe d'abord au déflationniste de défendre l'idée que la vérité n'est pas une propriété empirique et qu'elle ne joue aucun rôle causal-explicatif dans une description naturalisée du monde. L'usage logique assigné au prédicat de vérité ressemble alors fort à celui d'un second rôle. La portée de ce rôle n'apparaît véritablement que lorsque l'on délaisse la tâche de la justification logico-empirique des théories scientifiques pour s'atteler à l'explication de la façon dont le sujet de la science lui-même s'acquitte de cette tâche, c'est-à-dire lorsque l'on étudie le rôle de *l'attitude réflexive* dans l'élaboration de la connaissance. Ce que j'appelle l'attitude réflexive est l'attitude qui est la nôtre lorsque nous formons des pensées à propos d'autres pensées. Ces pensées à propos desquelles nous pensons dans l'attitude réflexive peuvent être nôtres, comme lorsque je pense que la conclusion à laquelle je suis parvenu est étonnante, mais je range également sous cette attitude les pensées que nous formons à propos des pensées d'autrui, comme lorsque je me demande si ce qu'affirme Quine à propos de l'indétermination de la traduction radicale est justifié ou que je note que Pierre a annoncé de la pluie pour demain. Ce qui est caractéristique de l'attitude réflexive est qu'elle engage la pensée sur deux niveaux de discours simultanément : un discours sur les choses, et un discours sur le discours sur les choses. L'attitude réflexive ainsi définie est donc une condition de possibilité à la fois de l'examen critique de ses propres pensées, et de l'apprentissage fondé sur le témoignage d'autrui.

Nous avons déjà eu l'occasion de suggérer dans la section précédente que le prédicat logique de vérité permettait, tel une ancre jetée du métalangage dans le langage objet, une forme de coordination des niveaux de discours simultanément engagés dans l'attitude réflexive.²⁵ Nous allons à présent préciser l'idée que cet usage logique du prédicat de vérité, et son inéliminabilité même, sont intimement liés à la possibilité de poursuivre les fins qui sont celles de la réflexion, tant dans l'examen critique de ses propres pensées que de celles d'autrui. Pour illustrer ce point je me concentrerai dans ce qui suit sur le cas de l'examen des pensées d'autrui et je laisserai à une autre occasion la discussion de la réflexion en première personne.

De ce que *Pierre affirme que le facteur est passé ce matin* il ne suit pas que *le facteur est passé ce matin*. Pour que je puisse apprendre de Pierre, c'est-à-dire atteindre la conclusion que le facteur est passé ce matin, il faut un principe supplémentaire à savoir : *Si Pierre affirme que le facteur est passé ce matin, alors le facteur est passé ce matin*. Or d'où tirer un tel principe ? Il n'y a que deux possibilités. On pourrait justifier ce conditionnel si, ayant entendu Pierre, nous savions en outre que le facteur est passé ce matin. Mais évidemment, si nous sommes en état de produire cette justification, nous n'avons pas besoin que Pierre nous apprenne que le facteur est passé ce matin. Pour expliquer la possibilité de l'apprentissage, c'est donc la seconde voie qui s'impose : il faut que nous tirions ce conditionnel d'autre part. Or d'où tirons-nous un tel principe ? Nous le tirons ordinairement d'un principe plus général, de quelque chose comme une théorie touchant la fiabilité de Pierre (une théorie fort modeste, qui peut n'établir sa fiabilité que pour ce qui concerne les envois postaux de ce matin !). Mais l'énoncé de la fiabilité de Pierre, afin que je puisse l'appliquer utilement, doit être une affirmation générale, au sens où

²⁵ Je fais allusion plus particulièrement à ce que nous avons dit à de la formulation du principe de réflexion, à savoir qu'elle résultait d'une exigence de solidarité intellectuelle vis-à-vis de notre pratique, ainsi qu'à l'opération de « pivot intellectuel » que nous avons évoquée peu après - un pivot suppose un point d'appui.

elle doit porter sur tous les énoncés d'une certaine gamme ouverte sous la description de laquelle se trouve tomber 'le facteur est passé ce matin'. Du fait de cette généralité, l'interrogation sur la fiabilité rend le recours à la vérité inéliminable : elle porte sur le point de savoir si *ce que Pierre affirme dans tel ou tel genre de circonstances est vrai*.

