

Denis Bonnay – Putnam est-il un traître ?¹

Le présent article examine la relecture de l'externalisme sémantique proposée par Putnam dans *Reason, Truth and History*², à travers la question des liens entre externalisme et réalisme. Dans quelle mesure l'externalisme sémantique participe-t-il d'une conception réaliste des rapports entre le langage et les choses ? Putnam doit-il revenir sur sa conception de la référence lorsqu'il met en place l'opposition entre réalisme métaphysique et réalisme interne ? Pour aborder ce problème, nous partirons dans la première section de l'accusation de trahison formulée par Devitt à l'encontre d'un Putnam qui aurait renoncé au réalisme. Les deux sections suivantes sont consacrées à une présentation du réalisme antérieur de Putnam, de la place qu'y tient l'externalisme sémantique, ainsi que du passage au réalisme interne. La quatrième section fait le point sur les reprises de l'externalisme dans *Reason, Truth and History*, qui sont examinées de plus près dans les sections suivantes, s'agissant de la critique de l'internalisme (section V), de la contribution de l'environnement (section VI), du projet d'une théorie physicaliste de la référence (section VII) et enfin de l'idée de chaînes causales menant à la référence (section VIII). La dernière section vise à mettre en lumière la cohérence de la reprise critique de l'externalisme, qui consiste en un certain sens à simplement rappeler le caractère théorique de la théorie sémantique.

I. L'accusation de Michael Devitt

La trahison dont il est question ici est le renoncement au réalisme dont Putnam se rendrait coupable en devenant le défenseur d'un réalisme interne qui ne serait qu'un constructivisme déguisé et en abandonnant le réalisme qu'il défendait dans ses écrits antérieurs. L'accusation est formulée par Michael Devitt dans *Realism and Truth*, ouvrage tout entier consacré à une défense du réalisme de sens commun³ et du réalisme

-
- 1 Ce travail trouve son origine dans un exposé présenté à la journée Putnam qui a eu lieu à l'Université Bordeaux Montaigne au printemps 2018. Je remercie le public qui y assistait pour ses questions et remarques, à la fois pertinentes et bienveillantes. Ce travail a également bénéficié de l'écoute exigeante des étudiants de l'Université Paris Nanterre lors d'un cours consacré à *Reason, Truth and History* en 2017-2018. Je remercie enfin les deux évaluateurs de l'article et le coordinateur du présent numéro pour leurs suggestions, qui ont permis de préciser des éléments importants.
 - 2 Notre propos restera donc volontairement centré sur cet ouvrage. Nous suivrons Putnam dans sa relecture de ses contributions antérieures, mais nous n'étudierons pas le devenir de ces idées postérieures à *Reason, Truth and History*, 1981.
 - 3 Le réalisme de sens commun considère que les objets matériels dépeints par nos sens existent objectivement, indépendamment de nos représentations mentales.

scientifique⁴. Après avoir explicité la nature du réalisme qu'il propose et les arguments en sa faveur, Devitt consacre les derniers chapitres de son livre à une critique des principales objections contre le réalisme. Celui de ces chapitres qui est consacré à Putnam est intitulé « Putnam le renégat ». Devitt y crédite le Putnam des *Philosophical Papers* (volumes 1 et 2) pour avoir inspiré sa propre défense du réalisme⁵, mais il y critique le Putnam de *Reason, Truth and History*, à la fois pour ses objections contre un réalisme franc adossé à une théorie correspondantiste de la vérité et pour la conception que Putnam se propose de substituer à ce réalisme franc, conception qui voit l'esprit et le monde concourir à constituer l'esprit et le monde, selon le slogan avancé dans la préface de *Reason, Truth and History*⁶.

II. Le réalisme de Putnam, première période

Les ingrédients du réalisme initial de Putnam sont de deux types, même si, comme nous allons le voir par la suite, ils gagnent à être considérés ensemble. Il y a d'abord une défense explicite du réalisme scientifique et, ensuite, le développement d'une nouvelle théorie de la signification, l'externalisme sémantique⁷, qui fait la part belle au monde dans la détermination de la référence de nos termes, indépendamment de notre capacité à saisir cette référence. Concernant le réalisme scientifique, Putnam argumente en faveur d'une interprétation réaliste à la fois des sciences empiriques et des mathématiques, sur la base du succès des premières et du rôle indispensable des secondes dans ce succès – voir par exemple l'article « Mathematical Truth »⁸ pour un tel développement parallèle. De manière plus spécifique, Putnam attaque la thèse de l'incommensurabilité de la référence, qui constitue un des éléments centraux du relativisme défendu par Kuhn ou Feyerabend⁹. Contre cette thèse, Putnam soutient que

4 Le réalisme scientifique considère que les théories scientifiques ne sont pas moins justes quant à ce qu'elles nous disent des entités inobservables que quant à ce qu'elles nous disent des phénomènes observables qu'elles visent à expliquer. à l'opposé, l'antiréalisme envisage les théories scientifiques dans une perspective instrumentale et voit dans les entités théoriques des fictions utiles.

5 Devitt M., *Realism and Truth*, 1984, p. 220.

6 *Reason, Truth and History*, 1981, p. xi (traduction française p. 9).

7 L'externalisme sémantique s'oppose à l'internalisme sémantique quant à la nature de la signification. Pour l'internaliste, la signification des mots est « dans les têtes », elle est ce que nous savons quand nous savons parler une langue. Pour l'externaliste, la signification des mots dépend aussi des circonstances extérieures de leur usage, pas seulement de ce que savent les locuteurs.

8 Putnam H., *Mind, Language and Reality*, 1975, chapitre 4.

9 La thèse d'incommensurabilité de la référence affirme que la référence des termes d'une théorie n'est déterminée qu'à l'intérieur de cette théorie, de sorte qu'on ne peut par exemple pas comparer la référence d'un terme comme « énergie » avant et après la révolution relativiste, et encore moins supposer que cette référence reste constante.

la référence de termes scientifiques comme « électricité » ou « masse » est trans-théorique, c'est-à-dire qu'elle ne change pas avec les théories scientifiques dans lesquelles ils figurent. Que la référence des termes ne change pas quand changent les théories permet de comparer celles-ci, dans la mesure où elles parlent bien, quoique de façon plus ou moins correcte, des mêmes choses. Dans « Explanation and Reference »¹⁰, Putnam dérive la stabilité de la référence des termes pour les grandeurs physiques de la théorie externaliste de la signification qu'il a commencé à développer quelques années plus tôt¹¹. Si la référence d'un terme comme « électricité » ne change pas lorsque l'on passe du cadre newtonien au cadre relativiste, c'est parce que la référence de ce terme n'est pas donnée par ce que la théorie scientifique que nous acceptons dit de l'électricité. Elle est déterminée par « une certaine chaîne causale, [qui renvoie à] une situation dans laquelle une *description* de l'électricité est donnée, une description qui est en général *causale*, c'est-à-dire une description qui distingue l'électricité comme *la* grandeur physique responsable de certains effets produits d'une certaine manière »¹². La référence du terme « électricité » étant ainsi fixée, elle est largement, sinon totalement, immune au changement scientifique. Même si notre conception de l'électricité change, variété de fluide hier, déplacement de particules chargées aujourd'hui, l'électricité est, hier comme aujourd'hui, ce qui dans le monde est responsable d'un certain nombre de phénomènes parmi lesquels la foudre, les chocs reçus au contact d'un poisson torpille ou la déviation de l'aiguille des boussoles. Cette immunité n'est cependant pas totale. Il se pourrait que l'on découvre que ces phénomènes ont en réalité des explications différentes, ou qu'il n'y a pas vraiment de grandeur physique correspondant à ce qu'on appelait l'électricité (mais bien sûr de telles découvertes ne seraient que des affirmations au sein d'une nouvelle théorie).

La théorie de la signification mobilisée pour rendre compte du caractère trans-théorique de la référence peut être considérée comme le second ingrédient à part entière du réalisme initial de Putnam. Parfois fusionnée avec les idées de Kripke sous le nom de « théorie causale de la référence » ou de « théorie de la référence de Kripke-Putnam », cette théorie comporte un versant négatif et un versant positif. Le versant négatif consiste à rejeter une approche internaliste de la signification, selon laquelle la signification d'un mot résiderait en totalité dans un sens cognitivement accessible aux locuteurs. Pour reprendre le slogan devenu célèbre de « Meaning and Reference » et de

10 Putnam H., *Mind, Language and Reality*, 1975, chapitre 11.

11 Putnam H., « Meaning and Reference », 1973.

12 Putnam H., *Mind, Language and Reality*, 1975, p. 200 (les italiques sont de l'auteur).

« The Meaning of “Meaning” »¹³, les significations ne sont simplement pas dans la tête. Certes, une partie de ce qui fait la signification d’un mot se trouve bien « dans la tête » – par exemple, la signification du terme « eau » accessible aux locuteurs compétents du français va comprendre le fait que « eau » fait référence à une substance qui est liquide dans des conditions normales de température et solide quand il fait très froid, ainsi que certaines représentations stéréotypiques que nous associons à cette substance comme le fait d’être inodore, ou d’être transparente à l’état liquide. Mais, dans une perspective externaliste, la signification du mot « eau » n’est pas épuisée par ce que nous savons de l’eau, et donc pas non plus par les représentations ou les sens que nous associons à ce terme. Le versant positif spécifique de la théorie externaliste de la signification vient quant à lui dire en quoi consiste la dimension externe de la signification, et plus précisément comment la référence des termes est fixée indépendamment de ce que les locuteurs ont dans la tête. Pour les termes d’espèces naturelles comme pour les termes de grandeur physique dont il a été question plus haut, la référence est d’abord rattachée de façon causale à certains éléments se trouvant dans l’environnement. Dans un premier temps, « or » fait référence à un ou des échantillons auxquels les premiers locuteurs à avoir utilisé le terme – ou éventuellement les experts reconnus par la communauté linguistique – se rapportent de manière indexicale (« voilà de l’or »). Le locuteur qui emploie le terme « or » ne sait pas forcément grand-chose de ces échantillons. Cependant, il est relié de façon causale à ceux-ci et c’est cette chaîne causale qui ancre la référence du terme. Insistant sur le fait que l’ancrage pertinent n’est pas forcément celui d’un premier locuteur mythique, Putnam décrit ainsi le cas où cette chaîne passe par des experts : « La “chaîne causale” est une chaîne de coopérations qui relie l’auditeur aux experts pertinents, tels que déterminés par la société¹⁴ ». Pour autant, « or » ne fait évidemment pas seulement référence aux échantillons identifiés lors du baptême initial ou reconnus par les experts, mais bien à toute substance de même nature que ces échantillons. La fixation de la référence se fait donc en deux temps : il y a d’abord l’ancrage de la référence qui lie le terme à un certain échantillon et ensuite l’extension de la référence qui étend la référence à tout ce qui dans le monde se trouve être identique à l’échantillon. Dans les termes de Michaelson et Reimer,

« La théorie causale de la référence de Putnam / Kripke pour les termes d’espèces naturelles [...] est similaire à la façon dont Kripke voit la référence des noms propres ; de fait, il s’agit plus ou moins d’une extension de cette

13 Putnam H., *Mind, Language and Reality*, 1975, chapitre 12.

14 Putnam H., *Mind, Language and Reality*, 1975, p. 287.

conception. La référence est initialement fixée par un baptême, soit par la perception soit par description d'échantillons d'une espèce naturelle particulière. Elle s'étend ensuite à tout ce dont la structure interne est identique à celle des échantillons¹⁵. »

Une théorie de la signification de ce genre apparaît intrinsèquement réaliste. Elle semble en effet présupposer un monde indépendant de nous, à la fois quant à son existence et sa structure, qui vient apporter sa contribution à la signification de nos termes en déterminant quelle est leur référence étant donné un usage indexical de ceux-ci. Dans « The Meaning of "Meaning" », cet engagement réaliste s'exprime à travers l'injonction associée à la théorie externaliste, « soyons réalistes »¹⁶. Putnam y contraste l'interprétation réaliste de la contribution du monde à la référence avec une approche internaliste anti-réaliste comme celle de Dummett¹⁷, pour laquelle « les notions de vérité et de référence ne sont pas disponibles à l'extérieur de la théorie »¹⁸. De même, « Meaning and Reference » se conclut par une invocation du monde « réel » :

« La théorie sémantique traditionnelle laisse de côté deux contributions à la détermination de la référence – l'apport de la société et l'apport du monde réel ; une meilleure théorie sémantique se doit de les inclure tous les deux¹⁹. »

L'externalisme de Putnam semble bien ainsi participer d'une compréhension réaliste de la signification, à l'opposé des conséquences antiréalistes de l'internalisme d'un Dummett. Il faut cependant noter que ces associations ne valent pas dans tout contexte philosophique²⁰. On peut être externaliste et constructiviste, en mettant l'accent sur la dimension sociale de l'externalisme (la référence est déferée aux experts, tout en restant partie prenante de la représentation du monde par la communauté linguistique). On peut être internaliste et réaliste, en attribuant au sens une nature et une visée objectives (comme chez Frege où la signification saisie par les locuteurs détermine la référence, l'une et l'autre étant indépendantes des représentations psychologiques).

15 Michaelson et Reimer, « Reference », 2019.

16 Putnam H., *Mind, Language and Reality*, 1975, p. 235.

17 Le propre de l'antiréalisme de Dummett est de déduire d'une théorie vérificationniste de la signification une révision des concepts sémantiques qui se retrouvent tous indexés à nos capacités (la vérité est analysée en termes de vérifiabilité, etc.).

18 Putnam H., *Mind, Language and Reality*, 1975, p. 236.

19 « Meaning and Reference », 1973, p. 711.

20 Cette remarque fait suite à un commentaire d'un des évaluateurs du présent article, que je remercie.

III. Le passage au réalisme interne

À partir du discours prononcé devant l'*American Philosophical Association* en décembre 1976 et publié sous le titre « Realism and Reason », Putnam revient sur la question du réalisme, cette fois pour interroger la nature même de ce réalisme. Il en distingue deux versions radicalement différentes. Le réalisme métaphysique se place en dehors de toute théorie pour poser l'existence d'un monde indépendant de nous et comprendre la vérité comme une relation de correspondance entre des symboles et des éléments de ce monde. C'est ce réalisme que Putnam va dorénavant rejeter, en cherchant à montrer qu'un tel point de vue extérieur adopté pour parler du monde, de la référence et de la vérité n'existe pas ou n'est pas réellement intelligible. Au réalisme métaphysique, Putnam oppose un réalisme interne qui assume de toujours devoir se placer à l'intérieur d'un cadre théorique pour parler du monde, de la référence, ou de la vérité. Pour le réalisme interne, le monde est toujours le monde tel que nous le concevons. La vérité est toujours la vérité telle qu'elle peut être connue, et c'est la raison pour laquelle elle doit être redéfinie en termes d'acceptabilité rationnelle idéalisée²¹. L'objectivité n'est pas fondée dans la relation mystérieuse que nous entretiendrions avec un monde vierge, mais dans l'entreprise de connaissance elle-même et les normes qui la guident.

Que devient alors le réalisme initial de Putnam, marqué par la défense du réalisme scientifique et la théorie externaliste de la signification ? S'agissant du réalisme scientifique, la question est abordée frontalement par Putnam dans un article de 1982 intitulé « Three Kinds of Scientific Realism ». Quand la question de la vérité des théories scientifiques est comprise en termes d'une vérité transcendante, le réalisme scientifique n'est qu'une déclinaison du réalisme métaphysique, et il faut selon Putnam rejeter celui-là avec celui-ci. Mais il y a pour Putnam une autre manière d'être réaliste pour la science, qui consiste à ne l'être pas moins que dans d'autres domaines. La thèse est alors simplement que les électrons existent tout autant que les chaises ou les sensations²². En ce sens-là, Putnam souscrit toujours au réalisme scientifique. Mais il s'agit d'un réalisme réinterprété à la lumière du réalisme interne et démarqué du réalisme métaphysique.

Qu'en est-il du réalisme associé à la théorie externaliste de la signification ? Est-il lui aussi préservé après réinterprétation ? Et si oui, en quoi cette réinterprétation consiste-

21 Il y aurait beaucoup à dire sur la capacité dont il est question lorsque l'on parle de la vérité telle qu'elle peut être connue, et l'acceptabilité rationnelle idéalisée risque fort d'excéder nos capacités réelles de connaissances (voir Dummett M., *The Logical Basis of Metaphysics*, 1991, chapitre 15 « Realism and the Theory of Meaning »).

22 Putnam H., « Three Kinds of Scientific Realism », 1982, p. 198.

t-elle ? Dans *Reason, Truth and History*, la manière dont les théories causales sont évoquées est ambivalente. D'un côté,

« Toute cette histoire complexe concernant la manière dont nous faisons référence à certaines choses en vertu du fait qu'elles sont reliées à nous par des "chaînes causales du type approprié", et à d'autres encore en vertu du fait qu'elles sont "de la même espèce" que les choses reliées à nous par des chaînes causales du type approprié [...] n'est pas tant fausse que vaine²³. »

La portée de la théorie causale semble pour le moins contestée. Pour autant, celle-ci ne serait pas « fausse », et de fait Putnam reprend dans le chapitre 5 l'explication qu'il a donnée de la sémantique des termes d'espèces naturelles pour l'appliquer au concept de rationalité, ce qui veut bien dire que la théorie n'est pas purement et simplement abandonnée. Comment le réalisme interne, grande thèse de la période dite intermédiaire de la philosophie de Putnam, accommode-t-il la théorie externaliste de la signification, acquis majeur de la première période et pilier du réalisme initial du même Putnam ? C'est la question que nous nous proposons maintenant d'examiner en détail, à travers une lecture des usages de l'externalisme dans *Reason, Truth and History*.