Ce petit exemple fait écho à ce que nous avons noté dans section 3 à propos de la façon dont l'emploi du prédicat de vérité faisait peser de façon oblique des considérations historiques touchant la parole de Platon sur des considérations touchant le climat athénien. On entrevoit maintenant les ressources épistémologiques ouverte par l'emploi d'un prédicat de vérité qui permet de faire peser logiquement sur mes pensées de premier ordre (concernant l'alimentation de la boîte-aux-lettres) le contenu enchâssé dans mes pensées de second ordre de type « Pierre croit que le facteur est passé ». La notion de vérité est logiquement nécessaire pour le travail intellectuel que nous menons dans l'attitude réflexive, non pas lorsque nous voulons nous interroger *directement* le contenu de ce qui est dit mais, au contraire, lorsque souhaitant faire l'économie de cet examen et cherchant à *apprendre* d'une source, nous nous interrogeons sur sa crédibilité.

Il est vrai que si Pierre ne m'avait pas annoncé que le facteur était passé, j'aurais tout simplement pu aller le vérifier par moi-même. La possibilité qui est la mienne de m'interroger sur la fiabilité de Pierre en vue de m'appuyer sur sa parole est donc certes utile, éventuellement indispensable en pratique, mais elle n'en demeure pas moins évitable en principe. Je voudrais pour conclure franchir encore une étape et montrer que l'existence de certaines situations, que l'on appelle parfois de dépendance épistémique, rendent ce genre de détour plus profondément indispensable encore.

Qu'est-ce en effet que la dépendance épistémique ? Disons qu'un sujet possède une méta-justification de sa croyance que p s'il possède une justification qu'il existe une justification de p sans posséder cette dernière. Une justification directe de p est au contraire une justification qui ne dépend pas logiquement de l'existence d'une *autre* justification de la *même* proposition. Lorsque Andrew Wiles justifie par une preuve mathématique le dernier Théorème de Fermat, sa justification est directe : il peut expliquer pourquoi l'équation $x^n + y^n = z^n$ n'a pas de solution entière pour $n > 2$, en rendre raison selon des lois ou des normes proprement mathématiques. Je n'en ai pour ma part qu'une méta-justification : il est écrit dans *Le Monde* que Wiles a démontré le Théorème de Fermat (observation), ce qui est écrit dans *Le Monde est vrai* (induction fondée sur des éléments variés), donc il existe une démonstration du théorème de Fermat et par conséquent l'équation $x^n + y^n = z^n$ n'a pas de solution entière pour $n > 2$. La meilleure justification que je suis capable de produire du dernier théorème de Fermat est tout à fait bonne, mais elle suppose l'existence d'une *autre* justification *de ce même* théorème comme condition logique de sa possibilité. La méta-justification que je possède, à la différence de la justification directe, ne me permet pas de comprendre pourquoi l'équation $x^n + y^n = z^n$ n'a pas de solution entière pour $n > 2$. Mais c'est aussi ce qui fait son prix. C'est parce qu'elle procède selon un détour qui ne mobilise aucune connaissance mathématique que cette justification peut être mienne : pour établir que l'équation $x^n + y^n = z^n$ n'a pas de solution entière pour $n > 2$, il n'est sans doute pas nécessaire d'avoir recours à la notion de vérité, mais *pour moi*, le détour par des

considérations sur la parole de Wiles ou sur la parole de ceux qui parlent des résultats de Wiles, *est* essentiel. Dans l'exemple précédent, je pouvais toujours me passer de Pierre pour aller voir directement ma boîte-aux-lettres. Mais je ne peux pas me passer d'Andrew Wiles pour justifier ma croyance que l'équation $x^n + y^n = z^n$ n'a pas de solution entière pour $n > 2$. Et pour faire peser sur mes croyances mathématiques mes considérations non mathématiques concernant Wiles, pour redescendre, pour ainsi dire, de mes considérations *sur* le discours que tiennent mes sources, aux mathématiques elles-mêmes, le prédicat de vérité, avec sa modeste fonction « logique » ou « décitationnelle » est nécessaire.²⁶ La possession d'un prédicat de vérité est donc une condition nécessaire de ce petit miracle épistémique par lequel nous parvenons à faire travailler pour notre compte les justifications chèrement acquises par d'autres, sans même les comprendre. Les avantages de l'entraide, en somme, à la fois sur le vol et le « travail honnête ».²⁷