IV. L'externalisme dans *Reason, Truth and History*

On peut identifier principalement quatre reprises de l'externalisme dans *Reason, Truth and History* : 1/ aux chapitres I et II, pour critiquer les illusions internalistes qui sont à l'origine de notre croyance dans la possibilité du scénario des cerveaux dans la cuve, 2/ au chapitre II encore, mais cette fois dans le cadre d'une critique de certaines pseudo-solutions au problème de la référence, 3/ au chapitre III, à nouveau dans le cadre d'une critique des conceptions réalistes de la référence, au sens du réalisme métaphysique cette fois, et 4/, plus ponctuellement, au chapitre I ou au chapitre V, pour rendre raison de l'écart entre ce que nous croyons savoir de ce dont nous parlons et ce à quoi nous faisons réellement référence.

Ces quatre reprises ont des statuts bien différents : les reprises 1/ et 4/ sont des reprises positives, l'externalisme étant mobilisé pour nous aider à nous déprendre de certaines illusions tandis que les reprises 2/ et 3/ sont des reprises critiques, l'externalisme étant convoqué pour rabaisser ses prétentions à résoudre le problème de la référence. Il s'agira alors pour nous de comprendre, à travers les reprises positives, ce qui est conservé (dans les paragraphes V et VI) et, à travers les reprises négatives, ce qui

23 Putnam H., *Reason, Truth and History*, 1981, p. 53 (traduction française p. 65 modifiée).

est abandonné ou mis à distance, afin de reconstruire la réinterprétation de l'externalisme qui accompagne implicitement le passage au réalisme interne (dans les paragraphes VII et VIII).

V. L'externalisme contre la détermination internaliste de la référence

Voyons d'abord la reprise 1/. Une des illusions constitutives du réalisme métaphysique est l'illusion de la détermination de la référence, qui consiste à supposer que nos termes ont une référence parfaitement déterminée, ancrée dans un monde indépendant de nous. Si tel était le cas, on pourrait donner un sens absolu et non épistémique à la vérité ou la fausseté des énoncés ou des propositions, comme le veut le réaliste. Contre cette illusion, Putnam affirme l'indétermination systématique de la référence, en s'appuyant notamment sur l'argument modèle-théorique qui permet de généraliser les exemples quiniens de relativité ontologique²⁴.

Si la référence était déterminée par ce que nous savons quand nous savons parler une langue, alors la référence serait bien sûr déterminée tout court. L'internalisme sémantique constitue pour Putnam une des causes possibles de l'illusion de la détermination de la référence. Si les significations sont dans la tête, alors, comme les cerveaux dans la cuve ont, pour ainsi dire, les mêmes choses dans la tête que moi, les cerveaux dans la cuve et moi visons bien les mêmes objets, de sorte que ceux-ci peuvent de façon cohérente penser, comme moi, qu'ils ne sont pas des cerveaux dans une cuve (la référence est sauve pour autant qu'elle est ce qui « hors de la tête » est visé et déterminé par ce que nous avons « dans la tête »).

Les arguments contre l'internalisme de « Meaning and Reference » et de « The Meaning of "Meaning" » peuvent être repris à l'identique. Il ne s'agit cependant pas cette fois de réfuter l'internalisme pour proposer une meilleure théorie de la signification, mais simplement de réfuter l'internalisme en tant que fondement, ou plutôt en tant qu'un des fondements possibles, de l'illusion de détermination de la référence. Pour le dire autrement, on ne retiendra pas cette fois que quelque chose d'autre doit venir déterminer la référence, mais plutôt que, sur cette base-là en tout cas, et contrairement à ce qu'on pouvait penser, la référence n'est pas déterminée.

24 Si Putnam présente explicitement cet argument comme une généralisation de l'argument quiniens en faveur de l'inscrutabilité de la référence, il ne parle pas lui-même volontiers d'indétermination ou d'inscrutabilité. Peut-être faut-il comprendre que parler ainsi reviendrait déjà à adopter une perspective extérieure sur la référence. L'argument de Putnam se présente moins comme un argument sceptique que comme une réfutation par l'absurde de la perspective métaphysique (voir l'article de David L. Anderson « What Is Realistic about Putnam's Internal Realism », 1992, pour une défense systématique de la lecture des arguments de Putnam contre le réalisme métaphysique comme des *reductio ad absurdum*).

C'est ainsi qu'est repris dans le chapitre I l'argument à propos des ormes et des bouleaux. Dans les articles antérieurs mettant en place l'externalisme sémantique, l'exemple était déjà mobilisé. Ayant une expertise forestière limitée, le concept que j'associe à « orme » et le concept que j'associe à « bouleau » sont tout à fait semblables, de sorte que ces concepts, pour autant que je les possède en tant que locuteur compétent du français, ne permettent pas à eux seuls de fixer la référence de « orme » et de « bouleau », qui bien sûr, elle, n'est pas la même. Putnam en conclut que les concepts n'ont pas davantage de capacité magique à faire référence que tout autre type de symboles ou de contenus mentaux :

« Les concepts ne sont pas des présentations mentales qui réfèrent de façon intrinsèque à des objets extérieurs pour la raison tout à fait décisive qu'ils ne sont pas du tout des présentations mentales. Les concepts sont des signes utilisés de certaines manières [...]»²⁵. »

La conclusion qui était celle de « Meaning and Reference » est rappelée en parenthèse :

« (Soit dit en passant, ceci montre que la détermination de la référence est sociale et non pas individuelle ; vous et moi déférons aux experts qui savent distinguer un bouleau d'un orme)²⁶. »

Cette parenthèse est révélatrice. Elle indique la suite de l'histoire qui était celle de « Meaning and Reference », s'agissant d'expliquer comment est déterminée la référence si elle ne l'est pas sur une base purement mentale. Cependant, dans le contexte de l'ouvrage, elle s'avère presque trompeuse : certes la détermination de la référence est sociale plutôt qu'individuelle, mais – ce sera une des principales leçons anti-réalistes de *Reason, Truth and History* – cette détermination est surtout toujours inachevée.

Le très célèbre argument de la terre jumelle, planète où le liquide appelé « eau » n'a pas la composition chimique de l'eau sur Terre, voit Oscar, un locuteur sur la Terre en 1750, et son jumeau sur la Terre jumelle se trouver dans le même état mental et pourtant faire référence à deux substances différentes. Cet argument est lui aussi repris, au début du chapitre II, pour conclure à nouveaux frais à l'impossibilité pour les cerveaux dans la cuve de concevoir leur sort :

« Une fois que l'on a compris que l'état mental (au sens individuel ou collectif) ne détermine pas la référence, il n'y a plus rien de surprenant au fait que les cerveaux dans la cuve ne puissent pas faire référence à des objets

25 Putnam H., *Reason, Truth and History*, 1981, p. 18 (traduction française p. 28-29 modifiée).

26 *Reason, Truth and History*, p. 18 (traduction française p. 29 modifiée).

extérieurs (bien qu'ils se trouvent dans le même état mental que nous) et qu'ils ne puissent donc pas dire ou penser qu'ils sont des cerveaux dans une cuve²⁷. »

Quelques alinéas plus haut, Putnam a envisagé l'objection selon laquelle la référence du terme « eau » sur la Terre ou la Terre Jumelle avant les progrès de la chimie serait en réalité identique, couvrant toute substance ayant les propriétés apparentes associées à l'eau. Il écarte cette objection au motif qu'elle impliquerait de renoncer au caractère trans-théorique de la référence. Il faudrait en effet admettre que, sur Terre comme sur la Terre Jumelle, la référence du mot « eau » n'est plus la même avant et après les progrès de la chimie, puisque « eau » ferait avant référence à tout liquide ayant les mêmes propriétés apparentes que l'eau, et après, sur Terre, à H₂O et, sur la Terre Jumelle, au composé y jouant le rôle de l'eau. Putnam ne renie donc pas la permanence de la référence, mais celle-ci est un élément mobilisé dans un argument contre l'internalisme, au lieu d'être établie sur la base de l'externalisme, de sorte qu'elle est moins démontrée sur la base de la théorie sémantique que convoquée au service de celle-ci.