C'est exactement de ce petit miracle que nous avons besoin pour faire vivre l'idéal d'autonomie de la raison dans un monde de dépendance épistémique radicale. Comme le notait déjà John Hardwig [HAR 90] il n'est plus rare que dans l'activité scientifique même, *aucun* des auteurs d'un travail ne possède de justification directe des résultats obtenus en collaboration. L'un a établi A, l'autre que A implique B, ils établissent ainsi ensemble que B sans que personne ne puisse seul rendre compte de B intégralement selon l'ordre des raisons. Chacun des coauteurs doit donc, pour justifier ses conclusions -ne serait-ce que devant lui-même-, mobiliser des justifications à l'effet que les résultats intermédiaires qui lui sont communiqués par les uns et les autres sont vrais. Et ce qui est vrai dans les sciences est peut-être plus vrai encore de notre condition épistémique ordinaire. Je dépends épistémiquement d'autrui de mille façons pour la justification d'un certain nombre de mes croyances les mieux ancrées ou les plus centrales pour l'orientation de mon action, et je dépends en particulier d'experts vis-à-vis desquels je suis dans une position d'asymétrie épistémique radicale, incapable de juger par moi-même des matières dont ils traitent. Il est vrai que la modernité philosophique était porteuse d'un double idéal épistémique d'autonomie *et* d'indépendance, - que l'on songe au sujet héroïque des *Méditations métaphysiques* et à son isolement inaugural -, mais il faut ici les distinguer.²⁸ Car le véritable sens de l'injonction à juger selon sa propre raison est d'abord celui d'une invitation à ne pas se le tenir pour dit, c'est-à-dire à ne pas former ses croyances sans examen rationnel. Or privé des moyens intellectuels de m'interroger sur les considérations géophysiques qui doivent établir la réalité du réchauffement climatique, je peux néanmoins m'interroger sur la vérité du discours des climatologues. Ce n'est pas se soustraire à l'injonction des lumières que de former des croyances sur la base de méta-justifications : ces justifications sont bien le produit d'une enquête rationnelle qui doit me permettre de retenir le témoignage de certains interlocuteurs et d'en disqualifier d'autres. L'idéal de la possession de justifications de toutes nos croyances n'est pas celui de la possession de justifications directes de toutes nos croyances. Le second est un idéal d'indépendance, celui d'un esprit qui serait capable d'embrasser en expert tous les

²⁶ Ou une notion de justification qui contient logiquement la notion de vérité, c'est-à-dire qui satisfait le principe de décitation. Voir la note 8 ci-dessus.

²⁷ Dans son *Introduction to Mathematical Philosophy*, Russell ironisait sur les avantages de la méthode de définitions par « postulats » (les définitions axiomatiques), avantages qu'il comparait à ceux « du vol sur le travail honnête ».

²⁸ Où l'on retrouve l'écho en épistémologie de considérations métaphysiques et politiques développées ailleurs par d'autres. Voir en particulier Alain Renaut [REN 89].

domaines de la connaissance. Il est communément admis que cet idéal doit être abandonné. Le premier, en revanche, est un idéal de non-sujétion et, pour le faire vivre, c'est ce que nous voulions montrer, il est *logiquement* nécessaire de faire le détour par la notion de vérité.

Il ne suffit pas, bien entendu, que je possède la notion de vérité pour être épistémiquement autonome. Encore faut-il que soient de surcroît réunies les conditions me permettant de justifier ma croyance dans la véracité de ce qui est rapporté dans *Le Monde* ou par les experts du GIEC, et non réduit à en faire le pari ou à en caresser l'espoir. Nous recourons pour cela à un ensemble de méta-données émises en permanence par la structure sociale : qualité de la formation des experts, réputation et reconnaissance institutionnelle, indices variés sur les vertus et les incitations de l'individu (la « transparence »), sur le contexte politique et social de l'intervention, etc. A y regarder de plus près, il apparaît alors que notre autonomie épistémique est conditionnée par la nature de certains arrangements sociaux comme l'existence d'une presse libre et plurielle, d'un système de formation ouvert, la publicité des lois ou la définition sincère de la raison sociale des institutions, et d'autres choses encore, soit ce que l'on pourrait appeler plus généralement, en s'inspirant d'une notion politique due à John Rawls, une *société épistémiquement bien ordonnée*²⁹. Dans un monde où ces conditions ne sont pas réunies les justifications épistémologiques que possèdent les citoyens pour leurs croyances sont affaiblies, et leur autonomie rationnelle réduite d'autant : le rapport à l'expert, alors, mais alors seulement, tend à se confondre avec le rapport au gourou. Lorsqu'à l'inverse elles sont réunies, le journal *Le Monde* et la parole du climatologue ne sont pas des obstacles mais des outils de la conquête de mon autonomie dans un monde complexe où la connaissance universelle est devenue impossible et la dépendance épistémique inévitable.³⁰ Ainsi apparaît-il que non seulement la possession d'un concept de vérité est nécessaire pour surmonter la dépendance épistémique mais que, plus profondément, l'idéal d'autonomie est solidaire de ce que l'on peut bien appeler un idéal de vérité. Et le déflationniste n'a rien à objecter à un tel constat ni à une telle formulation.³¹