De fait, dans *Meaning and the Moral Sciences*, Putnam avait présenté une explication de la permanence trans-théorique de la référence plus complexe que celle donnée dans « Explanation and Reference » :

« [Les arguments de Feyerabend en faveur de l'incommensurabilité de la référence] peuvent être bloqués en soutenant (comme je l'ai fait à plusieurs endroits, et comme l'a également fait Kripke) que les termes scientifiques ne sont pas synonymes de certaines descriptions. De plus, c'est un principe essentiel de la méthodologie sémantique que lorsque des locuteurs spécifient le référent d'un terme qu'ils utilisent à l'aide d'une *description* et que, à cause de croyances factuelles erronées entretenues par ces locuteurs, la description échoue à faire référence, nous devons supposer qu'ils accepteraient des reformulations raisonnables de leurs formulations [...]»²⁸. »

Putnam continue en prenant l'exemple de l'électron, qui, selon la mécanique quantique, ne vérifie pas exactement les propriétés que lui attribue le modèle de Bohr, mais qui les vérifie approximativement et qui est responsable des effets attribués par Bohr à l'électron. Il souligne que dans ce cas « le principe du bénéfice du doute exige que nous

27 Putnam H., *Reason, Truth and History*, 1981, p. 25 (traduction française p. 36).

28 Putnam H., *Meaning and the Moral Sciences*, 1978, p. 23-24.

considérations que Bohr fait référence à ces particules »²⁹. Théorie causale et principe de charité sont ainsi conjointement mobilisés pour assurer la permanence trans-théorique de la référence. La théorie causale permet de justifier le dépassement de la description. Les scientifiques s’y prennent d’une certaine manière pour décrire les particules, mais ces descriptions sont un moyen plutôt qu’une fin, cela ne préjuge pas du fait que ce à quoi ils font référence quand ils parlent de telle ou telle particule vérifie exactement les descriptions. Ils font référence à la cause de certains phénomènes qu’ils associent à l’usage des termes désignant ces particules, cause décrite à l’aide des lois reconnues par la théorie en usage. Mais la référence n’est pas pour autant magiquement fixée par cette assignation causale, il faut une identification trans-théorique de la référence, qui se fait du point de vue d’une nouvelle théorie, et qui recourt au principe de charité, rebaptisé pour l’occasion bénéfique du doute, afin d’isoler la référence des termes de l’ancienne théorie dans les termes de la nouvelle. Il n’y a donc pas de garantie automatique fournie par l’externalisme. Il faut qu’il soit en pratique possible de s’appuyer sur la préservation approximative des lois utilisées dans les descriptions scientifiques pour pouvoir appliquer le principe de charité et assurer la préservation de la référence.

On notera que le rôle des lois dans la permanence de la référence était déjà mis en avant par Putnam dans un article de 1962, « The Analytic and the Synthetic », qui précède le développement de l’approche externaliste. Dans cet article repris dans le volume 2 des *Philosophical Papers*, Putnam parle des termes théoriques comme de « law-cluster concepts »³⁰, des concepts associés à un faisceau de lois qui sont tels que « n’importe quelle loi particulière peut être abandonnée sans détruire l’identité du concept “faisceau de lois” en jeu »³¹. La conception de la permanence de la référence esquissée dans *Meaning and the Moral Sciences* réalise ainsi une forme de synthèse inattendue³² entre la perspective pré-externaliste de « The Analytic and the Synthetic » et la perspective externaliste de « Explanation and Reference ». La théorie causale légitime le genre de permanence que l’on peut attribuer de façon charitable sur la base d’un certain degré de préservation des lois.

29 *Meaning and the Moral Sciences*, 1978, p. 24.

30 Putnam H., *Mind, Language and Reality*, p. 50.

31 Putnam H., *Mind, Language and Reality*, p. 52.

32 La continuité entre l’externalisme de Putnam et son attachement initial à l’idée de concepts identifiés par des faisceaux de lois était déjà pointée par Åsa Wikforss (« Bachelors, Energy, Cats and Water. Putnam on Kinds and Kind Terms », 2013), qui souligne dans ce contexte l’influence de la critique quinienne de l’analyticité sur Putnam. Toute propriété venant caractériser un concept est négociable, ce qui place Putnam à distance de l’essentialisme à la Kripke.

La leçon, en tout cas, semble claire. L'externalisme n'a pas, ou plus, la force d'assurer comme de l'extérieur l'ancrage de la référence, rien ne se fait indépendamment des théories.

VI. La contribution externaliste à la référence

Les reprises de l'externalisme que nous avons examinées au paragraphe précédent avaient essentiellement une visée critique : il s'agissait, si l'on met de côté notre excursus concernant la permanence trans-théorique de la référence, de montrer comment la référence ne se fait pas. Il y a cependant également dans *Reason, Truth and History*, et pas seulement comme nous l'avons vu dans *Meaning and the Moral Sciences* où le tournant du réalisme interne est certes également déjà pris, une reprise substantielle de la théorie externaliste de la signification lorsqu'il s'agit, ponctuellement, d'éclairer la référence de tel ou tel terme. Au chapitre I, quand Putnam a établi que les cerveaux dans la cuve ne pouvaient pas faire référence aux objets du monde extérieur, il se demandait à quoi font alors référence les termes du français-de-la-cuve (la langue parlée par les cerveaux dans la cuve). Au chapitre V, quand il envisage une théorie de la rationalité idéale dépassant notre compréhension actuelle de celle-ci, il se demande comment la rationalité peut ne pas se réduire aux normes par lesquelles nous la caractérisons aujourd'hui. Dans les deux cas, la réponse est donnée par l'externalisme. Pour le premier cas, les mots du français-de-la-cuve font référence à ce qui est réellement à l'origine de leur usage :

« [...] "cuve" fait référence à des cuves dans l'image en français-de-la-cuve, ou à quelque chose de connexe (des impulsions électroniques ou des caractéristiques du programme), mais certainement pas à de véritables cuves, puisque le français-de-la-cuve n'a pas de liens causaux avec de véritables cuves³³. »

Et à propos de la rationalité, Putnam fait le parallèle avec l'or et rappelle que

« Nous sommes disposés à compter quelque chose comme appartenant à une certaine espèce même si nos tests *actuels* ne permettent pas à eux seuls de montrer qu'il s'agit bien d'un membre de l'espèce en question pour peu qu'il s'avère un jour que cette chose a la même nature essentielle que (ou, pour le

33 Putnam H., *Reason, Truth and History*, 1981, p. 14 (traduction française p. 25 modifiée).

dire de façon plus vague, est suffisamment similaire à) les exemples paradigmatiques (ou la grande majorité d'entre eux)³⁴. »

Cependant, l'appel à l'externalisme s'accompagne dans les deux cas de réserves qui peuvent sembler inattendues. À propos du français-de-la-cuve, Putnam ne s'engage pas quant à sa capacité à faire référence, laissant ouverte la possibilité qu'il n'y ait pas de référence du tout.

« Par conséquent, si nous sommes des cerveaux dans une cave, alors la phrase "Nous sommes des cerveaux dans une cuve" dit quelque chose de faux (si tant est qu'elle dise quelque chose)³⁵. »

À propos de l'or et de la rationalité³⁶, il suspend la possibilité d'un écart entre nos critères d'identification actuels et l'espèce à laquelle il est véritablement fait référence à l'existence de lois :

« Il y a de puissantes lois universelles auxquelles obéissent toutes les instances d'or, c'est ce qui rend possible de décrire l'or comme la substance qui se révélera obéir à ces lois quand nous les connaissons [...]»³⁷. »

À chaque fois, c'est la capacité du monde à garantir la référence qui semble en cause. Un externaliste pourrait juger ces réserves superflues. Les cerveaux dans la cuve sont reliés à une machine qui produit des représentations et nourrit l'usage du français-de-la-cuve. Il y a donc bien une chaîne causale à remonter, même si on ne sait pas exactement où il convient de s'arrêter. De même, il y a des instances paradigmatiques d'or ou de rationalité, et ces instances sont alors bien les instances paradigmatiques de quelque chose, quels que soient le nombre et la richesse des lois qui mettent cette chose en jeu. Mais cette manière de voir les choses relève sans doute déjà du réalisme métaphysique, puisqu'elle attribue une référence qui transcende la capacité de saisie, même idéalisée, des locuteurs. Du point de vue des cerveaux dans la cuve, il n'est pas sûr qu'une théorie idéale de leur sort qui identifie la cause informatique de leurs représentations fasse sens. S'agissant de l'or, il n'est pas sûr que parler de la substance sans avoir les moyens de l'identifier comme un élément chimique répondant à certaines lois fasse sens.

Il y a donc même dans les appels concrets à l'externalisme pour rendre compte de la

34 *Reason, Truth and History*, p. 103 (traduction française p. 119 modifiée).

35 *Reason, Truth and History*, 1981, p. 15 (traduction française p. 25 modifiée).

36 Putnam reste cependant prudent quant à l'existence de généralisations substantielles gouvernant notre concept de rationalité qui seraient analogues aux lois décrivant les propriétés d'un métal.