Conclusion : après la déflation

La conception logique de la vérité tient que tous les usages du prédicat de vérité sont dérivés de sa fonction logique. Cette fonction logique est celle accomplie en vertu du fait que le prédicat de vérité satisfait le principe d'équivalence. Le fait qu'une preuve d'une proposition soit identiquement une preuve de la vérité de cette proposition et réciproquement (quand cette

²⁹ [RAW 71] parle de « société bien ordonnée ». Dans une société bien ordonnée par des principes de justice, les citoyens acceptent les mêmes principes de justice ; ceux-ci gouvernent la structure de base de la société ; les citoyens le savent, chacun sait que tous le savent, et chacun cultive le sens correspondant de la justice. Dans la transposition épistémologique que je suggère, les membres de la société épistémiquement bien ordonnée acceptent les mêmes normes de justification, la société réalise des principes d'organisation sociale épistémologiquement vertueux acceptés, ces principes d'organisation et de justification sont de connaissance commune entre les citoyens, et chacun cultive le sens correspondant de la justification. A échelle réduite, la communauté scientifique tend à être un modèle de société épistémique bien ordonnée.

³⁰ Cette dialectique de l'indépendance et de l'autonomie est également au cœur de la philosophie politique et notamment des enjeux de la définition moderne de la liberté démocratique. Pour une mise en perspective, voir à nouveau [REN 89].

³¹ Je n'ai pas la place ici de développer cette idée d'idéal de vérité. En fidélité à la ligne de pensée développée ici, on pourrait le formuler comme un idéal de réflexivité ou, si l'on veut prendre les choses dans l'autre sens, de « décitabilité » : dans une société épistémiquement bien ordonnée le sujet cherche à réaliser les conditions qui permettent à autrui de le « dé citer » (et non seulement de le citer !).

proposition est donnée), loin de vider la notion de vérité de tout usage, assoit au contraire ses usages fondamentaux pour notre vie épistémique : il nous permet de faire travailler pour nous, via des méta-preuves, des preuves produites par d'autres, ou de faire peser sur nos croyances, en les corrigeant ou en les renforçant, les conclusions de notre propre interrogation sur la fiabilité de leur processus formation. On accordera que si, comme nous le soutenons, l'emploi de la notion de vérité est nécessaire à (la reconstruction rationnelle de) ces processus, il est un élément central de notre vie rationnelle.

Le temps est donc peut-être venu pour le déflationniste d'infléchir son discours. La vérité n'appartient pas au langage de la science ; mais elle est centrale dans le langage de la raison. La vérité n'est pas naturalisable ; mais la naturalisation de l'esprit (notamment de l'esprit scientifique) doit faire une place à la propriété de « posséder le concept de vérité »³². Si enfin posséder la notion de vérité est une condition nécessaire de notre activité réflexive et de notre autonomie épistémique, ne devient-il pas urgent, sans renoncer aux acquis des débats antérieurs sur la nature de la vérité, d'approfondir l'étude de ses *fonctions* et pourquoi pas, après la déflation, d'insister désormais sur l'importance de la notion logique de vérité et la valeur de ce qu'elle nous permet d'accomplir en première personne et en société ?³³