37 *Reason, Truth and History*, 1981, p. 104 (traduction française p. 120 modifiée).

référence de tel ou tel terme une méfiance à l'égard d'une application de celui-ci qui introduirait une conception transcendante de la référence. Cette méfiance est tout à fait dans le prolongement des remarques de la section précédente sur le caractère intra-théorique de la permanence trans-théorique de la référence. Pour mieux comprendre ce qui la justifie, il faut maintenant examiner les reprises critiques de l'externalisme, qui le confrontent à l'indétermination de la référence.

VII. Les relations causales et la détermination de la référence

Si l'externalisme est correct en tant qu'élément d'une théorie de la signification, il n'est pas correct en tant que justification d'une perspective réaliste au sens métaphysique sur la référence. Les reprises critiques de l'externalisme dans *Reason, Truth and History* ont précisément pour fonction de barrer le passage de l'un à l'autre. Celles-ci interviennent principalement à deux occasions. La première fois, c'est au chapitre II, après que Putnam a donné l'argument modèle-théorique en faveur de l'indétermination de la référence et alors qu'il examine des objections possibles à l'indétermination.

La thèse d'indétermination de la référence consiste à affirmer que les contraintes théoriques et opérationnelles supposées fixer la référence ne la déterminent pas de manière unique. La démonstration repose sur un argument modèle-théorique général montrant que l'ensemble des énoncés acceptés comme vrais, quel qu'il soit, sous-détermine la référence des termes figurant dans ces énoncés : il existera toujours des interprétations (hautement) non-standard compatibles avec la vérité des énoncés en jeu. Dans la mesure où l'acceptation des contraintes théoriques et opérationnelles fixant la référence peut être ramenée à l'acceptation de certains énoncés, il s'ensuit que la référence n'est pas fixée de manière unique par des contraintes théoriques et opérationnelles. Les objections à la thèse d'indétermination consistent alors à envisager d'autres manières dont la fixation de la référence pourrait avoir lieu, qui en garantiraient l'unicité. L'appel à des considérations darwiniennes est ainsi rejeté, seul ce que nous comptons comme vrai compte pour notre survie, de sorte que l'évolution peut bien sélectionner des croyances, elle ne sélectionnera pas la référence. C'est dans ce contexte qu'intervient également la critique des théories causales, envisagées comme une autre réponse possible quant à ce qui viendrait réduire le jeu – au sens de l'intervalle laissé entre deux pièces, Putnam parle de « slack »³⁸ – entre les conditions de la vérité et la référence.

38 *Reason, Truth and History*, 1981, p. 45.

Voyons cela de plus près. L'interprétation réaliste métaphysique de l'externalisme est attribuée à Field, en particulier en lien avec son article fameux de 1972, « Tarski's Theory of Truth ». Dans cet article, Field remet en question la portée philosophique de la définition tarskienne de la vérité, et conteste le fait qu'elle réussisse véritablement à rendre le concept de vérité fréquentable, ainsi que l'ont considéré des philosophes positivistes comme Carnap ou Popper. L'argument porte sur l'ultime chaînon de la définition tarskienne et la définition de la dénotation des termes. Field envisage deux lectures possibles de Tarski sur ce point. Selon la première, la dénotation est considérée comme une relation primitive, « John » dénote ce qu'il dénote, de sorte que la définition de la vérité dépend d'un concept sémantique, le concept de dénotation. Selon la seconde, la dénotation est fixée via traduction dans le métalangage dans lequel est couchée la définition de la vérité, « John » dénote Jean³⁹, et la relation de dénotation pour le langage-objet est fixée à l'aide d'une longue liste de ce genre. Field fait alors remarquer qu'aucune de ces deux lectures ne confère à la définition tarskienne la portée philosophique qu'on lui prête. Selon la première interprétation, le concept sémantique de vérité n'est éliminé qu'au profit d'un autre concept sémantique. Selon la seconde interprétation, le concept de dénotation n'est pas réellement analysé et la définition de la vérité n'a pas de portée explicative. Field conclut qu'une définition acceptable du point de vue physicaliste du concept de vérité doit faire davantage et incorporer une réduction physicaliste de la relation de dénotation :

« Je ne pense pas que Kripke, ou qui que ce soit d'autre, pense qu'il soit possible de développer des théories *purement causales* de la dénotation en tant que relation primitive [...] ; cependant cela ne doit pas nous faire oublier qu'il a suggéré un type de facteur intervenant dans la relation de dénotation qui renouvelle l'espoir d'expliquer la connexion entre le langage et les choses dont il parle. Il me semble que la possibilité d'une théorie de ce genre (je reste délibérément vague) est essentielle pour l'acceptabilité conjointe du physicalisme et du terme sémantique "dénote" [...] ⁴⁰. »

Autrement dit, les théories causales de la dénotation constituent pour Field l'embryon d'une réduction physicaliste de la relation de dénotation et, partant, de la notion de vérité. Si les théories causales l'intéressent, c'est parce qu'elles suggèrent la possibilité de briser le cercle des notions sémantiques en expliquant la dénotation d'un terme non

39 Ou alors « John » dénote John, selon la manière dont on envisage la traduction des noms propres.

40 Field H., 1972, « Tarski's Theory of Truth », p. 367.

pas à partir d'autres notions sémantiques (comme la vérité) ou à partir de certaines relations de dénotation primitives (comme dans la théorie Russellienne des noms propres logiques), mais à partir de notions utilisées pour expliquer les phénomènes physiques (comme la notion de causalité).

Putnam avait déjà critiqué Field dans *Meaning and the Moral Sciences*, en contestant l'exigence d'une explication de la relation de dénotation. L'objection était double. D'une part, Putnam mettait en doute la nécessité d'une théorie substantielle de la référence dans le cadre d'une définition de la vérité (objection que Putnam attribue à Stephen Leeds). Il soutenait que si l'objectif est simplement de rendre compte du pouvoir expressif de la notion de vérité⁴¹, la théorie tarskienne se suffit à elle-même. Peu importe alors que la dénotation des termes ne soit donnée qu'à travers une liste les traduisant dans le métalangage, il suffit de réussir à dériver les instances de la convention T. D'autre part, il mettait en doute la possibilité d'une théorie de la référence qui ne soit pas triviale. Si nous ne pouvons parler du monde que de l'intérieur de notre schème conceptuel, l'explication de la relation primitive de référence semble nécessairement passer par des équivalences triviales (« électron » fait référence aux électrons).

La perspective adoptée sur l'approche de Field dans *Reason, Truth and History* est différente. Il ne s'agit plus de s'interroger sur sa légitimité dans la perspective d'une théorie de la vérité (est-il nécessaire et possible de compléter la définition tarskienne de la vérité par une théorie physicaliste de la référence ?), mais sur son apport dans la perspective d'une défense du réalisme (une théorie physicaliste de la référence permet-elle de rendre raison d'une compréhension réaliste de la référence ?). Malgré les doutes formulés dans *Meaning and the Moral Sciences*, Putnam va supposer développée une théorie de la référence à la Field, et se demander si l'existence d'une telle théorie mettrait en échec l'indétermination de la référence, justifiant ainsi la conception réaliste métaphysique. L'argument modèle-théorique vise à établir qu'on ne peut pas sortir d'un cercle des notions sémantiques : on a besoin d'une compréhension réaliste de la relation de référence pour donner un sens réaliste métaphysique à la vérité, mais si la relation de référence n'est déterminée que par les énoncés acceptés comme vrais, alors nous n'avons pas à notre disposition une compréhension réaliste métaphysique de la

41 Le pouvoir expressif de la notion de vérité réside dans le fait qu'elle nous permet d'exprimer de manière finie ce qu'on ne pourrait sinon exprimer que de manière infinitaire. Ainsi, dire « tout ce que le Pape dit est vrai » permet d'exprimer de manière finie ce qu'on ne pourrait exprimer sans la notion de vérité que de manière infinitaire en disant « si le Pape a dit que la guerre doit cesser, alors la guerre doit cesser, et si le Pape a dit que les riches doivent partager les richesses, alors les riches doivent partager les richesses, etc. ».

référence. La promesse de l'analyse physicaliste est de briser ce cercle en fournissant un ancrage causal de la référence qui en autoriserait une compréhension réaliste. C'est donc le même trait d'une théorie « à la Field » qui est en jeu (une analyse physicaliste de la référence, sans recours circulaire à des notions sémantiques), mais au service d'un projet différent (non pas pour rendre acceptable la sémantique du point de vue physicaliste, mais pour justifier la perspective réaliste sur la référence). Une théorie de la référence formulée en termes physicalistes permet-elle d'échapper à l'indétermination de la référence ?

La réponse de Putnam est négative, et prend la forme d'un dilemme. Une théorie à la Field consiste à proposer une certaine relation R définie en termes physicalistes telle que

(1) x fait référence à y si et seulement si x est dans la relation R avec y ⁴².

Le dilemme porte sur l'interprétation de la relation R . Soit on considère qu'il s'agit d'une relation définie dans un certain métalangage physicaliste pour étudier la référence au sein du langage objet. Soit on considère qu'il s'agit d'une relation absolue constituant en quelque sorte la nature même de la référence, indépendamment de la théorie dans laquelle nous essayons de la saisir. Dans le second cas, on a simplement affaire à une pétition de principe réaliste, la relation de référence est contemplée comme de l'extérieur, d'un point de vue affranchi de toute théorie, qui pour Putnam n'existe tout simplement pas. Dans le premier cas, plus intéressant, l'interprétation de R est fixée par la théorie physicaliste dans laquelle elle s'inscrit. Elle est soumise à la même indétermination que celle qui vaut pour la relation de référence dans le langage-objet :

« À tout modèle admissible du langage objet correspondra un modèle de notre méta-langage dans lequel (1) vaut ; l'interprétation de " x est dans la relation R avec y " fixera l'interprétation de " x fait référence à y ". Mais il ne s'agira que d'une relation *au sein de chaque modèle admissible* ; cela ne permettra absolument pas de réduire le nombre de modèles admissibles⁴³. »

Autrement dit, on ne fixera pas davantage la relation de référence qu'elle ne l'était déjà car à une interprétation non-standard des termes du langage objet correspondra une interprétation non-standard de la relation R au sein d'un modèle de la métathéorie. L'argument est elliptique et il n'a rien d'évident. Ne se pourrait-il pas que la métathéorie en dise plus sur la référence que n'en disent les contraintes théoriques et opérationnelles formulées dans le langage objet ? N'est-ce pas après tout précisément

42 Putnam H., *Reason, Truth and History*, 1981, p. 45 (traduction française p. 57 modifiée).

43 *Reason, Truth and History*, 1981, p. 46 (traduction française p. 58 modifiée).

l'intérêt des théories causales que de nous dire des choses sur la référence qui vont au-delà de ce que savent les locuteurs ? Si la métathéorie en dit plus, le nombre de modèles admissibles sera bien réduit. On peut envisager deux réponses différentes. On peut concéder l'objection, corriger Putnam en accordant qu'il pourra bien y avoir réduction du nombre de modèles. Mais on répétera l'argument modèle-théorique pour la métathéorie. Celle-ci admet des modèles non-standard de sorte que, même si le nombre de modèles admissibles pour le langage-objet s'en trouve réduit, on n'arrivera jamais au modèle unique partie prenante de la conception correspondantiste du réalisme métaphysique. Une autre réponse possible, qui éviterait de corriger Putnam, consistera à supposer que l'on fait la théorie de la référence pour un certain langage dans ce langage, de sorte que la métathéorie est incluse dans la théorie du langage-objet, et que tout ce que pourrait dire la théorie causale de la référence serait déjà intégré aux contraintes théoriques. Dans ce cas, on aura bien une interprétation non-standard de *R* pour chaque interprétation non-standard de la référence compatible avec les contraintes théoriques et opérationnelles. Et dans tous les cas, l'important est que la référence du terme « relations causales » qui figure dans les théories causales n'est pas elle-même « collée à une relation déterminée par une colle métaphysique »⁴⁴.

La théorie causale de la référence se prête à une illusion réaliste métaphysique. Elle semble pouvoir justifier une approche correspondantiste de la référence en expliquant comment la référence des termes est ancrée dans le monde par des chaînes causales. L'illusion consiste à oublier que la théorie causale est bien une théorie et que cet ancrage ne peut être saisi que de l'intérieur d'un certain schème conceptuel. Il n'y a pas de correspondance qui soit établie entre les termes du langage d'un côté et un monde vierge de conceptualisation de l'autre. Il n'y a que le retour de notre théorie du monde sur une partie d'elle-même qui vient expliquer de l'intérieur comment notre langage fonctionne. Field ne disait rien d'autre⁴⁵ :

« La raison pour laquelle nous avons besoin d'une théorie de la vérité et d'une théorie de la référence primitive n'est pas qu'il faudrait arrimer notre schème conceptuel au réel de l'extérieur ; la raison est bien plutôt que, sans ces théories, notre schème conceptuel s'effondre de l'intérieur. Du point de vue de notre

44 Putnam H., « Models and Reality », 1980, p. 477/18.

45 La première phrase de la citation qui suit ne pourrait-elle pas avoir été écrite par Putnam ? Il y a quelque chose de curieux à prendre ici Field pour l'homme de paille du réalisme métaphysique.

théorie du monde, il serait extrêmement surprenant qu'il existe une connexion non-causale entre les mots et les choses⁴⁶. »

Devitt, qui voit pourtant dans les théories causales de la référence un moyen de rendre compte de la conception correspondantiste du réalisme, semble abonder :

« Derrière la position de Putnam, il y a pour une part un certain nombre d'évidences. Nous ne pouvons rien dire à propos de la relation entre le langage et le monde sans dire quelque chose – et donc sans utiliser le langage. Pour l'étude de quelque sujet que ce soit, on est emprisonné au sein du langage⁴⁷. »

En quoi réside alors exactement le désaccord entre le réalisme métaphysique et ses critiques concernant la théorie causale ? On l'a vu, il ne s'agit pas d'un désaccord sur la valeur de cette théorie, le Putnam que Devitt qualifie de renégat et Devitt lui-même sont d'accord pour dire qu'il s'agit sans doute d'un élément important pour expliquer la référence. Il s'agit d'un désaccord quant à sa portée. Pour Devitt, la théorie causale montre la possibilité du décalage réaliste entre notre théorie et le monde : la référence des termes est fixée par des relations causales et elle peut toujours être différente de ce que dit notre théorie. Il est certes légitime de demander comment est fixée la détermination de « relations causales », mais, pour Devitt, le réaliste dispose d'une réponse acceptable malgré son apparente circularité :

« “relations causales” est causalement relié d'une certaine manière B aux relations causales et à rien d'autre. “*relations causales*” est “*collé à une relation déterminée*” par des relations causales, pas par “une colle métaphysique”⁴⁸. »

Pour Devitt, Putnam se rend coupable d'une pétition de principe lorsqu'il demande au réaliste de justifier l'interprétation réaliste de la référence de « relations causales ». Il n'accepte pas qu'il faille à un moment s'arrêter dans nos explications. Pour Putnam, le réaliste n'a pas le droit de supposer que « relations causales » possède une référence déterminée car l'argument pour l'indétermination peut être répété. Le problème de la position de Devitt réside dans la circularité de la réponse : « relations causales » n'a une référence déterminée que si « relations causales » possède une référence déterminée. Le problème de la position de Putnam est que l'argument modèle théorique qu'il propose

46 Field H., « Tarski's Theory of Truth », 1972, p. 373.

47 Devitt M., *Realism and Truth*, 1984, p. 226.

48 *Realism and Truth*, 1984, p. 227.

de répéter à un niveau supérieur pour l'appliquer à « relations causales » suppose que la référence n'est fixée que par des contraintes théoriques et opérationnelles, ce qui est précisément ce que conteste le partisan de l'interprétation réaliste de la théorie causale. Tenter de trancher le désaccord entre Putnam et Devitt, ou montrer qu'il ne peut l'être, sortirait du cadre du présent article. Nous retiendrons simplement l'identification du désaccord : certains réalistes comme Devitt reprennent à leur compte une lecture pleinement réaliste de l'externalisme du Putnam de « Meaning and Reference », tandis que pour le Putnam de *Reason, Truth and History*, la théorie causale, correctement interprétée, ne permet pas de sortir de la perspective interne.