Bibliographie

- [ARA à paraître] Arana A., Panza M., Poggiolesi F., Wagner P., (éd.), *Précis de philosophie des mathématiques et de la logique*, Editions de la Sorbonne, Paris, à paraître.
- [DUM 91] Dummett M., *The Logical Basis of Metaphysics*, Duckworth, Londres, 1991.
- [ENG 89] Engel P., *La norme du vrai*, Gallimard, Paris, 1989.
- [FEF 62] Feferman S., «Transfinite Recursive progression of Axiomatic Theories», *Journal of Symbolic Logic* 27 (3), pp. 259-316, 1962.
- [FEF 91] Feferman S., «Reflecting on Incompleteness», *Journal of Symbolic Logic* 56 (1), pp. 1-49, 1991.
- [FIE 71] Field H., Tarski's Theory of Truth», traduction française « La théorie de la vérité de Tarski » dans Bonnay et Cozic (éd.) *Philosophie de la logique*, Vrin, Paris, 2010
- [FIE 99] Field H., «Deflating the conservativeness argument», *Journal of Philosophy* 96 (10), pp. 493-521, 1999.
- [FIE 01] Field H., *Truth and the Absence of Facts*, Oxford University Press, New York, 2001.
- [FRE 71] Frege G., « Der Gedanke » traduction française « La pensée » in *Ecrits logiques et philosophiques*, Seuil, Paris, 1971.
- [GAL 15] Galinon H., « Deflationary Truth: conservativity or logicity? », *Philosophical Quarterly* 65 (259), pp. 268-274 , 2015.
- [HAR 90] Hardwig J., « Epistemic dependence », *Journal of Philosophy* 82 (7)pp. 335-349.
- [KET 99] Ketland J., « Deflationism and Tarski's Paradise », *Mind* 108 (429), pp. 69-94, 1999.
- [PRO 13] Proust J., *The Philosophy of Metacognition*, Oxford University Press, New York, 2013.

³² C'est déjà le cas dans les études de psychologie centrées sur la métacognition, voir [PRO 13].

³³ Je remercie chaleureusement François Rivenc et Sébastien Gandon qui ont bien voulu lire une version préparatoire de ce texte.

- [PRO 18] Proust J., Fortier M., (éd), *Metacognitive diversity*, Oxford University Press, New York, 2018.
- [QUI 61] Quine W. O., *Word and Object*, Harvard University Press, Cambridge:MA, 1961, traduction française *Le Mot et la chose*, Champ Flammarion, Paris, 2010.
- [QUI 69] Quine W. O., « Ontological relativity », traduction française dans *Relativité de l'ontologie et autres essais*, Aubier, Paris, 2008.
- [QUI 70] Quine W. O., *Philosophy of logic*, Harvard University Press, Cambridge :MA, 1970, traduction française *Philosophie de la logique*, Aubier, Paris, 2008.
- [QUI 92] Quine W. O., *The Pursuit of Truth*, Harvard University Press, Cambridge: MA, 1992, traduction française *La Poursuite de la Vérité*, Seuil, Paris, 1993.
- [RAM 27] Ramsey F., «Facts and propositions», *Proceedings of the Aristotelians society* 7, pp.153-170, 1927.
- [RAM 91] Ramsey F., «On Truth», *Episteme* 16, pp. 1-16, 1991.
- [RAW 71] Rawls J., *A Theory of Justice*, Harvard University Press, Cambridge : MA, 1971, traduction française, *Théorie de la Justice*, Seuil, Paris, 2009.
- [REN 89] Renaut A., *L'ère de l'individu*, Gallimard, Paris, 1989.
- [RIV 98] Rivenc F., «Ce que Ramsey a vraiment dit, ou la théorie prophrastique de la vérité », *Philosophie* 57, p.16-50, 1998.
- [ROR 95] Rorty R., « Is Truth a Goal of Enquiry ? Davidson vs. Wright », *The Philosophical Quarterly* 45 (180), pp. 281-300, 1995.
- [SHA 98] Shapiro S., « Proof and Truth : Through Thick and Thin. » *The Journal of Philosophy* 95 (10), pp. 493-521, 1998.
- [TAR 35] Tarski A., « Der Wahrheitsbegriff in den Formaliesierten Sprachen », 1935, traduction française « *Le concept de vérité dans les langages formalisés* » in *Logique, sémantique, métamathématique*, Paris, Armand Colin, Paris, 1972, reprise dans Bonnay et Cozic (éd.), *Philosophie de la logique*, Paris, Vrin, 2010.
- [TEN 02] Tennant N., « Deflationism and the Gödel-Phenomena », *Mind* 111, pp; 551-582, 2002.