VIII. Externalisme et essentialisme

Putnam se livre à une seconde reprise critique de l'externalisme au chapitre III. Il s'agit à nouveau de montrer que celui-ci ne permet pas de justifier la conception correspondantiste de la référence propre au réalisme métaphysique. Putnam s'attache à relever un certain nombre de difficultés concernant la chaîne de relations causales supposée lier le terme et son référent. Pour commencer, que veut-on dire exactement quand on parle d'une chaîne du « type approprié », qu'est-ce qui fait que la chaîne qui part de « électron » pour aboutir aux manuels de physique n'est pas appropriée alors que celle qui part de « électron » pour aboutir à des particules l'est ? Ensuite, une telle chaîne causale existe-t-elle toujours ? Après tout, le terme « extraterrestre », s'il fait référence, fait référence à des créatures que nous n'avons jamais rencontrées. Le défenseur de l'externalisme doit restreindre le champ d'application à des termes de base, et expliquer que « extraterrestre » se laisse définir comme « créature intelligente vivant sur une autre planète », dont les termes seraient eux-mêmes justiciables d'une analyse causale. Mais même pour ces expressions, l'ancrage causal ne saurait rendre compte de l'ensemble de la référence. « Planète » fait référence à des planètes lointaines, à des planètes disparues, à des planètes à venir, avec lesquelles ni nous-mêmes ni aucun locuteur de notre communauté n'avons jamais été liés par une chaîne causale. La théorie causale doit donc fournir un principe d'extension de la référence pour suppléer la chaîne causale et attraper dans les filets de la référence les objets hors d'atteinte. Ce principe fait partie intégrante des théories causales « à la Kripke–Putnam », comme nous l'avons mentionné au début du présent article (section II). Il consiste à faire appel à une relation d'identité sortale (« être la même espèce de choses »). Le terme « planète » fait référence à la Terre, aux planètes du système solaire, ce sont sans doute les échantillons typiques

auxquels les locuteurs sont causalement reliés, et il fait référence aux choses « de la même espèce », aux autres objets qui sont comme la Terre et les autres planètes du système solaire, c'est-à-dire aux planètes.

C'est à propos de la relation d'identité ainsi convoquée que se pose la question du réalisme. Selon une lecture possible, l'identité en question renvoie à une sorte de nature des choses. Elle est purement et simplement donnée par le monde, dès que sont déterminés les échantillons caractéristiques. Une fois qu'on a dit que la Terre et ses voisines du système solaire sont des planètes, ce qui compte comme une planète serait automatiquement déterminé. Telle semble être la perspective de Kripke lorsqu'il analyse la référence des termes d'espèce naturelle dans *Naming and Necessity*. Pour Kripke, nous utilisons le mot « or » pour désigner une certaine espèce de choses : l'or est une substance, et une fois la référence associée à certains échantillons, le terme « or » fait référence à la substance exemplifiée par les échantillons. Certains passages de « The Meaning of "Meaning" » pourraient inviter à une lecture similaire :

« Selon la conception que je défends, quand Archimède disait que quelque chose était de l'or [...] il disait que cette chose possédait la même *structure cachée* d'ensemble (la même "*essence*" pour ainsi dire) que n'importe morceau normal d'or présent dans l'environnement⁴⁹. »

Mais essentialiser la relation d'identité sortale invoquée pour étendre la référence revient à céder à l'illusion métaphysique réaliste d'un monde constitué indépendamment de nous. Ce faisant,

« [...] on affirme tout bien considéré que le monde contient des Objets Auto-Identifiants, car c'est exactement ce qu'on dit quand on dit que c'est le *monde*, et non pas les sujets pensants, qui trie les choses en espèces⁵⁰. »

Pourtant, il serait excessif de considérer que le Putnam de « The Meaning of "Meaning" » tombait dans l'illusion critiquée par le Putnam de *Reason, Truth and History*. Les guillemets autour de « essence » dans la citation précédente de « The Meaning of "Meaning" » ont leur importance, ils indiquent que la notion est à prendre avec des pincettes. De fait, dans « The Meaning of "Meaning" » comme dans « Meaning and Reference », la relation d'identité assurant l'extension de la référence est présentée comme une relation théorique, ouverte à l'enquête scientifique et dépendant de nos intérêts. Passant de l'or à l'eau, voici l'explication proposée dans « The Meaning of

49 Putnam H., *Mind, Language and Reality*, 1975, p. 235.

50 Putnam H., *Reason, Truth and History*, 1981, p. 53 (traduction française p. 65 modifiée).

“Meaning” » de la relation « même_L » (où L est mis pour « liquide ») :

« Être de l’eau, par exemple, c’est être dans la relation même_L avec certaines choses. Mais quelle est cette relation ? x est dans la relation même_L avec y lorsque (1) x et y sont tous les deux des liquides, et (2) x et y partagent d’importantes propriétés physiques⁵¹. »

Les importantes propriétés physiques en question pourront typiquement inclure la structure moléculaire, mais celle-ci n’a pas de place à part : elle n’est supposée ni être plus importante que toute autre chose ni être donnée indépendamment de nos théories. Rassembler les approches externalistes de la signification de Kripke et de Putnam sous la bannière commune d’une « théorie causale de la référence de Kripke–Putnam » est trompeur. Comme l’explique Ian Hacking dans un article justement intitulé « Putnam’s Theory of Natural Kinds and their Names is Not the Same as Kripke’s », la théorie de la référence des termes d’espèce naturelle, parce qu’elle est externaliste, dépend de la question de savoir ce qu’est une espèce naturelle, et la réponse essentialiste de Kripke est très différente de la réponse pragmatique de Putnam.

IX. Conclusion

La théorie externaliste de la signification n’est pas abandonnée avec le passage au réalisme interne. Putnam en retient à la fois les aspects critiques (les significations ne sont pas dans la tête) et positifs (la référence est gagée sur des échantillons présents dans l’environnement). Que se passe-t-il alors ? Nous avons affaire à une clarification des implications de l’externalisme pour la question du réalisme, à la lumière de l’opposition mise en place entre réalisme métaphysique et perspective interne. Dans ses reprises critiques de la théorie causale, Putnam montre que celle-ci ne permet pas de défendre une compréhension réaliste métaphysique de la référence. L’ancrage causal de la référence ne justifie pas une conception correspondantiste de la référence liant nos expressions aux choses d’un monde vierge. Pourquoi ? Parce que la théorie de la signification n’a pas de privilège métaphysique : elle fait partie intégrante de notre théorisation du monde et ne nous donne pas accès à un point de vue extérieur sur le monde et ceux qui le pensent. Il y a, dans la perspective du réalisme interne, un danger propre à l’externalisme, qui consisterait à oublier que l’extériorité dont il est question n’est jamais celle, illusoire, d’un monde vierge dans lequel la pratique linguistique viendrait ancrer la référence, mais toujours celle du monde tel que nous le saisissons,

51 Putnam H., *Mind, Language and Reality*, 1975, p. 238-239.

avec nos concepts, nos théories, nos intérêts. De ce point de vue, la reprise critique des théories causales est plus systématique qu'il n'y paraît au premier d'abord. La théorie externaliste de la signification est articulée autour des deux moments distincts de l'ancrage indexical de la référence et de l'extension de la référence par similarité. Dans les deux cas, il s'agit pour Putnam de rappeler que ces deux moments sont bien des moments théorisés. Nous parlons de relations causales pour rendre compte de l'ancrage indexical, nous parlons d'identité sortale pour rendre compte de l'extension de la référence. Dans les deux cas, il ne s'agit pas de mystérieuses relations métaphysiques brutes. Les relations causales constituent une relation mobilisée au sein d'une certaine métathéorie sémantique, relation qui est donc susceptible d'interprétations non-standard. La relation d'identité sortale est une relation mobilisée au sein d'une certaine théorie scientifique, elle est donc dépendante de nos connaissances et de nos intérêts. Qu'il soit ou non un renégat du réalisme, Putnam ne serait donc pas un renégat de l'externalisme. Il se contenterait de le remettre à sa place, pour ainsi dire, c'est-à-dire au sein de notre théorie du monde et non pas en surplomb de notre activité théorique⁵².

Il semble cependant que subsiste une difficulté interprétative. Si l'externalisme est maintenu comme notre meilleure théorie de la signification, si sa reprise critique consiste simplement à nous prévenir contre les interprétations abusivement réalistes métaphysiques qui en seraient faites, pourquoi Putnam déclare-t-il que les théories causales de la référence sont vaines (« otiose »⁵³), comme nous l'avons rappelé dans la section III ? La réponse réside peut-être dans les conséquences de l'internalisation de la théorie de la signification. La lecture métaphysique réaliste de l'externalisme pouvait laisser penser que la théorie de la référence savait quelque chose de plus sur la référence des mots que ce que nous savons déjà. Ce que nous appelons « or » n'est peut-être pas un métal jaune, c'est la substance qui possède la même structure cachée que les échantillons présents dans notre environnement, et cela serait déjà déterminé par la théorie de la référence avant même que nous ayons découvert ce dont il s'agit. Dans la perspective du réalisme interne, la théorie de la signification est simplement partie prenante de notre théorie : elle nous dit que l'or est constitué des choses métalliques

52 On pourrait être tenté d'appliquer cette lecture plus largement, et analyser la réinterprétation d'autres concepts selon la perspective du réalisme interne. Maria Baghramian soutient ainsi à propos de la seconde période de Putnam que « les idées métaphysiques centrales de vérité, de référence et de représentation reçoivent une nouvelle jeunesse dans cette approche en perdant leur aura d'indépendance à l'égard de l'esprit et en étant traitées de manière interne ou à l'intérieur des limites d'un schème conceptuel » (« "From Realism Back to Realism": Putnam's Long Journey », 2008, p. 25).

53 Putnam H., *Reason, Truth and History*, 1981, p. 53.

partageant avec les échantillons disponibles autour de nous un certain nombre de propriétés physiques importantes. Mais c'est exactement la même chose que ce que nous dit la physique lorsqu'elle nous dit que l'or est l'élément chimique de numéro atomique 79 possédant trente-sept isotopes connus dont un seul est stable et correspond à l'or présent dans notre environnement. Les propriétés figurant dans cette description de l'or sont précisément les propriétés possédées, selon la théorie physique, par les échantillons qui constituent aujourd'hui pour nous la relation de similarité « être le même métal ». Ainsi, la théorie de la référence internalisée est bien redondante. Lorsqu'on l'applique à la référence d'un terme en particulier, elle nous dit la même chose que ce que nous dit notre théorie du monde des objets nommés par ce terme. C'est en ce sens que pour le Putnam de *Reason, Truth and History*, la théorie de la référence internalisée se ramène à des tautologies⁵⁴. « Or » fait référence à l'or. On peut certes déplier cet énoncé en disant que « or » fait référence à l'élément chimique caractérisé comme précédemment, mais ce dépliement ne nous mène pas au-delà de ce que nous savons (et comment le pourrait-il dans la perspective interne ?). Ces considérations renforcent la seconde interprétation proposée de la montée sémantique de l'argument modèle-théorique (voir plus haut, section VII) : si la métathéorie sémantique est déjà intégrée à notre théorie du monde, alors effectivement les modèles pour cette métathéorie ne sont pas plus informatifs que les modèles de notre théorie du monde, et la théorie de la référence aboutit à des tautologies.

En ce sens, la théorie de la référence est effectivement vaine (« otiose ») : elle ne nous dit rien de plus sur la référence de tel ou tel de nos termes que ce que nous savons déjà. Est-elle pour autant absolument vaine ? Après tout, cette interprétation déflationniste de la théorie de la référence dans la perspective du réalisme interne semble créer une difficulté herméneutique symétrique de la difficulté initiale. Si la théorie de la référence ne mène qu'à des tautologies, à quoi bon une telle théorie ? Et pourquoi Putnam continue-t-il de la mobiliser de façon positive dans *Reason, Truth and History*, comme nous l'avons pourtant vu dans la section VI ? Une réponse possible serait d'imaginer un Putnam wittgensteinien⁵⁵ nous invitant au terme du développement de la théorie de la référence à retirer l'échelle et à reconnaître son caractère tautologique et vide de sens.

54 *Reason, Truth and History*, 1981, p. 52 (traduction française p. 64).

55 La chose n'est pas si difficile à imaginer. Le premier chapitre de *Reason, Truth and History* esquisse une théorie des concepts fondés sur l'usage qui se réclame explicitement de Wittgenstein. Et l'idée selon laquelle les signes n'ont pas de signification intrinsèque est un des enseignements de Wittgenstein dans les *Recherches Philosophiques*, comme le rappelle Claudine Tiercelin en soulignant l'héritage pragmatique du réalisme interne (*Hilary Putnam, l'héritage pragmatiste*, 2013, chapitre 2).

Cette réponse semble cependant excessive en regard des usages tout à fait substantiels de la théorie de la référence que l'on peut trouver dans *Meaning and the Moral Sciences* et *Reason, Truth and History*. Une autre réponse possible, qui nous semble plus convaincante, consiste à distinguer un usage statique et un usage dynamique de la théorie de la référence. Pour un état donné de notre théorie du monde, l'analyse des relations causales et de la relation de similarité pertinentes pour la référence d'un terme ne nous mène effectivement qu'à ce que nous savions déjà. En revanche, lorsque nous passons d'une théorie du monde à une autre, la théorie de la référence acquiert une dimension explicative non triviale. Elle nous permet de comprendre la nature des considérations qui entrent en jeu dans la permanence de la référence et la manière dont intervient le principe de charité. C'est ce que nous avons vu dans la section V en présentant l'explication de la permanence trans-théorique de la référence dans la perspective du réalisme interne. On va adopter pour les termes de l'ancienne théorie une référence dans la nouvelle théorie compatible à la fois avec son rôle causal supposé dans les phénomènes visés et avec les propriétés décrites dans l'ancienne théorie. C'est, selon Putnam, comme cela que s'y prennent *de facto* les scientifiques lorsqu'il s'agit de reprendre les termes de l'ancienne théorie (on ne débaptise pas les électrons quand on passe du modèle de Bohr à la physique quantique, on considère que ce dont voulait parler Bohr, c'est des électrons au sens de la physique quantique). Mais s'y prendre de cette façon, c'est souscrire à une théorie externaliste de la référence. On considère que Bohr faisait référence aux choses qui ont les propriétés physiques importantes (extension) de ce qui explique les phénomènes qu'il attribuait aux électrons (ancrage). L'externalisme a donc bien au moins une vertu positive, celle de rendre compte de la façon dont s'applique le principe de charité lors du changement de théorie. Cette vertu explicative est tout à fait compatible avec le caractère tautologique de la théorie de la référence considérée de façon statique au sein de notre théorie du monde. Une telle interprétation est en outre renforcée par le fait que tous les usages positifs de la théorie causale de la référence dans *Meaning and the Moral Sciences* et *Reason, Truth and History*⁵⁶ correspondent à des usages dynamiques étudiant la référence au sein d'une théorie donnée du point de vue d'une autre théorie. À la lumière du réalisme interne, la théorie externaliste de la signification serait une théorie de la référence internalisée

56 Ceux que nous avons répertoriés au fil de cet article sont au nombre de trois : la référence de « électron » après le passage à la théorie quantique, la référence des termes du français-de-la-cuve du point de vue de notre théorie du monde, la référence de « rationalité » dans une conception dynamique de la rationalité. On pourrait y ajouter ce qui est dit de la référence de « eau » avant 1750 du point de vue de notre théorie chimique.

dont la vertu explicative se limiterait aux situations de réinterprétation d'une théorie dans une autre.

Bibliographie

- Anderson D. L., « What is Realistic about Putnam's Internal Realism ? », in *Philosophical Topics*, vol. 20, n° 1, 1992, p. 49-83.
- Baghramian M., « "From Realism Back to Realism": Putnam's Long Journey », in *Philosophical Topics*, vol. 36, n° 1, 2008, p. 17-35.
- Devitt M., *Realism and Truth*, Cambridge MA, Princeton University Press, 1984.
- Dummett M., *The Logical Basis of Metaphysics*, Cambridge MA, Harvard University Press, 1991.
- Field H., « Tarski's Theory of Truth », in *The Journal of Philosophy*, vol. 69, n° 13, 1972, p. 347-375.
- Hacking I., « Putnam's Theory of Natural Kinds and Their Names is Not the Same as Kripke's », in *Principia*, vol. 11, n° 1, 2007, p. 1-24.
- Kripke S., *Naming and Necessity*, Cambridge MA, Harvard University Press, 1980, traduction française par P. Jacob et F. Récanati, *La logique des noms propres*, Paris, Éditions de Minuit, 1982.
- Michaelson E. & Reimer M., « Reference », in E. N. Zalta (éd), *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*, en ligne (<https://plato.stanford.edu/archives/spr2017/entries/reference>).
- Putnam H., « Meaning and Reference », in *The Journal of Philosophy*, vol. 70, n° 19, 1973, p. 699-711.
- Putnam H., *Mathematics, Matter and Method, Philosophical Papers vol. 1*, Cambridge, Cambridge University Press, 1975.
- Putnam H., *Mind, Language and Reality, Philosophical Papers vol. 2*, Cambridge, Cambridge University Press, 1975.
- Putnam H., « Realism and Reason », in *Proceedings and Addresses of the American Philosophical Association*, vol. 50, n° 6, 1977, p. 483-498.
- Putnam H., *Meaning and the Moral Sciences*, Londres, Routledge & Kegan Paul Ltd, 1978.
- Putnam H., « Models and Reality », *The Journal of Symbolic Logic*, vol. 45, n° 3, 1980, p. 464-482. Réimprimé dans Putnam H., *Realism and Reason, Philosophical Papers vol. 3*, Cambridge, Cambridge University Press, 1983.
- Putnam H., *Reason, Truth and History*, Cambridge, Cambridge University Press, 1981 ; traduction française par A. Gerschenfeld, Paris, Les Éditions de Minuit, 1984.
- Putnam H., « Three Kinds of Scientific Realism », in *The Philosophical Quarterly*, vol. 32, n° 128, 1982, p. 195-200.
- Tiercelin C., *Putnam, l'héritage pragmatiste*, Paris, Presses Universitaires de France, 2002.
- Wikforss Å., « Bachelors, Energy, Cats and Water: Putnam on Kinds and Kind Terms », in *Theoria*, vol. 79, 2013, p. 242-261.